T.C. ANADOLU ÜNİVERSİTESİ YAYINI NO: 2590
AÇIKÖĞRETİM FAKÜLTESİ YAYINI NO: 1559

İSTATİSTİK

Yazarlar
Prof.Dr. Ahmet ÖZMEN (Ünite 1)
Prof.Dr. Fikret ER (Ünite 2)
Dr.Öğr.Üyesi Mahmut ATLAS (Ünite 3)
Dr.Öğr.Üyesi Atilla ASLANARGUN (Ünite 4)
Doç.Dr. Kadir Özgür PEKER (Ünite 5)
Prof.Dr. Emel ŞIKLAR (Ünite 6, 7)
Prof.Dr. Harun SÖNMEZ (Ünite 8)

Editörler
Prof.Dr. Ahmet ÖZMEN
Prof.Dr. Berat Fethi ŞENİŞ

ANADOLU ÜNİVERSİTESİ

ÖĞRENME TEKNOLOJİLERİ AR-GE BİRİMİ

Birim Yöneticisi
Doç.Dr. Alper Tolga Kumtepe

Kitap Hazırlama Grubu Sorumlusu
Öğr.Gör. Erdem Erdoğdu

Öğretim Tasarımcıları
Prof.Dr. Tevfik Volkan Yüzer
Öğr.Gör. Orkun Şen

Grafik Tasarım Yönetmenleri
Prof. Tevfik Fikret Uçar
Doç.Dr. Nilgün Salur
Öğr.Gör. Cemalettin Yıldız

Dil ve Yazım Danışmanı
Öğr.Gör.Dr. Sevgi Çalışır Zenci

Grafikerler
Aysun Şavlı
Ayşegül Dibek
Gülşah Karabulut
Hilal Özcan

Kapak Düzeni
Doç.Dr. Halit Turgay Ünalan

Dizgi
Kitap Hazırlama Grubu

İstatistik

E-ISBN
978-975-06-2693-7

ESKİŞEHİR, Ağustos 2018
2386-0-0-0-1609-V01

İçindekiler

Önsöz

Bu kitap, iktisadi ve idari bilimler lisans ve önlisans programları ile pek çok diğer sosyal bilimlerle ilgili programlarda verilen istatistik derslerinde işlenen konular temel alınarak oluşturulmuştur.

Sekiz üniteden oluşan bu kitabın birinci ünitesinde istatistiğin temel kavramları tanıtılmış, izleyen ünitelerde de sırasıyla betimsel istatistik, çıkarım, kestirim ve öngörü bilgilerinin üretilmesinde kullanılan bazı yöntem ve teknikler hem kuramsal düzeyde hem de örnekler üzerinde açıklanmıştır.

Kitapta yer alan konular belirlenirken sosyal bilim alanlarında araştırma yapanlar, yapılan araştırmalarda üretilen bilgilerden yararlananlar için yol gösterici bir kaynak olmasına özen gösterilmiştir.

Anadolu Üniversitesi'nin uzaktan öğretim hizmeti sunulan lisans ve önlisans programlarında kullanılmak amacıyla geliştirilen bu kitap, içerik tasarımı ile uygulamalı diğer bilim dallarının lisans ve önlisans öğrencileri ile araştırmacılar için de yararlı olacağını umuyoruz.

Editörler
Prof.Dr. Ahmet ÖZMEN
Prof.Dr. Berat Fethi ŞENİŞ

İSTATİSTİK

1

Amaçlarımız

Bu üniteyi tamamladıktan sonra;

- İstatistiğin konusu olan ve olmayan olayları tanımlayabilecek,
- İstatistik kavramının çeşitli tanımlarını yapabilecek,
- İstatistiğin temel kavramlarını açıklayabilecek,
- İstatistiksel bir araştırma sürecini tasarlayabileceksiniz.

Anahtar Kavramlar

- Yığın Olay
- Tipik Olay
- İstatistik
- Veri Kümesi
- Örneklem Değer
- Bilimsel Araştırma

İçindekiler

İstatistiğin Tanımı, Temel Kavramlar ve İstatistik Eğitimi

İSTATİSTİĞİN KONUSU NEDİR?

İstatistiğin konusu olan olayları açıklayabilmek için bir olayın kendi türünden olayları, ilgilenilen özellikler bakımından tam olarak temsil edip etmediğine, başka bir ifadeyle olayların ilgilenilen özelliklerinin ölçüm değerleri arasında değişkenlik ve belirsizlik olup olmadığına bakmak gerekir. Olayların özellikleriyle ilgili ölçüm değerlerindeki değişkenlik ölçüt alındığında, olaylar yığın olay ve tipik olay olarak sınıflandırılır.

K İ T A P

Bu sınıflandırma Özer Serper'in "Uygulamalı İstatistik 1" (Bursa: Genişletilmiş 5. Baskı, Ezgi Kitabevi, 2004.) adlı kitabından alınmıştır. Bu kitabın birinci bölümünde konuyla ilgili ayrıntılı bilgi bulabilirsiniz.

Yığın Olay

Özellikleri en az iki farklı sonuca sahip olaylar **yığın olaylar**dır.

İstatistik **yığın olaylar**la ilgilenir. Bu olaylar ilgilenilen özellikleri bakımından değişkenlik ve belirsizlik gösterirler. Aynı koşullar ve varsayımlar altında meydana gelen, özellikleri aynı sonuçların (ölçüm değerlerinin) yanında farklı sonuçları da alabilen olaylara yığın olay denir.

Yığın olayların incelenen özellikleri olaydan olaya değişkenlik gösterdiği için herhangi bir yığın olay ait olduğu kümedeki olayları tam olarak temsil edemez.

ÖRNEK 1

2011-2012 öğretim yılında Anadolu Üniversitesi'nin açıköğretim yapan fakültelerinde öğrenim gören öğrenciler;
öğretim yılı,
uygulanan öğretim sistemi,
cinsiyet,
başarı durumu
özellikleri bakımından araştırılmak isteniyor. Bu araştırmada açıköğretim yapan fakültelerde öğrenim gören öğrenciler ne tür olaydır, istatistiğin konusu olur mu?

Çözüm 1:

Öğrencilere araştırılmak istenen özelliklerle ilgili yöneltilen Tablo 1.1'deki sorular ve alınabilecek mümkün yanıtlar incelendiğinde öğrencilerin ilk iki soruya verdiği yanıtlar tümüyle aynıyken son iki soruya verdikleri yanıtlar farklı olabilecektir. İki ortak özelliğin yanında, iki farklı özelliğe sahip olan bu öğrenciler yığın olaydır ve istatistiğin konusu olurlar. *Öğrencileri yığın olay yapan neden, cinsiyet ve başarı durumu özelliklerinin değişkenlik gösterebilecek olmasıdır.*

Tablo 1.1 *Araştırmanın incelenen özelliklerini tanımlayan sorular ve mümkün yanıtları.*

Sorular	Mümkün Yanıtlar
Öğretim yılınız nedir?	2011-2012
Uygulanan öğretim sistemi nedir?	Açıköğretim
Cinsiyetiniz nedir?	Erkek () Kadın ()
Başarı durumunuz nedir?	0, 1, ..., 45, ..., 65, ... 78, ... 100 Veya Geçmez, Orta, İyi, Pekiyi

Türkiye'de faaliyette bulunan lojistik, sigorta, banka işletmelerinin her biri yığın olaydır. Çünkü ciroları, işgören sayıları, kârları ve benzeri özellikleri bakımından farklılıklar gösterirler. Benzer değerlendirme yapıldığında;

- Bir banka şubesinde görevliler tarafından zaman içinde düzenlenen fişlerin her biri,
- Bir tatil döneminde İstanbul- Antalya karayolunda meydana gelen trafik kazalarının her biri,
- Aylık enflasyon oranları ile ilgili araştırmalarda her bir ay yığın olaydır.

Örnek olarak yukarıda verilen yığın olaylara bakıldığında bu olayların birey, kurum, kuruluş, olay, nesne, işlem, zaman vb. olabildiğini görürüz. Özetle yığın olaylar canlı, cansız varlıklar, olaylar, işlemler olabilir. Süreç kararlılığı bulunmayan sosyal bilimler, mühendislik bilimleri, fen bilimleri ve sağlık bilimleriyle ilgili olaylar genellikle yığın olaylardır.

DİKKAT

İstatistiğin konusu olan olaylar topluluğunun bir başka özelliği de olaylardan ayrı ve farklı bir varlığa sahip olmamasıdır. Örneğin bir banka şubesindeki işgörenler birer yığın olay olarak tanımlandığında bu banka şubesi bir yığın olay olarak düşünülemez. Çünkü şubenin çalışanlarının dışında ayrı bir varlığı vardır. Fakat bir bankanın şubeleri birer yığın olay olarak değerlendirilebilir.

SIRA SİZDE 1

- **Türkiye'de faaliyette bulunan sigorta şirketleri hangi özellikleri bakımından araştırma konusu yapılırsa istatistiğin konusu olay olur?**
- **Benzinli otomobillerin hangi özellikleri incelenirse istatistiğin konusu olmaz?**

Tipik Olay

Aynı koşullar ve varsayımlar altında meydana gelen, incelenen özellikleri bakımından aynı sonuçları alan olaylara tipik olay denir.

ÖRNEK 2

2011-2012 öğretim yılında açıköğretim yapan fakültelere kayıtlı olan öğrenciler öğrenci olup olmama özelliği bakımından incelenecek olsun. Bu incelemede öğrenciler istatistiğin konusu olur mu?

Çözüm 2:

Araştırmaya konu olan özellik öğrenci olup olmama durumu olduğuna ve hakkında araştırma yapılacak olanların tamamı da açıköğretim yapan fakültelere kayıtlı öğrenciler olduğuna göre araştırmaya konu olan özellik aynı ölçüm değerini (öğrenciyim yanıtını) alacaktır. Bu değerlendirmeye göre 2011-2012 öğretim yılında açıköğretim fakültelerinde kayıtlı olan öğrenciler, öğrenci olup olmama özelliği bakımından **tipik olay**dır, istatistiğin konusu olmazlar.

Bir hastanın kan tahlili için aynı anda alınacak onlarca enjektör kanın tahlili, suyun deniz seviyesindeki kaynama noktasındaki sıcaklık düzeyini belirlemek için yapılacak, çok sayıda kaptaki suyun ısıtılması deneylerinde her bir kan tahlili ve suyun ısıtılması deneyi birer tipik olaydır. Çünkü bir enjektördeki kanın tahlil sonuçları diğer enjektörlerdeki kanın tahlil sonuçlarıyla bir kaptaki suyun ısıtılması deneyinin sonucu da onlarca kaptaki suyun ısıtılması deneyinin sonucuyla farklılık göstermez. Bu nedenle yapılan bir tahlilin ve deneyin sonucu bütün tahlillerin ve deneylerin sonucunu tam olarak temsil eder. Bu nedenle örnek verilen yukarıdaki olaylar tipik olaydır.

Tipik olayların ilgilenilen özellikleri tek bir ölçüm değerine sahiptir. Bu olayların özellikleri değişkenlik göstermez ve bu nedenle herhangi bir tipik olay ait olduğu olaylar kümesindeki bütün olayları tam olarak açıklar. İstatistik tipik olaylarla ilgilenmez. Ayrıca istatistik tek bir olayın bir özelliğinin alacağı bir ölçüm değeri ile de ilgilenmez.

SIRA SİZDE 2

- **Belirli bir banka şubesinde çalışan personel, çalıştığı işyeri türü bakımından araştırılmak isteniyor. Bu araştırmada şube çalışanları ne tür olaydır? Açıklayınız.**
- **Bir sigorta acentesi çalıştırdığı personel sayısı bakımından araştırılmak isteniyor. Bu araştırma için söz konusu sigorta acentası yığın olay mıdır? Açıklayınız.**

İSTATİSTİK NEDİR?

İstatistik değişik anlamlarda kullanılan, bunun sonucu olarak da farklı tanımları bulunan bir kavramdır. Quetelet isimli bir istatistikçi 19. yüzyılda kavramın yüzden fazla tanımının olduğunu belirlemiştir. Tarih boyunca toplumların, devletlerin istatistiğe duydukları gereksinime paralel olarak geliştirilmiş olan çeşitli tanımların üzerinde durmak yerine, bu kısımda istatistik kavramının yaygın olarak kullanılan üç farklı anlamı üzerinde durulmuştur.

Derlenen verilerin oluşturduğu kümeye **istatistik** denir.

Veri Kümesi Anlamında İstatistik

Tanımlanan belirli bir konuda belirli amaç için yığın olayların çeşitli özellikleriyle ilgili olarak derlenmiş olan ve bir anlam ifade eden rakam, sayı, simgelere veri; verilerin oluşturduğu topluluğa veri kümesi veya istatistik denir.

Veri kümesi anlamında istatistik kelimesi, daha çok kamu kurum ve kuruluşları olmak üzere her türlü kuruluş tarafından kendi görevleri, amaçları ve toplumun faydalanması için çeşitli konularda derlenen veriler kümesi anlamında kullanılmaktadır. Örneğin Eskişehir Turizm İstatistikleri dendiğinde Tablo 1.2'de gösterildiği gibi yıllar itibariyle Eskişehir'e gelen yabancı, yerli kişi sayıları, geceleme sayıları, ortalama kalış süreleri ve turistik tesislerin doluluk oranları vb. gibi konularda derlenmiş olan verilere genel olarak turizm istatistikleri adı verilmektedir. Pek çok kişi tarafından istatistiğin bu anlamda kullanımına ilişkin olarak, nüfus istatistikleri, dış ticaret istatistikleri, milli eğitim istatistikleri, sigorta istatistikleri, spor istatistikleri, enflasyon ve cari açık istatistikleri, döviz kuru istatistikleri sıkça kullanılan örnekler olarak gösterilebilir. Bu anlamda istatistik bir sayısal tanımlamadır.

Tablo 1.2
Eskişehir'de turistik tesislere giriş, geceleme sayısı, ortalama kalış süresi ve doluluk oranı.

Kaynak: *Eskişehir Sanayi Odası Dergisi ESO Dergi (Yıl 2, Sayı 3 Ocak-Şubat 2010).*

	TESİSE GELİŞ SAYISI			GECELEME SAYISI			ORTALAMA KALIŞ SÜRESİ			DOLULUK ORANI (%)		
Yıl	Yabancı	Yerli	Toplam	Yabancı	Yerli	Toplam	Yabancı	Yerli	Toplam	Yabancı	Yerli	Toplam
2000	2.070	60.024	62.094	7.780	76.449	84.229	3,8	1,3	1,4	2,91	28,61	31,53
2001	2.090	37.372	39.462	8.079	48.116	56.195	3,9	1,3	1,4	3,26	19,44	22,75
2002	1.963	45.783	47.746	4.858	58.110	62.968	2,5	1,3	1,3	1,95	23,34	25,29
2003	2.223	51.023	53.246	6.253	69.911	76.164	2,8	1,4	1,4	2,53	28,24	30,27
2004	2.609	60.151	62.770	9.206	94.704	103.910	3,5	1,6	1,7	3,74	38,45	42,19
2005	4.030	91.945	95.975	20.031	140.228	160.259	5,0	1,5	1,7	4,14	28,95	33,09
2006	3.693	103.556	107.249	10.251	157.235	167.486	2,8	1,5	1,6	2,13	32,60	34,72
2007	8.412	128.056	136.468	20.313	200.091	221.214	2,4	1,6	1,6	3,74	36,99	40,73
2008	9.062	129.774	138.836	19.657	195.904	215.561	2,2	1,5	1,6	3,50	34,87	38,36

İstatistiğin veri kümesi anlamında kullanılışı çok uzun bir geçmişe sahiptir. İnsanlar toplu hâlde yaşamaya geçip devletler kurmaya başladığında nüfus, vergi toplama, askere alma, tarım, ticaret vb. konularda veri toplamaya ve bunların kayıtlarını tutmaya başlamışlardır. Ülkemizde Cumhuriyet döneminden önce veri kümesi anlamında istatistik kelimesinin ihsai (sayımla ve sayımlarla ilgili) ve ihsaiyat (istatistik) sözcükleriyle kullanıldığı bilinmektedir. Cumhuriyetimizin ilk yıllarında olumsuz koşullara rağmen istatistiki verilerin geleceğin planlanmasındaki, gelişmelerin izlenebilmesindeki ve sağlıklı kararların alınabilmesindeki önemine inanıldığı için 1926 yılında Türkiye'nin istatistiklerini üreten bugünkü adı Türkiye İstatistik Kurumu (TÜİK) olan Devlet İstatistik Enstitüsü (DİE) kurulmuştur. TÜİK'in derleyip düzenlediği ve yayınladığı istatistikler kamu ve özel kuruluşlar ve karar alıcılar için önemli kaynak, yön gösterici durumundadır.

Yöntemler Topluluğu-Bilim Dalı Anlamında İstatistik

Veri kümesi anlamında istatistik günümüzde de yaygın olarak kullanılmakla birlikte, istatistik derlenen verilerin matematik temellerindeki ve olasılık kuramındaki gelişmeler sonucu, karar verme amacıyla da kullanılmaya başlanmıştır. İstatistikle ilgili artarak devam eden bu gelişmeler temelini matematikten alan yeni bir istatistik tanımı yapma ihtiyacını doğurmuştur. Bu ihtiyaç beraberinde istatistik bilim midir, yöntemler topluluğu mudur tartışmasını başlatmıştır.

İstatistik, yöntemler topluluğudur diyenler istatistiğin kendine özgü konusu, kuralları olmadığını, diğer bilimlerle ilgili bilimsel bilgi üretme süreçlerinde kullanılan önemli bir araç olduğu görüşündedir. Bu nedenle istatistik bir bilimsel araştırma sürecinde gerekli olan verilerin derlenmesi, düzenlenmesi, çözümlenmesi ve çözümleme sonunda elde edilen bilgilerin değerlendirilmesi amacıyla kullanılan teknik ve yöntemler topluluğudur şeklinde tanımlanmaktadır. Bu tanım bağlamında istatistik temel bir araçtır. Dersimizin adı yöntemler topluluğu anlamında istatistiğin karşılığıdır.

İstatistik bilim midir? Yöntemler topluluğu mudur? Konusundaki bilgi için Hüsnü Arıcı, İstatistik Yöntemler ve Uygulamalar (Ankara: Meteksan A.Ş., 2001) adlı kitaptan yararlanılmıştır. Konuyla ilgili bilgiyi 2-3. sayfalarda bulabilirsiniz.

KİTAP

Bir bilim dalı olarak istatistik, verilerin derlenmesi, düzenlenmesi, çözümlenmesi ve yorumlanması sürecinde kullanılacak teknik ve yöntemleri geliştiren, bu teknik ve yöntemleri uygulayan, kendine özgü kuralları olan bir bilim dalıdır.

Bir bilim dalı olarak istatistiğin bir başka tanımı için www.istatistikci.com İnternet adresinden erişebileceğiniz Suat Şahinler ve Özkan Görgülü, İstatistik adlı yazıya ulaşabilirsiniz.

İNTERNET

Bir bilim dalı olan istatistik diğer bilimler gibi yaşamın içinde var olan gerçekleri yöntem bilimi anlayışı içinde ortaya çıkaran ve kendi içinde çok güçlü teoriler üretmesinin yanında, uygulama yönüyle hemen hemen bütün birimlerin özellikle stokastik bilgilere dayalı bilimlerin gelişmesine katkı sağlayan bir bilim dalıdır. Bir bilim dalı olarak istatistik bu özellikleri nedeniyle katkı sağladığı bilimlerin de önünde öneme sahip bir bilim dalı durumundadır.

İstatistiğin yöntemler topluluğu mudur, yoksa bilim midir olduğu tartışması aslında önemli değildir. Önemli olan istatistiksel çözümlemenin sonuçlarıdır, istatistiğin işlevleridir. Bir istatistiksel çözümlemenin sonuçları betimleme, ilişki araştırma, karşılaştırma, tahminleme, kestirim ve öngörü şeklinde sayılabilir. Bu konular bu ünitenin bilimsel araştırma sürecinin aşamaları başlığının çözümleme altbaşlığı altında incelenmiştir.

Örneklem Değer Anlamında İstatistik

Tanımlanan yığın olaylar (birimler) kümesindeki *bütün olaylar* hakkında, ilgilenilen özellik bakımından 6 nolu ünitede ayrıntılı olarak açıklanacak nedenlerden ötürü, veri derlenemeyebilir. Bu durumda tanımlanan yığın olaylar arasından rassal olarak sınırlı sayıda olay seçilir. Seçilen olayların oluşturduğu topluluğa örneklem adı verilir. Örneklemde yer alan birimler üzerinden derlenen veriler kullanılarak hesaplanan özet değerlere istatistik denir. Bu anlamda istatistik örneklemin bir özelliğine ait tek bir değer olup farklı örneklemlerde değişik değerler alabilir.

BAZI TEMEL KAVRAMLAR

Her bilim dalında olduğu gibi istatistiğin de kendine özgü kavram ve sembolleri vardır. İzleyen ünitelerde ele alınan konuların kolayca anlatılabilmesi ve anlaşılabilmesi için bu kavramların bazılarının öncelikle bilinmesi gerekir.

Birim

Birimin Tanımı

Belirlenen bir tanım içinde yer alan bir veya daha fazla özelliği bakımından araştırmaya konu olan yığın olay niteliğindeki olayların her birine birim, istatistik birimi denir. Tablo 1.3'teki veri kümesi incelendiğinde futbolcunun oynadığı 5 maçın her biri bir birimdir. Başka bir ifadeyle yığın olaydır. Çünkü futbolcunun maçtan maça performans göstergesi olan gol sayısı, önsezi ve koştuğu mesafe değişik değerler almaktadır.

Tablo 1.3
Bir futbolcunun oynadığı maçlardaki performans istatistikleri.

Futbolcunun Oynadığı Maç No	Gol Sayısı	Önsezi	Koştuğu Mesafe
1	1	İyi	8000
2	0	İyi	9100
3	2	Orta	9500
4	1	Kötü	6800
5	3	İyi	12000

Bir olayın istatistiksel birim olabilmesi için sayılabilir, ölçülebilir olması, ortak özelliklerinin yanında farklı yönlerinin de olması, özelliklerinin en az iki ölçme düzeyine sahip olması gerekir. Özellikleri ölçülemeyen ya da sayılamayan olaylar istatistiksel anlamda birim sayılmaz. Örneğin koku, rüya istatistiksel anlamda birim olamaz.

Birim Türleri

Birimler farklı ölçütlere göre sınıflandırılmaktadır. Önemli sınıflandırmalar aşağıda incelenmiştir.

KİTAP

Birim kavramına ilişkin sınıflandırmalarla ilgili ayrıntılı bilgi için A. Fuat Yüzer, Embiya Ağaoğlu, Hüseyin Tatlıdil, Ahmet Özmen, Emel Şıklar, İstatistik, (Eskişehir, Anadolu Üniversitesi Yayını, 2011) isimli kitaptan yararlanabilirsiniz.

Maddi Birim-Maddi Olmayan Birim

Maddi bir varlığa sahip olan, elle tutulur ve gözle görülebilir, bu özellikleriyle en, boy ve yükseklik gibi boyutları olan birimler maddi birimdir. Bir üniversitenin binaları, öğretim elemanları maddi birimdir. Maddi bir varlığa sahip olmayan evlenme, boşanma, trafik kazası vb. gibi birimler maddi olmayan birimdir.

Devamlı Birim-Ani Birim

İstendiği zaman hakkında araştırma yapılabilecek olan birimler devamlı birimdir. Bir üniversitedeki öğretim elemanları, bir bankanın şubeleri, İstanbul Sanayi Odası'na kayıtlı olan iş yerleri devamlı birim için örnek olarak verilebilir.

Özellikleri ancak meydana geldiğinde ölçülebilen ya da sayılabilen birimler ani birimdir. Evlenme, trafik kazası, grev, bir hastanede yeni doğan bebekler, doğal afetler vb. nitelikteki birimler ani birim için örnek gösterilebilir.

Gerçek-Varsayımsal Birim

Maddi varlığa sahip olan ev, birey, sigorta acentası, mali kuruluş gibi birimler gerçek birim; gerçekten var olmayan fakat oluşturulabilecek birimler varsayımsal birimdir. Örneğin 25 kişiden oluşan bir topluluktan birbirinden farklı beşer kişilik oluşturulabilecek mümkün grupların her biri varsayımsal birimdir. Çünkü gerçekte bu birimler yoktur, tarafımızdan oluşturulmuş birimlerdir.

Doğal Birim-Doğal Olmayan Birim

Parçalandığı ya da birleştirildiğinde özelliğini yitiren birimlere doğal birim denir. Doğal olmayan birim ise birleştirildiğinde ya da parçalandığında özelliğini kaybetmeyen birimlerdir. Örneğin bir otomobil parçalandığında iki otomobil, iki otomobil birleştirildiğinde ise daha büyük bir otomobil olmayacağı yani özellikleri değişmeyeceği için otomobil doğal birimdir. Genellikle zaman, mekân ve uzaklık gibi birimler doğal olmayan birimdir. Örneğin ay bir zaman birimidir. Aylar birleştirilirse zaman birimi yıl olur, niteliği değişmez. Bir arsa iki parsele bölündüğünde yine arsadır. O hâlde yıl ve arsa doğal olmayan birimlerdir.

Evren

Evren Tanımı

Araştırmacı tarafından belirlenen bir tanım kapsamına giren, hakkında araştırma yapılması planlanan birimler topluluğuna **evren** (anakütle) denir. Daha kısa bir ifadeyle evren bir araştırmada ilgilenilmek istenen ve hakkında genellemelerin yapılacağı birimler topluluğudur.

Araştırma sonuçlarının genellenmek istendiği birimler topluluğuna **evren** denir.

Alan yazında birimlere ulaşılabilirlik kriter alındığında evrenle ilgili genel evren ve araştırma evreni sınıflandırması yapılır. Genel evren, tanımlanması kolay fakat birimlerine ulaşılabilirliğin zor, bazen de imkânsız olduğu evrendir ve soyut bir kavramdır. *Araştırma evreni ise birimlerine ulaşılabilirliğin mümkün olduğu evrendir.* Bilimsel araştırmalarda genellikle araştırma evreni tanımlanmaktadır. Bu nedenle yukarıda yapılan evren tanımı araştırma evreni anlamında bir tanımdır. Bu ünite ve diğer ünitelerde evren kavramı bu anlamda kullanılmıştır.

Evrenin hangi birimleri kapsayacağını, hangilerini kapsamayacağını, araştırmanın açık ve kesin olarak tanımlanan amacı belirler. Araştırma kimler, neler üzerinde yapılacaktır sorusunun yanıtı, araştırmanın evrenini tanımlar. Her araştırma için ayrı bir evren tanımı söz konusu olabilir.

Araştırma Amacı 1

Sigorta şirketlerini çalıştırdıkları işgören sayısı bakımından araştırmak.

Evren 1

Bu araştırmanın evreni açık ve kesin olarak tanımlanmamıştır. Araştırma hangi sigorta şirketleri üzerinde yapılacaktır, araştırılacak olan, dünyada faaliyette bulunan sigorta şirketleri midir? yoksa Eskişehir'de veya Türkiye'de faaliyette bulunan şirketler midir? belirli değildir. Yani evren tanımı soyuttur.

Araştırma Amacı 2

Eskişehir'de faaliyette bulunan banka şubelerini, mevduat tutarları ve işgören sayıları bakımından araştırmak.

Evren 2

Eskişehir il sınırları içindeki bütün yerleşim yerlerinde bugün itibariyle faaliyette bulunan kamu ve özel bütün banka şubelerinin oluşturduğu topluluk araştırmanın evrenini oluşturur.

Araştırma Amacı 3

X bisküvi fabrikasında üretilen bisküvi paketlerinin planlanan ağırlıkta üretilip üretilmediğini, paketleme sisteminin doğru çalışıp çalışmadığını belirlemek.

Evren 3

X bisküvi fabrikasında geçmişte üretilmiş, üretilmekte olan ve üretilecek olan bisküvi paketleri kümesi, araştırmanın evrenidir.

Evrenin sınırlarını belirlemek, genişletmek ya da daraltmak için araştırmanın amacıyla ilişkili olarak çeşitli ölçütler kullanılabilir. Bu ölçütlerden bazıları:

- Sektör sınıflandırması (sanayi, hizmet, tarım.)
- Zaman (2001 ve sonrası, 24 Ocak Ekonomik Kararları'ndan sonraki dönem.)
- Mekân (Anadolu Üniversitesi, Ege Bölgesi)

- Yaş Grupları (0-4, 5-15, ..., 40-50)
- Öğrenim Düzeyi (İlk, Orta, Yüksek)
- Öretim Üyesi Unvanı (Yrd. Doç. Dr., Doç. Dr., Prof. Dr.)
- Gelir Grupları (Açlık, Yoksulluk, Orta Gelir, Üst Gelir)
- Sosyal Sınıf (A, B, C, D)
- Cinsiyet (Erkek-Kadın)
- Çalışma Grupları (Beyaz Yakalı, Mavi Yakalı)
- İnanç Türü (Müslüman, Hristiyan, Musevi, ...)

Evrenin sınırlarını yukarıda belirlenen ölçütlerden uygun olanını belirleyip kullanırken üzerinde durulması gereken bazı hususları göz önünde bulundurmak gerekir. Bunlardan önemli olan bazıları araştırma bütçesi, araştırma süresi ve araştırma sonunda üretilecek olan bilginin önemi ve yeterliliğidir. Örneğin Türkiye'ye gelen yabancı turistlerin geliş nedenlerini konu alan araştırmada üretilecek bilginin niteliğiyle Antalya'ya gelen yabancı turistlerin geliş nedenlerini konu alan araştırmada üretilecek bilginin niteliği aynı değildir. Antalya'ya gelen yabancı turistler genellikle güneş, deniz, kum için gelirken ülkemizin pek çok coğrafyasına kültür, tarih, spor, kayak, rafting, yamaç paraşütü vb. amaçlarla gelmektedirler. Belirtilen nedenlerle nitelikli bilgi için evrenin daha geniş mekan ile sınırlandırılması uygun olur. Örneğin 2010 yılında ilk 500 içinde yer alan, ISO 9001 belgeli imalat işletmeleri üzerinde bir araştırma yapılması planlanmış ise evren mekân, üretim türü, zaman kriterleriyle sınırlandırılmış demektir.

Evren Türleri

Evreni oluşturan birimlerin nitelikleri kriter alındığında aşağıdaki türden sınıflandırmalar yapılabilir:

Gerçek Evren-Varsayımsal Evren

Birey, kurum, evlenme olayı, ölüm olayı gibi fiilen gözlenebilen birimlerin oluşturduğu evrenler gerçek evrenlerdir. Gerçekten var olmayan ve fakat oluşturulabilecek evrenlere varsayımsal evren denir. Kadrosunda 25 futbolcu bulunan bir futbol takımının birbirinden farklı futbolculardan oluşturulabilecek 11 kişilik mümkün maç kadrolarının oluşturduğu topluluk varsayımsal evrene örnek gösterilebilir.

Sonlu-Sonsuz Evren

Evrenler kapsadıkları birim sayısına göre sonlu-sonsuz evren olarak sınıflandırılır. Birimleri sayılabilen ve sayılabilir çoklukta olan evrenler sonlu evrendir. Sonlu evrendeki birimlerin sayısına evren hacmi (büyüklüğü) denir ve N simgesiyle gösterilir. Örneğin Türkiye'de faaliyette bulunan banka şubeleri topluluğu sonlu bir evrendir. Birimleri sayılamayacak kadar çok olan veya birimleri bir sürecin çıktıları niteliğinde olan evrenler sonsuz evrendir. Bir fırında üretilen ekmeklerin, bir su dolum tesisinde dolum yapılan şişelerin, Marmara Denizi'ndeki balıkların oluşturduğu topluluk sonsuz evren için örnek verilebilir.

Hazır-Hareket Halinde Evren

Her an incelemeye hazır olan ve devamlı birim olarak tanımlanan birimlerden oluşan evrenlere hazır evren denir. İnsan, bina, şirket, üniversite gibi birimlerden oluşan evrenler bu türdendir.

Özellikleri ancak meydana geldiğinde incelenebilen birimlerden oluşan evrenler hareket hâlindeki evrenlerdir. Doğum, evlenme, trafik kazası vb. gibi ani birimler belirli bir zamanda bir arada bulunamayacağı ancak meydana geldiklerinde incelenebilecekleri için hareket hâlindeki evrenleri oluştururlar.

Sürekli-Süreksiz Evren

Doğal birimlerin oluşturduğu evrenler sürekli, doğal olmayan birimlerin oluşturduğu evrenler ise süreksiz evrenlerdir.

Evren ile ilgili yukarıdaki sınıflandırmalar bağlamında 2 nolu araştırma değerlendirildiğinde Eskişehir'de faaliyette bulunan banka şubeleri gerçekten var oldukları için gerçek evren, sayıları sayılabilir çoklukta olduğu için sonlu evren, istenildiğinde incelenmeye hazır oldukları için hazır evren ve doğal birimlerden oluştuğu için de sürekli evrendir.

Bir canlı, cansız varlıklar ve olaylar topluluğunun evren olabilmesi için bu topluluğu oluşturan birimlerin öncelikle yığın olay niteliğinde birimler olması, özelliklerinin hem genel hem de özel tesadüfi sebeplerin etkisi altında olması gerekir. Ayrıca evrenin birimlerin toplamından farklı bir yapıya da sahip olmaması gerekir.

Değişken

Değişken Tanımı

Tanımlanan araştırmanın birimlerinin ilgilendiğimiz özelliklerine **değişken** denir. Tablo 1.3'te futbolcunun oynadığı maçların her biri birim olduğuna göre bu maçlarla ilgili ilgilendiğimiz özellikler gol sayısı, önsezi, koşma mesafesi birer değişkendir. Tablo 1.3'te görüldüğü gibi bu değişkenler futbolcunun oynadığı maçtan maça değişik değerler almaktadır. O hâlde, bir başka ifadeyle değişken tanımı birimden birime değişik değerler alabilen özelliklerdir şeklinde yapılır.

Birimden birime değişik değerler alabilen özelliklere **değişken** denir.

Bir özelliğin değişken sayılabilmesi için en az iki ölçme düzeyinin olması gerekir. Bir iş yerinde çalışanların cinsiyeti, yaşı, öğrenim durumu, medeni hâli, bir bankanın yıllık mevduat tutarı, kredi miktarı, kârı, şirketlerin hukuki şekli değişken için örnek verilebilir.

Değişken Türleri

Değişkenlerle ilgili yapılan en önemli sınıflandırma Şekil 1.1'de verilmiştir.

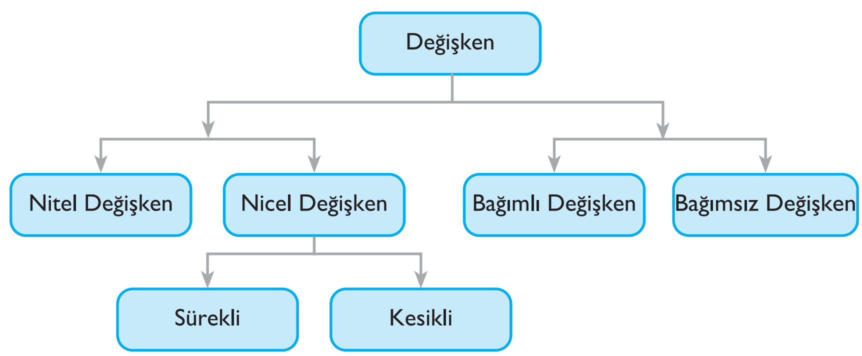
Nicel-Nitel Değişken

Ölçme veya sayma suretiyle ifade edilen değişkenlere **nicel,** sayılamayan veya ölçülemeyen bir niteliği tanımlayan değişkenlere **nitel** değişken denir.

Bir değişkenin mümkün sonuçları sayısal ise **nicel;** sayısal değil ise **nitel** değişkendir.

Şekil 1.1

Değişken sınıflandırması.



Tablo 1.3 incelendiğinde futbolcunun oynadığı maçlardaki koşma mesafesi ve gol sayısı nicel değişken, önsezi değişkeni ise nitel değişkendir. Gol sayısı değişkeninin alabileceği değerler (0, 1, 2, 3, ...) gibi sayma sayıları türünde, koştuğu mesafe ise (8000, 9100, 9500, 6800, 12000) gibi ölçüm değerleriyle ifade edilmiştir. Nicel değişkenlerin sürekli nicel değişken ve kesikli nicel değişkenler olmak üzere sınıflandırması yapılır. Bir nicel değişken ölçme süresiyle ifade ediliyorsa veya bir başka ifadeyle bir nicel değişken herhangi iki birimin aldığı değerler arasında teorik olarak pek çok ölçüm değeri alabiliyorsa bu değişken sürekli nicel değişkendir. Örneğin futbolcunun maçlardaki koşma mesafesi değişkeni sürekli nicel değişkendir. Çünkü futbolcunun birinci maçtaki koşma mesafesi olan 8000 metre ile ikinci maçtaki koşma mesafesi olan 9100 metre arasında ölçüm için en küçük uzunluk ölçüsü birimi kullanıldığını varsaydığımızda, teorik olarak çok sayıda ölçüm değeri vardır. Bir nicel değişken, futbolcunun maçlar itibariyle attığı gol sayısı örneğinde olduğu gibi sayma suretiyle ifade ediliyorsa bu değişken kesikli nicel değişkendir. Başka bir ifadeyle bir nicel değişken, herhangi iki birimde aldığı değerler arasında teorik olarak sınırlı sayıda değer alıyorsa bu değişken kesikli nicel değişkendir. Örneğin futbolcunun birinci maçtaki gol sayısı olan 1 ile beşinci maçtaki gol sayısı 3 arasında gol sayısı değişkeni sadece 2 değerini alabilir 1,75, 2,25 vb. ölçüm değerlerini alamaz.

DİKKAT

Bütün nitel değişkenler kesikli değişkendir.

SIRA SİZDE 3

- **Bir bankanın şubelerinin işgören sayısı, müşteri sayısı ne tür değişkendir? Tartışınız.**
- **Bir bankanın mevduat tutarı, kredi tutarı ne tür değişkendir? Tartışınız.**
- **Eskişehir Organize Sanayi Bölgesindeki şirketlerin hukuki şekli ne tür değişkendir? Tartışınız.**

Bağımlı-Bağımsız Değişken

Bilimsel araştırmalara konu olan her olayın bir nedeni vardır. Nedeni olmayan hiçbir olay yoktur. Bu nedenle araştırmalarda değişkenler arasında teorik olarak var olduğu düşünülen sebep-sonuç (nedensel) ilişkisinin olup olmadığının araştırılmasıyla ilgilenilir.

Değişkenler arasında teorik olarak var olan sebep-sonuç ilişkisinin yapısının konu alındığı araştırmalarda araştırmanın amacını tanımlayan değişken bağımlı değişken (sonuç değişkeni), bağımlı değişkeni etkileyen, bağımlı değişkendeki değer değişmelerine neden olan değişken/değişkenler ise bağımsız değişkendir. Bağımlı değişken araştırmacının denetiminde olmayan, bağımsız değişken ise araştırmacının denetiminde olan değişkendir. Örneğin bir işletmenin reklam harcamalarıyla satış tutarı arasında ilişki olup olmadığını konu alan bir araştırmada reklam harcamaları bağımsız değişken, satış tutarı ise bağımlı değişkendir. Çünkü işletme yönetimi, reklam harcamalarini istediği kadar artırıp azaltabilir.

Bağımlı değişkenin Y, bağımsız değişkenin X simgesiyle gösterildiğini varsayalım. X ile Y arasındaki nedensel ilişki X, Y'nin nedenidir şeklinde bir önermedir. Bu önermeye göre X bağımsız değişken, Y ise bağımlı değişkendir.

SIRA SİZDE 4

- **Öğrencilerin başarısını etkileyen bildiğiniz bağımsız değişkenleri sayınız.**
- **Bir malın fiyatı ile satış miktarı arasında ilişki olup olmadığını konu alan bir araştırmanın bağımlı değişkeni nedir?**

İlişki araştırmalarında değişkenler nicel veya nitel olabilir. Örneğin öğrencilerin başarısıyla barınma yeri arasında ilişki olup olmadığını konu alan araştırmada bağımsız değişken barınma yeri türüdür ve nitel değişkendir. Bu değişken evde, yurtta, pansiyonda şeklinde değerlere (ölçme düzeylerine) sahiptir. Başarı değişkeni ise bağımlı değişkendir ve istenirse nitel, istenirse de nicel olarak ifade edilebilir. Başarı değişkeni nitel olarak ifade edilmek istenirse (geçmez, orta, iyi, pekiyi), nicel olarak ifade edilmek istenirse 0-100 aralığında değerler alabilir.

Tam Sayım

Tanımlanan evrendeki bütün birimler üzerinden araştırmaya konu olan değişkenler itibariyle veri derleniyorsa yapılan işleme tam sayım denir. Tam sayım ancak sonlu evrenler için uygulanır. Genel nüfus sayımı, genel seçim tam sayım için önemli örneklerdir.

DİKKAT

Sonsuz evrenlere tam sayım uygulanamaz.

Örnekleme-Örneklem

Tanımlanan evrenin özelliklerini yansıtabilecek, bu evrenden belirli yöntemlerle *sınırlı sayıda birimin seçilmesi işlemine örnekleme,* örnekleme uygulaması sonucu seçilen sınırlı sayıda birimin oluşturduğu gruba ise *örneklem* denir.

Örnekleme konusu 6 nolu ünitede ayrıntılı olarak işlenecektir.

Parametre-İstatistik

Parametre evrenin özelliklerini tanımlayan değerlerin genel adıdır.

Tam sayım sonucu elde edilen veriler kullanılarak hesaplanan değerlere genel olarak **parametre,** örneklemdeki birimlerden derlenen veriler kullanılarak hesaplanan karakteristik değerlere ise genel olarak istatistik adı verilir.

Parametrenin genel gösterimi θ (theta) simgesiyle örneklem istatistiğinin genel gösterimi ise $\hat{\theta}$ simgesiyle yapılır. Evrenin özelliklerini tanımlayan önemli parametreler evren ortalaması (μ) ve evren standart sapması (σ) dır. Bu parametrelerin tahminleyicisi olan önemli istatistikler örneklem aritmetik ortalaması $\bar{X}$ ve örneklem standart sapması (s) dır. Yani, $\bar{X}$, μ'nün s ise σ'nın tahmini değeridir. Tahminleme konusunun işlendiği ünitede bu konuda ayrıntılı bilgi bulabileceksiniz.

BİLİM, BİLİMSEL ARAŞTIRMA VE İSTATİSTİK EĞİTİMİ

Bilim Nedir?

Bilim sistematik bilgiler kümesidir.

Bilim; araştırma sonucu kanıtlanmış, geçerliliği kabul edilmiş, nesnel, mantıklı, genellenebilir, düzenli ya da sistematik bilgiler bütünüdür. Bu tanım Milli Eğitim Ar-Ge Birimi Araştırma Yöntem ve Teknikleri isimli kitaptan alınmıştır.

KİTAP

Konusu sürekli değişim hâlinde olan bilimin sözlüklerde, ansiklopedilerde ve konuyla ilgili öteki kaynaklarda değişik tanımları yapılmıştır. Çeşitli bilim tanımlarını, Adnan Erkuş'un "Bilimsel Araştırma Sarmalı" (Ankara, Seçkin Yayıncılık San. Tic. A.Ş., 2005) isimli kitabında bulabilirsiniz.

Genel olarak sistematik bilgilerdir şeklinde tanımlanan bilimin tanımı Türk Dil Kurumu Sözlüğü'nde aşağıdaki gibi yapılmıştır:

Bilim;

- "Evrenin ya da olayların bir bölümünü konu olarak seçen deneysel yöntemlere ve gerçekliğe dayanarak yasalar çıkarmaya çalışan düzenli bilgi."
- "Genel geçerlilik ve kesinlik nitelikleri gösteren, yöntemli ve dizgisel bilgi."
- "Belli bir konuyu bilme isteğinden yola çıkarak belli bir ereğe yönelen bir bilgi edinme ve yöntemli araştırma süreci" dir.

Yukarıda yapılan bu bilim tanımları, bilimle uğraşanlar için yeterli olmayabilir. Dinamik bir süreç olan bilim ile ilgili yapılan tanımlara yeni tanımlar eklemek yerine bir bilim tanımında yer alması gereken temel nitelikleri açıklamaya çalışmak daha doğru olacaktır. Bilimin temel nitelikleri;

- Bilim olgusaldır. Yani bilim herkes tarafından gözlemlenebilir ve ölçülebilir gerçekleri konu alır. Gözlemlenemeyen ve ölçülemeyen hiçbir olgu (gerçek) bilimin konusu olamaz.
- Bilim nesneldir. Gözlemlenebilen, ölçülebilen ve ortak bir bilimsel dil olan matematik ve istatistik ile ilgilenen bilgi aynı koşullar altında, aynı yöntem ve araçlarla isteyen herkes tarafından yapılabilecek araştırmalarla geçerliliği ve doğruluğu kanıtlanmış bilgi niteliğindedir. Bu özellik bilgiyi özel olmaktan çıkarıp evrensel geçerliliği olan nesnel bilgi durumuna getirir. Bilimsel bilgi, bireyin kişisel görüşünden bağımsızdır.
- Bilim mantıksaldır. Araştırma sonuçlarının kendi içinde tutarlı olması gerekir.
- Bilim sürekli değişim içindedir. Bu değişim bilimin en önemli özelliğidir. Bilim hiçbir zaman tam doğrulara ulaştığını iddia etmez. Tekrarlanan araştırmalar sonunda mevcut bilgilerin doğruluğu veya yanlışlığı test edilir. Test sonucunda yanlış olduğu anlaşılan bilgiler ayıklanır, yerine yenileri konulur. Doğruluğu kanıtlanmış olanlar ise kesinliği olan "bilimsel yasa" niteliğini kazanır.

Bilimsel Araştırma

Bilimsel araştırma yeni bilgilere ulaşmak için gerçekleştirilen sistemli çabadır.

İnsanı doğrudan ya da dolaylı olarak fiziksel ve/veya düşünsel yönden rahatsız eden kararsızlık ve birden çok çözüm yolu olan problemlere araştırma problemi denir.

Bilimsel araştırma: Araştırma problemlerine güvenilir çözümler aramak amacıyla

- Verilerin sistemli olarak derlenmesi,
- Derlenen verilerin çözümlenmesi,
- Çözümleme sonunda elde edilen bilgilerin değerlendirilmesi ve yorumlanması,
- Bilgilerin ve yapılan değerlendirmelerin rapor edilmesi sürecidir.

Bilgi, Özbilgi

Günlük yaşamda aynı anlamda kullandığımız birbirleriyle bağlantılı fakat anlamları farklı olan bilgi (information) ve özbilgi (knowledge) kavramları vardır. Konuyla ilgili bazı çalışmalarda İngilizce ifadesi information olan kavram; haber, enformasyon olarak kullanılırken knowledge kavramı bilgi olarak Türkçeleştirilmektedir.

Bilgi Nedir?

Bilgi, verilerin işlenmiş hâlidir. Örneğin bir sınıfta bulunan öğrencilerin ortalama boy uzunluğu bilinmek istenirse öğrencilerin her birinin boy uzunluğu ölçüm değeri veri, öğrencilerin boy uzunluklarının ortalama değeri bilgidir. Örnekten görüldüğü gibi bilgiler verilere göre hacim olarak daha küçük fakat anlam ve kullanım değeri olarak daha büyüktür.

Özbilgi
Türkçe'de karşılığını bulamadığımız knowledge kelimesini özbilgi olarak Türkçeleştirebiliriz. Özbilgi, bilginin kişiye özgü anlamını ifade eder ve bilginin hacim olarak küçültülmüş ancak kullanım değeri çok artmış olan hâlidir.

Özbilgi ile ilgili ayrıntılı bilgi için Necmi Gürsakal'ın "Bilgisayar Uygulamalı İstatistik 1" (Bursa, Marmara Kitabevi Yayınları, 1997) isimli kitabından yararlanabilirsiniz.

K İ T A P

Örneğin yukarıda sözü edilen öğrencilerin ortalama boy uzunluğunun 170 cm olarak hesaplandığını varsayalım. 170 cm olan boy uzunluğu ortalamasını yani bilgiyi, bazıları öğrencilerin boylarının kısa bazıları ise uzun olduğu yönünde yorumlayabilir. Özetle *veri işlenirse bilgiye, bilgi yorumlanırsa özbilgiye dönüşmüş olur.*

Bilimsel Araştırma Sürecinin Aşamaları

Günlük yaşamda pek çok bilim dalında araştırmaya değer problemlerle karşılaşılır. Bu problemlere güvenilir çözümler aramak amacıyla bilimsel (istatistiksel) araştırmalar yapılmaktadır. Bu araştırmalar sistematik bir süreç izler. Bu süreç araştırmanın düşünme aşamasından sonuç aşamasına kadar yapılması gereken çalışmaları kapsar. Bu sürecin aşamaları aşağıdaki gibi sıralanabilir:

- Araştırma konusunun belirlenmesi; araştırmacının hangi konuyu araştıracağını açıklayan ifadedir; araştırma konusunun belirlenmesi ve netleştirilmesi araştırma sürecinin ilk ve en önemli aşamasıdır. Çünkü araştırma sürecinin izleyen aşamalarındaki başarı, konunun doğru ve net olarak tanımlanmasına bağlıdır. Bu bir anlamda gömleğin ilk düğmesinin yanlış iliklenmesine benzetilebilir. İlk düğme yanlış iliklenmiş ise diğerleri sıralı bile olsa gömlek yanlış iliklenmiş demektir.

Araştırma konusunun belirlenmesi hakkında bilgi için yararlı kaynak olarak Sosyal Bilimler ve Araştırma Yöntemleri SPSS Uygulamalı (R. Altunışık, R. Coşkun, E. Yıldırım, S. Bayraktar, Sakarya Kitabevi, 2001) kitaba bakabilirsiniz.

K İ T A P

Araştırma konusunun belirlenmesinin önemini A. Einstein "konunun belirlenmesi çözümünden daha önemlidir" ve J. W. Tukeye "Doğru soruya verilen yanlış cevap, yanlış soruya verilen kesin cevaptan daha değerlidir." sözleriyle belirtmişlerdir. Araştırma konusunun belirlenmesi aşamasında önce araştırma kimler, neler üzerinde yapılacaktır, araştırmanın temel ve alt amaçları nelerdir ve araştırmanın önemi nedir gibi sorular yanıtlanır. Araştırmacının bu sorulara vereceği doğru yanıtlarla araştırma konusu tam olarak belirlenmiş olur.

Araştırmanın kimler, neler üzerinde yapılacağının yanıtı araştırmanın evrenini belirler. Araştırmanın amaçlarının belirlenmesi ise araştırmanın değişkenlerinin neler olduğu, ne tür değişkenler oldukları ve bu değişkenlerin hangi ölçek türüyle ölçülmesinin uygun olacağının belirlenmesiyle mümkündür.

Daha sonra kaynak incelemesi yaparak konunun o ana kadar hangi yönleriyle, nasıl incelendiği, kaynakların konuya yakınlık ve önem dereceleri belirlenir. Bu aşamada son olarak araştırmanın, maliyeti, süresi ve gerekli olan diğer olanaklara sahip olup olmadığımızın değerlendirilmesi yapılır.

ÖRNEK

- Tanımlanan araştırmanın konusu:

A sınıfında öğrenim gören 100 öğrencinin sınıf başarısı ve başarıyı etkileyen faktörleri belirlemek olsun.

Araştırmada üretilmek istenen bilgiler (araştırmanın amaçları);

  - Öğrencilerin cinsiyet dağılımı nedir?
  - Öğrencilerin başarı dağılımı ve sınıf başarısı nedir?
  - Öğrencilerin derse devam sıklığı dağılımı nedir?
  - Öğrencilerin çalışma süresi dağılımı nedir?
  - Kız öğrencilerle erkek öğrencilerin başarıları arasında fark var mıdır?
  - Öğrencilerin çalışma süresi ve derse devam sıklığı arasında ilişki var mıdır?

- Araştırmanın Evreni:

A sınıfında öğrenim gören 100 öğrencinin oluşturduğu topluluk araştırmanın evrenidir. Sonlu evrendir ve hacmi N=100 birim (öğrenci) dir.

- Araştırmanın Değişkenleri:
  - Derse devam sıklığı,
  - Çalışma süresi,
  - Cinsiyet,
  - Başarı Puanı.
- Verilerin Derlenmesi:

Bu aşamada, önce tam sayım mı örnekleme mi yapılacağına karar verilir. Tam sayım uygulanacak ise evrendeki bütün birimler, örnekleme uygulanacak ise örneklemdeki birimler üzerinden araştırmanın amacını/amaçlarını tanımlayan değişken/değişkenlerle ilgili verilerin derlenmesinde kullanılacak uygun yöntem belirlenmeye çalışılır. Değişkenler hakkında doğru ve güvenilir veri derlenmemiş ise çözümleme amacıyla kullanılacak teknik ve yöntemler ne kadar güçlü olursa olsun güvenilir ve geçerli bilgilere ulaşmak mümkün olmayabilir. Veri derleme yöntemleri bu kitabın 2. ünitesinde ayrıntılı olarak incelenmiştir.

Veri derleme aşamasında;

- Değişkenlerin ölçülmesinde kullanılacak uygun ölçek türünün seçilmesi,
- Uygun veri derleme yöntem ve araçlarının seçilmesi,
- Verilerin bilgisayar ortamına girilmesi (kodlanması),
- Verilerde hata, aykırı değer ve eksik veri olup olmadığının denetlenmesi,
- Tam sayım mı yoksa örnekleme mi yapılacağı, örnekleme yapılacağına karar verilmiş ise uygulanacak örnekleme planının hazırlanması

işlemleri değerlendirilir.

DİKKAT

Neyi, nasıl ölçeceğinizi öğrenmeden verilerin çözümlenmesi aşamasına geçmeyiniz.

- Çözümleme:

Bu aşamada derlenen verilerden araştırma konusu olan evren hakkında bilgiler üretilir. Bu amaçla önce derlenen ve bilgisayar ortamına girilen veriler düzenlenir, tablolar ve şekiller hâlinde sunulur. Sonra daha üst nitelikli bilgilerin üretilmesi için uygun çözümleme yöntemi ve çözümlemelerin yapılabilmesi için geliştirilen hazır yazılımlar seçilir ve çözümleme yapılır.

Derlenen veriler için uygun çözümleme teknik ve yönteminin seçimi yapılırken aşağıdaki değerlendirmeler yapılır:

- Araştırmanın bağımlı ve bağımsız değişken sayısı nedir?
- Araştırmanın değişkenlerinin ölçümünde hangi ölçek türü kullanılmıştır?

- Araştırmada üretilecek bilginin türlerine bağlı olarak bu bilgileri üretebilecek ne tür geliştirilmiş teknik ve yöntemler vardır?

Bilimsel araştırmalarda üretilecek bilgiler, bir başka deyişle istatistiksel çözümlemenin sonuçları, *betimleme, ilişki araştırma, karşılaştırma, tahminleme, kestirim ve öngörü şeklinde sayılabilir.*

İstatistiksel çözümlemenin sonuçları ile ilgili bilgi için "Bilimsel Araştırma Yöntemi ve İstatistiksel Anlamlılık Sınamaları" (N. Çömlekçi, İstanbul, Bilim Teknik Yayınevi, 2001) isimli kitaptan yararlanılabilir.

KİTAP

- **Betimleme:**

Derlenen verilerin düzenlenmesi, tablo, grafik ve şekillerle gösterilmesi ve verilerin doğasında var olan özelliklerin profilinin çıkarılmasına betimleme, bu amaçla kullanılan teknik ve yöntemler ise betimsel istatistik olarak isimlendirilir. Örneğin üniversite öğrencilerinin istatistik dersinden beklentileri nedir, öğrencilerin Türkiye'nin ekonomik sorunlarına ilişkin görüşleri nedir, öğrencilerin kitap okuma alışkanlığı nedir sorularının yanıtlarını aramak betimleme yapmaktır, halihazırdaki durumun araştırılması demektir. Betimsel istatistikler verilerin sahip olduğu dağılımın merkezinde yer alan tek bir değerin hesaplanmasında kullanılan ölçülerdir veya tekniklerdir. Betimsel bilgi üretmek amacıyla kullanılan teknikler 2 ve 3 Nolu ünitelerde incelenmiştir.

Yukarıdaki örnek araştırmada tanımlanan değişkenlerle ilgili oluşturulan aşağıdaki soruların 100 öğrenciye sorulduğunu ve bu sorularla ilgili verilerin (yanıtların) derlendiğini varsayalım.

- Cinsiyetiniz nedir?
- Yılsonu başarı puanınız nedir?
- Bir dönem içinde derse devamsızlığınız nedir?
- Günlük ortalama ders çalışma süreniz nedir?

Bu verilerden üretilebilecek örnek betimsel bilgiler 100 öğrencinin 40 (%40) ı kız, kızların ortalama başarı puanı 75 puan, erkeklerin ortalama başarı puanı 65 puandır. Erkeklerin başarı puanları 10-85 puan aralığında değişmektedir. "Sınıfın ortalama devamsızlığı 10 saat olup erkeklerin 90 (%90)'ının hiç devamsızlığı bulunmamaktadır." "Kızların günlük serbest saat ortalama çalışma süresi 3 saattir." şeklinde olabilir. Çözümlemelerde görüldüğü gibi çözümleme amacıyla oran, aritmetik ortalama, sıklık vb. gibi betimsel teknikler kullanılmıştır.

Hemen hemen her araştırmada önce tanımlanan değişkenlerle ilgili betimsel bilgilerin üretilmesi gelenek olmuştur. Bu türden bilgiler istatistik kuramının gelişmesinde önemli kaynak durumundadır.

- **İlişki Araştırma:**

Değişkenler arasında kuramsal olarak var olan sebep-sonuç ilişkisinin yapısının ve ilişkinin derecesinin belirlenmesine ilişkin araştırmalara **ilişki araştırması** denir.

İlişki araştırması iki ya da daha fazla değişken arasında ilişki olup olmadığını incelemektir.

Bu bir anlamda bağımsız değişkenlerle bağımlı değişkenler arasında kuramsal olarak var olduğu düşünülen ilişkinin olup olmadığının araştırılmasıdır. Yukarıda ele alınan örnek araştırma probleminde, 100 öğrenciye dönemlik toplam ders devamsızlığı sorulduğunu ve devamsızlık ile başarı arasında ilişki araştırıldığını varsayalım. Derse devamsızlık ile başarı arasında ters yönde ilişki olduğu sonucu çıkabilir. Yani devamsızlık artarken öğrencinin başarısı azalabilir. Benzer ilişki başarıyla çalışma süresi arasında da incelenebilir. Çalışma süresi artarken başarının da arttığını yani başarıyla çalışma süresi arasında aynı yönlü ilişki olduğu bilgisine ulaşılabilir. Örneğin seyahat sıklığı ile okunan kitap sıklığı arasındaki ilişkinin araştırılması bir ilişki araştırmasıdır.

- **Karşılaştırma:**

Karşılaştırma araştırması, farklı grupları belirli değişkenler bakımından kıyaslamaktır.

İki veya daha fazla evren parametre değeri arasında anlamlı bir farkın olup olmadığının, fark varsa bu farkın tesadüfi bir fark olup olmadığı örneklem istatistiklerinden yararlanılarak araştırılıyorsa bu araştırma **karşılaştırma** bilgisi üreten bir araştırmadır.

Örnek alınan araştırmada kızların ortalama başarı puanı ile erkeklerin ortalama başarı puanları arasında fark olup olmadığı bilgisi bu bağlamda bir araştırma ile elde edilebilir.

- **Tahminleme, Kestirim ve Öngörü:**

Örneklem istatistiklerinden yararlanılarak bilinmeyen evren parametre değerlerinin belirlenmesi çalışmalarına tahminleme denir. Örneğin bir sınıfta kayıtlı olan öğrencilerin okuduğu yıllık ortalama kitap sayısını belirlemek amacıyla rassal olarak belirlenmiş 30 öğrencilik örneklem seçilsin ve bunların okuduğu ortalama kitap sayısı hesaplansın. Hesaplanan bu ortalama bilgisi kullanılarak *sınıfın tamamının* bilinmeyen ortalama kitap bilgisi belirlenmeye çalışılıyorsa yapılan araştırma tahminlemedir.

Kontrol altındaki bir *bağımsız değişkenin belirlenen bir değeri için* bağımlı değişkenin alabileceği değer tahminlenen bir ilişki modeli kullanılarak belirleniyorsa yapılan araştırma kestirim araştırmasıdır. Örneğin öğrencilerin devamsızlığı 10 saat olursa başarısı ne olur ya da çalışma süresi günlük ortalama 3 saat olunca başarısı ne olur türünden araştırmalar kestirim araştırmalarıdır.

Zamana bağlı bir bağımlı değişkenin *belirli bir dönemde alabileceği değerlerin belirlenmesi* çalışmasına öngörü araştırması denir. Örneğin bir firmanın dönem kârı, ülkenin cari açığı ve enflasyon oranı gelecek ay ne olacaktır sorusunun yanıtlandığı türden araştırmalar öngörü araştırmalarıdır.

- Araştırma Raporunun Yazılması

Araştırma raporunun yazılması bilimsel araştırma sürecinin son aşamasıdır. Araştırmacı elde ettiği bilgileri bu bilgilere gereksinimi olanlarla paylaşması gerekir. Araştırma raporunun yazılması bu amaç için bir araçtır. Araştırma raporu yazılırken bilimsel yazım kurallarına uyulmalıdır.

İstatistik Eğitiminin Önemi

İstatistik eğitiminin iki yönü vardır: Teorik (Matematiksel) İstatistik Eğitimi, Uygulamalı İstatistik Eğitimi. İstatistik eğitiminin önemi ile ilgili açıklamaların bu ikili ayrım esas alınarak yapılması uygun olur.

Teorik İstatistik Eğitiminin Önemi

Teorik istatistik eğitimi istatistiğin matematik ve olasılık kuramındaki gelişmelere bağlı olarak kendine özgü kuram, teknik ve yöntemlerin geliştirilmesini konu alan eğitimdir. Bu anlamda istatistik kendi alanında teorik bir bilim dalıdır.

Bilindiği gibi bilimin en önemli özelliği devamlı bir değişim içinde olduğu gerçeğidir. Bilim hiçbir zaman kesin doğrulara ulaştığını iddia etmez fakat kesin doğrulara ulaşmak uğraşısı içindedir. Teorik istatistikçinin görevi olan bu çaba daima diğer bilim dallarında araştırma yapan ve yapacak olan bütün araştırmacılar için daha doğru, daha güvenilir bilgi üretilmesine imkân verecek teknik ve yöntem arayışı yönündedir. Bu arayışta, geliştirilen yeni teknik ve yöntemlerle elde edilen araştırma bilgileri öncekilerle karşılaştırılır.

Uygulamalı İstatistik Eğitiminin Önemi

Uygulamalı istatistik eğitimi teorik istatistikçiler tarafından geliştirilmiş olan istatistiksel teknik ve yöntemlerin pek çok bilim dalıyla ilgili gerçek yaşamda karşılaşılan araştırma problemlerine nasıl uygulanacağını konu alan bir eğitimdir. Bu eğitim, istatistiğin üreticileri (araştırmacılar) ve tüketicileri (kullananlar) için önemlidir. İstatistiğin önemi nedir denildiğinde alan yazında genellikle istatistiğin bir araç olarak önemi üzerinde durulmaktadır.

İstatistik eğitimi;

- Bilgi, çağı yakalamak ve bilgi toplumu olma yarışında öne geçmek için gereklidir. Hebrert G. Wells'in "İstatistiksel düşünce, gün gelecek tıpkı okur yazar olmak gibi iyi yurttaş olmanın en gerekli ögelerinden olacaktır." şeklindeki ifadesi istatistik eğitiminin önemini özetleyen bir ifadedir. Sözü edilen gün gelmiştir; o gün bilgi çağıdır.

Bilgisayar ve iletişim teknolojilerindeki hızlı gelişmelerin ve bunların yaygınlaşması sonucu oluşan bilgi toplumu; ekonomik, sosyal, kültürel, sayısal, yönetsel vb. gibi pek çok değişkenle ilgili derlenen, bilgi ve iletişim teknolojileri ile erişilen verileri işleyen, bilgi üreten, bilgi ağlarına bağlanan, hazır bilgilere erişen, bu bilgileri paylaşan toplum, bilgi toplumu olarak tanımlanır. Sürekli olarak üretilen bilginin yerel ve/veya küresel bilgi ağlarında iletilebilir hale gelmesi, bu üretimden kazanç elde etmeyi önemli hâle getirmiştir. Bu nedenle bilgi; iş gücü, sermaye, toprak gibi geleneksel üretim faktörlerinin önünde öneme sahip olmuştur.

Yücel, İ. H. (Kasım 2011). Bilim Teknoloji Politikaları ve 21. Yüzyılın Toplumu, Ankara, Devlet Planlama Teşkilatı, http://ekutup.dpt.gov.tr/bilim

İNTERNET

Bilginin önemini algılayan ülkeler, kurumlar ve kuruluşlar gereksinim duydukları konularla ilgili bilgileri üretmek ve yeni teknolojiler geliştirmek amacıyla kendi olanaklarına göre araştırma-geliştirme (Ar-Ge) faaliyetlerine başlamışlardır. Yeni bilgilerin, ürünlerin üretildiği ve teknolojilerin geliştirildiği Ar-Ge sürecinde gerekli olan verilerin nasıl derleneceğinin, derlenen ve üretilmiş bilgilere nasıl ulaşılacağının, bunların nasıl paylaşılacağının ve kullanılacağının bilinmesi gerekir. Bu da araştırma sürecinin hemen hemen her aşamasında görev alan araştırma ekibinin istatistik konusunda bilgili olmasını, eğitim almasını zorunlu kılar. Bir toplumun bilgi toplumu olup olmadığının belirlenmesinde kullanılan kriterlerden biri Ar-Ge harcamalarının GSMH içindeki oranıdır. Bilgi ve teknoloji üretmeyen, bilgi toplumu olmamış ülkelerde bu oran %1'in altındadır. Türkiye 2010 yılı verilerine göre bu oran %0.84 olarak bulunmuştur.

TÜİK, 2010 Yılı Bilimsel ve Teknolojik Araştırma Geliştirme Faaliyetleri Anketlerinin Sonuçları, Ankara, 2010, http://www.tubitak.gov.tr/sid/468/pid/0/cid/25430/index.htm;jsessionid=78F91D43475E83243B8BD3B15EAB99D0

İNTERNET

- Araştırma problemi tanımlayabilmek ve araştırma yapabilmek için, tanımlanacak araştırma problemiyle ilgili alanda uzmanlık, araştırma yöntem bilimi ve yeterli düzeyde istatistik eğitimi alınmalıdır. Araştırmanın başlayabilmesi için merak ve bilme isteği, olmazsa olmaz koşuldur.

- Toplumun bilgiye güven duymasını sağlamak için gereklidir.

Pek çok ulusal ve uluslararası kurum, kuruluş, sivil toplum örgütü, derledikleri verileri ve ürettikleri bilgileri zaman zaman toplumla paylaşmaktadırlar. Toplumun bu veri ve bilgilerin doğruluğunu sorgulayabilmesi ve varsa bilgi kirliliğine karşı korunabilmesi için istatistik eğitimin yaygınlaştırılması gerekir.

- İstatistiksel çözümleme sonuçlarını doğru okuyabilmek için gereklidir.

İstatistiksel çözümleme doğru uygulansa bile, elde edilen istatistiksel bilgileri doğru anlama ve yorumlama çok zor olabilir. Bu zorluk, alan uzmanlığı ve yeterli seviyede istatistiksel okur-yazarlığı kazanarak aşılabilir. Ancak bazen istatistiklere işimize geldiği yönden bakma alışkanlığı ile gerçekler saptırılabilmektedir. Böylece istatistik yalan söyleme aracı olarak da kullanılabilmektedir.

İstatistik eğitimi topluma doğru yolu göstermek için rehberdir. Klasik karar alma alışkanlıklarını bırakmak, etkin karar alabilmek için gereklidir.

- İstatistik eğitiminin önemli faydalarından biri de belirsizlik ortamında en doğru karar alma yöntemini olmasıdır. Ekonomi ve iş dünyasında uygulanacak her faaliyetin bir karara dayandığı ve her gün çok sayıda kararın verildiği düşünülürse, karar verme sürecinin etkin yönetilmesinin önemi ortaya çıkar. *Veri ve bilgi temelli karar verme, kararların etkinliklerinin artırılması için önemlidir.*

Küreselleşmenin etkisiyle artan rekabet koşulları, girdi maliyetlerindeki artışlar, tüketici beklentilerindeki değişmeler, yerel ve küresel krizler, kârların azalması, veri ve bilgi temelli karar vermeyi zorunlu kılmıştır. Bu durum iş dünyasının örgüt yapılarının hemen hemen bütün kademelerinde çalışanları istatistik eğitimi almaya yöneltilmiştir. Toplam kalite yönetimi uygulamasına geçen iş yerlerinde işgörenlere yeterli ve gerekli içerikte istatistik eğitiminin veriliyor olması bunun önemli bir göstergesidir.

- İstatistik eğitimi çevremizde olup bitenleri anlama ve bunları başkalarına anlatmada ve iletişim kurup etkileşmede ortak bir dil kazandırmaktadır. Bu dile sahip olmayanlar arasındaki bilgi alışverişinin etkili olması beklenemez. Bu nedenle basit, gerekli içerikte bir istatistik eğitimi bütün çalışanlar için verilmelidir.
- İstatistik eğitimi kişisel gelişmemiz için gereklidir. Bu eğitim saygınlık kazandırır ve entelektüel haz verir.
- İstatistiksel çözümİstatistik eğitimi daha iyi, sistemli ve inandırıcı düşünebilmek ve daha nesnel, kesin ve kararlı konuşabilmek için alınmasında fayda olan bir eğitim olarak düşünülebilir.

İstatistiğin önemi ile ilgili açıklamaları Sir Francis Galton'un "Üzerinde yazılacak büyük bir konu var: İstatistik ancak anlatımımın titizlikten ödün vermeden, hiçbir yönünü eksik bırakmadan kolayca anlaşılacak biçimde sunmakta yetersiz kalacağını hissediyorum." sözüyle tamamlamak istiyoruz.

Özet

AMAÇ 1 *İstatistiğin konusu olan ve olmayan olayları tanımlamak.*

İstatistiğin konusu yığın olaylardır. Bu olaylar aynı tanım içinde yer alan diğer olayları tam olarak temsil edemez. Çünkü araştırmaya konu olan özellikler olaydan olaya farklı ölçüm değerleri alır. Aynı koşullar altında meydana gelen ve ilgilenilen özellikleri bakımından aynı sonuçları alan olaylar tipik olay olarak tanımlanır. Tipik olaylar istatistiğin konusu olmazlar.

AMAÇ 2 *İstatistik kavramının çeşitli tanımlarını yapmak.*

İstatistik kelimesinin yaygın olarak kullanılan 3 anlamı vardır: veri kümesi, yöntemler topluluğu ve örneklem değeri.

Belirli bir konuda belirli bir amaç için **derlenen verilerin oluşturduğu kümeye** istatistik denir. İstatistik kelimesi ikinci olarak **yöntemler topluluğu** anlamında kullanılır. Bu anlamda istatistik belirli bir konuda belirli bir amaç için yapılacak araştırma için gerekli olan verilerin derlenmesi, düzenlenmesi, çözümlenmesi ve çözümleme sonucu elde edilen bilgilerin yorumlanarak rapor edilmesi amacıyla kullanılan teknik ve yöntemler topluluğudur. Dersimizin adı bu anlamda kullanılmaktadır. İstatistik kelimesi son olarak, **örneklemden elde edilen verilerin kullanılmasıyla hesaplanan sayısal değerler** olarak kullanılmaktadır.

AMAÇ 3 *İstatistiğin temel kavramlarını açıklamak.*

Her bilim dalında olduğu gibi istatistiğin öğretilmesinde de önce temel kavramlar gelir. Bu kavramlar birim, evren, değişken, tam sayım, örnekleme, örneklem ve parametre gibi sayılabilir. İstatistiğin bu temel kavramlarının tanımları bu ünitenin 3. kısmında yapılmıştır. Bu tanımların ve tanımlar arasındaki bağlantının iyi anlaşılması diğer ünitelerin kolayca anlaşılmasını sağlayacaktır.

AMAÇ 4 *İstatistiksel bir araştırma sürecini tasarlamak.*

Bilgisayar ve iletişim teknolojilerindeki gelişmeler, bilgi toplumuna geçişi hızlandırmıştır. Bilgi toplumunun en önemli göstergesi araştırma-geliştirmeye GSMH'dan ayrılan paydır. Araştırma yapmak, bilgi üretmek, teknoloji geliştirmek demektir. Bir bilimsel araştırma konusunun belirlenmesi, verilerin derlenmesi, çözümlenmesi ve üretilen bilgilerin değerlendirilerek rapor edilmesi aşamalarından oluşur.

Kendimizi Sınayalım

1. Aşağıdakilerden hangisi istatistik birimi olarak **alınamaz?**
- a. Coğrafi Bölge
- b. Doğum
- c. Koku
- d. Boykot
- e. Aile

2. Aşağıdakilerden hangisi ani birimdir?
- a. Okul
- b. Bina
- c. Deprem
- d. Sanayi Kuruluşu
- e. Otomobil

3. Aşağıdakilerden hangisi sürekli bir değişkendir?
- a. Medeni hâl
- b. Cinsiyet
- c. Bir okuldaki öğrenci sayısı
- d. Üniversitedeki bina sayısı
- e. Hanehalkı geliri

4. Aşağıdakilerden hangisi sürekli değişken değildir?
- a. Boy uzunluğu
- b. Ağırlık
- c. Hanehalkı geliri
- d. Çocuk sayısı
- e. Tüketilen su miktarı

5. İstatistik biriminin sahip olduğu özelliklere ne ad verilir?
- a. Değişken
- b. Veri
- c. Parametre
- d. İstatistik
- e. Örneklem

6. Birimlerle ilgili aşağıdaki ifadelerden hangisi **yanlıştır?**
- a. Maddesel bir varlığa sahip birimler sürekli birimlerdir.
- b. Canlı ve cansız varlıklar birer istatistik birimidir.
- c. Olay ya da fiil biçiminde ortaya çıkan birimler ani birimdir.
- d. Nitelikleri açısından bir bütün olma özelliğini gösteren birimlere doğal birim adı verilir.
- e. Bir birimin gerçek birim olabilmesi için mutlaka maddesel bir varlığa sahip olması gerekmez.

7, 8 ve 9. soruları aşağıdaki açıklamayı kullanarak cevaplayınız.
A bisküvi fabrikasında üretilen bisküvi paketlerinin, planlanan ağırlıkta olup olmadığı araştırılmak isteniyor.

7. Bu araştırmada incelenen değişken nedir?
- a. Fabrika
- b. Üretilen bisküviler
- c. Bisküvi paketleri
- d. Bisküvi paketlerinin ağırlığı
- e. Herhangi bir bisküvi paketinin ağırlığı

8. Bu araştırmada birim nedir?
- a. Bisküvi üreten fabrikalar
- b. Bisküvi üreten A fabrikası
- c. A fabrikasında üretilen bisküvi paketlerinin her biri
- d. Bisküvi paketlerinin tamamı
- e. Bisküvi üreten fabrikaların her biri

9. A fabrikasında üretilen her bir bisküvi paketi için aşağıdaki seçeneklerden hangisi **söylenemez?**
- a. Sürekli birim
- b. Gerçek birim
- c. Maddesel birim
- d. Varsayımsal birim
- e. Doğal olmayan birim

10. Verilerin derlenmesi, düzenlenmesi, çözümlenmesi ve sonuçların yorumlanması amacıyla kullanılan yöntemler topluluğuna ne ad verilir?
- a. İstatistik
- b. Parametre
- c. Örneklem
- d. Örnekleme
- e. Değişken

Kendimizi Sınayalım Yanıt Anahtarı

1. c	Yanıtınız yanlış ise "Birim Tanımını ve Birim Türleri"ni yeniden gözden geçiriniz.
2. c	Yanıtınız yanlış ise "Devamlı Birim-Ani Birim Sınıflandırması"nı yeniden gözden geçiriniz.
3. e	Yanıtınız yanlış ise "Sürekli-Kesikli Değişken ve Nicel Değişken" tanımlarını yeniden gözden geçiriniz.
4. d	Yanıtınız yanlış ise "Sürekli-Kesikli Değişken" tanımlarını yeniden gözden geçiriniz.
5. a	Yanıtınız yanlış ise "Değişken" tanımını yeniden gözden geçiriniz.
6. e	Yanıtınız yanlış ise "Birim Türleri"yle ilgili açıklamaları yeniden gözden geçiriniz.
7. d	Yanıtınız yanlış ise "Değişken" tanımını yeniden gözden geçiriniz.
8. c	Yanıtınız yanlış ise "Birim ve Evren" tanımlarını yeniden gözden geçiriniz.
9. d	Yanıtınız yanlış ise "Birim Türleri"ni yeniden gözden geçiriniz
10. a	Yanıtınız yanlış ise "İstatistiğin" kelime anlamlarını yeniden gözden geçiriniz.

Sıra Sizde Yanıt Anahtarı

Sıra Sizde 1

- Türkiye'de faaliyette bulunan sigorta şirketleri, yıllık gelirleri, işgören sayıları, ürettikleri poliçe sayısı vb. özellikleri bakımından aralarında farklılıklar bulunacağı için birer yığın olaydır ve istatistiğin konusu olur.
- Benzinli otomobiller yakıt türü özelliği bakımından incelenirse istatistiğin konusu olmaz.

Sıra Sizde 2

- Belirli bir banka şubesinde çalışan personelin çalışma yeri türü ilgili banka şubesidir. Bütün çalışanlar bu özellik bakımından farklılık göstermezler. Dolayısıyla belirli bir banka şubesinde çalışan personel çalıştığı iş yeri türü özelliği bakımından tipik olaydır, istatistiğin konusu olmaz.
- Burada tek bir sigorta acentesi tek bir özelliği bakımından incelenmek isteniyor. Bu durum, araştırma ve istatistiğin konusu olmaz. Buna göre tek bir sigorta acentesi yığın olay olarak değerlendirilmez.

Sıra Sizde 3

- Bir bankanın şubelerinin işgören sayısı ve müşteri sayısı kesikli nicel değişkendir. Çünkü bu değişkenler sayma suretiyle ifade edilebilir. Bu değişken, örneğin 156 kişi, 75 kişi vb. gibi değerler alır.
- Mevduat tutarı ve kredi tutarı ölçme suretiyle ifade edildiğinden sürekli nicel değişkenlerdir. Çünkü T180656, T956518 gibi değişik değerlerle ölçülür.
- Şirketlerin hukuki şekli nitel bir değişkendir ve kesiklidir çünkü anonim şirket, limited şirket, kolektif şirket vb. şeklinde ifade edilir.

Sıra Sizde 4

- Öğrencilerin başarısını etkileyen değişkenler barınma yeri türü, çalışma süresi, derse devam sıklığı şeklinde sayılabilir.
- Bu araştırmada, söz konusu malın fiyatı bağımsız değişken satış miktarı ise bağımlı değişkendir. Fiyat, satışlar üzerinde etkilidir.

Yararlanılan ve Başvurulabilecek Kaynaklar

Arıcı, H. (2001). **İstatistik Yöntemler ve Uygulamalar,** (13. Baskı), Ankara, Meteksan A.Ş.

Çömlekçi, N. (2001). **Bilimsel Araştırma Yöntemi ve İstatistiksel Anlamlılık Sınamaları,** İstanbul, Bilim Teknik Yayınevi.

Çömlekçi, N. (2005). **Temel İstatistik İlke ve Teknikleri,** (4. Baskı), İstanbul, Bilim Teknik Yayınevi.

Işığıçok, E. (2009). **İstatistiksel Bakış Ekonomi-İş Dünyasına-Siyasete,** (2. Baskı), Bursa, Marmara Kitabevi Yayınları.

Serper, Ö. (2004). **Uygulamalı İstatistik 1,** (5. Baskı), Bursa, Ezgi Kitabevi.

Yıldızbakan, A. K. (2001). **İstatistik ve Ormancılıktaki Önemi,** Tarsus, Doğu Akdeniz Ormancılık Araştırma Müdürlüğü DOA Dergisi.

Yücel, İ. H. (1997). **Bilim Teknoloji Politikaları ve 21. Yüzyılın Toplumu,** Ankara, Devlet Planlama Teşkilatı.

Yüzer, A. F. (Ed.), (2011). **İstatistik,** (8. Baskı), Eskişehir, Anadolu Üniversitesi Yayını.

İSTATİSTİK

2

Amaçlarımız

Bu üniteyi tamamladıktan sonra;

- Verileri derleyebilecek,
- Derlenen verileri düzenleyerek tablolarla göstererek betimsel bilgi üretebilecek,
- Derlenen ve düzenlenen verilerden grafikle göstererek betimsel bilgi üretebileceksiniz.

Anahtar Kavramlar

- Veri Derleme
- Veri Derleme ve Derleme Yöntemleri
- Gözlem, Görüşme, Anket
- İstatistiksel Seriler
- Dağılma Serileri
- Zaman Serileri
- Mekân Serileri
- İstatistiksel Serilerin Grafiksel Çözümlemesi

İçindekiler

İstatistik → Veri Derleme, Düzenleme ve Grafiksel Çözümleme

- GİRİŞ
- DEĞİŞKENLERİN ÖLÇÜLMESİ
- VERİ DERLEME VE VERİ DERLEME YÖNTEMLERİ
- İSTATİSTİKSEL SERİLER
- İSTATİSTİKSEL SERİLERİN GRAFİKSEL ÇÖZÜMLEMESİ

Veri Derleme, Düzenleme ve Grafiksel Çözümleme

GİRİŞ

İstatistiksel araştırmaların en önemli aşamalarından biri verilerin derlenmesi aşamasıdır. Verilerin derlenmesi zahmetli, uzun süren ve dikkatli olunması gereken bir süreçtir. Verilerin doğru kaynaktan doğru araçlarla toplanması büyük önem arz etmek ile beraber, verilerin özetleyici tablolar veya uygun grafikler ile sunulması da çok önemlidir. Sunum ve raporların yüksek kaliteli grafik, şema ve tabloları içermeleri beklenmektedir. Her ne kadar yazılı raporda yer alan bilgiler ve rakamlarla ifade edilen istatistikler can alıcı olsa da bu bilgilerin çarpıcı grafikler ile eşleştirilmesi sonucunda rapordan istenen etki miktarı da artmaktadır. Yalnızca rapor ve sunumlarda değil bilimsel makalelerde, gazete yazılarında ve dergi incelemelerinde de sıklıkla istatistiklerin grafikler yardımıyla sunumuna rastlanılmaktadır. İstatistiksel analizlerde veriler, farklı formlarda ortaya çıkabilmektedir. Veriler harita, fotoğraf, MR görüntüsü, ders notu, deney sonucu, gelen otobüsün rengi, üzerinizdeki kazağın materyali olabilir. İstatistik açısından verinin doğasının anlaşılması büyük önem taşır çünkü farklı veri türleri için farklı istatistiksel analizlerin uygulanması söz konusudur. Verilerin doğru anlaşılması doğru tablo ve grafik kullanımını da kolaylaştıracaktır.

Bu ünitede öncelikle değişkenlerin nasıl ölçüleceği ve ölçme düzeylerinin önemi vurgulanarak farklı ölçek türleri ele alınacak, verilerin derlenmesi aşamasında kullanılan teknikler açıklanarak değişkenler için bazı tablolama ve grafik teknikleri sunulacaktır.

DEĞİŞKENLERİN ÖLÇÜLMESİ

Planlanan bir bilimsel araştırma sürecinde değişkenler belirlendikten sonra sıra bu değişkenin ölçülmesine gelir. Değişkenlerin ölçülmesinde kullanılabilecek ölçekler Psikolog Stevens (1946) tarafından dört başlık altında toplanmıştır. Ünitenin izleyen kesiminde öncelikle ölçme tanımlanacak daha sonra da Stevens tarafından önerilen ölçek türleri incelenecektir.

Ölçme Tanımı

Araştırmalarda üzerinde çalışılan *evrenin birimlerinin sahip olduğu özelliklerin anlamlı rakam, sayı ve simgeler ile ifade edilmesi işlemine ölçme adı verilir*. Ölçme işlemi sonucu elde edilen bu anlamlı rakam, sayı ve simgelere *veri* adı verilir. Ölçme işlemi öyle yapılmalıdır ki bu rakam, sayı ve simgeler ölçülen özellik içinde ortaya çıkan ilişkileri yansıtabilmelidir. Ölçmede üç aşama vardır: Bunlar *ölçülecek bir niteliğin olması, niteliğin gözlemlenebilir olması ve amaca uygun sayı ve semboller ile gösterilebilir olmasıdır*. Ölçme doğrudan ölçme ve dolaylı ölçme olmak üzere iki ana başlıkta ele alınabilir. Doğrudan ölçme işleminde ölçme konusu değişken kendisi ile aynı türden bir araç ile ölçülür. Örneğin, sıcaklığın Fahrenhayt, kapasitenin metreküp ile ölçülmesi gibi. Dolaylı ölçme işleminde ise ölçmeye konu olan değişken doğrudan gözlemlenmez ancak kendisi ile ilgili olduğu bilinen başka değişkenler aracılığı ile ölçülür, zeka testleri bu duruma örnek olarak gösterilebilir.

Ölçme işleminde *ölçen, ölçülen* ve ölçüm amacı ile kullanılan *araçlar* söz konusudur. Örneğin, hastanın ateşini ölçerken; ölçüm işlemini yapan *hemşire ölçen*, kullanılan *termometre ölçüm aracı* ve *ölçülen de hastanın ateşidir*. Ölçmede ölçen, ölçüm amacı ile kullanılan araç ve ölçülen arasında tutarlı bir ilişkinin olması gerekir. Dikkatsiz bir hemşirenin veya bozuk bir derecenin ölçüm işleminde kullanılması ölçümün hatalı olmasına neden olur. Sosyal bilimlerde, doğal bilimlerde olduğu gibi bir değişkenin diğer faktörlerin etkisinden soyutlanarak laboratuvar ortamında ölçme ve deney yapma olanağı çok azdır. Örneğin, enflasyonu ölçmek istediğimizi düşünelim. Fiyatlar genel seviyesinin yükselmesi anlamına gelen enflasyonu ölçmek için derece veya metre gibi araçlara sahip olmamız gerekli değildir. Enflasyon değişkenine ilişkin ölçüm yapılırken ölçümü yapanlar insan ve ölçülen ise enflasyon sepetinde yer alan mal ve hizmetlerin fiyat değişimidir. Örneklerden de görüldüğü gibi, ölçme işleminde işin içine değişkenin yapısına bağlı olarak çok sayıda ölçen ve ölçüm yöntemi ortaya çıkmaktadır. Bu nedenle sosyal bilimlerde ölçme doğal bilimlere göre daha zordur (Serper ve Gürsakal, 1989).

Ölçek Türleri

Psikolog Stevens (1946) bilimsel çalışmalarda yapılabilecek tüm ölçümler için aşağıdaki dört farklı ölçek türünün kullanılabileceğini belirtmiştir. Bu ölçek türleri; sınıflayıcı ölçek, sıralayıcı ölçek, eşit aralıklı ölçek ve oranlı ölçektir. Bu sıralama ölçek türlerinin ürettiği bilginin niteliğine göre düşük nitelikten yüksek niteliğe doğru yapılan bir sıralamadır.

Sınıflayıcı ölçek, en basit düzeyli ölçek türüdür. Nitel değişkenlerin ölçülmesi amacı ile kullanılır. Bu ölçek türünde, değişkenin ölçme düzeyleri kategorilerdir. Kategorilerin temsili için harfler, kelimeler ve hatta sayılar sadece etiketleme amaçlı olarak kullanılır. Sınıflayıcı ölçekte değişkenin aldığı sonuçlar kategorilere ayrılmakla beraber kategoriler arasında doğal bir sıralama söz konusu değildir. Örneğin, cinsiyet nitel değişkenini ele alalım. Bu değişkenin kadın ve erkek olmak üzere iki kategorisi vardır. Cinsiyet değişkeninin sonuçları olan kadın ve erkek kategorilerinin doğal bir sıralaması bulunmamaktadır. Kimi araştırmacılar kadın bilgisini ilk sırada verebilecekleri gibi kimi diğer araştırmacılar erkek bilgisini ilk sırada verebilirler. Benzer bir şekilde bir otoparkta yer alan araçların renk değişkeni de sınıflayıcı ölçekle ölçülmektedir. Bu değişkenin alabileceği değerler, kırmızı, mavi, sarı, yeşil vb. değerleri almakta ve herhangi bir rengin önce söylenmesini gerektirecek doğal bir sıralama bulunmamaktadır. Sınıflayıcı ölçekte kategoriler karşılıklı ayrık olma özelliği taşır. Bu özellik, sınıflama işlemi yapılırken belirli bir kategoriye atanan birimin bir başka kategoriye de atanmasının mümkün olmaması anlamına gelir. Her birim yalnızca bir kategori içerisinde yer alabilir. Örneğin bir otoparkta yer alan araç hem kırmızı hem de yeşil renkli olarak sınıflanamaz.

Sıralayıcı ölçek, sınıflayıcı ölçeğe göre bir üst seviye nitelikte bilgi üreten ölçek türüdür. Sınıflayıcı ölçekte olduğu gibi nitel değişkenlerin kategorilere ayrılarak sayılması işlemini içermekle birlikte, bu ölçek türünde ölçüm sonuçlarının *doğal bir sıralaması söz konusudur*. Örneğin, bir güzellik yarışmasında 10 adet aday olduğunu düşünelim. Yarışma sonucunda jüri kararı ile adaylar ilk 3 şeklinde sıralanırlar. Yarışmanın kazananı 1 numaralı sırayı alırken sonrakiler ikinci ve üçüncü sırayı alırlar. Burada dikkat edilmesi gereken nokta sıralamaya girenlerin yarışmaya katılan bireyler oldukları ve bu bireylerin güzellik yarışması kriterlerine göre sıralandıklarıdır. Yarışmadaki ölçüm sonucuna göre bir sıralama olmakla beraber birinci ile ikinci arasındaki güzellik farkının ölçümlenmesi veya bu farkın matematiksel olarak ikinci ve üçüncü arasındaki fark ile karşılaştırılabilmesi mümkün değildir. At yarışlarında yarış sonucunda oluşan sıralama, sıralayıcı ölçek için örnek olarak verilebileceği gibi, bir mağaza çalışanlarının unvan sıralamaları, dereceleri, tütün kalitesi sıralayıcı ölçeğe örnek olarak verilebilir. Sıralayıcı ölçekte de değişkenin kategorileri karşılıklı ayrıktır.

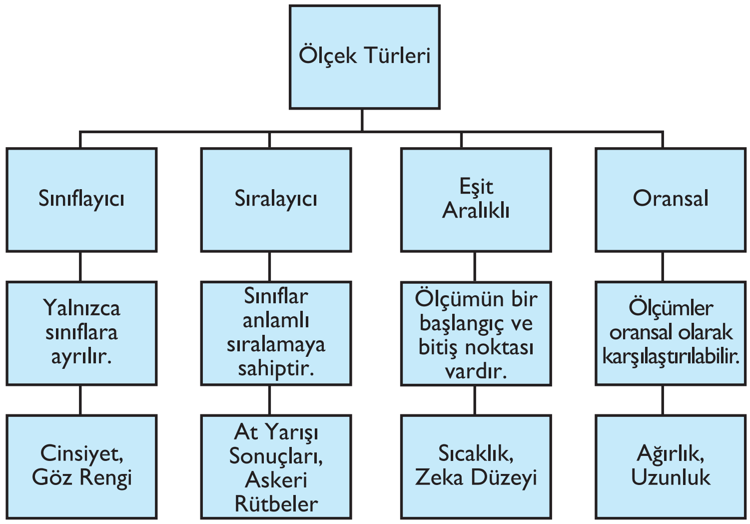
Eşit Aralıklı ölçek, sıralayıcı ölçeğin tüm özelliklerini içermek ve ürettiği bilgileri üretmekle birlikte, birimler arasında özellik farkları matematiksel olarak belirlenebilir. Bu ölçek, nicel değişkenlerin ölçümünde kullanılır. Ölçümün belirli bir başlangıç noktası ve bitiş noktası olmakla birlikte, bir de ölçü birimi bulunmaktadır. Eşit aralıklı ölçek, sayısal olarak ifadelerin sıralanabilmesine olanak vermektedir. Özellikler arasındaki eşit farklılıklar, eşit ölçme düzeyleri ile temsil edilebilir. Her ne kadar eşit aralıklı ölçekte ilgilenilen değişken, matematiksel sonuçlar vermekte olsa da kullanılan ölçüm için belirli bir yokluk anlamına gelmeyen sıfır ölçme düzeyi bulunabilir. Örneğin, hava sıcaklığını düşünelim. Bu değişken nicel ölçme düzeylerine sahiptir ve yokluk anlamına gelmeyen 0 değeri alabilir. Buradaki 0 ölçme düzeyi havada sıcaklığın olmadığı anlamına gelmez. Bu ölçek ile yapılan ölçümler matematiksel işlemler için uygun olmakla beraber, oran hesaplamaları için uygun değildir.

Oranlı ölçek, en üst düzeyde bilgi üreten ölçek türüdür. Bu ölçekte sıfır başlangıç noktası tüm ölçüm araçlarında aynı anlamı taşımaktadır. Örneğin, bir varlığın ağırlığı için "sıfırdır" ifadesi kullanıldığında ölçüm metrik türüne bakmaksızın, bu varlığın olmadığı anlamına gelir. Sıfır kilogram ve sıfır gram aynı özelliği tarif etmektedir. Eğer "Banka hesabımda hiç param yok." ifadesini kullanırsanız paranın cinsinin Avro veya Türk lirası olması aynı anlama gelecektir, bir başka ifadeyle sizin o an için parasız olduğunuz gerçeğini gösterir. Oranlı ölçekte, ölçüm sonuçları daha önce ele aldığımız üç ölçek türünün de özelliklerini içermektedir. Ama en büyük üstünlüğü yokluk anlamına gelen belirli bir sıfır ölçme düzeyi olması, dolayısıyla ölçme düzeyleri arasında oransal analizler yapılabilmesine olanak vermesidir. Aralıklı ölçek, ölçümler arasında eşit aralıklar olması, sıralama olması ve oranların hesaplanabilmesi özelliklerini taşımaktadır.

Buraya kadar ele alınan ölçek türleri Şekil 2.1'de bir tablo yardımıyla özetlenmektedir. Araştırmacıların çalışmalarında tanımladıkları değişkenlerin hangi ölçek türüyle ölçüleceğini belirlemesi, araştırmada kullanılacak yöntem ve teknik türünün seçiminde üretilecek bilginin niteliğinde önemli rol oynar.

Şekil 2.1

Ölçek türleri ve ölçek türleri için örnekler



ÖRNEK 1

Bir araştırmada aşağıda verilen değişkenler hakkında bilgi bir soru kağıdı yardımıyla toplanmak istenmektedir. Araştırmada kullanılan değişkenleri inceleyerek hangi ölçek türlerine göre ölçüm yapıldığını belirleyiniz.

1. Cinsiyetiniz ( ) Kadın ( ) Erkek
2. Yaşınız ....
3. Medeni Durumunuz ( ) Bekar ( ) Evli
4. En son mezun olduğunuz eğitim kurumu seviyesi nedir?
( ) İlköğretim ( ) Lise ( ) Üniversite ( ) Master veya Doktora
5. Kilonuz (kg olarak) ....
6. Boyunuz (cm olarak) ....

Bu araştırmada birimler 6 farklı değişken itibari ile incelenmek istenmektedir. Birinci soruda yer alan cinsiyet değişkeninin iki kategorisi vardır, bu iki kategorinin doğal bir sıralaması yoktur, dolayısı ile bu değişkenin ölçümünde sınıflayıcı ölçek kullanılmıştır. Bu değişken nitel bir değişkendir. İkinci soruda, yaş sorulmaktadır. Yaş nicel bir değişkendir ve belirli bir başlangıç noktası vardır, sıfır değeri anlamlıdır, oranlı ölçektir. Medeni durum nitel bir değişkendir, verilen kategoriler için bir sıralama söz konusu değildir, sınıflayıcı ölçektir. En son mezun olunan eğitim kurumu seviyesi olan dördüncü değişken ise sıralayıcı ölçektir. Çünkü eğitim seviyeleri birer birer atlanır. İlköğretim'den mezun olunmadan üniversite mezunu olunması mümkün değildir. Dolayısı ile kategorilerin doğal bir sıralaması bulunmaktadır. Beşinci değişken olan kilo değişkeni de belirli bir sıfır başlangıç noktasına sahip nicel bir değişkendir, dolayısı ile oranlı ölçektir. Altıncı değişken olan boy değişkeni sürekli nicel bir değişkendir ve oranlı ölçek kullanılmaktadır.

VERİ DERLEME VE VERİ DERLEME YÖNTEMLERİ

Veri Derleme Tanımı

Araştırmacı kişi veya kuruluşların araştırmalarında tanımladıkları evrendeki veya örneklemdeki birimlerin özelliklerini *uygun ölçek türü kullanarak ölçmesine* veri derleme denir.

Araştırmacıların sağlıklı sonuçlar ve yorumlamalara ulaşabilmeleri, doğru, hatasız ve yeterli sayıda veriyi derlemelerine bağlıdır. Başka bir ifadeyle ilgilenilen değişkenlere ilişkin sağlıklı veriler derlenememiş ise çözümleme yöntemi, tekniği ve kuramsal altyapı ne kadar güçlü olursa olsun doğru bilgilere ulaşmak mümkün olmaz.

Veri derlemede dikkat edilmesi gereken ilkeler;

- Veri derlemenin amacının açıkça ifade edilmesi,
- Amaca uygun olarak hangi değişkenlere ilişkin verilerin derleneceğine karar verilmesi, derlenecek verilerin birbirleriyle ve araştırmanın genel amaçlarıyla ilişkisinin incelenmesi,
- Verilerin nasıl, kimler tarafından, nerede, ne zaman ve hangi şartlar altında derleneceğinin belirlenmesi,
- Verilerin nasıl saklanacağı ve gerektiğinde kullanıcıların yararına nasıl sunulacağının (tablo, grafik vb.) tespit edilmesi

olarak sayabiliriz (Serper ve Gürsakal, 1989).

Veri Derleme Yöntemleri

Değişkenlerle ilgili verilerin derlenmesi amacıyla kullanılacak yöntemler hakkında çeşitli sınıflandırmalar yapılmaktadır.

Veri derleme yöntemlerinin sınıflandırılmasına ilişkin daha ayrıntılı bilgi için "Uygulamalı İstatistik I" (Serper, 2004) eserinden yararlanabilirsiniz.

KİTAP

Bilimsel araştırma süreçlerinde genellikle kullanılan sınıflandırma, birinci elden veri derleme ve ikinci elden veri derleme sınıflandırmasıdır. Bu nedenle aşağıda bu veri derleme yöntemleri ile ilgili açıklamalarla yetinilmiştir. Verilerin derlenmesi aşamasına gelen bir araştırmacının *"verilerin doğru olması", "verilerin güvenilir olması", "verilerin kullanılabilir olması", "verilerin yararlı olması"* ve *"verilerin eksiksiz olması"* özelliklerini göz önüne alması gerekmektedir.

Birinci Elden Veri Derleme Yöntemleri

Araştırmacılar, araştırmalarında belirledikleri değişkenlerle ilgili bilgiler üretirken genellikle kendi derledikleri verileri kullanma eğilimindedirler. Çünkü farklı amaçlarla aynı değişkenler hakkında daha önceden derlenen verilerin ne şekilde hangi araçlar kullanılarak derlendikleri konusunda güvenilir bilgilere sahip olamayabilirler ya da veriler güncelliğini yitirmiş olabilir.

Araştırmacının araştırma kapsamına aldığı değişkenlerle ilgili ihtiyaç duyduğu özgün verileri uygun araçlar kullanarak kendisi derliyorsa veya derletiyorsa bu verilere birincil veriler, yapılan işleme de birinci elden veri derleme denir. Veri derlemede dikkat edilmesi gereken ilkelere uyulduğu sürece, birincil kaynak verilerin en güvenilir veriler olacakları düşünülmelidir. Bunun nedeni olarak araştırmacıların verilerin hangi koşullar altında derlendiğini, nasıl yorumlanacağını ve ortaya çıkan sonuçların problemin çözümüne olan katkılarını diğer kişilerden daha çok bilen kişiler olmaları gösterilebilir.

Birinci elden veri derleme amaç olduğunda başlıca üç veri derleme yönteminden biri kullanılır. Bunlar; Gözlem, Görüşme ve Anket başlıkları altında toplanabilir.

Gözlem

Bir nesnenin, olayın veya bir gerçeğin, niteliklerinin öğrenilmesi amacıyla dikkatli ve planlı olarak ele alınmasına "*gözlem*" adı verilir. Gözlemcinin, gözlenen olay veya grubun içinde yer aldığı duruma *katılımlı* gözlem, gözlemcinin gözlenen olay veya grubun içinde yer almadığı duruma ise *katılımsız* gözlem adı verilir.

Katılımlı gözlem durumunda, gözlemci gözlenen durumun içinde yer alır. Böyle bir durumun faydası, olayların dışarıdan bakıldığında farklı, içeriden bakıldığında ise daha farklı olabileceği gerçeğini gözönüne almasıdır. Böylece en özgün veriye ulaşılabilir.

Katılımsız gözlemde, gözlemci gözlenen duruma doğrudan katılmayarak dışarıdan izler, bilinçli bir şekilde gözlemlenen nesneyle ilişki veya etkileşime girmez. Katılımsız gözlem çeşitlerinden olan ve araştırmada yer alan birimlerin kendi doğal ortamlarında ve doğal ortaya çıkış şekillerinde gözlemlenmesine doğal gözlem adı verilir. Bu tür durumlarda gözlemci herhangi bir durumu kontrol altında tutmaz ve olayların kendi doğal sürecinde ortaya çıkmasına izin verir, gözlemlenen durumların üzerinde bir etkisinin olmaması için özel çaba sarfedilir. Bir başka katılımsız gözlem çeşidi ise canlandırmadır. Bu tür durumlarda araştırmacı, gözlemlenecek durumu kontrollü bir ortamda oluşturur ve deneklere neler yapmaları gerektiğini söyler. Canlandırma gözlem tekniği özellikle düzenli olarak ortaya çıkmayan durumları gözlemlemek için kullanılır. Canlandırma gözlem tekniğinde kişisel rol oynama, grup rol oynama ve oyun oynama gibi farklı yöntemler uygulanır (Altunışık vd. 2010).

Görüşme

Görüşme iki veya daha fazla birey arasında belli bir amaca yönelik sözlü iletişim yoluyla veri derleme yöntemidir. Görüşmenin içeriği araştırmanın amaçları ve araştırma sorularına bağlı olarak oluşturulur. Görüşme sırasında araştırmacı, verileri ya yazılı olarak ya da bir ses/görüntü kaydedici yardımıyla kayıt altına alır.

Görüşme değişik ölçütlere göre sınıflandırılabilir. Bunlar, biçimsel, yarı-biçimsel ve biçimsel olmayan görüşmelerdir. Biçimsel görüşmeler, daha önceden belirlenmiş ve bir standart dahilinde ortaya çıkan soruların cevaplanmasından oluşan görüşmelerdir. Görüşmeyi yapan görüşmeci veya araştırmacı soruları bireylere okur ve bireylerin cevaplarını kayıt altına alır. Nüfus sayımlarında sayım memurlarının hane bireyleri ile yaptıkları görüşmeler, biçimsel görüşmelerdir. Yarı-biçimsel görüşmelerde, görüşmeci kaba hatlarıyla bir yol haritasına sahiptir ancak cevaplayıcının ilgi ve bilgisine göre genel çerçeve içerisinde değişik sorular sorarak konunun farklı boyutlarını ortaya çıkarmaya çalışır. Biçimsel olmayan görüşmelerde, genel bir alanda var olan bilginin açığa çıkarılması amaçlanır. Üzerinde çalışılan alandaki genel anlayışın belirlenmesine katkıda bulunulur. Bu tür görüşmelerde önceden hazırlanmış sorular bulunmaz ancak görüşmecinin konunun hangi boyutlarını ortaya çıkarmaya çalıştığını bilmesi ve görüşmeyi bu zemin üzerinde ilerletmeye çalışması gerekir. Bu görüşmelerin biçimsel olmamasının nedeni ise görüşmenin içeriğini cevaplayıcının üzerinde durulan konuyu anlayış ve yorumlayış biçiminin oluşturmasıdır.

Görüşmelerin en zayıf yanı çok uzun bir süre alabilmeleridir. Diğer bir zayıf yan ise bireysel temasların fazla olmasından dolayı subjektif ve önyargısal durumların ortaya çıkabilmesidir. Örneğin görüşmecinin dış görünüşü görüşülen kişi üzerinde olumsuz bir etki yaratabilir ve birey görüşme için iş birlikçi olmaktan vazgeçebilir. Kimi durumlarda ise bireyler görüşmeciyi memnun etme tavrı takınabilirler. "Görüşme, balık avlamak gibidir. Dikkatli bir hazırlık, çok fazla sabır ve yeterince araştırma gerekir" (Altunışık vd. 2010).

Anket

Cevaplayıcının daha önceden belirlenmiş bir sıralamada ve yapıda oluşturulan sorulara karşılık vermesiyle verilerin derlenmesi yöntemine anket adı verilir. Soruların bir araya geldiği forma anket formu adı verilir. Anketler, tüm bireylerden aynı sorular için cevaplar beklendiğinden çok fazla sayıda kişiye uygulanabilir. Ayrıca modern istatistik yazılımları anket formlarının hızlı bir şekilde işlenmesi ve yorumlanmasına olanak vermektedir. Anketler, araştırmacıların ihtiyaç duyacağı ve araştırma sorularına cevap bulabileceği soruları içermelidir. Eğer ihtiyaç duyulan tüm bilgilere ilişkin sorular anket formunda yer almaz ise aynı bireylerden tekrar bilgi alınması çok ama çok zor olacaktır. Anketlerde cevaplayıcılar kendilerine verilen anket formunu okur ve gerekli yerlere görüşlerini işaretleme veya açıklama yaparak belirtirler. Görme engelli cevaplayıcılara tarafsız bir kişi anket formunun okunması ve doldurulmasında yardımcı olabilir. Anket yoluyla veri derlenmesinin istendiği durumlarda araştırmacının izleyen süreci izlemesi gerekecektir: Anket süreci, *problemin ifade edilmesi, evrenin ve örneklemin belirlenmesi, anket formunun düzenlenmesi, anket formunun az sayıda birey üzerinde geçerliliğinin sorgulanması*, anket formunun girişinde yer alacak *açıklama metninin* (çalışmanın tanımı, amacı, sonuçların nasıl kullanılacağı, formun doldurulmasında tarih sınırlaması olup olmadığı vb bilgiler) *hazırlanması, anket formunun bireylere sunulması, takip çalışması* ile *formun hedeflenen bireylere ulaşması* ve *geri dönüşün sağlanması* ile sonuçların analizi aşamalarından oluşur.

Veri derleme yöntemlerine ilişkin daha ayrıntılı bilgiler için "Sosyal Bilimlerde Araştırma Yöntemleri," (Altunışık vd. 2010) isimli eserden yararlanabilirsiniz.

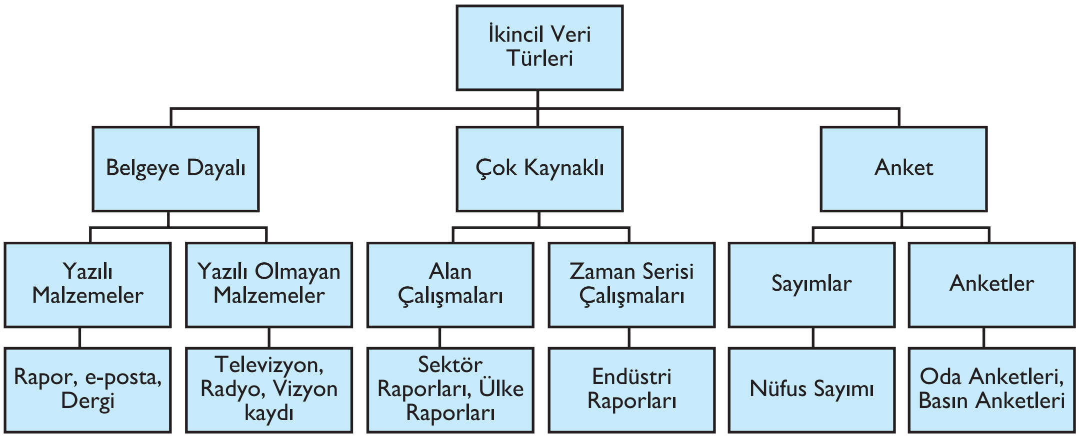
KİTAP

İkinci Elden Veri Derleme Yöntemleri

Araştırmacı kişi veya kuruluş, araştırmasında başka kişi veya kuruluşların daha önce kendi amaçları için derlemiş olduğu verileri kullanıyorsa yapılan veri derlemeye ikinci elden veri derleme adı verilir. İkincil veriler sınıflandırılırken farklı yazarlar değişik yöntemler kullanmıştır. Ancak Saunders ve diğerlerinin (1997) oluşturduğu ve Şekil 2.2'de yer alan sınıflandırma yeterince açıklayıcı olmaktadır.

Şekil 2.2

İkincil veri türleri



Kaynak: *Altunışık vd. 2010*

İkinci elden verilerle çalışılırken araştırmacıların üzerinde durması gereken unsurların başında, kişi veya kurumlar tarafından derlenen verilerin nasıl derlendiğinin bilinmesi gelmektedir. Ayrıca verinin nasıl kayıt altına alındığı, derlendiği zamandan araştırmacının ihtiyaç duyduğu döneme kadar değişikliğe uğrayıp uğramadıkları da bilinmelidir. Her ne kadar araştırmanın niteliğine göre değişecek olsa da ikincil verilerin kullanımında zaman kavramına dikkat edilmeli ve çok eski dönemlerde derlenen veriler yerine güncel derlemelerin kullanılmasına dikkat edilmelidir. İkincil verilerin kulanılması araştırmacılara büyük kazanımlar sağlarlar. Örneğin, araştırmacı ikinci elden veri derleme yöntemini seçtiğinde, birinci elden veri derlemeye göre maliyet, zaman ve iş gücünden tasarruf edebilir. Önceden herhangi başka bir amaç için derlenmiş veri, araştırmacının da kullanımına uygun ise araştırmanın veri derleme aşamasında geçecek olan süre çok kısalacaktır. Örneğin, bir araştırmacı belirli bir şehirde ikamet edenlerin sahip oldukları toplam cep telefonu sayısını (abone sayısını) belirlemek istiyorsa bu amaçla gerekli olan veriler GSM servis sağlayıcılarından elde edilirse ikinci elden veri derlenmiş olur. İkincil veriler derlendiğinde araştırmacının elde ettiği bu verilerin kendi araştırmasına uygunluğunu da son derece dikkatli bir şekilde sınaması gereklidir. Kimi durumlarda ise araştırmacılar, ikincil verileri araştırmadan veriyi birinci elden toplama yöntemini izlemektir, oysa öncelikle ikincil kaynakların taranması ve mevcut bilgilerin elde edilmesi çok daha uygundur. İkinci elden veriler derlenirken aynı değişkene ilişkin farklı kaynaklardan yararlanma durumu olabilir. Bu durumda kaynak seçimi önem arz eder.

KİTAP

Veri derleme yöntemlerine ilişkin daha ayrıntılı bilgiler için "Araştırma Yöntemleri" (Serper ve Gürsakal, 1989) isimli eserden yararlanabilirsiniz.

Veri Derleme Araçları

Veri derleme amacı ile kullanılacak araçlar maddi ve beşeri araçlar olmak üzere sınıflandırılırlar. Gözlemlenecek birimlerin özelliklerini ölçmeye yarayan aletler maddi araçları oluşturur. Örneğin, ağırlık değişkeninin ölçümünde kullanılan terazi, baskül; hava sıcaklığı değişkeninin ölçümü için kullanılan termometre ve anket formları maddi araçlardır. Veri derlemede birimleri ve özelliklerini ölçen kişiler veri derlemenin beşeri araçlarıdır. Örneğin, anketör, gözlem memuru, görüşmeci veri derlemenin beşeri araçları olarak sayılabilirler.

Veri Derleme Hataları

Veri derleme hazırlıkları hangi düzeyde olursa olsun, yine de bazı hatalar işlenebilir. Bu hataların kaynağı veri derlemeyi düzenleyenler, cevaplayıcılar ve derlemenin beşeri araçları olabilir.

Derlemeyi düzenleyenlerden dolayı ortaya çıkan veri derleme hatalarına derlemenin uygunsuz bir zamanda yapılması ve evren tanımının hatalı yapılması gösterilebilir. Üniversite öğrencilerine ilişkin bir araştırmanın yaz tatilinde planlanması ve uygulanması bu tür sakıncalara bir örnek olarak gösterilebilir.

Cevaplayıcılardan kaynaklanan hatalar içinde soruların yeterince anlaşılamaması, sorular anlaşılsa bile bilgisizlik, ilgisizlik vb. gibi nedenlerle yanlış cevap vermeleri gösterilebilir.

Beşeri araçlardan kaynaklanan hatalara örnek olarak ise dikkatsizlik, yorgunluk ve objektif olamama nedenleri gösterilebilir.

Veri derleme hatası kimden kaynaklanırsa kaynaklansın, kapsam hatası ve vasıf hatası işlenmiş olabilir. Kapsam hatası için örnek olarak tanımlanan evrene dahil olan birimlerin eksik veya fazla sayılması gösterilebilir. Örneğin, nüfus sayımında bazı kişilerin sayılmaması gibi. Vasıf hatası ise veri derlemede birimlerin incelenen özelliklerinin ölçümlenememesi veya yanlış ölçümlenmesi gösterilebilir. Örneğin, bazı kişilerin yaşları doğru öğrenilemeyebilir, evliler bekâr olarak yazılabilir.

Veri derleme hatalarını, etkileri bakımından inceleyecek olursak rassal hata ve sistematik hata sınıflaması yapılabilir. Rassal hatalar denkleşen hatalar olup birimlerin incelenen özelliklerini yüksek ve düşük gösterme yönünde etki yapan hatalardır. Örneğin, bankalara kredi kartı başvurusunda bulunanların gelir bildirimlerini yüksek göstermeleri, kredi yurtlar kurumuna yurt başvurusunda bulunan öğrencilerin hane halkı gelirini düşük göstermesi sistematik hataya neden olur. Bu hatalar değişkenlerin ölçüm düzeylerinde ya hep küçük ya da hep büyük olma etkisi yaratır.

Eşit aralıklı ölçek ile oranlı ölçek arasındaki temel farklılık nedir? Veri derleme araçlarını sınıflandırınız ve birer örnek veriniz.

SIRA SİZDE 1

İSTATİSTİKSEL SERİLER

Yukarıda açıklanan veri derleme yöntemlerinin uygulanması ile derlenmiş olan ve elde ediliş sırasına göre sunulmuş veriler ham verilerdir. Bu verilerden bilgi üretmek zor, zaman alıcı ve bazen de imkânsızdır. Ham verilerden istenen bilgilerin kolayca üretilebilmesi için bu verilerin düzenlenmesi ve tablolarla sunulması gerekir. Ham verileri, zaman ve mekân değişkenleri ile maddi bir değişkenin ölçme düzeylerine göre sıralanmış olarak gösteren sayı dizilerine "seri" denir. Seriler; dağılma serileri, zaman serileri ve mekân serileri olarak sınıflandırılabilir. Serileri oluşturan sayılardan her birine terim adı verilir.

Dağılma Serileri

Gözlem değerlerini maddi bir değişkenin ölçme düzeylerine göre sınıflandırılması ile oluşturulan serilere dağılma serisi adı verilir. Dağılma serilerinin birinci sütununda gözlem değerleri ikinci sütunda ise bu gözlem değerlerinin sıklıkları (frekansları) yer alır. Bu serilere frekans dağılımı adı da verilmektedir. Dağılma serileri nicel değişkenler için nicel dağılma serileri, nitel değişkenler için ise nitel dağılma serileri adını alır.

Aslında dağılma serisi niteliğinde olmayan fakat yapılacak açıklamalarda kolaylık sağladığı ve ilerideki kısımlarda incelenecek konulara temel oluşturulacağı için, önce basit serilerle ilgili bilgi verilecektir.

Basit Seri

Araştırmacılar gözlem veya deney sonucu derledikleri verileri derledikleri sıraya göre bir not defterine veya bilgisayar dosyasına kayıt edebilir. Kayıt edilen, düzensiz olan ve ham veri adı verilen bu verilerin belirli bir kritere göre sıralanması ile oluşturulan seriye **basit seri** adı verilir. Nicel veriler için basit seri oluşturulurken genellikle kullanılan kriter verilerin küçük değerliden büyük değerliye veya büyük değerliden küçük değerliye doğru sıralamasının yapılmasıdır. Nitel veriler için basit seriler örneğin, unvan düzeyi, hizmet yılı, eğitim düzeyi ve benzeri kriterler kullanılarak oluşturulur. Görüldüğü gibi, hem nicel hem de nitel veriler basit seri olarak gösterilmektedir.

Gözlem değerlerinin küçükten büyüğe veya büyükten küçüğe doğru alt alta sıralanması ile oluşturulan serilere **basit seri** denir.

Tablo 2.1'de, bir kargo şirketinin büyük bir şehir merkezindeki şubesine taşıma amaçlı olarak müşteriler tarafından bırakılan 30 adet kargo paketinin ağırlıklarının (gram) küçükten büyüğe doğru sıralanması ile oluşturulan basit seri sunulmuştur.

Tablo 2.1
30 adet kargo paketinin ağırlıkları (gram)

53	53	59	60	60	60	66	66	74	74
77	77	77	81	81	81	81	84	84	89
89	90	90	90	90	94	94	94	95	95

Tablo 2.1 incelendiğinde, gün içerisinde taşıma amaçlı getirilen paketlerden en ağır olanının 95 gram, en hafif olanın ise 53 gram ağırlığa sahip olduğu bilgisi kolayca elde edilebilir. Bazı gözlem değerleri birden fazla tekrarlanmaktadır. Verilerin bu sunum şekli, hangi gözlem değerlerinin kaç kez tekrarlandığı bilgisinin belirlenmesi için veri sayısı çok olduğunda zor olabilir, zaman alabilir. Tekrar eden gözlemlerin tekrar sayılarının kolayca belirlenebilmesi için basit seri yeni bir düzenlemeye tabi tutulursa, bu yeni düzenleme ile ortaya çıkan seriye dağılma serisi adı verilir. Dağılma serileri nicel dağılma serileri ve nitel dağılma serileri şeklinde sınıflandırılmaktadır.

Nicel Dağılma Serileri

Evrendeki veya örneklemdeki birimlerin nicel değişkenin çeşitli ölçme düzeylerine göre dağılışını gösteren serilere nicel dağılma serileri denir. Nicel dağılma serileri, frekans serileri (sınıflandırılmış seriler) ve gruplandırılmış seriler şeklinde düzenlenebilir.

Frekans Serisi

İlgilenilen bir nicel değişkenin aldığı değerleri bir sütunda ve bu değerlerin frekanslarını diğer sütunda göstermek sureti ile oluşturulan serilere **frekans serileri** denir.

Tanımlanan bir araştırmada ilgilenilen bir X nicel değişkeni için derlenen verilerde tekrar eden gözlemlerin kaç kez tekrarlandığını gösterecek şekilde, basit serinin yeniden düzenlenmiş hâline **frekans serisi** adı verilir. Tablo 2.2'de, düzenlenen örnek frekans serisi gösterilmiştir. Bu düzenlemeden görüldüğü gibi tabloda iki adet sütun yer alır. Birinci sütun da birbirinden farklı olarak ortaya çıkan gözlem değerleri küçük değerliden büyük değerliye doğru sıralı olarak (birden fazla tekrarlanan terimler bir kez yazılırlar) yer alırken, ikinci sütunda ise birinci sütunda alt alta sıralanan değerlerin sayma yoluyla elde edilen tekrar sayılarını gösteren frekanslar yazılır. Frekans serilerinde birinci sütun X_i ile sembollenir, frekansların yer aldığı ikinci sütun ise n_i ile sembollenir.

Örneğin, Tablo 2.1'de sunulan kargo paket ağırlıkları basit serisinde bazı terimler birer kez tekrarlanırken, bazıları 2, 3 ve 4 kez tekrarlanmaktadır. Serideki en yüksek frekansın değeri 4'tür ve bu seride 81 değerine sahip 4 terim olduğu anlamına gelir. Bu frekansları tüm birbirinden farklı değerler için sayarak frekans serisini oluşturalım. Paketlerin ağırlığı değişkeninin aldığı X_i değerlerini birinci sütunda ve her X_i değerinin tekrarlanma sayısı (frekans) n_i sütununda gösterilerek Tablo 2.2'de yer alan frekans serisi, bir başka ifadeyle nicel sınıflandırılmış frekans serisi oluşturulmuş olur.

Tablo 2.2
30 adet kargo paketi ağırlıkları frekans serisi

Ağırlık (X_i)	Frekans (n_i)
53	2
59	1
60	3
66	2
74	2
77	3
81	4
84	2
89	2
90	4
94	3
95	2
Toplam	30

Frekans serilerinden yararlanarak hangi verinin bir başka ifadeyle hangi gözlem değerinin kaç kez tekrarlandığı bilgisinin üretilmesi, basit seriye göre daha kolay olur. Tablo 2.2'de yer alan frekans serisini incelersek 74 gram ağırlığa sahip kargo paketi sayısının 2 paket, 81 gram ağırlığındaki paket sayısının 4 olduğu bilgisi kısa zamanda üretilebilir.

Frekans serileri bazen her gözlem değerine karşılık gelen frekansın toplam frekans içindeki oranı alınarak da düzenlenebilir. Bu şekilde düzenlenen serilere oransal frekans serisi adı verilir. Tablo 2.2'de yer alan verilerden yararlanarak oransal frekans serisinin nasıl düzenlendiği Tablo 2.3'te sunulmuştur.

Tablo 2.3
30 adet kargo paketi ağırlıkları oransal frekans serisi

Ağırlık (X_i)	Frekans (n_i)	Oranlar
53	2	2/30=0,067
59	1	1/30=0,033
60	3	3/30=0,100
66	2	2/30=0,067
74	2	2/30=0,067
77	3	3/30=0,100
81	4	4/30=0,133
84	2	2/30=0,067
89	2	2/30=0,067
90	4	4/30=0,133
94	3	3/30=0,100
95	2	2/30=0,067
Toplam	30	

Tablo 2.3'ten yararlanarak örneğin, kargo paketlerinden %6,70'inin 53 gram ağırlığa sahip olduğu ve ağırlığı 81 gram olan paket oranının %13,3 olduğu bilgisi üretilebilir.

Nicel frekans serilerinde birbirinden farklı olarak ortaya çıkan birim sayısının artması durumunda, verileri frekans serisi şeklinde sunulması zorlaşır. Bu tür durumlarda tablonun daha da özet hale getirilmesi istenebilir. Bu amaçla gruplandırılmış seri oluşturulur.

Gruplandırılmış Seri

Bir değişkenin birbirine yakın ölçme düzeylerini bir araya getirmeye gruplama adı verilir. Frekans serileri basit serilere göre bilgi üretme amacıyla büyük kolaylık getirmekle birlikte, derlenen veri sayısı ve farklı sonuç sayısı arttıkça basit seriye göre üstün olan frekans serisi de yetersiz olabilir. Veri sayısı çok çok büyük olduğunda ve birbirinden farklı çok fazla sonucun ortaya çıkması söz konusu olduğunda, verilerin kolay ve daha anlaşılır gösterimi amacıyla, verilerin gruplandırılmış seri olarak düzenlenmesi bilgi üretme bağlamında önemli kolaylıklar sağlayacağı düşünülür. Gruplandırılmış serilerde de frekans serilerindeki gibi iki sütun yer alır. Tablo 2.4'ten de görülebileceği gibi ilk sütun grupları gösterirken ikinci sütun frekans değerlerini gösterir. Nicel verilerin gruplandırılmış seri olarak gösteriminde her bir satır, frekans serilerinden farklı olarak, tek bir sonuç değeri ile değil bir değer aralığı ile temsil edilir. Bu değer aralıklarına sınıf adı verilir. Her değer aralığına (sınıfa) düşen birim sayılarının sayma yoluyla elde edilmesi ve frekans sütununa yazılmasıyla gruplandırılmış seri oluşturulur.

Tablo 2.4
30 adet kargo paketi ağırlıklar verisi için gruplandırılmış frekans serisi

Ağırlık Sınıfları	Frekans
50 - 60	3
60 - 70	5
70 - 80	5
80 - 90	8
90 - 100	9
Toplam	**30**

Örneğin, paket ağırlığı sürekli nicel değişkendir, bu değişkene ilişkin Tablo 2.2'deki frekans serisinin gruplandırılmış seri olarak düzenlenmiş hâli Tablo 2.4'te verilmiştir.

Tablo 2.4'ün birinci sütununda 50 - 60, 60 - 70 ve benzeri şekilde oluşturulmuş 5 sınıf, sütunda ise her sınıf aralığına düşen gözlem sayısı, frekanslar gösterilmiştir. Sınıfları tanımlayan değer aralığının küçük olan değerine sınıf alt sınır değeri, büyük olan değerine ise sınıf üst sınır değeri adı verilir. Örneğin, birinci sınıfın alt sınır değeri 50 iken bu sınıfın üst sınır değeri 60'dır. Her sınıfın üst sınır değeri ile alt sınır değeri arasındaki fark **sınıf aralığı** olarak tanımlanır ve h ile gösterilir. Örneğin, Tablo 2.4'te birinci sınıfın sınıf aralığı, 60 - 50 = 10 olarak hesaplanır.

Her sınıfın üst sınır değeri ile alt sınır değeri arasındaki fark **sınıf aralığı** olarak tanımlanır ve h ile gösterilir.

DİKKAT

Araştırmacının elinde basit seri veya frekans serisinin olmayıp yalnızca gruplandırılmış seri olması durumunda, gruplandırılmış serinin her sınıfında yer alan birimlerin gerçek değerlerinin ne olduğu bilgisine ulaşılamaz. Yalnızca tanımlanan sınıf aralığında değerler alan birimler olduğu frekanslar sayesinde bilinir.

Sürekli nicel değişkenlere ilişkin gruplandırılmış seri oluşturulurken oluşturulacak sınıfların üst sınır değeri izleyen sınıfın alt sınır değeri olarak yazılır. Sınıfların üst sınır değerleri yazılı değere karşı gelen gözlem değerini kapsamaz, en yakın değeri kapsar. Örneğin, Tablo 2.4'te 50 - 60 sınıfının üst sınır değeri 60 - 70 sınıfının alt sınır değeri olarak alınmıştır. Bu tablodaki birinci sınıf, bu sınıfın üst sınırı olan 60'a eşit olan gözlem biriminin değerini kapsamaz en yakın gözlem değeri olan 59,99 değerine sahip birimi kapsar. Başka bir ifadeyle 60 gram ağırlığındaki paket birinci sınıfta değil ikinci sınıfta sayılacak demektir. Sınıfların alt sınır değerleri, yazılı değerleri kapsar. Yani 60 gram ağırlığındaki paket ikinci sınıfta sayılacaktır.

Gruplandırılmış serilerde h ile gösterilen sınıf aralıkları Tablo 2.4'teki gibi bütün sınıflarda eşit (h=10) olabileceği gibi kimi sınıflarda farklı da olabilir. Bütün sınıfların sınıf aralıkları eşit ise bu gruplandırılmış seri eşit aralıklı gruplandırılmış seri, bir sınıfın dahi sınıf aralığı diğer sınıfların sınıf aralığından farklı ise bu seriye eşit olmayan sınıf aralıklı gruplandırılmış seri adı verilir. Örneğin, Tablo 2.4'te yer alan gruplandırılmış seri, eşit aralıklı gruplandırılmış seri iken bu serinin Tablo 2.5'teki gibi düzenlenmiş hâli eşit olmayan sınıf aralıklı gruplandırılmış seri olacaktır. Tablo 2.5'te ikinci sınıfın sınıf aralığı 20 iken diğer sınıfların sınıf aralıkları 10'a eşittir.

Tablo 2.5
30 adet kargo paketi ağırlıklar verisi için gruplandırılmış frekans serisi (gram cinsinden)

Ağırlık Sınıfları	Frekans	h_i (sınıf aralığı)
50 - 60	3	60-50=10
60 - 80	10	80-60=20
80 - 90	8	90-80=10
90 - 100	9	100-90=10
Toplam	**30**	

Tablo 2.6
İşçilerin çocuk sayısı değişkeni gruplandırılmış serisi.

Çocuk Sayısı	Frekans
0 - 2	20
3 - 5	65
6 - 8	20
9 - 11	12
12 - 14	3
Toplam	**120**

Kesikli nicel değişkenlere ilişkin gruplandırılmış seriler düzenlenirken birbirini izleyen sınıfların alt sınır değerleri önceki sınıfların üst sınır değerine 1 ilave edilerek belirlenir ve her sınıfa karşılık gelen frekanslar, frekans sütununda yer alır. Tablo 2.6'da kesikli bir nicel değişkene ilişkin örnek gruplandırılmış seri verilmiştir. Kesikli nicel değişkenlere ilişkin gruplandırılmış serilerde belirlenen sınıflar alt ve üst sınır değerleri de dahil olmak üzere bu değer aralığındaki bütün gözlem değerlerini kapsar. Tablo 2.6'da bir fabrikada çalışan işçilerin çocuk sayısı değişkeni için dağılımları, gruplandırılmış seri yardımıyla gösterilmiştir. Çocuk sayısı değişkeni kesikli nicel bir değişkendir.

Örneğin, Tablo 2.6. incelendiğinde 0, 1 ve 2 çocuğa sahip aileler gruplandırılmış serinin birinci sınıfında; 3, 4 ve 5 çocuğa sahip aileler gruplandırılmış serinin ikinci sınıfında sayılırlar.

Gruplandırılmış serilerde de frekans serilerine benzer olarak sınıfların, toplam frekansa göre oransal dağılımları dikkate almak amacı ile oransal frekanslar, sınıf frekansı toplam frekansa bölünerek hesaplanabilir. Tablo 2.5'te yer alan gruplandırılmış seri için oranlar hesaplanarak Tablo 2.7'de sunulmuştur.

Tablo 2.7
30 adet kargo paketi ağırlıklar verisi için oransal frekans serisi

Ağırlık Sınıfları	Frekans	Oranlar
50 - 60	3	3/30=0,10
60 - 80	10	10/30=0,33
80 - 90	8	8/30=0,27
90 - 100	9	9/30=0,30
Toplam	**30**	

Tablo 2.7'den yararlanarak örneğin, taşınma amaçlı olarak teslim edilen kargo paketlerinin %33'ünün ağırlıkları 60 ile 80 gram arasında olduğu ve kargoya verilen paketlerin %27'sinin ise 80 ile 90 gram arasında olduğu bilgisi üretilebilir.

Tablo 2.8'de işçilerin çocuk sayıları değişkenine ilişkin ve Tablo 2.6'da sunulan gruplandırılmış seri için oranlar hesaplanarak oransal frekans serisi sunulmuştur.

Tablo 2.8
İşçilerin çocuk sayısı değişkeni oransal frekans serisi

Çocuk Sayısı	Frekans	Oranlar
0 - 2	20	20/120=0,17
3 - 5	65	65/120=0,54
6 - 8	20	20/120=0,17
9 - 11	12	12/120=0,10
12 - 14	3	3/120=0,03
Toplam	**120**	

Tablo 2.8'den yararlanarak örneğin, 6, 7 veya 8 çocuğa sahip işçi oranının %17 olduğu ve işçilerin %54'ünün ise çocuk sayısının 3 ile 5 arasında (3, 4, ve 5 çocuklu işçiler) olduğu bilgisi üretilebilir.

KİTAP

Gruplandırılmış serilere ilişkin daha ayrıntılı bilgileri "Uygulamalı İstatistik I" (Serper, 2004) eserinden yararlanarak öğrenebilirsiniz.

Nitel Dağılma Serileri

Evrendeki veya örneklemdeki birimlerin bir nitel değişkenin ölçme düzeylerine (kategorilerine) göre dağılımını gösteren serilere nitel dağılma serileri denir. Nitel dağılma serilerinde iki sütun yer alır. İlk sütunda ilgilenilen değişkenin kategorileri yer alırken ikinci sütunda ilgili kategorilerin frekansları gösterilir. Ölçümlenen değişken sınıflayıcı ölçek düzeyine sahip ise ilk sütunda yer alan kategoriler araştırmacının uygun bulduğu bir düzene göre alt alta sıralanırlar. Ancak değişkenin ölçümünde sıralayıcı ölçek kullanılmış ise ilk sütunda değişkenin ölçme düzeyleri olan kategoriler doğal sıralamalarına göre alt alta sıralanırlar. Örneğin, Tablo 2.9'da bir fabrikada çalışmakta olan işçilerin cinsiyet değişkenine göre dağılımları gösterilmiştir. Cinsiyet değişkeni nitel bir değişkendir ve ölçme düzeyi ise sınıflayıcı ölçektir.

Tablo 2.9'dan yararlanarak örneğin, toplam 195 işçinin 75 tanesi kadın iken 120 tanesi erkektir bilgisi üretilir.

Cinsiyet	Frekans
Kadın	75
Erkek	120
Toplam	**195**

Tablo 2.9
İşçilerin cinsiyete göre dağılımı frekans serisi

Nitel değişkenin sıralayıcı ölçek düzeyinde ölçümlendiği duruma örnek olarak, bir anaokulunda öğrencilerin kendilerine verilen pastanın lezzeti değişkeni ele alınabilir. Tablo 2.10'da lezzet değişkenine ilişkin görüşlerin ölçüm düzeyleri 'mükemmel'den 'çok kötü'ye doğru sıralanmıştır. Her ölçme düzeyine karşılık gelen frekanslar da Tablo 2.10'da sunulmuştur.

Tablo 2.10'a göre, pastanın lezzeti hakkında fikir bildiren 50 öğrenci bulunmaktadır. Bu 50 öğrenciden 10 tanesi pastanın lezzetini mükemmel olarak nitelerken 4 tanesi çok kötü olarak nitelemiştir. Dikkat edilirse sıralayıcı ölçek düzeyi ile ölçümlenen bir değişken sözkonusu olduğunda, Tablo 2.10'un ilk sütunu ilgili değişkenin doğal bir sıralaması olarak verilmiştir. Tablo 2.10'da sütunda 'en iyi'den (mükemmel) 'en kötü'ye (çok kötü) doğru bir sıralama benimsenmekle birlikte, istenirse 'en kötü'den 'en iyi'ye doğru bir sıralama da söz konusu olabilir.

Pasta lezzeti	Frekans
Mükemmel	10
İyi	18
Orta	12
Kötü	6
Çok kötü	4
Toplam	**50**

Tablo 2.10
Anaokulu öğrencilerinin görüşlerine göre pasta lezzet dağılımı

Hatırlanırsa bir değişkenin birbirine yakın ölçme düzeylerini bir araya getirilmesine gruplama adı verilmektedir. Nitel değişkenler ile çalışırken kimi durumlarda nitel değişken için ortaya çıkan farklı kategoriler belli ana başlıklar altında incelenebilir. Örneğin, tanımlanan bir araştırmada araştırmacı araştırmaya katılan birimlerin meslek gruplarına göre dağılımını belirlemek isteyebilir. Böyle bir dağılımda ortaya çıkan meslek grubu sayısı çok fazla olabilir. Araştırmacı meslek gruplarını tek tek yazmak yerine bunları ana başlıklar altında da toplayabilir. Örneğin; mesleği polis, öğretim üyesi ve benzeri olan kişiler kamu personeli olarak sınıflanırken, mesleği avukat, kırtasiyeci ve benzeri olanlar serbest meslek olarak sınıflandırılabilirler. Görüldüğü gibi her sınıflandırma bir grup oluşturmaktadır. Dolayısıyla nitel değişkenlerin farklı düzeylerinin araştırma için anlamlı gruplar altında toplanarak frekans serisi oluşturulduğunda ortaya çıkan seriye gruplandırılmış nitel seri adı verilir. Tablo 2.11'de bir GSM firmasının A şehrindeki abonelerinin meslek gruplarına göre dağılımı gruplandırılmış nitel serisi sunulmuştur. Tablo 2.11'e göre bu şehirdeki 161677 GSM abonesinin 37385 tanesi serbest meslek sahibidir.

Meslek Grupları	Frekans
Kamu Çalışanı (Polis, Öğretim Üyesi vb.)	60214
Serbest Meslek Çalışanı (Avukat, Esnaf vb.)	37385
Özel Sektör Çalışanı (Sanayi İşçisi, Maden İşçisi vb.)	64078
Toplam	**161677**

Tablo 2.11
GSM şirketi abonelerinin mesleklere göre gruplandırılmış serisi

Hatırlanacağı gibi, araştırmacılar kimi durumlarda ele alınan değişkenin sonuçları için frekansları kullanmak yerine frekansların toplam frekansa göre oranlarını kullanma eğilimi taşıyabilirler. Nicel değişkenlerde olduğu gibi nitel değişkenler için oluşturulan serilerde de bu oranlar hesaplanabilir. Oranlara ihtiyaç duyulduğunda, nitel frekans serisine oran sütunu eklenir. Her satır veya kategori frekansı toplam frekansa bölünerek ilgili satır veya kategorinin toplam içerisindeki oranı belirlenir. Tablo 2.12'de daha önceden ele alınan anaokulu öğrencilerinin kendilerine sunulan pastaya yönelik lezzet sınıflamaları için oranlar hesaplanarak gösterilmiştir.

Tablo 2.12
Anaokulu öğrencilerinin görüşlerine göre pasta lezzet dağılımı ve oranlar

Pasta lezzeti	**Frekans**	**Oran**
Mükemmel	10	10 / 50=0,20
İyi	18	18 / 50=0,36
Orta	12	12 / 50=0,24
Kötü	6	6 / 50=0,12
Çok kötü	4	4 / 50=0,08
Toplam	**50**	

Tablo 2.12'den yararlanarak örneğin, anaokulundaki öğrencilerin %20'sinin pastayı mükemmel olarak değerlendirdiği, öğrencilerin %36'sının da pastayı iyi olarak değerlendirdiği bilgisi üretilir. Tablo 2.12'ye göre öğrencilerin %8'i de pastayı çok kötü olarak değerlendirmiştir.

ÖRNEK 2

Bir kargo firmasının yerel şube sorumlusu bir gün boyunca şubelerini ziyaret eden 50 müşterinin kargo için yaptıkları ödemeleri (₺ olarak) kayıt altına almıştır. Kayıt edilen ödemeler Tablo 2.13'te verilmiştir.

Tablo 2.13
50 müşterinin kargo için yaptıkları ödeme miktarı değişkeni gruplandırılmış serisi

Ödeme Miktarı	**Frekans**
25 - 35	4
35 - 45	10
45 - 55	11
55 - 65	18
65 - 75	4
75 - 85	3
Toplam	**50**

Oluşturulan bu gruplandırılmış serinin değişkenini tanımlayınız, değişkenin ölçme düzeyini belirtiniz, gruplandırılmış seride oranları hesaplayınız. Gruplandırılmış seriye dayalı olarak ödeme miktarları hakkında genel bilgiler veriniz.

Bu araştırmada kargo için müşteriler tarafından yapılan ödeme, araştırma değişkenidir. Bu değişken sonuçları sayısal değerlerdir. Dolayısıyla araştırma değişkeni nicel bir değişkendir. Ele alınan değişkenin ölçüm değerleri ₺ cinsindendir ve değişken ölçüm sonuçları için anlamlı bir "sıfır" noktası vardır. Bu değişken sürekli nicel bir değişkendir. Gruplandırılmış seri için oranlar her sınıf frekansı toplam frekansa bölünerek hesaplanır, hesaplanan oran değerleri Tablo 2.14'te sunulmuştur.

Ödeme Miktarı	Frekans	Oran
25 - 35	4	4 / 50 = 0,08
35 - 45	10	10 / 50 = 0,20
45 - 55	11	11 / 50 = 0,22
55 - 65	18	18 / 50 = 0,36
65 - 75	4	4 / 50 = 0,08
75 - 85	3	3 / 50 = 0,06
Toplam	**50**	

Tablo 2.14
Ödeme değişkeni için gruplandırılmış seri ve oranlar

Tablo 2.14'e göre ödeme yapan müşterilerin %36'sı 55 ile T*65 arası bir ödeme yapmaktadır. En az ödeme frekansı 75 - 85 sınıfı için gerçekleşmiştir ve 3 değerine eşittir, bu sınıfın genel içerisindeki oranı ise %6'dır.*

Zaman Serisi

Zamana bağlı bir değişkenin zaman değişkeninin uygun ölçme düzeylerine göre aldığı değerlerin alt alta (kronolojik olarak) sıralanmasıyla oluşturulan serilere zaman serisi denir.

Zaman serileri iki sütun halinde oluşturulur. Birinci sütunda zaman değişkeninin ölçme düzeyleri yer alır. Bu ölçme düzeyleri gün, hafta, ay, çeyrek yıl, mevsim, yıl ve benzeri gibi olabilir. İkinci sütunda ise ilgilenilen bir değişkenin zaman değişkeninin bu ölçme düzeylerine karşı gelen değerleri sıralanır. Zaman serileri ilgilendiğimiz serinin gözlem değerleri üzerinde etkili olan zaman serisi bileşenlerinin belirlenebilmesi ve bu değişkene ilişkin öngörülerin yapılabilmesi için oluşturulur. Zaman serisi bileşenleri dört başlık altında incelenebilir. Bunlar *trend, mevsimsel dalgalanma, konjonktürel dalgalanma ve düzensiz dalgalanmalardır.*

Trend, zaman serisi gözlem değerlerinde uzun zaman dönemi içerisinde artma ya da azalma yönündeki genel eğilim olarak tanımlanır. Eğilim, artma yönünde ise *artan trend*, azalma yönünde ise *azalan trend* adını alır. Trend, zamana bağlı değişkenin zaman içerisinde ekonomik, sosyal, kültürel, teknolojik, nüfus artışı ve benzeri gibi faktörlerin etkisi ile ortaya çıkan bir genel eğilimdir.

Mevsimsel dalgalanma, coğrafi anlamda mevsimin etkisiyle zamana bağlı değişkenin aldığı değerlerin birbirini izleyen yılların aynı aylarında veya aynı mevsimlerinde periyodik ve devri olarak bir artma bir azalma şeklindeki genel eğilimidir. Yaz ayarında ülkemize gelen turist sayısının artması, kış aylarında azalması buna örnek gösterilebilir. Bir alışveriş merkezinin cumartesi günleri satışlarının artması, hafta içi günlerde azalması şeklindeki eğilim de bu anlamdaki bir dalgalanma için örnek teşkil edebilir. Bu kez dalgalanma mevsime bağlı değildir ancak mevsimsel dalgalanmada olduğu gibi periyodik ve devri özellik taşır. Bu örnekte sözü edilen türden dalgalanmaya düzenli dalgalanma adı verilmektedir.

Zamana bağlı değişkenin gözlem değerlerinde devri fakat periyodik olmayan türden dalgalanmalara konjonktürel dalgalanma adı verilir. Zaman içerisinde ilgilendiğimiz serinin gözlem değerleri üzerinde ekonominin yükselme ve daralma durumunda meydana gelen periyodik olmayan devri nitelikte artma ve azalma yönündeki dalgalanmalara konjonktürel dalgalanma adı verilir. Bu dalgalanmaya dünyada yaşanan finansal kriz nedeniyle üretimin düşmesi, işsizliğin artması örnek olarak gösterilebilir.

Tablo 2.15
Türkiye 2010 yılı aylara göre gerçekleşen doğum sayısı zaman serisi

Kaynak: *Türkiye İstatistik Kurumu 2010 Yıllığı*

Aylar	Doğum Sayısı (2010)
Ocak	117412
Şubat	93410
Mart	101903
Nisan	97849
Mayıs	101953
Haziran	107591
Temmuz	111721
Ağustos	111187
Eylül	108296
Ekim	102121
Kasım	99718
Aralık	85809
Toplam	**1238970**

Düzensiz dalgalanma, zamana bağlı değişkenin gözlem değerleri üzerinde yangın, deprem, sel baskını, grev, lokavt ve benzeri gibi tesadüfi sebeplerin etkisiyle meydana gelen dalgalanmalardır. Gözlem değerleri üzerinde etkili olan bu faktörler birbirlerinin etkilerini yok edebilir.

Örneğin, Tablo 2.15'te Türkiye İstatistik Kurumu istatistiklerine göre 2010 yılı için Türkiye'deki aylara göre doğum sayıları zaman serisi yer almaktadır. Seri zaman değişkeninin ay ölçme düzeyi için oluşturulduğudan aylık zaman serisi adı verilir. Tablo 2.15'ten yararlanarak, Türkiye'de en çok doğumun Ocak ayında, en az doğumun ise Aralık ayında gerçekleştiği bilgisi üretilebilir.

Mekân Serisi

İlgilenilen değişkenin sahip olduğu değerlerin mekâna göre sınıflandırılması ile elde edilen seri, mekân serisi olarak adlandırılır. Mekân serilerinde değişkenin ortaya çıkan sonuçları mekân değişkeninin ölçme düzeyleri itibari ile gösterilir. Bu serilerin ilk sütununda mekân değişkeninin ölçme düzeyleri ikinci sütunda ise bu ölçme düzeyine karşılık gelen değerlere yer verilir. Mekân serilerinde ilk sütunda yer alan mekân değişkenin ölçme düzeyleri ülkeler, şehirler, istatistiki bölge birimleri ve benzeri gibi kategoriler kullanılabilir.

Örneğin, Tablo 2.16'da Türkiye'de 2010 yılında gerçekleşen 1238970 doğumun, Türkiye İstatistik Kurumu istatistiki bölge birimleri sınıflamasına göre dağılımını gösteren mekân serisi yer almaktadır.

Tablo 2.16
Türkiye 2010 yılı gerçekleşen doğum sayısının istatistiki bölgelere göre dağılımı mekân serisi

Kaynak: *Türkiye İstatistik Kurumu 2010 Yıllığı*

İstatistiki Bölge	Doğum Sayısı (2010)
TR1 İstanbul	211088
TR2 Batı Marmara	35898
TR3 Ege	126678
TR4 Doğu Marmara	99395
TR5 Batı Anadolu	105815
TR6 Akdeniz	163969
TR7 Orta Anadolu	63283
TR8 Batı Karadeniz	60877
TR9 Doğu Karadeniz	33980
TRA Kuzeydoğu Anadolu	49291
TRB Ortadoğu Anadolu	83300
TRC Güneydoğu Anadolu	205396
Toplam	**1238970**

Tablo 2.16'dan yararlanarak, 2010 yılı içerisinde Türkiye'de en çok doğum sayısının İstanbul'da gerçekleştiği; İstanbul'u Güneydoğu Anadolu istatistik bölgesinin izlediği söylenebilir. Tablo 2.16'ya göre en az doğum sayısı ise Doğu Karadeniz istatistik bölgesine aittir.

Birikimli Seri

Bir seride *belirli bir değerden daha küçük* veya *belirli bir değerden daha büyük* değerli kaç tane gözlem değeri (terim) olduğu bilgisini üreten serilere birikimli seriler adı verilir. Birikimli seriler frekans dağılımında her sınıfın frekansına kendinden önceki veya sonraki frekansların eklenmesi ile oluşturulurlar. Birikimli seriler küçükten büyüğe veya büyükten küçüğe doğru oluşturulabilirler. Belirli bir değerden daha küçük terim sayısını gösteren birikimli frekanslara *-den az* birikimli frekansları, belirli bir değerden daha büyük terim sayısını gösteren birikimli frekanslara da *-den çok* birikimli frekansları adı verilir. Birikimli seriler hem frekans serileri hem de gruplandırılmış seriler için oluşturulabilir. Birikimli serilerde 3 adet sütun yer alır; İlk sütunda sınıflar veya gözlem değerleri yer alır, ikinci sütunda her sınıfın veya gözlem değerlerinin frekansları yer alırken üçüncü sütunda ise istenen birikimli frekans değerleri (-den az veya -den çok) bulunur.

-den az birikimli frekans değerleri hesaplanırken işlemlere serinin ilk satırından başlanır. -den az sütununun birinci satırında ilgili satırın frekans değeri yer alır. İkinci satırda ise birinci satırda yer alan -den az değerinin ikinci satır frekansı ile toplamı yer alır. Bu toplama işlemine serinin son satırı da oluşturuluncaya kadar devam edilir. *İşlemin doğruluğunun kontrolü için toplam frekans ile -den az sütununun son satır değerinin aynı olup olmadığına bakılmalıdır.*

-den çok birikimli frekans değerleri hesaplanırken işlemlere serinin en son satırından başlanır. -den çok sütununun en son satırında ilgili satırın frekans değeri yer alır. Daha sonra son satırdan bir önceki satıra geçilir ve son satırın frekansı ile bir önceki satırın frekansı toplanarak bu satıra yazılır. Benzer şekilde ilk satıra ulaşılıncaya kadar toplama işlemine devam edilir. *İşlemin doğruluğunun kontrolü için toplam frekans ile -den çok sütununun ilk satır değerinin aynı olup olmadığına bakılmalıdır.*

Birikimli frekans serilerini oluşturulmasını bir örnek yardımıyla inceleyebiliriz. Örneğin, Tablo 2.17'de ikinci el araç satımı yapan bir firmanın son bir ay içerisinde gerçekleştirdiği ikinci araç satışlarının, satılan aracın değeri değişkenine göre dağılımını gösteren frekans serisi yer almaktadır.

Tablo 2.17
İkinci el araç satış fiyatları değişkenine göre araç satış miktarları frekans serisi

Araç Değeri (₺)	Frekans
15000	5
17000	15
18500	10
20500	8
25000	6
28500	3
Toplam	**47**

Tablo 2.17'de yer alan frekans serisi için -den az ve -den çok birikimli frekans değerleri hesaplanabilir. Bu amaç ile frekans serisine -den az ve -den çok sütunları eklenir. -den az sütununun ilk satırında, bu satırın frekans değeri olan 5 yer alır. İkinci satır ise ilk satır frekansı olan 5'e ikinci satır frekansı olan 15'in eklenmesi ile bulunan 20 değerine sahiptir. Benzer şekilde diğer satırlar da oluşturulur. -den çok sütununun son satırında, bu satırın frekans değeri olan 3 yer alır. Son satırdan bir önceki satırda ise, son satır frekansı olan 3'e bir önceki satır frekansı olan 6'nın eklen-

mesi ile bulunan 9 değeri yer alır. Benzer şekilde ilk satırda toplam frekansa ulaşılıncaya kadar -den çok satırları oluşturulur.

Tablo 2.18
İkinci el araç satış fiyatları değişkenine göre araç satış miktarları birikimli serileri

Araç Değeri (₺)	Frekans	-den az	-den çok
15000	5	5	5+42=47
17000	15	15+5=20	15+27=42
18500	10	10+20=30	10+17=27
20500	8	8+30=38	8+9=17
25000	6	6+38=44	6+3=9
28500	3	3+44=47	3
Toplam	**47**		

Tablo 2.18'in -den az sütunundan yararlanarak ₺20500 ve altında fiyata satılan araç sayısının 38 olduğu, ₺17000 ve altında fiyata satılan araç sayısının ise 20 olduğu söylenebilir. Tablo 2.18'in -den çok sütunundan yararlanarak ₺20500 ve üzerinde fiyata satılan araç sayısının 17 olduğu, ₺17000 ve üzerinde fiyata satılan araç sayısının ise 42 olduğu söylenebilir.

Gruplandırılmış serilerde de birikimli seri oluşturulabilir. Gruplandırılmış serilerde birikimli seri oluşturma işlemleri, frekans serileri ile aynı olmakla birlikte ele alınan değişkenin kesikli veya sürekli olması durumuna göre -den az ve -den çok serilerinin yorumlamalarında dikkatli olunması gerekir. Sürekli nicel değişkenlerde sınıf alt sınırları yazılı değerleri kapsarken sınıf üst sınırları yazılı değerleri değil bu değere en yakın gözlem biriminin değerini kapsamaktadır. Kesikli nicel değişkenlerde hem sınıf alt sınırları hem de sınıf üst sınırları yazılı değerleri kapsar. Yorumlamalarda bu durum göz önüne alınmalıdır.

Sürekli nicel değişken için gruplandırılmış serilerde birikimli serilerin doğru yorumlanmasına örnek olması için daha önce Tablo 2.13'te verilen kargo ödeme miktar (₺ olarak) değişkeni gruplandırılmış serisi ile -den az ve -den çok birikimli frekansları Tablo 2.19'da gösterilmiştir.

Tablo 2.19
Müşterilerin kargo ödeme miktarı değişkeni için gruplandırılmış ve birikimli frekans serileri

Ödeme Miktarı (₺)	Frekans	-den az	-den çok
25 - 35	4	4	4+46=50
35 - 45	10	10+4=14	10+36=46
45 - 55	11	11+14=25	11+25=36
55 - 65	18	18+25=43	18+7=25
65 - 75	4	4+43=47	4+3=7
75 - 85	3	3+47=50	3
Toplam	**50**		

-den az sütununun ilk satırı bu satırın frekansı olan 4 ile temsil edilirken ikinci satır ilk satır -den az değeri olan 4 ile ikinci satır frekansı 10'un toplamı olan 14 ile temsil edilmektedir. -den çok sütununun son satırında bu satırın frekansı olan 3 yer alırken son satırdan bir önceki satırda ise son satır frekansı olan 3'e bir önceki satır frekansı olan 4'ün eklenmesi ile bulunan 7 değeri yer alır. Tablo 2.19'da yer alan -den az ve -den çok değerleri yorumlanırken, değeri ₺65'den daha az ödeme yapan müşteri sayısı -den az sütunu yardımıyla 43 kişi olarak tespit edilir. Burada dikkatli olunması gereken nokta bu 43 müşteri arasında tam olarak ₺65 ödeme yapan müşterilerin bulunmadığıdır. Tablo 2.19'un -den çok sütununa göre değeri ₺45'den daha çok ödeme yapan müşteri sayısı 36'dır. Tam olarak ₺45 ödeme yapan müşterilerde bu 36 müşteri arasındadır.

Kesikli nicel değişken için gruplandırılmış serilerde birikimli serilerin doğru yorumlanmasına örnek olması için daha önce Tablo 2.6'da verilen işçilerin çocuk sayısı değişkeni gruplandırılmış serisi ile -den az ve -den çok birikimli frekansları Tablo 2.20'de gösterilmiştir.

Çocuk Sayısı	Frekans	-den az	-den çok
0 - 2	20	20	20+100=120
3 - 5	65	65+20=85	65+35=100
6 - 8	20	20+85=105	20+15=35
9 - 11	12	12+105=117	12+3=15
12 - 14	3	3+117=120	3
Toplam	**120**		

Tablo 2.20
İşçilerin çocuk sayısı değişkeni için gruplandırılmış ve birikimli seriler

Frekans serileri ve sürekli nicel değişken gruplandırılmış seriler için yürütülen birikimli seri hesaplamalarında yapılan işlemler, kesikli nicel değişken birikimli serileri oluşturulurken de kullanılır. Kesikli değişken gruplandırılmış serileri için birikimli seriler yorumlanırken gruplandırılmış seride yer alan sınıfların yazılı alt ve üst sınır değerlerinin sınıf aralığı içerisinde yer alan birimlerin değerlerini temsil etmeleridir. Örneğin Tablo 2.20'ye göre, 6 - 8 sınıfı içerisinde 6, 7 ve 8 çocuklu işçiler bulunmaktadır. Dolayısıyla bu sınıfa karşılık gelen -den az frekans değeri yorumlanırken, "bu tabloya göre seride 8 ve 8'den daha az çocuğa sahip işçi sayısı 105'dir" ifadesi kullanılmalıdır. Eğer bu sınıf sürekli bir değişken için oluşturulmuş olsaydı hatırlanırsa kullanılması gereken ifadenin "Bu tabloya göre 8'den daha az çocuğa sahip işçilerin sayısı 105'dir." şeklinde söylenmesi gerekirdi. 6-8 sınıfının -den çok değerine göre ise bu veri setinde 6 ve 6'dan daha çok çocuğa sahip işçi sayısı 35'dir.

DİKKAT

Kesikli nicel değişkenler için düzenlenmiş gruplandırılmış serilerde sınıf alt ve üst sınır değerleri yazılı değerleri kapsamaktadır.

İSTATİSTİKSEL SERİLERİN GRAFİKSEL ÇÖZÜMLEMESİ

İstatistiksel bilgilerin serilerin yanında görsel olarak sunulması da mümkün olmaktadır. Gözlem değerlerinin matematiksel ve bilimsel temellere sahip resim, şekil veya çizgilerle gösterimine grafik adı verilir. Pek çok kişi, derlenen verilerin kendilerine seriler şeklinde sunulması yerine, grafiklerle sunulması durumunda daha kolay, daha ayrıntılı ve net bilgiyi elde edebildiklerini belirtmektedirler. Grafikler olayların doğasını seriler ve tablolar yardımıyla değil, resimler, şekiller veya çizgilerle karşı tarafa aktarır, ayrıca sunumun teknik detaylarından çok ana fikir üzerinde yoğunlaşılmasına yardımcı olurlar. Grafikler oluşturulurken açık ve anlaşılır olmalarına gayret gösterilmeli ve hedef kitleyi yanıltıcı özellikler taşımamalıdır. Grafikte yer alan bütün bileşenlerin doğru şekilde ölçeklenip etiketlendiğine dikkat edilmelidir. Eğer grafik bir başka kaynaktan elde edildi ise orijinal kaynak bilgisine de grafikte yer verilmelidir.

İstatistiksel verilerin ve analiz sonuçlarının grafiksel gösterimi için çok sayıda farklı teknik bulunmaktadır. Bu ünitenin izleyen kısımlarında bazı grafik teknikleri üzerinde durulmuştur.

Nicel Değişkenlerin Grafiksel Gösterimi

Nicel veri çalışmalarında çeşitli deney ya da ölçümler ile elde edilmiş rakamlar ile çalışılır. Nicel verinin grafiksel gösterimi için sıklıkla kullanılan grafik teknikleri; histogram, frekans poligonu ve birikimli frekans poligonu'dur.

Histogram

Frekans serilerinin grafiksel gösteriminde sıklıkla kullanılan teknik histogram grafiğidir. Alanları frekanslar ile gösterilen dörtgenlerin yanyana sıralanması ile ortaya çıkan kapalı şekile histogram adı verilir (Pearson, 1895). Histogram, çeşitli sınıf ya da aralıklarda yer alan frekansların sütunlar ile temsil edilmesidir. Histogram, sürekli nicel değişkenler ile ilgili betimsel bilgi üretmek amacıyla kullanılır.

Gruplandırılmış seriler için histogram çizerken frekans dağılımının düşey ekseninde frekans değerleri ya da oransal frekanslar, yatay eksende ise dağılımın sınıfları yer alır ve sınıf aralığı bir birim uzunluk olarak tanımlanır. Gruplandırılmış serinin her sınıfı, histogramda bir sütun ile temsil edilir.

Bir gruplandırılmış serinin histogramının çizilebilmesi için aşağıdaki adımlar izlenir:

- Histogramı çizilecek gruplandırılmış serinin sınıf aralıklarının eşit mi, yoksa farklı mı olduğu araştırılır.
- Sınıf aralıkları eşit ise
  - Gruplandırılmış serinin sınıf aralığı bir birim uzunluk olarak tanımlanır,
  - Her sınıf için tabanı bir birim uzunluk, yüksekliği o sınıfın frekansına eşit olan dörtgenler çizilir. Çizilen dörtgenlerin oluşturduğu kapalı şekil histogram olarak adlandırılır.

Örneğin, Tablo 2.21'de bir firmada çalışan 50 şoförün aylık katettiği mesafe (km olarak) ile ilgili dağılım gruplandırılmış seri olarak düzenlenmiştir.

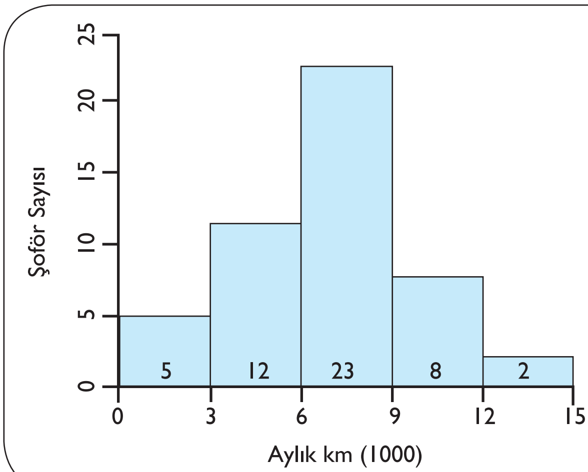
Tablo 2.21
Şoförün aylık katettiği mesafenin dağılımı

Aylık km (1000)	Şoför sayısı	Sınıf Aralığı (h)
0 - 3	5	3
3 - 6	12	3
6 - 9	23	3
9 - 12	8	3
12 - 15	2	3
Toplam	**50**	

Tablo 2.21'de yer alan serinin histogramını çizelim.

- Tablo 2.21 incelendiğinde bütün sınıfların sınıf aralıklarının birbirine eşit ve 3 (1000 km) olduğu görülmektedir.
- Sınıf aralığı h=3 birim uzunluk olarak tanımlanır. Böylece herhangi bir sınıf için çizilecek dörtgenin taban uzunluğu bir birim uzunluk olarak tanımlanmış olur.
- Her sınıf için tabanı bir birim uzunluk, yüksekliği o sınıfın frekansına eşit olan birbirine bitişik dörtgenler çizilir. Meydana gelen şekil histogramdır ve Şekil 2.3'te gösterilmiştir.

Bu şekilde; her sınıf için çizilen dörtgenin alanı o sınıfın frekansına eşit birim karedir. Örneğin birinci sınıfa ait dörtgenin alanı 5 birim kare olarak açıklanır. Şekil 2.3'te gösterilen histogramın alanı toplam frekansa eşittir. Bu seride toplam frekans 50'dir. Şekil 2.3 incelendiğinde örneğin, 3000 km'den az yol yapan şoför sayısı 5; 3000 ve daha fazla, 6000 km'den daha az yol yapan şoför sayısı 12'dir, şoförlerin 23 tanesi 9000 km'den daha az, 6000 km ve daha fazla km yol yapmıştır ve benzeri betimsel bilgiler üretilmiş olur.



Şekil 2.3
Aylık katedilen km (1000) miktarlarına göre şöfor sayıları için histogram

- Sınıf aralıkları eşit değilse;
  - Herhangi bir sınıfın sınıf aralığı, bir birim uzunluk olarak tanımlanır. Genellikle işlem sayısını azaltmak için en çok tekrarlanan sınıf aralığı bir birim uzunluk olarak tanımlanır.
  - Sınıfların sınıf aralıkları, birim uzunluk olarak alınan sınıf aralığına bölünürek her sınıfın sınıf aralığının kaç tane birim alınan sınıf aralığına karşı geldiği hesaplanır (ayarlama oranı). Sınıfların birim sınıf aralığına karşı gelen aralık sayıları hesaplanır.
  - Sınıfların gerçek frekansları, birim sınıf aralığına karşılık gelen aralık sayılarına bölünerek ayarlanmış frekanslar hesaplanır.
  - Her sınıf için tabanı birim sınıf aralığına karşı gelen aralık sayısı olmak üzere, yüksekliği o sınıfın ayarlanmış frekansına eşit olan dörtgenler çizilerek histogram adı verilen grafik çizilmiş olur.

Örneğin, bir banka yöneticisinin oluşturduğu son bir ayda bankasına kumbarasını getiren çocukların kumbaralarındaki paralarının dağılımı Tablo 2.22'de verilmiştir.

Para Miktarı (₺)	Frekanslar	Sınıf Aralığı	Ayarlama oranı	Ayarlanmış Frekans
50 - 60	80	60 - 50 = 10	10 / 10 = 1	80 / 1 = 80
60 - 70	60	70 - 60 = 10	10 / 10 = 1	60 / 1 = 60
70 - 90	40	90 - 70 = 20	20 / 10 = 2	40 / 2 = 20
90 - 110	40	110 - 90 = 20	20 / 10 = 2	40 / 2= 20
Toplam	**220**			

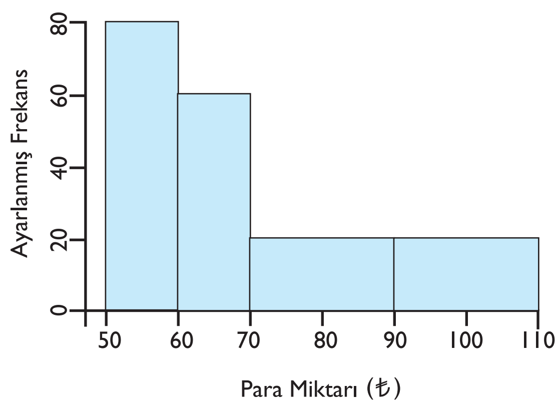
Tablo 2.22
Kumbara içinden çıkan para miktarı değişkeni için gruplandırılmış seri ve ayarlanmış frekans

Tablo 2.22'de yer alan gruplandırılmış serinin grafiğini çizelim.

- Tablo 2.22 incelendiğinde, 1 ve 2 numaralı sınıfların sınıf aralıkları 10; 3 ve 4 numaralı sınıfların sınıf aralıkları ise 20'dir. Bu bilgilere göre bu gruplandırılmış seri eşit olmayan aralıklı gruplandırılmış seridir. Eşit olmayan gruplandırılmış serilerin histogramını çizebilmek için serinin ayarlanmış frekanslarının hesaplanması gerekir.
- Tablo 2.22 incelendiğinde, 10 ve 20 sınıf aralıklarının ikişer kez tekrarlandıkları görülmektedir. 10 ve 20 sınıf aralıkları eşit sayıda olduğu için histogram grafiğinin çiziminde en çok tekrarlanan sınıf bulunmamaktadır. Böyle durumlarda en küçük değerli sınıf aralığı, birim uzunluk olarak tanımlanır. Dolayısıyla bu seri için birim uzunluk değeri 10 alınmıştır.
- Tablo 2.22'nin dördüncü sütununda gösterildiği gibi her sınıfın sınıf aralığı temel olarak alınan sınıf aralığı olan 10'a bölünür. Bulunan 1 değerleri ilgili sınıf aralığının temel alınan sınıf aralığına eşit olduğunu, 2 değerleri ise bu sınıfların sınıf aralıklarının temel alınan sınıf aralığının 2 katı olduğu anlamına gelir.
- Sınıfların gerçek frekansları Tablo 2.22'nin dördüncü sütununda gösterilen ayarlama oranlarına bölünür, böylece elde edilen bölme sonuçları tablonun son sütununda yer alan ayarlanmış frekanslar olarak yer alır.
- Her sınıf için tabanı ayarlama oranı sayısı genişliğinde, yüksekliği ayarlanmış frekanslar ile gösterilen dörtgenler çizilerek histogram adı verilen Şekil 2.4'teki grafik çizilmiş olur.

Şekil 2.4

Kumbara içinden çıkan para miktarı değişkeni için ayarlanmış frekanslara göre histogram.



Şekil 2.4 incelendiğinde ilk iki sınıfın sınıf aralığının 10 üçüncü sınıfın ise 20 değerine eşit olduğu görülebilir. Bu grafik yardımıyla örneğin, 80 çocuğun kumbarasından ₺50'den fazla ₺60'den az para çıktığı, 40 çocuğun kumbarasından ₺90'den az ₺70'den fazla para çıktığı betimsel bilgileri üretilebilir.

Frekans Poligonu

Histogram oluşturan dörtgenlerin tepe orta noktalarının Şekil 2.5'te görüldüğü gibi bir çizgi yardımıyla birleştirilmesiyle oluşturulan grafiğe frekans poligonu adı verilir. Bir serinin histogramının kapalı alanı ile aynı serinin frekans poligonunun (eğrinin altındaki) alanı birbirine eşittir. Bu alan *toplam frekans* ile gösterilir. Şekil 2.5'te yer alan frekans poligonunun alanı toplam frekans sayısı olan 30 birim karedir. Frekans poligonunun yatay eksendeki kesişim noktaları, gruplandırılmış serinin ilk sınıfından önce ve son sınıfından sonra sıfır frekanslı iki adet sanal sınıfın oluşturulması ile belirlenir.

Ağırlık Sınıfları	Frekans	Sınıf Ortası	Koordinat İkilileri
40 - 50	0	45	(45,0)
50 - 60	3	55	(55, 3)
60 - 70	5	65	(65, 5)
70 - 80	5	75	(75, 5)
80 - 90	8	85	(85, 8)
90 - 100	9	95	(95, 9)
100 - 110	0	105	(105,0)
Toplam	**30**		

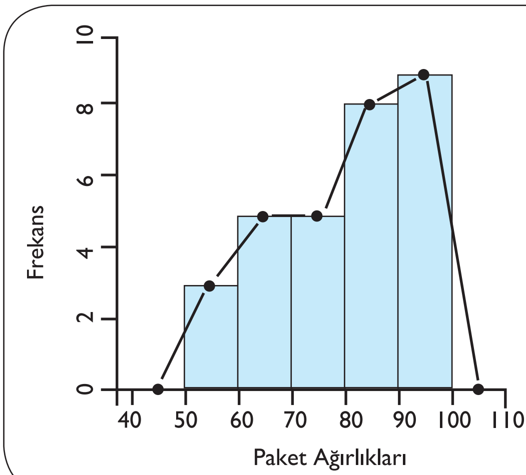
Tablo 2.23
Paket ağırlıkları gruplandırılmış serisi ve frekans poligonu değerleri

Tablo 2.4'te sunulan paket ağırlıkları gruplandırılmış serisi Tablo 2.23'te yeniden sunulmuştur. Frekans poligonunun oluşturulmasında kolaylık sağlaması bakımından tabloya sınıf orta noktası ve poligon için kullanılacak koordinat ikilileri de eklenmiştir.

Frekans poligonunun çizilebilmesi için sınıf orta noktaları yatay eksende ve ilgili frekans değerleri düşey eksende birbirleri ile çizgiler ile bağlanacak şekilde grafiklenir. Tablo 2.23'ün son sütununda gerekli veri eşleşmelerine yer verilmiştir. Şekil 2.5'te Tablo 2.23'te verilen gruplandırılmış seri ve sınıf orta noktalarına göre çizilen frekans poligonu verilmiştir.

Şekil 2.5'te yer alan frekans poligonuna göre gruplandırılmış seride ele alınan değişkenin değeri arttıkça sınıflarda yer alan birim sayısı da artış göstermektedir.

Şekil 2.5



Tablo 2.20'de yer alan gruplandırılmış seri için frekans poligonu

Birikimli Frekans Poligonu

Araştırmacılar, incelenen seride belirli bir değerden daha küçük değerli veya belirli bir değerden daha büyük değere sahip kaç tane gözlem değeri (birim) olduğu bilgisine ihtiyaç duyabilir. İhtiyaç duyulan bu bilgiler, birikimli frekans serilerinin düzenlenmesiyle elde edilebilir. Buna göre, belirli bir değerden daha küçük değere veya büyük değere sahip kaç tane gözlem değeri (birim) olduğu bilgisini üretmeye yarayan serilere birikimli frekans serileri, bu serilerin grafiksel gösterimine birikimli frekans poligonu denir.

Birikimli frekanslar -den az ve -den çok olmak üzere iki şekilde düzenlenirler. -den az frekanslar ve bunların grafiksel gösterimi olan -den az birikimli frekans poligonu, bir seride belirli bir değerden daha küçük değere sahip kaç tane gözlem değeri (birim) olduğu bilgisini üretirken -den çok frekanslar ve bunların grafiksel gösterimi olan -den çok birikimli frekans poligonu, bir seride belirli bir değerden daha büyük değere sahip kaç tane gözlem değeri (birim) olduğu bilgisini üretmektedir. -den az frekans poligonu sürekli artan bir yapıya sahip iken -den çok frekans poligonu sürekli azalan bir yapıya sahiptir.

Birikimli frekans poligonu oluşturmak için önce istenen birikimli -den az veya -den çok birikimli frekans serisi oluşturulur. Eğer -den az birikimli frekans poligonu çizilecek ise ilk sınıfın alt sınırı olan değerin frekansı sıfır kabul edilerek poligon çizimine başlanır. Daha sonra sınıf üst limitleri ile ilgili frekanslar koordinat sisteminde eşleştirilerek çizgiler ile birleştirilir. Eğer -den çok birikimli frekans poligonu çizilecek ise ilk sınıfın alt sınırı ile -den çok birikimli frekansının ilk sınıfı için belirlenen değer eşleştirilerek poligon çizimine başlanır. Daha sonra izleyen sınıfların alt limitleri ile ilgili -den çok frekanslar koordinat sisteminde eşleştirilerek çizgiler ile birleştirilir. Son olarak son sınıfın üst sınır değeri için "sıfır" frekans değeri ataması yapılır.

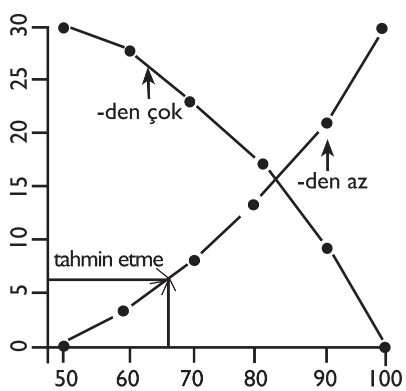
Şekil 2.6'da 30 adet paket ağırlık verisi için birikimli frekans poligonları verilmiştir. Poligonların çizimi için gerekli değerler ise Tablo 2.24'te yer almaktadır.

Tablo 2.24
Paket ağırlıkları gruplandırılmış serisi ve birikimli frekans poligonu hesaplamaları

Ağırlık Sınıfları	Frekans	-den az	-den az ikilileri	-den çok	-den çok ikilileri
			(50, 0)		
50 - 60	3	3	(60, 3)	3 + 27 = 30	(50, 30)
60 - 70	5	5 + 3 = 8	(70, 8)	5 + 22 = 27	(60, 27)
70 - 80	5	5 + 8 = 13	(80, 13)	5 + 17 = 22	(70, 22)
80 - 90	8	8 + 13 = 21	(90, 21)	8 + 9 = 17	(80, 17)
90 - 100	9	9 + 21 = 30	(100, 30)	9	(90, 9)
					(100, 0)
Toplam	**30**				

Şekil 2.6

Tablo 2.21'de yer alan gruplandırılmış seri için birikimli frekans poligonları



Birikimli frekans poligonlarından faydalanarak araştırmacılar gruplandırılmış serinin herhangi bir değerinin frekansını tahmin edebilirler. Bu amaç ile istenen değerden -den az veya -den çok grafiğine kadar ilerlenir ve bu çizgiler ile çakışılan noktadan frekans eksenine dönülür. Örneğin yatay eksende 70 değerinden yukarıya doğru ilerlendiğinde -den az eğrisi ile kesişen noktanın frekansı 70'den daha az birim sayısını verecektir. Şekil 2.6'ya göre 70'den daha az birim sayısı 8'dir. Eğer yatay eksende yer alan 70 değerinden yukarıya doğru -den çok eğrisine kadar ilerlenir ise 70'den daha çok birim sayısının 22 olduğu belirlenebilir. Dolayısıyla yatay eksende yer alan değişkenin herhangi bir değeri için bu değerden daha az veya daha çok kaç adet birim olduğu tespit edilebilir. Örneğin, Şekil 2.6'da 65 değerine karşılık gelen frekans değeri olarak -den az serisine göre 5,5 değeri tespit edilmiştir. Bir başka değişle bu veri setinde 65'den daha az birim sayısı 5,50 (ya da yaklaşık 6) olarak tahmin edilebilir.

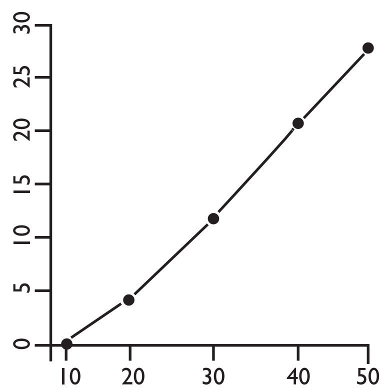
ÖRNEK 3

Tablo 2.25'te verilen gruplandırılmış seri için -den az birikimli frekans poligonunu çiziniz.

Uzunluk Sınıfları	Frekans	-den az
10 - 20	4	4
20 - 30	8	4 + 8 = 12
30 - 40	9	12 + 9 = 21
40 - 50	6	21 + 6 = 27
Toplam	**27**	

Tablo 2.25
Uzunluk değişkeni için gruplandırılmış ve birikimli seri

Tablo 2.25 incelendiğinde üzerinde çalışılan değişkenin sürekli bir değişken olan uzunluk olduğu görülebilir. Tablo 2.23, veri seti için oluşturulan gruplandırılmış seriye ek olarak birikimli frekans serisini de içermektedir. Birikimli frekans poligonu -den az serisi için çizileceğinden grafikte kullanılacak veri ikilileri sırasıyla (10; 0), (20; 4), (30; 12), (40; 21), (50; 27) olacaktır. Bu ikililer kullanılarak oluşturulan birikimli frekans poligonu Şekil 2.7'de verilmiştir.



Şekil 2.7
Tablo 2.25'te yer alan gruplandırılmış seri için -den az birikimli frekans poligonu

SIRA SİZDE 2

50 birimlik bir gruplandırılmış seri aşağıda verilmiştir. -den çok birikimli frekans poligonunu çiziniz.

Sınıflar	Frekans
100 - 110	5
110 - 120	9
120 - 130	17
130 - 140	11
140 - 150	8
Toplam	**50**

Nitel Değişkenlerin Grafiksel Gösterimi

Sütun Grafiği

Üzerinde çalışılan değişken veya değişkenler kategorik sonuçlara sahip olabilir. Örneğin nitel değişken olan cinsiyet değişkenin iki farklı kategorisi vardır. Bu kategoriler kadın ve erkektir. Bir araba parkında yer alan araçların renklerinin dağılımı ile ilgili bir araştırmada da nitel değişken ile çalışılmaktadır. Görüldüğü gibi, nitel değişkenlerin sonuçları sayısal değerler almaz. Bu değişkenlere ilişkin seriler bir niteliği tanımlayan kategoriler ve her kategoriye karşılık gelen frekanslar veya oransal frekanslarla ifade edilirler.

Örneğin, Tablo 2.26'da bir anaokulunda bulunan 50 öğrencinin kendilerine sunulan bir pastanın lezzetine ilişkin görüşleri ile ilgili frekans serisi ve oransal frekans serisi verilmiştir. Bu tabloya göre öğrenciler pastanın lezzetini 5 farklı kategoride derecelendirmektedir. Burada kullanılan değişken, pasta lezzeti değişkeni nitel bir değişkendir ve ölçme düzeyi ise sıralayıcı ölçektir.

Tablo 2.26
Anaokulu öğrencilerinin görüşlerine göre pasta lezzet dağılımı frekans serisi ve oranlar

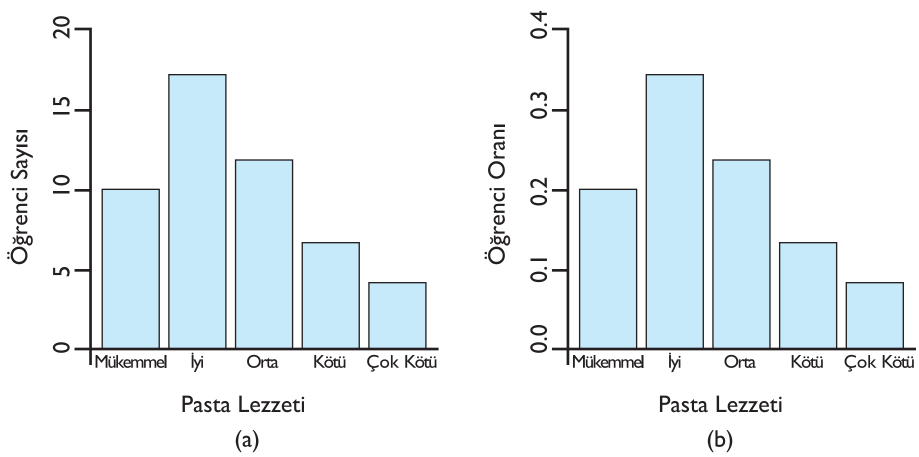
Pasta Lezzeti	**Frekans**	**Oran**
Mükemmel	10	10 / 50 = 0,20
İyi	18	18 / 50 = 0,36
Orta	12	12 / 50 = 0,24
Kötü	6	6 / 50 = 0,12
Çok Kötü	4	4 / 50 = 0,08
Toplam	**50**	

Tablo 2.26'ya göre öğrencilerden 18 tanesi pastayı iyi olarak nitelemiştir. Pastanın lezzetini mükemmel olarak niteleyen öğrenci sayısını, toplam öğrenci sayısı içindeki payı %20'dir, çok kötü olarak niteleyen öğrencilerin oranı ise %8'dir.

Nitel bir değişkenin frekans dağılımının sütun grafiği çizilirken, önce bu dağılımın yatay ekseninde nitel değişkenin kategorileri yan yana gelecek ve birbirinden ayrık şekilde etiketlenir, daha sonra her kategori için yüksekliği frekanslar veya oransal frekanslar ile gösterilen sütunlar çizilir.

Örneğin, Şekil 2.8 (a) ve (b)'de, Tablo 2.26'da verilen nitel frekans ve oransal frekans serileri için çizilen sütun grafikleri verilmiştir.

Şekil 2.8
Anaokulu öğrencilerinin pasta lezzeti için sütun grafikleri



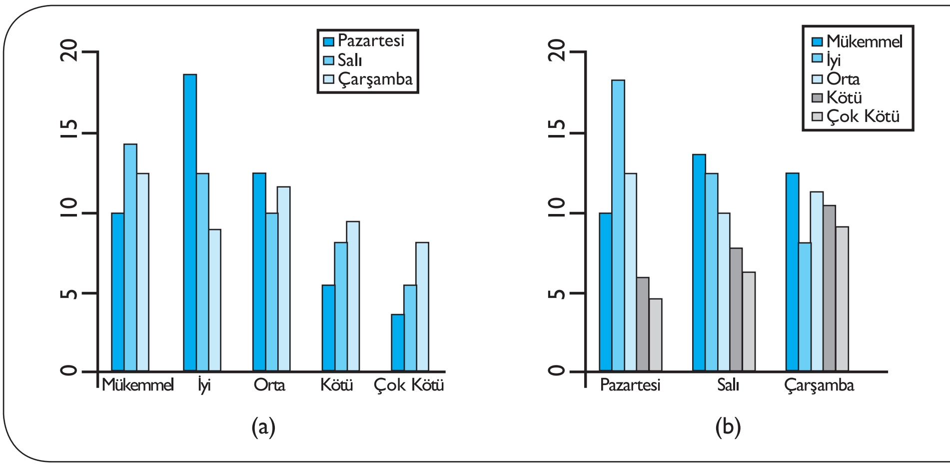
Şekil 2.8 (a) ve (b) incelendiğinde, 50 anaokulu öğrencisinden 10 tanesi (%20 'si) kendilerine sunulan pastanın lezzetini mükemmel olarak değerlendirirken %36'sı bir başka deyişle 18 öğrenci iyi olarak değerlendirmiştir. Pastanın lezzetini çok kötü olarak değerlendiren öğrenci sayısı 4 (%8) olmuştur.

İlgilenilen nitel değişken için bazen birden fazla özellik için değer elde edilebilir. Bu tür durumlarda eşleştirilmiş seriler oluşturulur. Örneğin, Tablo 2.26'da yer alan anaokulu öğrencilerinin kendilerine sunulan pastanın lezzetine ilişkin görüşlerini, 3 gün arka arkaya verilen pastaların lezzetine ilişkin görüşlere genişletebiliriz. Tablo 2.25'te haftanın ilk 3 gününde sunulan pastaların lezzetine ilişkin anaokulu öğrencilerinin görüş dağılımları yer almaktadır.

Tablo 2.27
Anaokulu öğrencileri haftanın ilk 3 günü pasta lezzet eşleştirilmiş serisi

Pasta Lezzeti	Pazartesi	Salı	Çarşamba
Mükemmel	10	14	12
İyi	18	12	9
Orta	12	10	11
Kötü	6	8	10
Çok Kötü	4	6	8
Toplam	**50**	**50**	**50**

Tablo 2.27'de yer alan seri için hem satırlara hem de sütunlara göre sütun grafiği çizilebilir. Satırlara göre sütun grafiği oluştururken öncelikle yatay eksende satırları oluşturan nitel değişken ölçme düzeyleri etiketlenir. Daha sonra yatay eksende yer alan her bir nitel değişken düzeyi için Tablo 2.27'nin sütunlarında yer alan farklı özellik/kategorilere göre sütunlar çizilir. Örneğin, sütun grafiğinin mükemmel ölçme düzeyi için yan yana bitişik 3 adet sütun (pazartesi, salı ve çarşambayı temsil etmek üzere) ilgili frekansları ile orantılı yükseklikte (10, 14, 12) çizilirler. Benzer şekilde diğer ölçme düzeyleri içinde sütun grafiği grupları oluşur. Aynı teknikle sütunlar bir nitel değişkenin ölçme düzeyi olarak ele alınarak da çizim gerçekleştirilebilir. Bu sütun grafikleri Şekil 2.9 (a) ve (b)'de sunulmuştur.



Şekil 2.9
Anaokulu öğrencileri haftanın ilk 3 günü pasta lezzet eşleştirilmiş serisi sütun grafikleri

Şekil 2.9 (a) ve (b)'ye göre öğrencilerin farklı günlerde kendilerine sunulan pastalara ilişkin görüşleri hem günlere göre hem de pastanın lezzet düzeylerine göre ayrı ayrı değerlendirilebilir. Örneğin, Şekil 2.9 (b)'ye göre öğrenciler çarşamba günü yedikleri pastanın lezzet düzeyi hakkında farklı görüşlere sahip olmakla beraber, hemen hemen her düzeyde eşit sayıda öğrenci olduğu görülmektedir. Bu da araştırmacıya çarşamba günü verilen pastanın lezzet düzeyi hakkında öğrencilerin bir birlik içinde olmadığını gösterirken pazartesi günü verilen pasta da ise öğrencilerin büyük bir çoğunluğunun hem fikir oldukları, pastanın lezzetini beğendikleri söylenebilir.

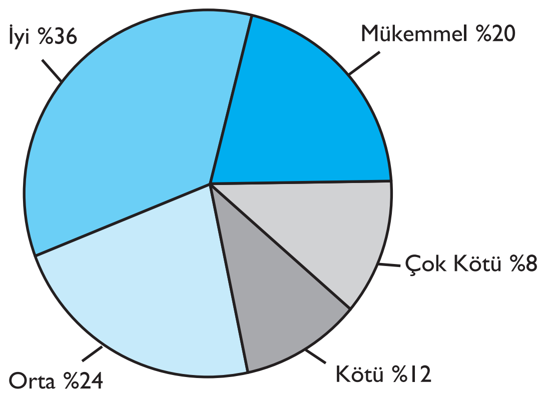
Pasta Grafiği

Pasta grafiği, ilgilenilen nitel değişkenin ölçme düzeylerinin toplam birim sayısına göre paylarını göstermek için bir dairenin dilimlere ayrılması ile oluşturulan grafiktir. Öncelikle nitel frekans serisinde, nitel değişkenin ölçme düzeylerinin toplam içerisindeki oranları, ilgili ölçme düzeyi frekansının toplam frekansa bölünmesi ile elde edilir. Daire her dilim alanı hesaplanan oran ile orantılı olacak şekilde dilimlere ayrılır. Dilimlerin büyüklüğü belirlenirken ilgili ölçme düzeyi için hesaplanan oranların 360 derece ile çarpılmasıyla dairede kapladıkları dilim payları belirlenir. Elle çizimlerinden çok bilgisayar yardımıyla oluşturulmaları beklenir. Şekil 2.10'da Tablo 2.26'da verilen anaokulu öğrencileri pasta lezzeti nitel değişkeni için oluşturulan pasta grafiği yer almaktadır.

Şekil 2.10'da yer alan pasta grafiğine göre en büyük dilim %36'lık bir oran ile nitel değişkenin iyi ölçme düzeyi için gerçekleşirken bu oran değerini %24'lük bir oran ile orta düzeyi izlemektedir.

Şekil 2.10

Anaokulu öğrencilerinin görüşlerine göre pasta lezzet dağılımı pasta grafiği



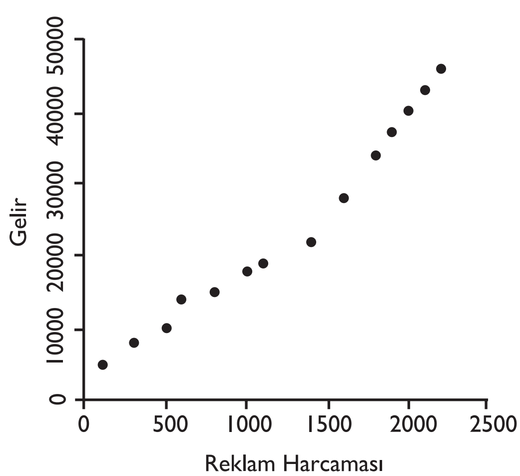
Eşleştirilmiş Verilerin Grafiksel Gösterimi

Serpilme Grafiği

İki nicel değişken için derlenen verilerin bir koordinat sistemi içerisinde eşlenmiş olarak gösterimine serpilme grafiği adı verilir. İki nicel değişken arasında var olabilecek ilişkiyi açıklamak için kullanılan çok faydalı bir grafiksel gösterimdir. Bir serpilme grafiğinde ele alınan bir değişken yatay eksende, diğer değişken ise düşey eksende yer alır. Koordinat sistemi içerisinde araştırmada yer alan her bir terimi temsil etmek üzere, ilgili terimin yatay eksende yer alan birinci değişkene göre aldığı değer ile düşey eksende yer alan ikinci değişkene göre aldığı değer, bu terimin grafik üzerindeki koordinatlarını belirler. Terimlerin her iki değişken için aldıkları değerlere göre ortaya çıkan koordinat de-

Şekil 2.11

Firma reklam harcamaları (₺) ile ürünün satış geliri serpilme grafiği



ğerleri, bir koordinat sistemi içerisinde noktalar veya farklı semboller yardımıyla işaretlenir. Koordinat sistemi içerisinde yer alan bu sembollerin dağılımına (saçılımına) bakılarak iki değişken arasındaki ilişki araştırılır.

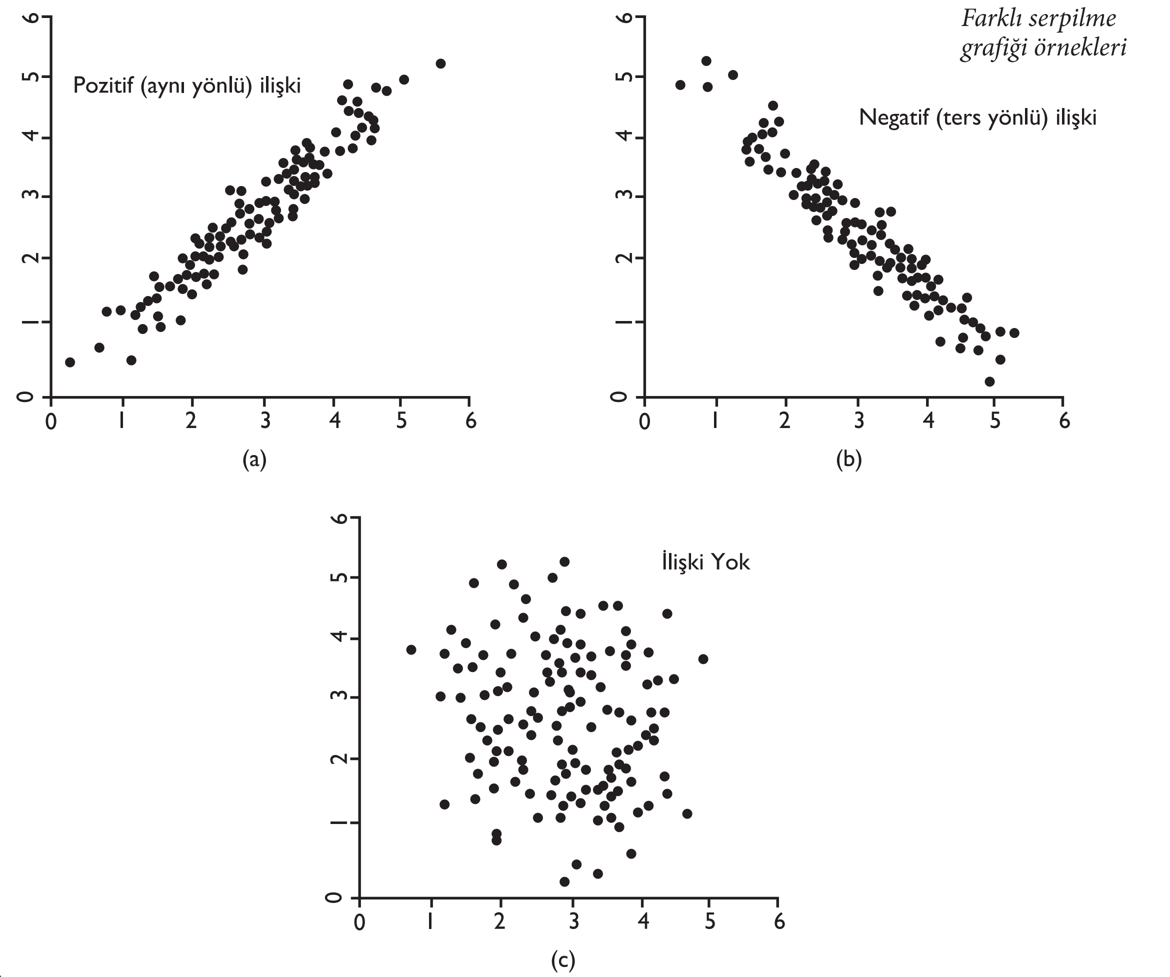
Şekil 2.11'de bir firmanın belirli bir ürün için reklam harcamaları (₺) ile bu ürünün satış geliri (₺) arasındaki ilişkiyi araştırmak üzere oluşturulan serpilme grafiği yer almaktadır.

Şekil 2.11 incelendiğinde reklam harcamaları arttıkça üründen elde edilen gelirin de arttığı gözlemlenmektedir.

İki değişken arasında ilişki araştırılırken ortaya çıkabilecek farklı serpilme grafikleri ve bu grafiklerin değişkenler arasındaki ilişkiler bakımından anlamları Şekil 2.12'de sunulmuştur. Şekil 2.12 (a)'ya göre yatay eksende yer alan değişkenin değerleri arttıkça düşey eksendeki değişkenin değerleri de artmaktadır, bu tür durumlarda iki değişkenin ilişkisi için pozitif (aynı yönlü) ilişki denilir. Şekil 2.12 (b)'ye göre yatay eksende yer alan değişkenin değerleri arttıkça düşey eksendeki değişkenin değerleri azalmaktadır, bu tür durumlarda iki değişkenin ilişkisi için negatif (ters yönlü) ilişki denilir. Şekil 2.12 (c)'de ise hem yatay hem de düşey eksendeki değişkenlere göre gözlemlenen terimlerin rassal olarak dağıldıkları, belirli bir yöne eğilim veya toparlanmayı göstermedikleri için ilgilenilen iki değişken arasında ilişki yoktur denilir.

Şekil 2.12

Farklı serpilme grafiği örnekleri



Zaman Serisinin Grafiksel Gösterimi

Zamanla ilişkili bir değişkenin, zaman değişkeninin uygun gelen ölçme düzeylerine göre aldığı değerleri kronolojik olarak sıralayarak oluşturulan seriye zaman serisi adı verilmektedir. Zaman serileri kartezyen grafik yardımıyla gösterilir. Kartezyen grafiğin yatay ekseninde, zaman değişkenin ölçme düzeyleri yer alırken düşey ekseninde zaman ile ilişkilendirilen araştırma değişkeni yer alır. Kartezyen grafikte zaman değişkeninin uygun ölçme düzeyleri ile zamanla ilişkilendirilen değişkenin aldığı değerler eşleştirilerek bir nokta veya sembol çizilir. Daha sonra bu nokta veya semboller çizgi yardımıyla birleştirilir. Zaman serisi grafiği yardımıyla zaman serisinin durağan, artan, azalan veya mevsimsel dalgalanmaya sahip olup olmadığına dair bilgiler üretilebilir.

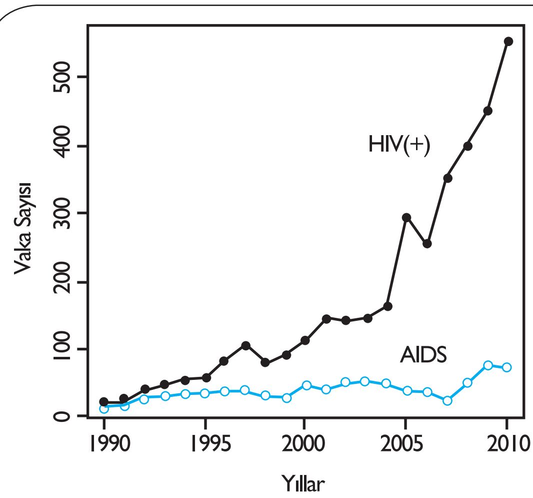
Örneğin, Tablo 2.28'de T.C. Sağlık Bakanlığı istatistiklerinden Türkiye'de yıllara göre HIV/AIDS vaka sayılarının zaman serileri yer almaktadır. Şekil 2.13'te bu veriler için oluşturulan zaman serisi bulunmaktadır. Zaman serisi grafiğinin yatay ekseninde yıllar yer alırken düşey eksende ise vaka sayıları yer alacaktır. Tablo 2.28, bileşik bir zaman serisidir. Dolayısı ile Şekil 2.13'te her iki özellik içinde zaman serisi grafikleri yer almaktadır. Dolayısıyla iki serinin zaman içerisindeki durumları da karşılaştırılabilir.

Tablo 2.28
Türkiye HIV/AIDS vaka sayıları zaman serileri

Kaynak: *T.C. Sağlık Bakanlığı istatistikleri*

Yıllar	AIDS	HIV (+)
1990	14	19
1991	17	21
1992	28	36
1993	29	45
1994	34	52
1995	34	57
1996	37	82
1997	38	105
1998	29	80
1999	28	91
2000	46	112
2001	40	144
2002	48	142
2003	52	145
2004	47	163
2005	37	295
2006	35	255
2007	24	352
2008	49	401
2009	75	453
2010	70	557

Şekil 2.13 incelenirse AIDS vaka sayısı çok az da olsa artan trende sahip zaman serisi iken HIV (+) vaka sayısı ile karşılaştırıldığında trend olmayan bir zaman serisi gibi gözükmektedir. HIV (+) vaka sayısı zaman serisi ise artan trende sahip zaman serisi olarak ortaya çıkmaktadır. Grafik yardımıyla son yıllardaki HIV (+) vaka sayısındaki hızlı artış net bir şekilde görülebilmektedir. Her iki seride de dönemden döneme küçük de olsa alçalma ve yükselme gözlemlenmekle beraber genel eğilim artış yönündedir.



Şekil 2.13

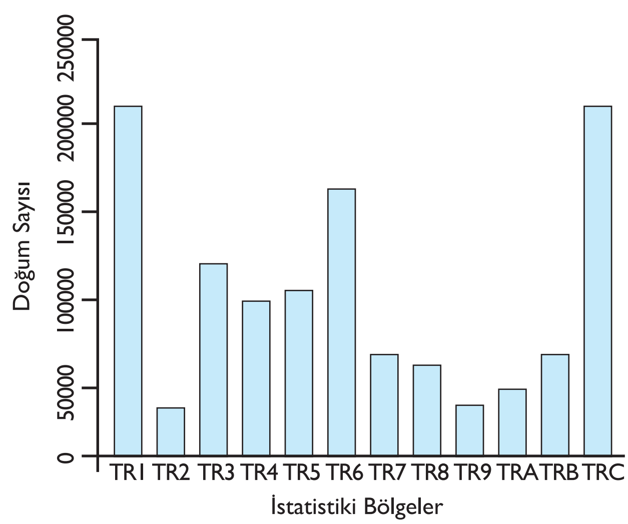
Türkiye HIV/AIDS vaka sayıları zaman serilerinin grafikleri

Kaynak: *T.C. Sağlık Bakanlığı istatistikleri*

Mekân Serisinin Grafiksel Gösterimi

İlgilenilen değişkenin sahip olduğu değerlerin mekâna göre sınıflandırılması ile elde edilen mekân serilerinin grafiksel gösteriminde sütun grafiğinden faydalanılır. Grafik yardımıyla üzerinde çalışılan değişkenin aldığı değerlerin mekân değişkenine göre nasıl bir değişkenlik gösterdiği gözlemlenebilir. Mekân değişkeninin ölçme düzeylerine karşılık gelen seviyeler ya da kategoriler yatay eksende, daha sonra etiketlenen her kategori için frekans ile orantılı yükseklikte sütunlar çizilir.

Örneğin, Şekil 2.14'te Türkiye İstatistik Kurumu istatistiki bölge birimleri sınıflamasına göre dağılımları Tablo 2.16'da verilen doğum sayısı mekân serisi için oluşturulan sütun grafiği yer almaktadır.



Şekil 2.14

Türkiye 2010 yılı gerçekleşen doğum sayısının istatistiki bölgelere göre dağılımı mekân serisi sütun grafiği

Kaynak: *Türkiye İstatistik Kurumu 2010 Yıllığı*

Şekil 2.14'ten faydalanarak en çok doğumun TR1 istatistiki bölgesinde olduğu bilgisi üretilebilir. Bu istatistiki bölgeyi TRC ve TR6 bölgeleri izlemektedir. En az doğumun gerçekleştiği istatistiki bölge ise TR9'dur, ayrıca TR2 istatistiki bölgesi de düşük doğum sayısına sahiptir.

Özet

AMAÇ 1

Verileri derlemek.

Verilerin doğru kaynaktan ve doğru teknikler yardımıyla derlenmesi beklenir. Veri derleme sürecinde araştırma sonuçlarını olumsuz etkileyebileceği düşünülen bütün koşulların en iyi şekilde değrlendirilmesi ve en doğru ve güvenilir bilgiye ulaşılmasına yardımcı olacak verilerin derlenmesi gerekir. İstatistiksel teknikler yardımıyla toplanan veriler uygun bir şekilde düzenlenmelidir. Çoğu zaman araştırmacı verilerini ya bir deney sırasında verilerin ortaya çıkış sırasına göre kayıt altına alabilirken veriler ikincil kaynaklar yardımı ile de derlenebilir. Verilerin araştırmacının kullanabileceği hâle dönüştürülmesi, düzenlenmesi, frekans serilerinin oluşturulması gerekir.

AMAÇ 2

Derlenen verileri düzenleyerek tablolarla göstererek betimsel bilgi üretmek.

Verilerin doğru tablolar ile gösterilerek, tabloyu inceleyen herkes tarafından doğru ve eksiksiz anlaşılması gerekir. Günümüz teknolojisi kullanılarak oluşturulan tabloların çok daha göze hitap eden ve kolay anlaşılır bir yapıda olmasının sağlanması çok da zor değildir. Tabloların doğru ve eksiksiz bir şekilde oluşturulmaları, gerekli olduğu durumlarda kaynakların açıkça belirtilmesi gerektiği de unutulmamalıdır.

AMAÇ 3

Derlenen ve düzenlenen verilenden grafikle göstererek betimsel bilgi üretmek.

Veriler frekans serisi gösterimine ek olarak şekiller, grafikler yardımıyla da temsil edilebilirler. Grafiklerin kullanılması ile yığın bilgi bir anda okuyucu ya da hedef kitleye görsel bir yardımcı yoluyla sunulurlar. Bu görsel yardımcılar grafik olarak adlandırılırlar. Değişken tipine uygun grafik tekniğinin seçilmesi çok önemlidir. Değişken tipi ile kullanılacak grafik tipinin iyi eşleştirilmesi, grafikten beklenen olumlu faydaların artmasına neden olacaktır. Kimi durumlarda grafikler, araştırmacının sayısal tabloları incelerken göremediği veya gözden kaçırdığı kimi özelliklerinde belirlenmesinde fayda sağlayabilir.

Kendimizi Sınayalım

1. Aşağıda verilen seçeneklerden hangisinde yer alan değişken nitel bir değişkendir?

a. Metre olarak uzunluk
b. Ton olarak ağırlık
c. Cinsiyet
d. Yaş
e. Seyahat süresi

2. Aşağıda verilen seçeneklerden hangisinde yer alan değişken sürekli bir değişkendir?

a. Bir otoparktaki araç sayısı
b. Bir sınıftaki öğrenci sayısı
c. Bir evdeki oda sayısı
d. Öğrencinin boyu
e. Bir kutudaki misket sayısı

3. Aşağıda verilen seçeneklerden hangisinde yer alan değişken kesikli bir değişkendir?

a. Evinizin işyerinize olan km cinsi uzaklığı
b. Cebinizdeki metal para sayısı
c. Masanızın alanı
d. Yeni doğan bebek ağırlığı
e. Bir kolinin ağırlığı

4. Aşağıda verilen seçeneklerden hangisinde veri toplama kaynağı **bulunmamaktadır?**

a. Anket
b. Öğrenci kayıt bilgileri
c. Cep telefonu veri tabanı
d. Hane halkı tüketici anketi
e. Karadeniz'deki hamsi balığı sayısı

5. İki nicel değişken için derlenen verilerin bir koordinat sistemi içerisinde eşlenmiş olarak gösteriminde kullanılan grafik aşağıdakilerden hangisidir?

a. Histogram
b. Serpilme diyagramı
c. Sütun grafiği
d. -den az grafiği
e. -den çok grafiği

6. Üzerinde çalışılan özelliğin sonuçları sayısal değerlere sahip değilse araştırma değişkeni için ne ad verilir?

a. Nitel değişken
b. Nicel değişken
c. Sürekli değişken
d. İkili değişken
e. Sonsuz değişken

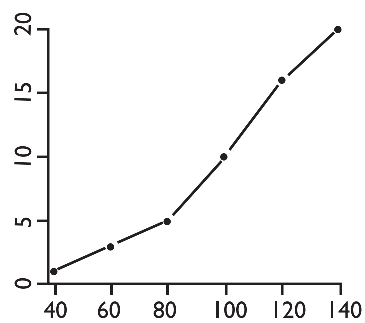
7. İlgilenilen nitel değişkenin ölçme düzeylerinin toplam birim sayısına göre paylarını göstermek için kullanılan grafiğe ne ad verilir?

a. Pasta grafiği
b. Serpilme grafiği
c. Sütun grafiği
d. Histogram
e. Mekân grafiği

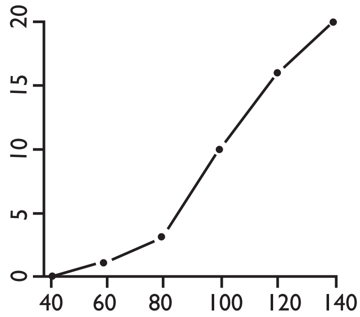
8. Tabloda verilen gruplandırılmış seri için çizilen -den az birikimli frekans poligonu hangisidir?

Sınıflar	Frekans
40 - 60	1
60 - 80	2
80 - 100	7
100 - 120	6
120 - 140	4
Toplam	**20**

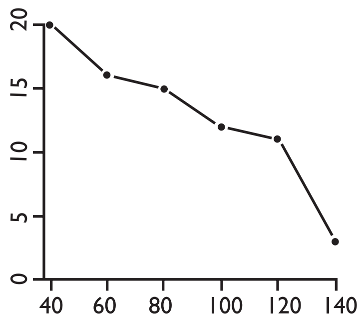
a.



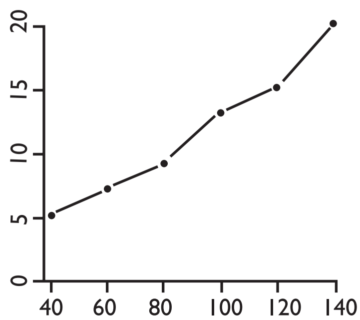
b.



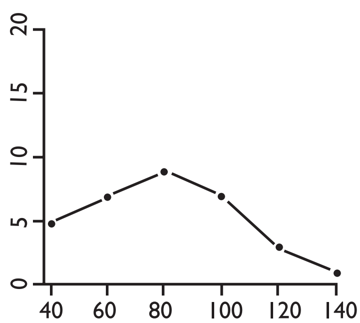
c.



d.



e.



9. Tabloda verilen gruplandırılmış seriye göre değeri 100'den az birim sayısı nedir?

Sınıflar	Frekans
40 - 60	1
60 - 80	2
80 - 100	7
100 - 120	6
120 - 140	4
Toplam	**20**

a. 1
b. 2
c. 3
d. 7
e. 10

10. Aşağıda verilen gruplandırılmış seride 50 ile 60 arasında değere sahip birimlerin oranı nedir?

Sınıflar	Frekans
20 - 30	4
30 - 40	8
40 - 50	14
50 - 60	10
60 - 70	7
Toplam	**43**

a. 0,093
b. 0,186
c. 0,326
d. 0,233
e. 0,163

Kendimizi Sınayalım Yanıt Anahtarı

1. c	Yanıtınız yanlış ise "Değişkenlerin Ölçülmesi" başlıklı konuyu yeniden gözden geçiriniz.
2. d	Yanıtınız yanlış ise "Değişkenlerin Ölçülmesi" başlıklı konuyu yeniden gözden geçiriniz.
3. b	Yanıtınız yanlış ise "Değişkenlerin Ölçülmesi" başlıklı konuyu yeniden gözden geçiriniz.
4. e	Yanıtınız yanlış ise "Veri Derleme" başlıklı konuyu yeniden gözden geçiriniz.
5. b	Yanıtınız yanlış ise "Serilerin Grafiksel Gösterimi" başlıklı konuyu yeniden gözden geçiriniz.
6. a	Yanıtınız yanlış ise "Değişkenlerin Ölçülmesi" başlıklı konuyu yeniden gözden geçiriniz.
7. a	Yanıtınız yanlış ise "Serilerin Grafiksel Gösterimi" başlıklı konuyu yeniden gözden geçiriniz.
8. b	Yanıtınız yanlış ise "Birikimli Frekans Poligonu" başlıklı konuyu yeniden gözden geçiriniz.
9. e	Yanıtınız yanlış ise "Birikimli Frekans Serisi" başlıklı konuyu yeniden gözden geçiriniz.
10. d	Yanıtınız yanlış ise "Birikimli Frekans Serisi" başlıklı konuyu yeniden gözden geçiriniz.

Sıra Sizde Yanıt Anahtarı

Sıra Sizde 1

- Eşit aralıklı ölçeğin sıfır (0) olan ölçme düzeyi yokluk anlamına gelmez. Oysa oranlı ölçekte sıfır (0) ölçme düzeyi yokluk anlamına gelir.
- Veri derlemenin araçları beşeri ve maddi araçlar şeklinde sınıflandırılırlar. Anketör beşeri araçtır, tansiyon ölçme aleti ise maddi araçtır.

Sıra Sizde 2

Öncelikle birikimli frekanslar -den çok değerlerine göre hesaplanır. Daha sonrada veri ikilileri belirlenir. İzleyen tabloda -den çok frekans değerleri sunulmuştur.

Sınıflar	Frekans	-den çok
100-110	5	45 + 5 =50
110-120	9	36 + 9 =45
120-130	17	19 + 17 =36
130-140	11	11 + 8 =19
140-150	8	8
Toplam	**50**	

Grafiği çizmek için gereken veri ikilileri (100;50), (110;45), (120;36), (130;19), (140;8), (150,0) olacaktır. Bu değerler kullanılarak oluşturulan grafik şekilde sunulmuştur.

Yararlanılan Kaynaklar

Ackoff, R. L. (1989). "**From Data to Wisdom**", Journal of Applies Systems Analysis, Volume 16, p 3-9.

Altunışık, R., Coşkun, R., Bayraktaroğlu, S. ve Yıldırım E. (2010). **Sosyal Bilimlerde Araştırma Yöntemleri**, Gazi Kitabevi, Ankara.

Dalgard, P. (2002). **Introductory Statistics with R**, Springer, Germany.

Er, F. (2003). **Açıklayıcı Veri Analizi**, Kaan Kitabevi, Eskişehir.

Freund, J.E. (1974). **Modern Elementary Statistics**, Prentice-Hall Int., London.

Friendly, M. (2000). **Visualizing Categorical Data**, SAS Publishing, SAS, USA.

Gültekin, M. (2001). **Öğretimde Planlama ve Değerlendirme**, Anadolu Üniversitesi Açıköğretim Fakültesi Yayını No: 716, Eskişehir.

Gürsakal, N. (2000). **Sosyal Bilimlerde Araştırma Yöntemleri**, Vipaş Yayınları, Bursa.

Huber, P.J. (2011). **Data Analysis What Can Be Learned from the Past 50 Years**, Wiley, USA.

Lind, D., Marchal, W.G. ve Wathen, S.A. (2005). **Statistical Techniques in Business And Economics**, McGraw-Hill Irwin, USA.

Myatt, G. (2007). **Making Sense of Data**, Wiley InterScience, USA.

Pearson, K. (1895). **"Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogenous Material"**, Phil. Trans. R. Soc. Lond. A 186, 343-414, İngiltere.

R development Core Team (2011). **R: A Language and Environment for Statistical Computing**, Vienna, Austria.

Serper, Ö. ve Gürsakal, N. (1983). **Araştırma Yöntemleri Üzerine**, Filiz Kitabevi, İstanbul.

Serper, Ö. ve Gürsakal, N. (1989). **Araştırma Yöntemleri**, Filiz Kitabevi, İstanbul.

Serper, Ö. (2004). **Uygulamalı İstatistik 1**, Ezgi Kitabevi, Bursa.

Stevens, S.S. (1946). **"On the Theory of Scales of Measurement"**. Science 103 (2684): 677-680.

Stevens, S.S. (1951). Mathematics, measurement and psychophysics. In S.S. Stevens (Ed.), **Handbook of Experimental Psychology** (pp. 1-49). Wiley, New York.

Sönmez, H. (2006). **R yazılımı ile İstatistiksel Analiz**, Kaan Kitabevi, Eskişehir.

Sönmez, H. (2009). **Biyoistatistik**, T.C. Anadolu Üniversitesi Yayınları, Eskişehir.

Sullivan, M. (2005). **Fundamentals of Statistics**, Pearson Prentice Hall, USA.

Sümbüloğlu, V. ve Sümbüloğlu, K. (1998). **Sağlık Bilimlerinde Araştırma Yöntemleri**, Hatiboğlu Kitabevi, Ankara.

Tekin, H. (1993). **Eğitimde Ölçme ve Değerlendirme**, Yargı Yayınları, Ankara.

Özmen, A. (2000). **Uygulamalı Araştırmalarda Örnekleme Yöntemleri**, Anadolu Üniversitesi Yayınları, No: 1257, Eskişehir.

Wallgren, A., Wallgren, B., Persson, R., Jorner, U. ve Haaland, J. (1996). **Graphing Statistics & Data**, Sage Publications, UK.

Yüzer, A.F. (2009). **İstatistik**, Anadolu Üniversitesi Yayınları, Eskişehir.

İSTATİSTİK

3

Amaçlarımız

Bu üniteyi tamamladıktan sonra;

- Ortalamaları sınıflayıp tanımlayabilecek,
- Aritmetik ve diğer duyarlı ortalamaları hesaplayıp açıklayabilecek,
- Medyan ve mod ortalamaları hesaplayıp yorumlayabilecek,
- Standart sapmayı hesaplayıp yorumlayabilecek,

Anahtar Kavramlar

- Ortalama
- Aritmetik ortalama
- Standart sapma
- Dağılma ölçüleri
- Homojenlik (türdeşlik)
- Medyan (ortanca)
- Mod
- Değişim katsayısı
- Asimetri ölçüsü
- Basıklık ölçüsü
- Gözlem değeri
- Değişkenlik ölçüsü

İçindekiler

Ortalamalar, Değişkenlik ve Dağılma Ölçüleri

GİRİŞ

Bu ünitede sayısal veri kümelerinin önemli karakteristiklerini özetleyen ölçüler incelenecektir. Bu ölçüler, tanımlayıcı istatistiklerdir. Bu istatistikler, örneğin; Borsa'da işlem gören bir hisse senedinin yıllık ortalama getirisi nedir? Bu senedin yıl içindeki getiri değişkenliği, dalgalanması nedir? Bir okuldaki öğrencilerin başarı dağılımı nedir? vb. gibi bilgilerin üretilmesi amaçları için kullanılırlar. Bu ünitede incelenecek teknikler, 2-nolu ünitede incelenen tekniklere göre daha nesnel bilgi üretme imkânı vermektedir.

ORTALAMALAR

İstatistik serileri ve grafikler ait oldukları olay hakkında bilgi vermekle birlikte bu bilgilerin nesnelliği düşüktür. Bu nedenle gözlem değerlerinden oluşan veri kümesinin, hangi değer etrafında kümelendiğini gösteren, veri kümesinin özelliğini özetleyen daha nesnel bilgi niteliğindeki tek değerin hesaplanmasına imkân veren ölçülere gereksinim vardır. İşte istatistikte bir seriyi temsil etmeye yarayan bir tek sayısal değere **"Ortalama"** denir. Bir başka ifadeyle ortalama, gözlem değerlerinin merkezinde yer alan ve bu değerleri özetleyebilen sayısal bir değerdir. Bu değer, büyüklüklerine göre sıralanmış bir serinin merkezinde yer alma eğiliminde olduğunda ortalamalarla ilgili genel gösterim;

Ortalama, serinin en küçük gözlem değeri ile en büyük gözlem değeri arasında yer alan, hesaplanabilir bir değerdir.

$X_{min} \leq \text{Ortalama} \leq X_{max}$ şeklinde yazılır.

Bu koşul tüm ortalama türleri için geçerlidir. Burada, X_{min} serideki en küçük değeri, X_{max} ise serideki en büyük değeri gösterir.

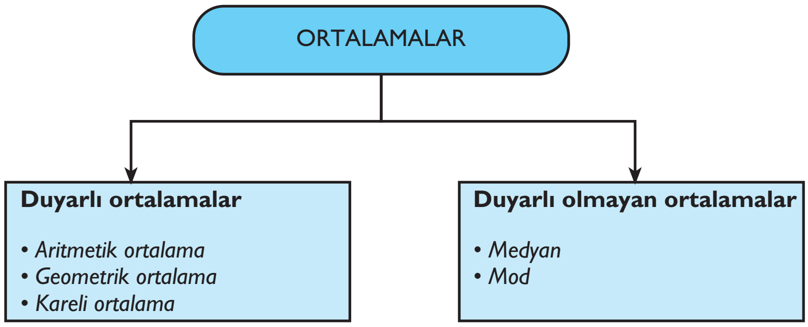
Ortalama değer hesaplamada amaç, seri değerlerini temsil eden, bir değer hesaplamaktır. Hesaplanan ortama değer seri değerlerine büyüklük itibariyle benziyorsa bu ortalama temsilidir denir. Hesaplanan herhangi bir ortalamanın temsili olabilmesi için, serideki gözlem değerlerinin çoğuna yakın değer alması gerekir. Bir seride, **homojenlik (türdeşlik)** arttıkça hesaplanan ortalamanın temsil kabiliyeti artar, homojenlik azaldıkça hesaplanan ortalamanın temsil kabiliyeti azalır. Homojenlik (türdeşlik), gözlem birimlerinin incelenen özellik bakımından bir birine benzer olmasıdır.

Homojenlik, gözlem birimlerinden derlenen verilerin değerce bir birine benzemesi demektir.

Temsili ortalama hesaplamak amacıyla geliştirilmiş olan ortalamalar, duyarlı ve duyarlı olmayan ortalamalar başlıkları altında toplanabilir. Hesaplanan ortalama değer, serideki herhangi bir gözlem değerinde meydana gelen değişiklikten etkileniyorsa ortalama hesaplamak amacıyla kullanılacak ortalama türü, *Duyarlı (Hassas) ortalamalar,* etkilenmiyorsa *Duyarlı (Hassas) olmayan ortalamalar*dır. Bu ünitede incelenecek olan duyarlı ve duyarlı olmayan ortala türleri aşağıdaki Tablo 3.1'de verilmiştir:

Tablo 3.1

Ortalama türleri



SIRA SİZDE 1

Duyarlı ile duyarlı olmayan ortalamalar arasındaki temel fark nedir?

Duyarlı Ortalamalar

Bu grupta yer alan ortalamaların ortak özellikleri;

- Ortalama hesaplanırken serinin bütün gözlem değerleri dikkate alınır.
- Gözlem değerlerinde meydana gelebilecek değişiklikten hemen etkilenirler.

Duyarlı ortalamalar içinde en çok bilinen aritmetik ortalama olmakla birlikte, geometrik ortalama ve kareli ortalama da bu grup içinde yer almaktadır.

Aritmetik Ortalama

Ortalaması hesaplanacak serideki gözlem değerlerinin hepsi eşit önemde ise "Basit Aritmetik Ortalama" *farklı önemde ise* "Tartılı Aritmetik Ortalama" söz konusudur.

Basit Aritmetik Ortalama

Ortalamalar içinde en yaygın biçimde kullanılan ortalama, aritmetik ortalamadır ve ortalama denildiğinde ilk akla gelen ortalama türüdür. Aritmetik ortalama, gözlem değerlerinin cebirsel toplamının gözlem sayısına bölünmesi ile hesaplanır.

Dizilerde aritmetik ortalamanın hesaplanmasında, tanımlanan evren (ana kütle), N-hacimli sonlu bir evren olduğunda ve tamsayım sonucu derlenen veriler X_1, X_2, X_3, ..., X_n olarak gösterildiğinde hesaplanacak aritmetik ortalama, evren aritmetik ortalamasıdır. m (mü) simgesiyle gösterilir ve aşağıdaki formül geçerlidir:

Evrenin aritmetik ortalaması: $\mu = \dfrac{\sum_{i=1}^{N} X_i}{N}$ formülüyle hesaplanır.

Burada, $\sum_{i=1}^{n} X_i$: gözlem değerlerinin (seri değerlerinin) cebirsel toplamını,

N : gözlem sayısını(evren hacmini) gösterir.

Aritmetik ortalamada, n-hacimlik örneklemden elde edilen $X_1, X_2, X_3, ..., X_n$ veri kümesi için hesaplanacak olursa, $\overline{X}$ *(X-ortalama veya X-bar diye okunur)* simgesiyle gösterilir, örneklem aritmetik ortalaması olarak adlandırılır ve aşağıdaki formül yardımıyla hesaplanır:

Örneklem aritmetik ortalaması: $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

ÖRNEK 1

Bir işyerinde çalışan sekiz işçinin, aylık kazançları (₺) aşağıda verilmiştir. Bu işçilerin aylık ortalama kazançlarını hesaplayalım.

X_i $X_1 = 640$ $X_2 = 800$ $X_3 = 860$ $X_4 = 980$ $X_5 = 1120$ $X_6 = 1160$ $X_7 = 1560$ $X_8 = 1680$ $\sum_{i=1}^{8} X_i = 8800$	İş yerinde çalışan toplam sekiz işçinin aylık kazanç değişkeni incelenmiş ve tam sayım uygulanmıştır. Evren hacmi N = 8 işçidir ve gözlem değerleri $\sum_{i=1}^{8} X_i = 8800$ ₺/ay olduğundan evren aritmetik ortalaması: $\mu = \dfrac{\sum X}{N} = \dfrac{8800}{8} = 1100$₺/ay olarak bulunur.

Bu sonuca göre, bu iş yerinde çalışanların aylık ortalama kazançları ₺1100'dir. ₺1100 sıralı gözlem değerlerinin merkezine yakın bir yerdedir ve gözlem değerlerine büyüklük itibariyle benzemektedir. Hesaplanan aritmetik ortalama değeri temsilidir.

ÖRNEK 2

Örnek 1 yeniden ele alınarak bu işyerinden dört çalışandan oluşan bir örneklem oluşturulur ve aylık ortalama kazanç hesaplanmak istenirse hesaplanacak ortalama örneklem ortalaması olup $\overline{X}$ simgesiyle gösterilir ve $\overline{X} = \dfrac{\sum_{i=1}^{4} X_i}{n}$, eşitliği ile hesaplanır.

Farz edelim ki örnekleme 1, 3, 4 ve 6 nolu işçiler seçilmiş olsun. Bu işçilerin aylık kazanç verileri, aşağıdaki gibi derlenmiş olur:

X_i 640 860 980 1160 $\sum_{i=1}^{4} X_i = 3640$	Çalışan işçilerden dört işçinin aylık kazançları gözlenmiştir. Bu nedenle n = 4'dir. Örneklem aritmetik ortalaması: $\overline{X} = \dfrac{\sum X}{n} = \dfrac{3640}{4} = 910$ ₺/ay olarak hesaplanır.

Bir istatistik serisinde, diğer gözlem değerlerine göre, aşırı derecede küçük veya aşırı derecede büyük bir değer varsa buna **uç değer** denir.

Bu sonuca göre, örneklemdeki işçilerin aylık ortala kazançları T910'dir. *Bu değer örnekleme başvurulduğu zaman bilinmeyen m değerine eşit, m değerinden büyük veya küçük olabilir.*

Bazen bir seride, çok küçük ve/veya çok büyük gözlem değerleri bulunabilir. Böyle gözlem değerlerine **uç değer** denir.

ÖRNEK 3

Bir köyde yaşayan aileler arasından seçilen 6 ailenin aylık gelirleri (₺) aşağıdaki gibi olsun. Bu ailelerin aylık gelir ortalaması nedir?

X_i
200
300
400
500
600
8.000
10000

Bu seri için aritmetik ortala hesaplar isek;

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{10000}{6} = 1666,67 \text{ ₺ bulunur.}$$

Hesaplanan ortalama değer, beşinci ve altıncı gözlem değerinin arasında bir değerdir. Gözlem değerlerinin merkezinde değildir. Ayrıca

$\overline{X}$ = 1666,67 değeri serinin ilk beş gözlem değerinden çok büyük, altıncı gözlem değerinden de çok küçüktür. Bu sonuca göre bu özellikteki seriler için aritmetik ortalama, temsili bir ortalama türü değildir. Uç değer içeren seriler için uygun ortama türü ilerleyen kesimlerde incelecektir.

Aritmetik ortalamayla ilgili buraya kadar olan açıklamalarda evren için m, örneklem için sembolleri kullanıldı. İzleyen kısımlarda hem genel kabul görmüş uygulama nedeniyle hem de anlatım ve gösterim kolaylığı sağladığı için aritmetik ortalama hesabında örneklem aritmetik ortalaması ($\overline{X}$) esas alınacaktır.

Frekans serilerinde aritmetik ortalama, frekans serisi şeklinde düzenlenen bölünme serilerinin aritmetik ortalaması, gözlem değerlerinin frekanslarla(sıklıklar) çarpımları toplamının frekanslar toplamına oranı olarak tanımlanır.

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i n_i}{\sum_{i=1}^{n} n_i} \quad \text{formülü ile hesaplanır.}$$

Burada, $\sum_{i=1}^{n} X_i n_i$: gözlem değerleri ile bu gözlem değerlerine karşı gelen frekansların çarpımlarının cebirsel toplamını,

$\sum_{i=1}^{n} n_i$: frekanslar toplamını gösterir.

ÖRNEK 4

Bir işletmenin kargo ile gönderdiği kolilerin ağırlıklarının dağılımı aşağıda verilmiştir. Kargo ile gönderilen kolilerin ağırlıklarının aritmetik ortalaması nedir?

Çözüm: Aritmetik ortalaması hesaplanacak seri, bir frekans serisi olarak düzenlenmiştir. Öncelikle gözlem değerleri ile frekansları tek tek çarpılacak, sonra toplanacaktır.

X_i	n_i	$X_i . n_i$
10	1	10.1 = 10
20	2	20.2 = 40
30	4	30.4 = 120
40	2	40.2 = 80
50	1	50.1 = 50
	10	300

Bu türden düzenlenmiş serilerin aritmetik ortalaması;

$$\overline{X} = \frac{\sum_{i=1}^{10} X_i n_i}{\sum_{i=1}^{10} n_i} = \frac{300}{10} = 30 \text{ Kg}$$

olarak bulunur.

$\overline{X}$ = 30 Kg değeri, X_i gözlem değerlerinin tam ortasında yer alan ve en yüksek frekansa sahip olan, gözlem değerine karşı gelen özet değerdir. $\overline{X}$ = 30 Kg. değeri seriyi temsil eden bir ortalama değerdir.

Gruplandırılmış serilerde aritmetik ortalama; ikinci ünitede ifade edildiği gibi, gözlem değerlerinin sayısı çok fazla olduğunda, verilerin gerek sunum kolaylığı gerekse hesaplama kolaylığı nedeniyle veriler gruplandırılmış seri şeklinde düzenlenir. Gruplandırılmış serilerin aritmetik ortalamasını hesaplayabilmek için önce seri frekans serisine dönüştürülür. Sonra da frekans serisi için aritmetik ortalama hesaplamada kullanılan formül uygulanır.

ÖRNEK 5

Bir ilde faaliyette bulunan 100 firmanın 2010 yılında ödedikleri vergi tutarlarının dağılımı aşağıda verilmiştir. Firmaların ödedikleri 2010 yılı ortalama vergi tutarı nedir?

Çözüm: Firmaların 2010 yılı vergi tutarları gruplandırılmış seri olarak düzenlenmiştir. Ortalama vergi tutarı aşağıdaki adımlar izlenerek hesaplanabilir:

- Gruplandırılmış seri frekans serisine dönüştürülür. Dönüştürme her sınıfın orta noktası hesaplanarak yapılır. Bu değer, sınıf alt ve üst sınırları yardımıyla hesaplanır. Örneğin birinci sınıf: 100 - 200 sınıfı için, (100 + 200) / 2 = 150 (alt ve üst sınırların aritmetik ortalaması) şeklinde hesaplanır. Benzer işlem diğer sınıflar için de yapılırsa aşağıdaki serini üçüncü sütununda yer alan X_i değerleri ele edilmiş olur.

Gruplar (Bin ₺)	n_i	X_i	$X_i . n_i$
100-200	7	150	1050
200-300	18	250	4500
300-400	25	350	8750
400-500	30	450	13500
500-600	20	550	11000
	100		38800

- Frekans serileri için uygulanan aritmetik ortalama hesaplama yaklaşımı uygulanır.

$$\bar{X} = \frac{38800}{100} = ₺\,388 \text{ bin bulunur.}$$

Bu bilgi 100 firmanın 2010 yılı ortalama vergi tutarlarının ₺388 bin olduğu şeklinde yorumlanır.

SIRA SİZDE 2

Gruplandırılmış seri, frekans serisine nasıl dönüştürülür?

Aritmetik Ortalamanın Özellikleri

Aritmetik ortalamanın özellikleri burada diziler için gösterilmiştir:

- Gözlem değerlerinin aritmetik ortalamadan olan sapmalarının matematiksel (cebirsel) toplamı sıfırdır: $\Sigma\,(X_i - \bar{X}\,) = 0$
- Aritmetik ortalamanın, gözlem sayısı ile çarpımı gözlem değerleri toplamını verir: $n\,\bar{X} = \Sigma\,X_i$
- Gözlem değerlerinin aritmetik ortalamadan olan sapmalarının kareleri toplamı $\Sigma\,(X_i - \bar{X}\,)^2$ minimumdur. Yani gözlem değerlerinin, aritmetik ortalama değerinden, farklı c - gibi başka bir değerden sapmalarının karesinden daima daha küçüktür. Bu özellik,

$$c \neq \bar{X} \text{ için } \Sigma\,(X_i - \bar{X}\,)^2 < \Sigma\,(X_i - c)^2 \text{ şeklinde gösterilir.}$$

- Aritmetik ortalama hesabında serideki tüm gözlem değerleri kullanılır.
- Aritmetik ortalama iki veya daha çok kütleyi karşılaştırabilmek için çok yararlı bir ölçüdür.
- Alt ve/veya üst limiti belirsiz olan (açık uçlu) serilerde aritmetik ortalama hesaplanmaz.

Bu özelliklerin ilk üçünü basit bir örnek üzerinde gösterelim.

ÖRNEK 6

Bir lojistik firmasının kış aylarında (üç ay) yol yardımı isteyen araç sayısı aşağıdaki gibi olsun:

X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$	$X_i - 4$	$(X_i - 4)^2$
3	-2	4	-2	4
4	-1	1	-1	1
8	3	9	3	9
15	0	14	0	14

Aritmetik ortalama: $\bar{X} = \frac{15}{3} = 5$ iken

- $\Sigma\,(X_i - \bar{X}\,) = 0$
- $n\bar{X}\;(3.5) = \Sigma\,X_i = 15$
- $\Sigma\,(X_i - \bar{X}\,)^2 = 14$ minimumdur. Burada c = 4 aritmetik ortalamadan farklı bir değer olarak alınmıştır.

Görülebileceği gibi $\Sigma\,(X_i - \bar{X}\,)^2 = 14 < \Sigma\,(X_i - c)^2 = 17$ olduğu açıktır.

Tartılı Aritmetik Ortalama

Bir istatistik serisindeki bazı gözlem değerlerinin önem derecesi farklılık gösterebilir. Ortalama hesabında bu önem derecelerinin hesaba katılmasına imkân veren ortalama, tartılı aritmetik ortalamadır. Gözlem değerlerine verilen önem (tartı, ağırlık), i'inci gözlem değeri için t_i sembolüyle gösterilir. Tartılı aritmetik ortalama ise $\overline{X}_t$ sembolü ile gösterilir.

Diziler için tartılı aritmetik ortalama,

$\overline{X}_t = \dfrac{\sum X_i t_i}{\sum t_i}$ formülü ile hesaplanır.

ÖRNEK 7

Kredili sistemde öğrenim gören bir işletme bölümü öğrencinin almış olduğu dersler, derslerin kredileri değerleri ve öğrencinin derslerdeki başarı puanları (kredili sistemde 4 üzerinden) aşağıda verilmiştir. Öğrencinin ortalama başarı puanı nedir?

Ders	Başarı notu	Kredi değeri
İstatistik	2 (CC)	3
Ticaret Hukuku	4 (AA)	4
Muhasebe Sis.	3 (BB)	3
İşletme Org.	1 (DD)	2
Atatürk İlk. İnk. Tar.	4 (AA)	4

Çözüm: Önce derslerin her birine eşit önem veren basit aritmetik ortalamayı, sonra tartılı aritmetik ortalamayı hesaplayalım ve sonuçları birlikte değerlendirelim.

X_i	t_i	$X_i t_i$
2	3	6
4	4	16
3	3	9
1	2	2
4	4	16
14	16	49

Basit aritmetik ortalama;

$\overline{X} = \dfrac{\sum X_i}{n} = \dfrac{14}{5} = 2{,}8$ puan olur.

Tartılı aritmetik ortalama;

$\overline{X}_t = \dfrac{\sum X_i t_i}{\sum t_i} = \dfrac{49}{16} = 3{,}06$ puan olur.

Görüldüğü gibi tartılı aritmetik ortalamada, ortama değer yüksek kredili derslerin başarı puanına daha yakın çıkmıştır.

Gözlem değerlerinin önemi birbirinden farklı ise ortalama nasıl hesaplanır?

SIRA SİZDE 3

Frekans serilerinde tartılı aritmetik ortalama hesabında; gözlem değerleri frekanslarla ve ayrıca tartılarla çarpımları toplamı, frekanslarla tartıların çarpımları toplamına bölünür.

$$\overline{X}_t = \frac{\sum X_i n_i t_i}{\sum n_i t_i}$$

ÖRNEK 8

A işletmelerindeki saat başı ücret ve işçi sayısı aşağıda verilmiştir. Bu bilgilere dayanarak saat başı ortama ücreti belirleyiniz.

A işletmesi:

Saat Başı Ücret (₺)	İşçi Sayısı	İşçi Kıdemleri
5	20	3
8	12	6
10	8	10

Çözüm: İşçilerin bir işletme için önemi, kıdem olarak ortaya çıkmaktadır. Çünkü bir işletme, ancak iyi bir işçi ile uzun yıllar çalışmak isteyecektir.

A işletmesi için tartılı aritmetik ortalama:

X_i	n_i	t_i	$X_i n_i$	$X_i t_i n_i$	$n_i t_i$
5	20	3	100	300	60
8	12	6	96	576	72
10	8	10	80	800	80
	40		276	1676	212

$$\bar{X}_t = \frac{1676}{212} = ₺\,7,90$$

A işletmesi için saat başı ortalama ücret ₺7,90 olarak bulunur.

Gruplandırılmış serilerde tartılı aritmetik ortalama hesabında önce her sınıfın sınıf ortası bulunarak seri frekans serisine dönüştürülür, böylece frekans serisi için geçerli olan;

$$\overline{X}_t = \frac{\sum X_i n_i t_i}{\sum n_i t_i}$$ formülünden yararlanılır.

Tartılı aritmetik ortalamanın çok geniş uygulama alanları vardır. *Ortalamaların ortalaması, oranların ortalaması ve bileşik indeks* hesabında etkin olarak kullanılmaktadır.

ÖRNEK 9

70 gözlem değerinden oluşan üretim serisi, gruplandırılmış olarak verilmiştir. Verilen gruplandırılmış seriye dikkatli bakılırsa sınıf aralıkları farklıdır. İlk dört sınıfın sınıf aralığı 10 iken sonraki dört sınıfın sınıf aralığı 25'tir.

Gruplar	n_i
10 - 20	3
20 - 30	5
30 - 40	17
40 - 50	15
50 - 75	12
75 - 100	14
100 - 125	3
125 - 150	1

Genel olarak gruplandırılmış serilerin aritmetik ortalama hesabında, *sınıf aralıklarının eşit olması* arzu edilir. Ancak gerçek yaşamda sıkça, sınıf aralıkları farklı olan gruplandırılmış seriler ile karşılaşılır. Böyle serilerin aritmetik ortalaması hesaplanırken seri, sınıf arkalıkları eşit alt *kısmi serilere* ayrılır. Verilen seride de iki farklı sınıf aralığı olduğundan, öncelikle seri, iki kısmi seriye ayrılacak, sonrada ortalama hesabına geçilecektir.

I. Kısmi seri				***II. Kısmi seri***			
Gruplar	n_i	X_i	$X_i n_i$	Gruplar	n_i	X_i	$X_i n_i$
10 - 20	3	15	45	50 - 75	12	62,5	750
20 - 30	5	25	125	75 -100	14	87,5	1225
30 - 40	17	35	595	100 -125	3	112,5	337,5
40 - 50	15	45	675	125 -150	1	137,5	137,5
	40		1440		30		2450,0

$n_1 = 40$ $\qquad\qquad$ $n_2 = 30$

$$\overline{X}_1 = \frac{1440}{40} = 36 \qquad\qquad \overline{X}_2 = \frac{2540}{30} = 84,67$$

Buraya kadar yapılanlar basit aritmetik ortalama hesaplamasıdır. Verilen gruplandırılmış serinin aritmetik ortalaması, kısmi serilerin frekansları tartı kabul edilerek hesaplanacaktır.

$n_1 = t_1 = 40$ $\qquad\qquad$ $n_2 = t_2 = 30$

$\overline{X}_1 = 36$ $\qquad\qquad$ $\overline{X}_2 = 84,67$

$$\overline{X_t} = \frac{\overline{X}_1 t_1 + \overline{X}_2 t_2}{t_1 + t_2} = \frac{36(40) + 84,67(30)}{40 + 30} = 56,86$$

Verilen gruplandırılmış serinin aritmetik ortalaması 56,86 olarak hesaplanmıştır.

Oranları ortalaması hesaplanırken kesrin paydasında bulunan değer tartı olarak kabul edilir. Bunu bir örnekle açıklamaya çalışalım:

ÖRNEK 10

Eskişehir'in Odunpazarı ve Tepebaşı belediye bölgelerinde yapılan bir araştırmada, seçilen 200 ailenin su tüketimleri hakkında bilgi toplanmıştır. Odunpazarı belediye bölgesinden seçilen 80 aile için iki aylık su tüketim ortalaması 40m^3 *olarak hesaplanmıştır. Bu bölgedeki ailelerin %40'ı ortalamanın üzerinde su tüketimi yapmaktadır. Tepebaşı belediye bölgesinden seçilen 120 aile için iki aylık su tüketimi ortalaması 30m*^{3} *olarak hesaplanmıştır. Bu bölgedeki ailelerin %20'si ortalamanın üzerinde su tüketimi yapmaktadır. Bu bilgiler ışığında;*

a. 200 ailenin iki aylık ortalama su tüketimi nedir?

b. İki bölgedeki ortalama su tüketiminin üzerinde tüketim yapanların oranları ortalaması nedir?

Çözüm:

a. Bu örnekte iki bölgenin ortalaması verilerek 200 ailenin tamamı için ortalama hesabı istenmektedir. *Ortalamaların ortalaması tartılı ortalama ile hesaplanacağı* daha önce belirtilmişti. Böylece;

$\bar{X}_i$	t_i	$\bar{X}_i \cdot t_i$	
$\bar{X}_i = 40$	80	3200	$\overline{X_t} = \dfrac{\sum \overline{X_i} t_i}{\sum t_i} = \dfrac{6800}{200} = 34m^3$
$\bar{X}_i = 30$	120	3600	
	200	6800	Aile başına ortalama su tüketimi 34 m^3'tür.

b. Bu şıkta ise ortalama üzerinde su tüketenlerin *oranlarının ortalaması* istenmektedir.

X_i	t_i	$X_i \cdot t_i$	
0,40	80	32	$\overline{X_t} = \dfrac{\sum \overline{X_i} t_i}{\sum t_i} = \dfrac{56}{200} = 0,28$
0,20	120	24	
	200	56	Aile başına ortalama su tüketimi %28'dir.

Geometrik Ortalama

Geometrik ortalama GO simgesiyle gösterilir. Geometrik ortalama ile ilgili *iki tanım* yapılmaktadır:

i. Geometrik ortalama; gözlem değerleri çarpımının gözlem sayısına eşit dereceden köküne eşit olan ortalamadır.

$$G.O. = \sqrt[n]{(X_1).(X_2)...(X_n)}$$

ÖRNEK 11

Aynı bölgede bulunan dört kömür madeninde rezervler 2001 yılından bu güne kadar sırasıyla 3, 2, 4 ve 6 kat(misli) artış göstermiştir. Buna göre dört kömür madeni için rezerv artış oranlarının geometrik ortalaması nedir?

Çözüm: Gözlem değerlerine dikkatle bakıldığında gözlem sayısı az ve gözlem değerleri büyük sayılar değildir. Geometrik ortalama formülü kullanıldığında,

$$G.O. = \sqrt[4]{(3).(2).(4).(6)} = \sqrt[4]{144} = (144)^{1/4} = 3,46 \text{ olarak bulunur.}$$

Buna göre dört kömür madeninde rezerv artış ortalaması 3,46 misli olmuştur.

Gözlem sayısı az ve gözlem değerleri küçük değerlerden oluştuğunda, geometrik ortalamanın bu tanımı bağlamında, hesaplanması kolaydır. Ancak gözlem sayısı çok ve gözlem değerleri büyük sayısal değerlerden oluşuyorsa bu tanım kapsamında geometrik ortalamayı hesaplamak zordur. Bu durumda geometrik ortalama ikinci tanım esas alınarak hesaplanabilir.

ii. G.O. gözlem değerlerinin logaritmalarının aritmetik ortalamasının anti-logaritmasına eşit ortalamadır. Bu tanıma göre geometrik ortalama

$$\log .G.O \frac{\sum \log Xi}{n}$$

G.O.= Anti-log(GO) bulunur.

Frekans ve gruplandırılmış serilerde geometrik ortalama,

$$\log .G.O \frac{\sum (ni \log Xi)}{\sum ni}$$ formülüyle hesaplanır.

G.O.= Anti-log (GO) bulunur.

ÖRNEK 12

Eskişehir'de faaliyet gösteren 120 lojistik işletmesi için bir yıllık taşıma kapasitesi(ton) gruplandırılmış seri aşağıda verilmiştir. Ortalama taşıma miktarını Geometrik ortalama ile hesaplayalım.

Çözüm: Öncelikle gruplandırılmış seri frekans serisine dönüştürülür. Sonra Xi değerlerinin logaritmaları ($\log X_i$) alınır. Son olarak da $\log X_i$'ler, frekanslarla çarpılıp, toplamları alınır.

Gruplar	n_i	X_i	logX	$n_i \log X_i$
600- 800	3	700	2,84509	8,53529
800-1000	7	900	2,95424	20,67969
1000-1200	11	1100	3,04139	33,45532
1200-1400	22	1300	3,11394	68,50675
1400-1600	40	1500	3,17609	127,04365
1600-1800	24	1700	3,23045	77,53077
1800-2000	9	1900	3,27875	29,50878
2000-2200	4	2100	3,32221	13,28887
	120			378,54912

Hesaplanan değerler formülde yerine konarak logaritmik geometrik ortalama hesaplanır.

$$\log .G.O = \frac{\sum (ni \log Xi)}{\sum ni} = \frac{378,54912}{120} = 3,154576$$

Henüz verilen gruplandırılmış serinin geometrik ortalamasına ulaşılamamıştır. Ulaşabilmek için hesaplanan logaritmik ortalamanın anti-logaritmasının alınması gerekir. Buradan,

G.O.= anti-log (3,154576) = 1427,5 (ton) olarak bulunur.

Geometrik ortalamanın iki temel uygulama alanı vardır: *Birincisi,* gözlem değerleri yaklaşık aynı oranda değişen yüzde, indeksler ve oransal verilerin ortalama hesabında kullanılır. *İkincisi ise;* nüfus, milli gelir ve ekonomik serilerde kullanılır. İncelenen *seride sıfır veya negatif değer varsa* geometrik ortalama kullanılmaz.

Kareli Ortalama

Kareli ortalama KO simgesiyle gösterilir ve

$$KO = \sqrt{\frac{\sum Xi^2}{n}}$$

formülü yardımıyla hesaplanır. Formülden de anlaşılabileceği gibi kareli ortalama, *gözlem değerlerinin karelerinin aritmetik ortalamasının kareköküne* eşit olan ortalamadır.

Kareli ortalama bazı istatistiksel işlemlerde, özellikle değişkenlik ölçülerinin hesaplanmasında kullanılan bir ortalamadır. Bunun yanında istatistik serisinde negatif işaretli terim varsa kullanılabilecek tek ortalama türü kareli ortalamadır.

Dizilerde kareli ortala hesaplanırken gözlem değerlerinin kareleri alınıp toplanmakta ve sonuç gözlem sayısına bölündükten sonra karekökü alınmaktadır.

ÖRNEK 13

Aşağıda verilen dizi için kareli ortalamayı hesaplayınız. Bu dizi için aritmetik ortalamanın uygun olup olmadığına karar veriniz.

X_i	X_i^2	
-3	9	Kareli ortalama,
-2	4	
1	1	$KO = \sqrt{\frac{32}{5}} = 2{,}53$ olarak
3	9	
3	9	bulunur.
	32	

Çözüm: Öncelikle her bir gözlem değerinin karesi alınır.

Verilen dizinin kareli ortalaması 2,53'dür. Dizi için aritmetik ortalama uygun değildir. Çünkü hem gözlem değerleri toplamı sıfırdır hem de dizide negatif gözlem değeri vardır.

Frekans ve gruplandırılmış serilerde kareli ortalama hesabında, karesi alınan gözlem değerlerinin, frekansları ile çarpımları toplamı, frekanslar toplamına bölünerek karekökü alınmaktadır.

ÖRNEK 14

On ailenin sahip olduğu çocuk sayıları aşağıda frekans serisi olarak düzenlenmiştir. Frekans serisinin kareli ortalamasını hesaplayınız. Aynı seri için aritmetik ortalamasının hangi değerler arasında yer alacağını belirleyiniz.

Çözüm: Kareleri alınan gözlem değerleri frekansları ile çarpılır.

X_i	n_i	X_i^2	$X_i^2 n_i$
1	1	1	1
2	3	4	12
3	4	9	36
4	1	16	16
5	1	25	25
	10		90

Kareli ortalama,

$$KO = \sqrt{\frac{90}{10}} = \sqrt{9} = 3$$

bulunur.
Aritmetik ortalamanın bulunacağı aralık ise,
$1 < \bar{X} < 3$ olacaktır.

Aritmetik ortalama için alt ve üst limitini belirlemek için duyarlı ortalamalar arasında olan ilişkiden yararlanılır. Söz konusu ilişki,

$$X_{min} < G.O < \bar{X} < K.O < X$$

şeklinde formüle edilebilir.
Duyarlı ortalamalar arasındaki söz konusu bu ilişkiyi, bir örnek üzerinde gösterebiliriz.

ÖRNEK 15

100 üretim işletmesinin(aynı ürünü üreten) aylık üretim miktarları(Bin Ton) frekans serisi olarak düzenlenmiş ve aşağıda verilmiştir. Verilen serinin;
a. Aritmetik ortalamasını,
b. Geometrik ortalamasını,
c. Kareli ortalamasını hesaplayınız.
d. Hesapladığınız duyarlı ortalamalar arasındaki ilişkiyi belirleyiniz.

X_i	n_i	$X_i n_i$	$\log X_i$	$n_i \log X_i$	X_i^2	$X_i^2 n_i$
10	10	100	1,00000	10,00000	100	1000
20	15	300	1,30102	19,51544	400	6000
30	35	1050	1,47712	51,69924	900	31500
40	25	1000	1,60205	40,05149	1600	40000
50	15	750	1,69897	25,48455	2500	37500
	100	3200		146,75072		116000

Çözüm:
a. Aritmetik ortalama hesaplandığında;

$$\bar{X} = \frac{\sum X_i n_i}{\sum n_i} = \frac{3200}{100} = 32 \rightarrow 32.000$$ Ton olarak bulunur.

b. Geometrik ortama hesaplandığında;

$$\log GO = \frac{146,75072}{100} = 1,46750$$

GO= anti-log(1,4675)
GO= 29,3431 → 29.343,1 Ton olarak bulunur.

c. Kareli ortalama hesaplandığında;

$$KO = \sqrt{\frac{116000}{100}} = 34,05 \rightarrow 34.050 \text{ Ton olarak bulunur.}$$

d. Hesaplanan duyarlı ortalamalar arasındaki ilişki,

$X_{min} < G.O < \bar{X} < K.O < X_{max}$'den,

$X_{min} < G.O = 29{,}34 < \bar{X} = 32 < K.O = 34{,}05 < X_{max}$

olarak bulunmuş olur. Görüleceği gibi aynı seri için hesaplanan duyarlı ortalamalardan en küçüğü geometrik, sonra aritmetik ve en büyüğü de kareli ortalama olarak hesaplanmıştır. Bu durum, bir anlamda geometrik ortalamanın diğerlerine göre daha az duyarlı olduğunu da göstermektedir.

Duyarlı (Hassas) Olmayan Ortalamalar

Duyarlı ortalamalar seriyi oluşturan gözlem değerlerinin herhangi birinde meydana gelen aşırı değişiklikten etkilenen ortalamalardır. Oysa seriyi oluşturan gözlem değerlerinden bazıları, uç değer alabilir. Böyle serilerin ortalama hesabında, duyarlı olmayan ortalamaları kullanmak gerekir. Duyarlı olmayan ortalamalar medyan ve mod olmak üzere iki başlık altında incelenecektir.

Medyan

Küçük değerden büyük değere doğru sıralanmış gözlem değerlerinden oluşan seriyi, *gözlem sayısı bakımından iki eşit kısmi seriye ayıran değer,* medyan ortalamadır. Dizilerde ve frekans serilerinde medyan hesabında bir formül yoktur. Ancak medyanın serideki kaçıncı gözlem değeri olduğunu veya hangi gözlem değerleri arasında bulunduğunu belirlemek için basit bir formül yazılabilir.

Dizilerde medyan hesabı; gözlem değerleri küçükten büyüğe doğru sıralanmış bir seride, tam ortadaki değer medyandır. Medyan hesaplama sürecinin adımları;

- Derlenen veriler küçükten büyüğe sıralanır.
- Hesaplanacak medyan ortalama değerine karşı gelecek gözlem değeri sırası, şöyle belirlenir. Eğer gözlem sayısı *(n) tek* sayıda ise $\left(\frac{n+1}{2}\right)$ hesaplanır, bu değer medyan ortalamanın sıra numarasını gösterir ve $\left(\frac{n+1}{2}\right)$'inci sıradaki gözlem değeri medyan ortalamadır.

ÖRNEK 16

Anadolu Üniversitesi spor salonuna iki ay süre ile devam eden 5 üyenin ağırlık (kg) kayıpları aşağıdaki gibidir. Bu ağırlıkların medyan ortalaması nedir?

X_i: 10, 5, 19, 8, 3

Çözüm: Veriler incelendiğinde küçük değerliden büyük değerliye bir sıralamanın olmadığı görülmektedir. Bu nedenle, önce veriler küçükten büyüğe doğru sıralanır.

<table>
<tr><td>

X_i
3
5
8
10
19

</td><td>

Burada 5 adet gözlem değeri olduğuna göre n = 5'tir.
n - tek sayıdır. Ortadaki gözlem değerinin sıra numarası
$\frac{n+1}{2} = \frac{5+1}{2} = 3$ olarak hesaplanır. Buna göre **üçüncü sıradaki** gözlem değeri medyan olacaktır. Üçüncü sıradaki gözlem değeri 8 olduğuna göre, medyan = 8 kg'dır.

</td></tr>
</table>

Söz konusu beş üyenin verdikleri ağırlıkların medyan ortalaması 8 kg olarak bulunmuştur. Üyelerin yarısı 8 kg veya daha az ağırlık kaybetmiş iken diğer yarısı da 8 kg veya daha çok ağırlık kaybetmişlerdir.

- Eğer serideki gözlem sayısı *(n) çift* sayıda ise sıra sayısı bakımından tam ortadaki bir yerine iki değer bulunur. Bu iki gözlem değerinin aritmetik ortalaması medyan ortalamadır.

ÖRNEK 17

Kimyasal ürün sektöründe faaliyet gösteren 6 İşletmenin 2010 yılında Türk Standartları Enstitüsünden (TSE) aldıkları buluş onayı (patent) sayıları aşağıdaki gibidir. Bu 6 işletme için onaylanan patent sayılarının medyan ortalamasını hesaplayınız.

İşletme	Onaylanan Patent Sayısı
A	257
B	239
C	97
D	385
E	50
F	249

Çözüm: Veriler incelendiğinde, küçük değerliden büyük değerliye bir sıralamanın olmadığı görülmektedir. Aşağıda veriler küçükten büyüğe doğru sıralanmıştır.

<table>
<tr><td>

X_i
50
97
239
249
257
385

</td><td>

Burada 6 adet gözlem değeri olduğuna göre n= 6'dır.
n - çift sayıdır. Tam ortaya düşen iki gözlem değeri;
n/2=6/2=3'üncü sıradaki değer ile (n/2)+1=3+1=4'üncü sıradaki değerdir. Böylece medyan 3'üncü ve 4'üncü gözlem değerlerinin aritmetik ortalaması olacaktır.
X_3 = 239 ve X_4 = 249 olmak üzere;
$Med = \frac{239+249}{2} = 244$ buluş (patent)

</td></tr>
</table>

Böylece 2010'da kimyasal üretim sektöründe faaliyet gösteren işletmeler için TSE onayı alan patent sayısı ortalaması 244 olarak bulunmuştur. Seri gözlem değerlerine dikkatle bakıldığında 50'den başlayıp 385'te son bulmaktadır. Medyan uç değerlerden etkilenmeyen bir ortalamadır. Bu nedenle medyan, aşırı küçük ve/veya aşırı büyük uç değerler bulunan serilerde, diğer ortalamalara tercih edilen bir ortalama türüdür. Bunun yanında basit bir sıralama ile bulunması mümkün olduğundan, pratik bir ortalamadır.

Frekans serisinin medyan ortalaması, izleyen işlem sırası ile hesaplanır.

- Serideki gözlem değerleri küçükten büyüğe doğru sıralanır
- "Den az" frekansları hesaplanır.
- Gözlem sayısının *(n) tek sayı ve çift* sayı olması durumuna göre dizilerde açıklanan işlemler ile medyan ortalama hesaplanır.

ÖRNEK 18

Eskişehir'de faaliyet gösteren 100 un fabrikası günlük üretim miktarları (ton) itibari ile incelenmiş ve aşağıdaki frekans serisi oluşturulmuştur. Bu frekans serisinin medyan ortalaması nedir?

X_i	n_i	"den az"
100	7	7
200	18	25
300	25	50
400	30	80
500	12	92
600	8	100

Gözlem değerleri küçükten büyüğe doğru sıralanmıştır. Gözlem sayısı **n=100 - çift sayı**dır. n/2=100/2=50 hesaplanır. Böylece 50'inci ve 51'inci sıradaki gözlem değerlerinin ortalaması;

$$Med = \frac{X_{50} + X_{51}}{2} = \frac{300 \quad + \quad 400}{2} = 350 \text{ Ton}$$

medyan değeridir:

Serinin medyanı 350 ton'dur. Buna göre, Eskişehir'de bulunan 100 un fabrikasından 50 tanesinin günlük un üretim miktarı 350 ton veya 350 ton'dan daha azdır. Diğer 50 un fabrikasının günlük un üretim miktarı ise 350 ton veya 350 ton'dan daha çoktur.

Gruplandırılmış serilerde medyan hesaplanırken her sınıftaki gözlem değerlerinin, o sınıf içerisinde eşit aralıklarla dağıldığı varsayılır. Bu varsayımdan hareketle gruplandırılmış seride sınıf aralıklarının farklı olmasının veya açık uçlu sınıf bulunmasının medyan hesabında önemi yoktur. Gruplandırılmış serilerde medyan hesabında bir değişiklik yapılmaksızın aşağıdaki formül kullanılır:

$$Med = l_a + \frac{\frac{n}{2} - n_a}{n_m} c_m$$

Formülde;

l_a : medyan sınıfının alt limiti,
n : serideki gözlem sayısı,
n_a : medyan sınıfına kadar olan sınıfların frekansları toplamı,
n_m : medyan sınıfının frekansı,
c_m : medyan sınıfının sınıf aralığıdır.

Bu formülü kullanabilmek için öncelikle medyan sınıfının belirlenmesi gerekir. Medyan sınıfı, "den az" frekansları yardımıyla belirlenir ve (n/2)'inci gözlem değerinin içinde bulunduğu sınıf, medyan sınıfı olarak kabul edilir.

ÖRNEK 19

50 erkek öğrencinin ağırlıkları ölçülerek aşağıdaki gruplandırılmış seri elede edilmiştir. Medyan ortalama ağırlık nedir?

Çözüm: Öncelikle *"den az"* frekanslar hesaplanarak, n/2 yardımıyla medyan sınıfının belirlenmesi gerekir.

Ağırlık Grupları	n_i	"den az"
60 - 63	1	1
63 - 66	2	3
66 - 69	13	16
69 - 72	20	36
72 - 75	11	47
75 - 78	3	50
	50	

$\frac{n}{2}=\frac{50}{2}=25$ olduğundan 25'inci gözlem değerinin içinde bulunduğu sınıf, medyan sınıfı olacaktır. Bu durumda;
Medyan sınıfı = **69 - 72**'dir.
$l_a = 69$
$n_a = 1+2+13 = 16$
$n_m = 20$
$c_m = 3$

$$Med = 69 + \frac{25-16}{20} \times 3 = 70,35 \text{ kg olarak bulunur.}$$

50 erkek öğrencinin ağırlıklarının medyan ortalaması 70,35 kg olarak bulunmuştur. Bu 50 öğrencinin yarısın ağırlığı 70.35 kg'dan az, diğer yarısının ise daha fazladır.

SIRA SİZDE 4

Gruplandırılmış seride medyan hesabında medyan sınıfı nasıl belirlenir?

Medyan seriyi iki eşit kısma bölen ortalama bir değerdir. Ancak küçükten büyüğe doğru sıralanmış seriyi dört, on ve yüz eşit kısmi seriye bölen ortalamalarda bulunmaktadır. Bunların genel adı *"kantil"* ortalamalardır.

KİTAP

Kantil ortalamalar konusunda daha geniş bilgi, Prof Dr. Özer SERPER'in İstatistik I (Ezgi Kitabevi, Bursa, 2004) kitabından bulunabilir.

Mod

Mod, seride en çok tekrarlanan gözlem değeri olarak tanımlanır. Mod, şimdiye kadar incelenen ortalamalar arasında en az hassas (duyarlı) olanıdır. Serideki çok büyük ve çok küçük gözlem değerlerinin etkisi altında kalmaz. Mod iktisadi olaylarda, kalite kontrolünde ve özellikle ayakkabı ve hazır giyim üretiminde yaygın kullanıma sahip bir ortalamadır.

Mod ve moda sözcüklerinin benzerliği dikkatinizi çekmiştir.

Modun hesaplanması kolaydır, dizilerde ve frekans serilerinde hesap makinesi kullanılmadan belirlenebilir.

ÖRNEK 20

Eskişehir-Bursa arasındaki yol üzerinde sekiz farklı marka otomobilin bir saatlik sürede kat ettikleri mesafeler (km) aşağıda verilmiştir. Mod ortalama nedir?

68, 69, 71, 73, 74, 75, 77, 81

Veriler incelendiğinde gözlem değerlerinin aldığı değerler itibariyle hepsi birbirinden farklıdır. Bu nedenle söz konusu bu serinin modu yoktur. Mod hesaplanabilmesi için bir değerin gözlem sayının öne çıkması gerekir. Örneğin söz konusu sekiz araçtan ikisi aynı mesafeyi kat etseydi

68, 69, 71, 73, 74, 74, 77, 81

bu durumda diğerlerinden daha çok gözlenmiş olan 74 mod olurdu. Mod = 74 km'dir.

ÖRNEK 21

Bir sürücü kursunda otomobil kullanmayı öğrenerek sürücü belgesi alan 10 sürücünün, son üç yılda aldıkları ceza sayıları aşağıdaki gibidir. Buna göre mod ortalama nedir?

3, 2, 0, 0, 2, 3, 3, 1, 0, 4

Çözüm: Öncelikle gözlem değerleri küçükten büyüğe sıralandığında;

X_i
0
0
0
1
2
2
3
3
3
4

Seride; bir adet 4, biri adet 1, iki adet 2 ve üçer adet 0 ve 3 vardır. Böylece seri için 0 ve 3 olmak üzere iki mod ortaya çıkar ki bu iki değer seriyi özetlemekten uzaktır. Çünkü buna göre, sürücü kursunun verdiği eğitimin kaliteli mi yoksa kalitesiz mi olduğu anlaşılamamaktadır. Bu haliyle seri için tek bir mod değer belirlenemez. Ancak bir karar verebilmek için gözlem sayısı artırılmalıdır. Ayrıca seri frekans serisi şeklinde düzenlense en yüksek frekans (3) iki kez yer alımıştır.

X_i	n_i
0	**3**
1	1
2	2
3	**3**
4	1

Frekans serilerinde mod, en yüksek frekansa sahip gözlem değeri olarak tanımlanır. Serinin frekanslar sütunundaki en yüksek frekans belirlendikten sonra, bunun karşısındaki gözlem değeri mod ortalama kabul edilir.

ÖRNEK 22

Eskişehir'de kredi kartı kullananlardan seçilen 220 kişinin aylık kredi kartı harcama miktarları gözlenmiş ve aşağıdaki frekans serisi düzenlenmiştir. Verilen frekans serisinin mod ortalaması nedir?

X_i	n_i
1000	10
2000	22
3000	54
4000	83
5000	31
6000	20

Seride en yüksek frekans 83 olduğuna göre, buna karşı gelen gözlem değeri 4000 mod'dur.
Mod = ₺4000 olarak bulunur.

Eskişehir'de kredi kartı kullananların aylık ortalama harcama miktarı T4000'dir.

Gruplandırılmış serilerde mod hesaplanırken, medyan hesabında olduğu gibi formül kullanılır. Sınıf aralıklarının eşit olması hâlinde, mod hesabında izlenen işlem sırasına göre hesaplanır.

- Öncelikle seride en yüksek frekansa sahip sınıf, *mod sınıfı* olarak kabul edilir.
- Mod sınıfı belirlendikten sonra izleyen formül uygulanarak mod ortalama hesaplanır.

$Mod = l_a + \frac{\Delta_1}{\Delta_1 + \Delta_2}.c_m$	**Formülde;** l_a : mod sınıfının alt limiti, D_1 : mod sınıfının frekansı ile bir önceki sınıfın frekansı arasındaki fark D_2 : mod sınıfının frekansı ile bir sonraki sınıfın frekansı arasındaki fark c_m : mod sınıfının, sınıf aralığı

Sınıf aralıkları farklı olan gruplandırılmış serilerde mod hesaplayabilmek için, önce frekansların ayarlanması gerekir.

DİKKAT

ÖRNEK 23

Eskişehir sanayi bölgesinde faaliyet gösteren 120 işletmenin geçen iki aylık sürede tükettikleri elektrik enerjisi miktarları aşağıda gruplandırılmış seri olarak verilmiştir. Mod enerji tüketimini hesaplayalım.

Enerji Tüketimi Mik. Grupları (kws)	İşletme Sayısı
1500-2000	16
2000-2500	27
2500-3000	38
3000-3500	20
3500-4000	14
4000-4500	5

Çözüm: Verilen gruplandırılmış seride sınıf aralıkları eşit ve en yüksek frekansa(38) sahip tek sınıf olduğuna göre, mod hesabı yapılabilir.

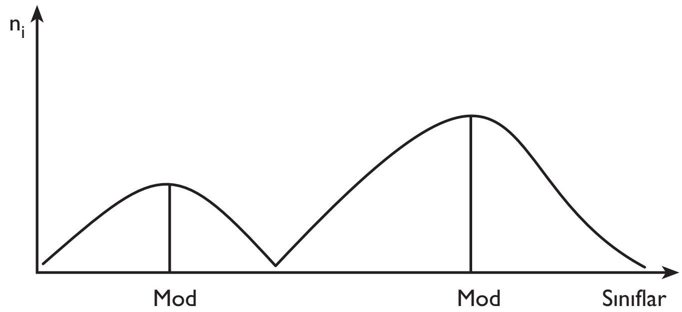
Mod sınıfı = 2500 - 3000 l_a = 2500 D_1 = 38 - 27 = 11 D_2 = 38 - 20 = 18 c_m = 500	$MOD = 2500 + \frac{11}{11+18}.500$ MOD = 2689,655 kws olarak bulunur.

Mod hesabında, mod sınıfından bir önceki ve bir sonraki sınıf hesaba katılır. Dolayısıyla mod, uç değerlerden etkilenmez. Bunun yanında en yüksek frekansa sahip olsa bile serideki ilk sınıf veya son sınıf mod sınıfı kabul edilmez.

Mod hesaplama sürecince ilgili istatistik serisinin grafiği Şekil 3.1'de görüldüğü gibi birden çok noktada maksimum yapıyorsa mod hesaplanmaz. Söz konusu şekilde iki tepe noktası bulunmaktadır. Bu tür serilerde mod hesaplanabilmesi serini tek maksimumlu yapıya dönüştürülmesi gerekir. Bu dönüştürme işlemi, sınıf birleştirmesiyle veya gözlem sayısını artırarak yapılabilmektedir.

Şekil 3.1

Çift maksimumlu seri



SIRA SİZDE 5

Bir serinin mod'unun hesaplanmasında aranan temel özellik nedir?

Son olarak ortalama hesabında, aritmetik ortalama, medyan mod vb. den hangisinin daha iyi bir ortalama olduğu söylenemez. Bunlardan her biri farklı durumlar için en iyi ortalama olabilir. Ortalamalar içinde en yaygın kullanıma sahip olan aritmetik ortalamadır, bunu medyan izler. Medyan, uç değerler içeren seriler için en iyi ortalamadır. Mod ise kolay hesaplanmakla birlikte pratik yaşamda kullanımı azdır.

DEĞİŞKENLİK VE DAĞILMA ÖLÇÜLERİ

Ortalamalar (aritmetik. ort., medyan, mod vb.) bir seriyi özetlemek için gerekli olmakla birlikte, tek başlarına yetersiz kalırlar. *Seriyi oluşturan gözlem değerlerinin birbirlerinden ve aritmetik ortalamadan ne derece uzaklaştıkları,* değişkenlik ölçüsü ile belirlenir. Birçok durumda olduğu gibi, aşağıdaki bazı durumlarda seriyi oluşturan gözlem değerlerinin değişkenlik derecesinin bilinmesi büyük önem taşır.

- Bir işletmede çalışanların verimliliğinin değişkenliği düşünülürse verimlilik ortalaması artırmak amaç olabileceği gibi, yavaş ve hızlı çalışanları ayırmakta özel amaç olabilir. Bu kişisel farklılıkları belirlemek için verimliliğinin değişkenlik ölçüsüne ihtiyaç vardır.
- Bir yatırımcı, ortalama getirilerin yanında, değişkenlik ölçülerini bilmesi hâlinde yatırım seçeneklerinin taşıdığı riski görerek daha sağlıklı karar verebilecektir.
- Bir yönetici, işletmesi için uygulanan pazarlama stratejilerinden en uygun olanını, değişkenlik ölçüleri kullanarak seçebilir.
- İki ayrı ülke, kişi başına düşen milli gelir itibari ile aynı ortalamaya sahip olabilir. Bu durumda iki ülkede refah seviyesi aynı gibi görünmekle birlikte çok farklı da olabilir. Ülkelerden birinde bireyler birbirlerine yakın gelire sahip iken diğer ülkede bireylerin gelirleri arasında uçurum olabilir. Sağlıklı bir kıyaslama yapabilmek için yine değişkenlik ölçülerine ihtiyaç vardır.

Aynı aritmetik ortalamaya sahip iki seri, bütünüyle farklı dağılabilir. Örneğin, iki işletmede çalışanların yaşları aşağıdaki gibi olsun:

İşletme-1 : 47,	38,	35,	40,	36,	45,	39
İşletme-2 : 70,	33,	18,	52,	27,		

Bu iki işletmede çalışanların yaşlarının aritmetik ortalaması hesaplandığında ikisinin de aynı ($\overline{X}_1$ =40 ve $\overline{X}_2$ =40) olduğu görülür. Eğer bu iki işletme çalışanlarının yaşları, yalnız ortalamalar dikkate alınarak karşılaştırılırsa, ortalama yaş aynı olduğundan çalışanların aynı yaşlarda olduğu söylenebilir. Oysa serilere dikkatle bakılırsa: İşletme-1'de orta yaşlarda işçiler çalışıyor gözükürken İşletme-2 için aynı şeyler söylenemez.

Değişkenlik ölçüleri hesaplanmasında bütün gözlem değerleri dikkate alınabileceği gibi, bunlardan çok az bir kısmına dayanarak da ölçü hesaplanabilmektedir.

Değişkenlik ölçüleri ne işe yarar?

SIRA SİZDE 6

Değişim Aralığı

Değişkenlik ölçüleri içinde en basit ve en kolay hesaplanan değişim aralığı: serinin en büyük gözlem değeri ile en küçük gözlem değeri arasındaki fark olup aşağıdaki gibi formüle edilir:

$$D.A. = X_{max} - X_{min}$$

Değişim aralığı daha çok, kalite kontrolünde arka arkaya çekilen 5-6 birimlik örneklerin değişkenliğini belirlemek amacıyla kullanılır.

ÖRNEK 24

Bir maden işletmesinin altı aylık üretim miktarı aşağıdaki gibi gerçekleşmiştir. Değişim aralığını hesaplayalım:

X_i (ton)	
200	Gözlem değerleri 200 ile 1000 arasında yer almaktadır. Bu nedenle değişim katsayısı:
300	
400	
500	D.A.= $X_{max} - X_{min}$ = 1000 - 200 = 800 ton
700	olarak bulunur.
1000	

Standart Sapma ve Varyans

Bir seriyi oluşturan gözlem değerlerinin aritmetik ortalamadan sapmalarının kareli ortalaması **standart sapma** olarak tanımlanır. Değişkenlik ölçüleri içinde en yaygın ve etkin kullanıma sahip olan standart sapmadır. Standart sapma serideki gözlem değerlerinin tümüne dayanan, dolayısıyla matematiksel bakımdan en kuvvetli değişkenlik ölçüsüdür.

Standart sapma daha çok istatistiksel analizlerde kullanılan bir ölçü olup özellikle normal dağılım fonksiyonunun en önemli unsurudur.

Bir seriyi oluşturan gözlem değerlerinin birbirine yakınlığı standart sapma ile ifade edilir. Genel olarak seriyi oluşturan gözlem değerleri, hesaplanan ortalamaya yakın değerlerden oluşuyorsa standart sapma değeri küçük, uzak değerlerden oluşuyorsa standart sapma değeri büyük olur.

Standart sapma genellikle örnek kütle için s-(küçük-s) ile evren için σ-(küçük sigma) ile gösterilir. Standart sapmanın karesine varyans denir ve örnek için s^2, ana kütle için σ^2 ile gösterilir.

Dizilerde standart sapma hesaplanmasında;

- Öncelikle serini aritmetik ortalaması ($\overline{X}$) hesaplanır.
- Hesaplanan aritmetik ortalama, her bir gözlem değerinden çıkarılır ve bu farkların karelerinin toplamı alınır. $\sum_{i=1}^{n}(X_i - \overline{X})^2$
- İzleyen formüller uygulanarak standart sapma ve varyans hesaplanır.

$$s = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n}} \text{ varyans ise; } s^2 = \frac{\sum (X_i - \overline{X})^2}{n} \text{ dır.}$$

Bu formüllerdeki standart sapma ve varyans örnek içindir. Evren için standart sapma ve varyans formülleri yazılırken $\overline{X}$ yerine evren aritmetik ortalaması μ ve n yerine evren gözlem sayısı N konur. Buna göre evren standart sapması formülü;

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{N}} \text{ varyans ise; } \sigma^2 = \frac{\sum (X_i - \mu)^2}{N} \text{ dir.}$$

DİKKAT

Standart sapma formülünün paydasında yer (n) yerine bazı kaynaklarda (n-1) kullanılmaktadır. Ancak gözlem sayısı çok olduğunda bunun önemi yoktur.

ÖRNEK 25

Aşağıda altı iş makinesinin saatlik hafriyat miktarları (ton) verilmiştir. Verilen dizinin standart sapmasını hesaplayalım.

X_i	$(X_i - \overline{X})$	$(X_i - \overline{X})^2$
7	-2,5	6,25
7	-2,5	6,25
9	-0,5	0,25
10	0,5	0,25
11	1,5	2,25
13	3,5	12,25
57		27,50

- Öncelikle aritmetik ortalama hesaplanır;

$$\overline{X} = \frac{\sum X_i}{n} = \frac{57}{6} = 9{,}5 \text{ ton}$$

- Sonra $\sum_{i=1}^{n}(X_i - \overline{X})^2 = 27{,}50$ bulunur,
- Formül uygulanarak, standart sapma hesaplanır.

$$s = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n}} = \sqrt{\frac{27{,}50}{6}} = 2{,}14 \text{ ton}$$

Diziyi oluşturan gözlem değerlerinin, aritmetik ortalama 9,5 ton'dan ortalama olarak 2,14 ton'luk sapma gösterdikleri söylenebilir.

Hesaplanan standart sapma küçüldükçe değişkenliğin azaldığı, dolayısıyla gözlem değerlerini bir birine daha çok yaklaştığı söylenebilir. Örneğin, bir serinin $\overline{X}$ = 75 ve s = 0 olsun bu serinin hiç değişkenliği yok demektir. Ancak bu, gözlem değerlerinin birbirine eşit olması durumunda görülür. Oysa gerçek yaşamda bu tür serilerle pek karşılaşılmaz. Yine de *standart sapmanın sıfıra yakın değer alması değişkenliğin az olduğunu göstermektedir.*

Frekans ve gruplandırılmış serilerde standart sapma hesaplanması:

- Öncelikle serini aritmetik ortalaması ($\overline{X}$) hesaplanır.
- Hesaplanan aritmetik ortalama, her bir gözlem değerinden çıkarılır ve bu farkların toplamı alınır. $\sum_{i=1}^{n}(X_i - \overline{X})^2 n_i$
- İzleyen formüller uygulanarak standart sapma ve varyans hesaplanır.

$$s = \sqrt{\frac{\sum (X_i - \overline{X})^2 . n_i}{\sum n_i}} \text{ variyans ise; } s^2 = \frac{\sum \left[(X_i - \overline{X})^2 . n_i \right]}{\sum n_i} \text{ dır.}$$

Bu formüllerdeki standart sapma ve variyans örnek içindir. Evren standart sapma ve varyans formülleri yazılırken $\overline{X}$ yerine evren aritmetik ortalaması μ ve n yerine evren gözlem sayısı N konur.

ÖRNEK 26

On pasta fırınında aylık işlenen un miktarı (ton) aşağıda frekans serisi olarak düzenlenmiştir. Frekans serisinin standart sapmasını hesaplayınız.

X_i	n_i	$(X_i n_i)$	$(X_i - \overline{X})$	$(X_i - \overline{X})^2$	$(X_i - \overline{X})^2 n_i$
10	1	10	-42	1764	1764
20	2	40	-32	1024	2048
50	3	150	-2	4	12
70	2	140	18	324	648
80	1	80	28	784	784
100	1	100	48	2304	2304
	10	520			7560

$$\overline{X} = \frac{\sum Xini}{\sum ni} = \frac{520}{10} = 52 \text{ ton}$$

Pasta fırınları ayda ortalama 52 ton un kullanmaktadırlar. Tabidir ki bazıları daha az, bazıları ise daha çok un kullanmaktadır. Her bir pasta fırının kullandığı un miktarının bu ortalamadan (52) ne kadar sapma gösterdiği yukarıdaki serinin dördüncü sütununda (X_i - $\overline{X}$) hesaplanmıştır. Bu sapmaların bazıları negatif sayı olarak görünmektedir. Ortalama sapmayı hesaplayabilmek için kareleri alınır ve ortalama sapma formül yardımıyla hesaplanır.

$$s = \sqrt{\frac{7560}{10}} = \sqrt{756} = 27{,}5 \text{ ton olarak hesaplanır.}$$

Bunun anlamı, seriyi oluşturan gözlem değerlerinin aritmetik ortalamadan olan sapmalarının ortalaması 27,5 ton'dur.

Standart sapmanın ortalamalarla olan ilişkisine dayanarak hesaplanması: Standart sapma aritmetik ortalama ile kareli ortalamadan yararlanılarak da hesaplanabilir. Burada standart sapma, kareli ortalamanın karesi ile aritmetik ortalamanın karesi arasındaki farkın kareköküne eşit olur:

$$s = \sqrt{K^2 - \overline{X}^2}$$

Özellikle aritmetik ortalama tamsayı çıkmadığında elle veya hesap makinesi kullanılarak yapılan standart sapma hesabında kolaylık ve çabukluk sağlar.

ÖRNEK 27

80 üniversite öğrencisinin geçen ay kütüphanede geçirdikleri süreler, gruplandırılmış seri olarak aşağıda verilmiştir. Serinin standart sapmasını duyarlı ortalamalardan yararlanarak hesaplayalım:

Gruplar (saat)	n_i	X_i	$X_i n_i$	$X_i^2 n_i$
10 - 14	8	12	96	1152
14 - 18	28	16	448	7168
18 - 22	27	20	540	10800
22 - 26	12	24	288	6912
26 - 30	4	28	112	3136
30 - 34	1	32	32	1024
		80	1516	30192

$$\overline{X} = \frac{\sum Xini}{\sum ni} = \frac{1516}{80} = 18,95 \text{ saat} \qquad K^2 = \frac{\sum Xi^2 ni}{\sum ni} = \frac{30192}{80} = 377,4 \text{ saat}$$

$$s = \sqrt{K^2 - \overline{X}^2} = \sqrt{377,4 - (18,95)^2} = \sqrt{18,2975} = 4,2775 \text{ saat}$$

Gruplandırılmış seriyi oluşturan gözlem değerlerinin, aritmetik ortalamadan olan uzaklıkları ortalaması 4,2775 saattir. Öğrenciler geçen ay ortama 18,95 saatlerini kütüphanede geçirmişlerdir. Öğrencilerin kütüphanede geçirdikleri sürelerin bu ortalamadan olan uzaklıklarının ortalaması 4,2775 saat olduğu görülmektedir.

Değişim Katsayısı

İki seri **değişim katsayısı** ölçüsü ile kıyaslandığında değişim katsayısı küçük olan serinin değişkenliğinin, diğerine göre daha az olduğu söylenir.

Değişim katsayısı, bir serinin standart sapmasının aritmetik ortalamaya bölünüp 100 ile çarpılması sonucu elde edilen oransal bir ölçüdür. Değişim katsayısı;

Örnek için, D.K.= $\frac{s}{\overline{X}} \cdot 100$ Evren için, D.K.= $\frac{\sigma}{\mu} \cdot 100$

formülleri ile hesaplanır. Değişim katsayısı, gözlem değerleriyle hiçbir ilgisi olmadığından, gözlem değerlerinin büyüklüğünden etkilenmez. Bu nedenle daha önce incelenen değişkenlik ölçülerinin taşıdıkları sakıncaları taşımaz. Değişim katsayısının iki yaygın kullanım alanı vardır:

- Ölçü birimleri farklı olan iki veya daha çok serinin değişkenliğini karşılaştırmada kullanılır (kg. ile ₺ gibi).
- Ölçü birimleri aynı ancak gözlem sayıları ve büyüklükleri farklı olan serilerin değişkenliğini karşılaştırmakta kullanılır. Özellikle küçük ve büyük işletmelerin çeşitli yönlerden kıyaslanması imkânını sağlar.

SIRA SİZDE 7

Değişim katsayısının kullanım amacı nedir?

ÖRNEK 28

A ülkesinde kişi başına milli gelir 7.500 dolar ve standart sapma ise 1.500 dolardır. B ülkesinde kişi başına milli gelir ise 6.000 euro ve standart sapma 600 euro'dur. Bu bilgelere göre hangi ülkede gelir daha dengeli dağılmaktadır.

$\overline{X}_A = 7500$ Dolar $\qquad \overline{X}_B = 6.000$ Euro

$S_A = 1500$ Dolar $\qquad S_B = 600$ Euro

$$\mathrm{DK_A} = \frac{s_A}{\overline{X}_A} \cdot 100 = \frac{1500}{7500} \cdot 100 = \%20$$

$$\mathrm{DK_B} = \frac{s_B}{\overline{X}_B} \cdot 100 = \frac{600}{6.000} \cdot 100 = \%10$$

$\mathrm{DK_A} = \%20 > \mathrm{DK_B} = \%10$ buna göre B ülkesinde fertlerin gelirinin birbirine yakın olduğu, dolayısıyla B ülkesinde milli gelirin A ülkesine göre daha dengeli dağıldığı söylenebilir.

Asimetri Ölçüleri

Bir serinin ortalaması ve değişkenlik ölçüsünün bilinmesi kadar, dağılım şeklinin bilinmesi de önemlidir. Seriler incelendiğinde bazılarının ortadaki gözlem değeri etrafında toplandıklarını, bazılarının yüksek değerlere eğimli olduğunu, bazılarının ise küçük değerlere eğimli oldukları gözlenebilir. Simetriden uzaklaşarak bir yana çarpık (eğik) olan serilerin bu özellikleri asimetri ölçüleri ile belirlenebilir.

Ortalamalara Dayalı Asimetri Ölçüleri

Serinin simetrik olup olmadığı ortalamaların (aritmetik ortalama, medyan ve mod) arasındaki ilişkiden yararlanılarak da belirlenebilir.

i. Eğer; = $\overline{X}$ = Med = Mod ise seri simetriktir.

ÖRNEK 29

Bir sınıfın boy uzunlukları incelenmiş ve 165 cm ile 185 cm arasında değerler alan boy değişkeni için aşağıdaki ortalamalar hesaplanmış olsun (Şekil 3.2).

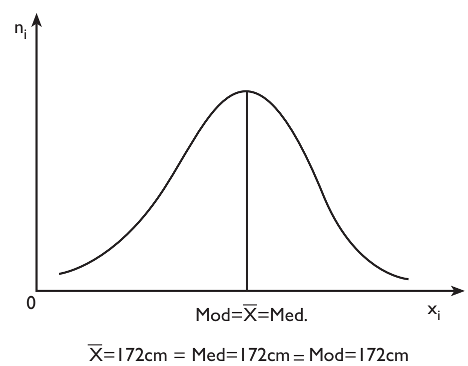
$\overline{X}$ = 172 cm Med = 172 cm Mod = 172 cm

ii. Eğer; Mod < Med < $\overline{X}$ ise seri asimetriktir.

Simetrik olmayan serilere asimetrik seri denir. Asimetrisi sağa çarpık serilerde medyan, daima mod ile aritmetik ortalama arasında değer alır.

Şekil 3.2

Tek maksimumlu simetrik eğri

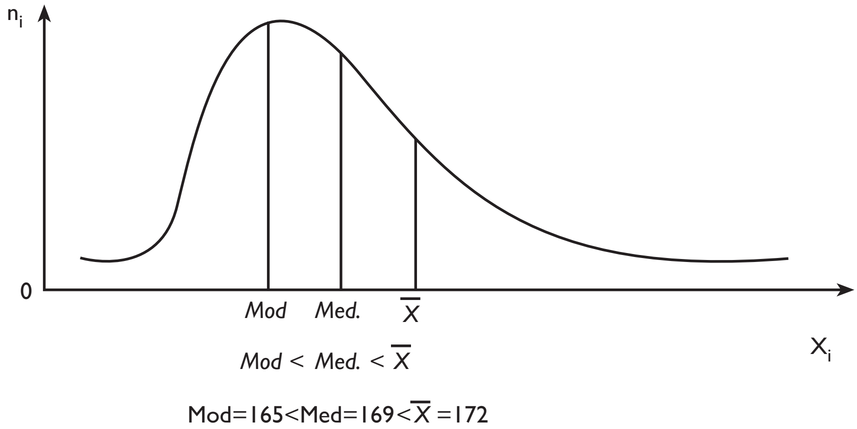


ÖRNEK 30 *Bir sınıfın boy uzunlukları incelenmiş ve aşağıdaki ortalamalar hesaplanmış olsun (Şekil 3.3).*

$\overline{X}$ = 172 cm Med = 168 cm   Mod = 165 cm

Şekil 3.3

Asimetrisi sağa çarpık eğri



iii. Eğer; $\overline{X}$ ***< Med < Mod ise seri asimetriktir***

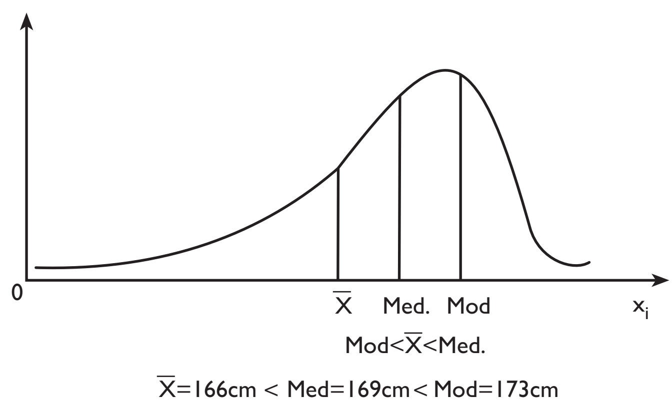
Asimetrisi sola çarpık serilerde medyan, daima aritmetik ortalama ile mod arasında değer alır.

ÖRNEK 31 *Bir sınıfın boy uzunlukları incelenmiş ve aşağıdaki ortalamalar hesaplanmış olsun (Şekil 3.4).*

$\overline{X}$ = 166 cm Med = 169 cm   Mod = 173 cm

Şekil 3.4

Asimetrisi sola çarpık eğri



$\overline{X}$ = 166 cm < Med = 169 cm < Mod=173 cm

Ortalamalara dayanan asimetri ölçüleri basitliği nedeniyle kullanılmasına rağmen güvenilir ölçüler değillerdir.

Özet

AMAÇ 1

Ortalamaları sınıflayıp tanımlamak.

Bir istatistik serisini temsil etmek, özetlemek, amacıyla geliştirilmiş olan ortalamalar, duyarlı ve duyarlı olmayan ortalamalar başlıkları altında toplanabilir. Hesaplanan ortalama değer, serideki herhangi bir gözlem değerinde meydana gelen değişiklikten etkileniyorsa ortalama hesaplamak amacıyla kullanılacak ortalama türü, ouyarlı (hassas) ortalamalar, etkilenmiyorsa ouyarlı (hassas) olmayan ortalamalardır.

AMAÇ 2

Aritmetik ve diğer duyarlı ortalamaları hesaplayıp açıklamak.

Aritmetik ortalama, gözlem değerleri toplamının gözlem sayısına bölünmesi ile bulunur. Duyarlı ortalamalar içinde en çok bilinen aritmetik ortalama olmakla birlikte; tartılı aritmetik ortalama, geometrik ortalama ve kareli ortalama da bu grup içinde yer almaktadır. Bu ortalamalar seriler için formüllerle kolaylıkla hesaplanabilir.

AMAÇ 3

Medyan ve mod ortalamaları hesaplayıp yorumlamak.

Küçük değerden büyük değere doğru sıralanmış gözlem değerlerinden oluşan seriyi, gözlem sayısı bakımından iki eşit kısmi seriye ayıran noktadaki seri değeri Medyan ortalamadır. Medyan uç değerlerden etkilenmeyen bir ortalamadır. Bu nedenle medyan, aşırı küçük ve/veya aşırı büyük uç değerler bulunan serilerde, diğer ortalamalara tercih edilen bir ortalama türüdür. Mod, şimdiye kadar incelenen ortalamalar arasında en az hassas (duyarlı) olanıdır. Serideki çok büyük ve çok küçük gözlem değerlerinin etkisi altında kalmaz. Mod iktisadi olaylarda, kalite kontrolünde ve özellikle ayakkabı ve hazır giyim üretiminde yaygın kullanıma sahip bir ortalamadır.

AMAÇ 4

Standart sapmayı hesaplayıp yorumlamak.

Bir seriyi oluşturan gözlem değerlerinin aritmetik ortalamadan sapmalarının kareli ortalaması standart sapma olarak tanımlanır. Seriyi oluşturan gözlem değerlerinin birbirine yakınlığı standart sapma ile ifade edilir. Genel olarak seriyi oluşturan gözlem değerleri, hesaplanan ortalamaya yakın değerlerden oluşuyorsa standart sapma değeri küçük, uzak değerlerden oluşuyorsa standart sapma değeri büyük olur.

Kendimizi Sınayalım

1. Bir istatistik serisindeki gözlem değerlerini özetleyen tipik değere, genel olarak ne ad verilir?

a. Değişken
b. Ortalama
c. Sapma
d. Homejenlik
e. Dağılma

2. Hesaplanan ortalamanın temsil kabiliyetini artıran en önemli unsur nedir?

a. Dağılım
b. Ortalama türü
c. Değişkenlik
d. Tesadüfilik
e. Homojenlik(türdeşlik)

3. Aşağıdakilerden hangisi duyarlı ortalama **değildir?**

a. Aritmetik
b. Tartılı Aritmetik
c. Geometrik
d. Kareli
e. Medyan

4. Aşağıda verilen frekans serisinin aritmetik ortalaması nedir?

X_i	n_i
2	2
4	3
5	4
7	1

a. $\bar{X} = 3{,}4$
b. $\bar{X} = 3{,}5$
c. $\bar{X} = 4{,}3$
d. $\bar{X} = 5{,}12$
e. $\bar{X} = 5{,}2$

5. Aşağıda verilen gruplandırılmış serisinin aritmetik ortalaması nedir?

Gruplar	n_i
0 - 2	2
2 - 4	10
4 - 6	6
6 - 8	2

a. $\bar{X} = 2{,}4$
b. $\bar{X} = 3{,}5$
c. $\bar{X} = 3{,}8$
d. $\bar{X} = 4{,}0$
e. $\bar{X} = 4{,}3$

6. İşçiler için saat başı ortalama ücreti, tartılı aritmetik ortalamayla hesaplayınız?

Saat başı ücret (krş)	işçi sayısı	işçi kıdemi (yıl)
1500	20	3
1 800	12	6
11000	8	10

a. $\bar{X}_t = 1690$
b. $\bar{X}_t = 1700$
c. $\bar{X}_t = 1690{,}45$
d. $\bar{X}_t = 5186{,}79$
e. $\bar{X}_t = 1801{,}40$

7. Aşağıda verilen gruplandırılmış serinin standart sapmasını $s = \sqrt{K^2 - \bar{X}^2}$ 'den yararlanarak hesaplanırsa ne elde edilir?

Gruplar	n_i
0 - 2	4
2 - 4	10
4 - 6	15
6 - 8	6
8 - 10	5

a. s = 1,815
b. s = 2,278
c. s = 3,015
d. s = 3,228
e. s = 3,512

8. Yıllık ortalama üretim 5.000 adet ve standart sapma 750 adet iken değişim katsayısı nedir?

a. D.K. = %15
b. D.K. = %25
c. D.K. = %5
d. D.K. = %10
e. D.K. = %7,5

9. Bir portföy yöneticisi ₺60'lık A hisse senedinden 200 hisse, ₺250'lık B hisse senedinden 100 hisse ve ₺65'lık C hisse senedinden 200 hisse satın almıştır. Satın alınan hisse başına ödenen ortalama fiyat nedir?

a. $\bar{X}$ = ₺60
b. $\bar{X}$ = ₺65
c. $\bar{X}$ = ₺100
d. $\bar{X}$ = ₺140
e. $\bar{X}$ = ₺250

10. Bir hava alanında gecikmeyle (rötarlı) kalkan uçakların gecikme süreleri ve sayıları gruplandırılmış seri olarak verilmiştir.

Gruplar (dakika)	Uçak Sayıları
0 - 10	15
10 - 20	16
20 - 30	14
30 - +	5

Buna göre, medyan ortalama gecikme süresi nedir?

a. Med = 15 dakika
b. Med = 15,25 dakika
c. Med = 16 dakika
d. Med = 16,25 dakika
e. Med = 20 dakika

Kendimizi Sınayalım Yanıt Anahtarı

1. b	Yanıtınız yanlış ise "Ortalamalar" başlıklı konuyu yeniden gözden geçiriniz.
2. e	Yanıtınız yanlış ise "Ortalamalar" başlıklı konuyu yeniden gözden geçiriniz.
3. e	Yanıtınız yanlış ise "Duyarlı Ortalamalar" başlıklı konuyu yeniden gözden geçiriniz.
4. c	Yanıtınız yanlış ise "Aritmetik Ortalama" başlıklı konuyu yeniden gözden geçiriniz.
5. c	Yanıtınız yanlış ise "Aritmetik Ortalama" başlıklı konuyu yeniden gözden geçiriniz.
6. d	Yanıtınız yanlış ise "Tartılı Aritmetik Ortalama" başlıklı konuyu yeniden gözden geçiriniz.
7. b	Yanıtınız yanlış ise "Standart Sapma" başlıklı konuyu yeniden gözden geçiriniz.
8. a	Yanıtınız yanlış ise "Değişim Katsayısı" başlıklı konuyu yeniden gözden geçiriniz.
9. c	Yanıtınız yanlış ise "Tartılı Aritmetik Ortalama" başlıklı konuyu yeniden gözden geçiriniz.
10. d	Yanıtınız yanlış ise "Medyani" başlıklı konuyu yeniden gözden geçiriniz.

Sıra Sizde Yanıt Anahtarı

Sıra Sizde 1
Hesaplanan ortalama değer, serideki herhangi bir gözlem değerinde meydana gelen değişiklikten etkileniyorsa ortalama hesaplamak amacıyla kullanılacak ortalama türü, duyarlı (hassas) ortalamalar, etkilenmiyorsa duyarlı (hassas) olmayan ortalamalardır.

Sıra Sizde 2
Gruplandırılmış seriyi frakans serisine dönüştürebilmek için bir sınıftaki gözlem değerlerinin, o sınıf içinde eşit aralıklarla dağıldıkları varsayımı kabul edilir. Dönüştürme her sınıf için aritmetik ortala hesaplayarak yapılır.

Sıra Sizde 3
Bir istatistik serisindeki gözlem değerlerinin önem derecesi her zaman aynı değildir. Ortalama hesabında bu önem derecelerinin hesaba katılması ile tartılı aritmetik ortalama hesaplanabilir.

Sıra Sizde 4
Gruplandırılmış seride medyan sınıfı, "den az" frekansları yardımıyla belirlenir ve (n/2)' inci gözlem değerinin içinde bulunduğu sınıf, medyan sınıfı olarak kabul edilir.

Sıra Sizde 5
Modun hesaplanabilmesi için serinin tek maksimumlu olması gerekir. Bir serinin grafiği düşünüldüğünde birden çok yerde maksimum yapıyorsa mod hesaplanamaz.

Sıra Sizde 6
Seriyi oluşturan gözlem değerlerinin birbirlerinden ve aritmetik ortalamadan ne derece uzaklaştıkları, değişkenlik ölçüsü ile belirlenir.

Sıra Sizde 7
Değişim katsayısının iki yaygın kullanım alanı vardır:
i. Ölçü birimleri farklı olan iki veya daha çok serinin değişkenliğini karşılaştırmada kullanılır (kg. ile ₺ gibi).
ii. Ölçü birimleri aynı ancak gözlem sayıları ve büyüklükleri farklı olan serilerin değişkenliğini karşılaştırmakta kullanılır.

Yararlanılan ve Başvurulabilecek Kaynaklar

Atlas, M.(2006), **İstatistik I(Çözümlü Örnekler),** Ak Ofset, Eskişehir.

Çömlekçi, N.(1998), **Temel İstatistik,** Bilim Teknik Yayınevi, İstanbul.

Gürsakal, N.(2007), **Betimsel İstatistik,** Nobel Yayın Dağıtım, Ankara.

Kazmier, L.J. and Pohl, N.F.(1987), **Basic Statistics for Business and Economics,** McGraw-Hill Book Co., New York.

Mann, P.S.(1995), **Statistics for Business and Economics,** J. Wiley, New York.

Newbold, P.(2001), **İşletme ve İktisat için İstatistik,** Çev: Ümit Şenesen, Literatür Yayıncılık, İstanbul.

Serper, Ö.(2004), **Uygulamalı İstatistik 1(Genş. 5. Baskı),** EzgiKitapevi, Bursa.

4

Amaçlarımız

Bu üniteyi tamamladıktan sonra;

- Endeks kavramlarını tanımlayabilecek,
- Endeks kavramlarını kendi içinde türlerine ayırabileceksiniz.

Anahtar Kavramlar

- Mekân endeksi
- Zaman endeksi
- Sabit esaslı endeks
- Değişken asaslı endeks
- Basit endeks
- Bileşik endeks
- Laspeyres endeksi
- Paasche endeksi
- Fisher endeksi
- Tüketici fiyatları endeksi (TÜFE)
- Üretici fiyatları endeksi (ÜFE)

İçindekiler

Endeksler

ENDEKSLER

Endeks, bir değişken veya değişkenler grubunun, zamana veya mekâna göre aldığı değerlerin oransal değişimlerini gösteren istatistiksel bir ölçüdür. Bu ölçü yardımıyla hesaplanan sayısal değerlerin oluşturduğu seriye de endeks serisi denir.

Endeks, zaman ve mekân serilerinin gözlem değerlerinde meydana gelen değişmeleri gösteren oransal bir ölçüdür.

Endekslerin hesaplanmasında ve yorumlanmasında, biri kıyaslanan diğeri temel (baz) alınacak bir değere gereksinim vardır. Oransal değişmeleri gösteren, endeks adı verilen ve I simgesiyle ifade edilen bu ölçünün payında kıyaslanan değer, paydasında da temel (baz) alınan değer yer alır. Kıyaslanan değer X_i ve temel devre değeri de X_0 ile gösterilirse endeksin genel gösterimi,

$$I = \frac{X_i}{X_0} \cdot 100 \tag{4.1}$$

şeklinde yazılır. Bu (4.1) eşitliğindeki 100 sayısı, (X_i / X_0) oransal ifadesinin ürettiği bilginin yorumlama niteliğini kolaylaştırmak ve değişimi % (yüzde) olarak ifade edebilmek amacıyla yazılır. Eğer oransal değişim (X_i / X_0) çok çok küçük bir değer ise bu durumda genel olarak 1000 ile çarpılır ve

‰ (binde) değişim olarak yorumlanır.

Örneğin; $I = \frac{X_i}{X_0} = \frac{10}{1250} = 0,008$

endeks değeri, yüzde değişim olarak,

$$I = \frac{X_i}{X_0}.100 = \frac{10}{1250}.100 = {}^0\!/_0\, 0,8$$

hesaplanır ve bulunan endeks değeri için değişim "yüzde 0,8" olarak okunur.

Eğer, aynı endeks değeri, binde değişim olarak bulunacaksa

$$I = \frac{X_i}{X_0}.1000 = \frac{10}{1250}.1000 = {}^0\!/_{00}\, 8$$

olarak hesaplanır ve bulunan endeks değer için değişim "binde 8" olarak okunur.

ENDEKS TÜRLERİ

Endeksler, çalışmanın amacına, değişken türüne ve sayısına göre sınıflandırılırlar.

Bu ünitede, uygulamada yaygın olarak kullanılan bazı endeks türleri izleyen kısımlarda incelenmiştir.

Mekân ve Zaman Endeksleri

Endeksi hesaplanacak seri; mekân serisi ise hesaplanacak endekse "mekân endeksi", zaman serisi ise "zaman endeksi" denir.

Mekân Endeksi

Mekân endeksi ilgilenilen değişkenin aldığı değerlerinin, mekândan mekâna değişkenin ortalama değerine göre değişiminin oransal bir göstergesidir.

Mekân endeksi, mekân serisinin gözlem değerlerinde mekândan mekâna değişiminin oransal bir göstergesidir. Bu endeksin hesaplanmasında, temel devre değeri olarak serinin aritmetik ortalaması alınır. Hesaplanan endeks değeri, bu ortalama değere göre yorumlanır.

$$I_i = \frac{X_i}{\overline{X}} \cdot 100 \tag{4.2}$$

$\overline{X}$: ilgili değişkenin aritmetik ortalama değeridir.

ÖRNEK 1

Ocak-2012 ayına ilişkin, beş farklı şehirdeki sinema bilet fiyatları aşağıdaki tabloda verilmiştir. Mekân endekslerini hesaplayınız ve yorumlayınız.

Tablo 4.1
Beş farklı şehirdeki Ocak-2012 ayına ilişkin sinema bilet fiyatları

Şehir	Fiyat (₺)
Ankara	12
Bursa	15
Çankırı	11
Denizli	10
Eskişehir	17

Çözüm:

Bu serinin mekân endekslerini hesaplayabilmek için önce verilen serinin aritmetik ortalamasının bulunması gerekir. Hatırlanacağı gibi, aritmetik ortalama, gözlem değerlerinin toplamının, gözlem sayısına oranıdır. Serinin aritmetik ortalaması

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{12+15+11+10+17}{5} = ₺13$$

olarak bulunur.

Mekân ile ilgili hesaplanacak endeks değeri için, (4.2) formül kullanılır.

Örneğin, Ankara ili için endeks değeri,

$$I_{Ankara} = \frac{X_{Ankara}}{\overline{X}} \cdot 100 = \frac{12}{13} .100 = 92,31$$

Bursa ili için endeks değeri,

$$I_{Bursa} = \frac{X_{Bursa}}{\overline{X}} \cdot 100 = \frac{15}{13}.100 = 115,38$$

bulunur.

Sinema bilet fiyatı, tabloda verilen beş ilin ortalama fiyatına göre,

Ankara'da: 92,31-100 = -7,69 ; %7,69 daha düşüktür,

Bursa'da: 115,38-100 = 15,38 ; %15,38 daha yüksektir şeklinde yorumlanır.

SIRA SİZDE 1

Tablo 4.1'deki sinema bilet fiyatlarını kullanarak Eskişehir için endeks değerini hesaplayınız ve bulduğunuz değeri yorumlayınız.

Zaman Endeksi

Zaman endeksi, incelenen zaman serisinde, ilgilenilen değişkenin (veya değişkenlerin) gözlem değerlerinin zaman içindeki değişiminin oransal bir ifadesidir. Zaman endeksi hesaplanırken herhangi bir zaman dönemine ilişkin gözlem değeri temel devre değeri olarak alınır ve diğer gözlem değerlerindeki değişme bu sabit dönem değerine göre hesaplanır. Bunun yanında, her yeni endeks hesaplanırken temel devre değeri değiştirilebilir. Birinci durumda sabit esaslı endeks, ikinci durumda değişken endeks söz konusu olur.

Zaman endeksi, zaman serisinde, ilgilenilen değişkenin (veya değişkenlerin) zaman içindeki değişiminin oransal bir göstergesidir.

Ayrıca, zaman endeksi bir tek değişken için hesaplanabileceği gibi birden fazla değişkenin birlikte değişimi olarak da hesaplanabilir.

İzleyen kısımlarda incelenecek endeks türleri, zaman serileri için hesaplanacak endeks türleridir.

Sabit ve Değişken Esaslı Endeksler

Sabit Esaslı Endeksler

Zaman serisinin gözlem değerlerinin, belirlenen sabit bir dönem değerine (temel devre değeri) göre değişimleri önem kazandığında **sabit esaslı endeks** kullanılır.

Sabit esaslı endeksler, zaman serisindeki dönemlere ilişkin değişken değerlerinin, belirlenen sabit bir dönem değerine (temel devre değeri) göre değişimlerini gösteren oransal bir göstergesidir.

Sabit esaslı endeks (S.E.E.), $I_{i/0}$ simgesiyle gösterilir ve aşağıdaki eşitlik yardımıyla hesaplanır:

$$I_{i/0} = \frac{X_i}{X_0} \cdot 100 \tag{4.3}$$

Burada;

X_i= zaman serisinin endeks değeri hesaplanacak gözlem değerini

X_0= zaman serisinin belirlenen temel devre değerini

gösterir.

ÖRNEK 2

A lojistik şirketinin, 2005 - 2011 yıllarına ilişkin kamyon sayıları aşağıdaki tabloda verilmiştir. 2005 yılı temel devre olmak üzere 2006 ve 2007 yıllarına ilişkin sabit esaslı endeks değerlerini hesaplayınız ve yorumlayınız.

Tablo 4.2
A lojistik şirketinin, 2005 - 2011 yıllarına ilişkin kamyon sayıları

Yıllar	(Kamyon Sayıları)
2005	47
2006	65
2007	25
2008	52
2009	58
2010	68
2011	64

Çözüm:

Verilen tabloda, 2005 yılı temel devre olduğu için bu yıla 0 ve izleyen yıllara da sırasıyla 1, 2, 3, 4, 5 ve 6 devre numaraları verilir ve (4.3) eşitliği uygulanır. Buna göre; 2006 yılı için S.E.E. değeri,

$$I_{1/0} = \frac{X_1}{X_0} \cdot 100 = \frac{65}{47} \cdot 100 = 138,29$$

$$138,29 - 100 = 38,29$$

olarak hesaplanır. Fark pozitif bir değer olduğu için “2006 yılındaki kamyon sayısı, 2005 yılındaki kamyon sayısına göre % 38,29 artmıştır.” şeklinde yorumlanır.

$$I_{2/0} = \frac{X_2}{X_0} \cdot 100 = \frac{25}{47} \cdot 100 = 53,19$$

$$53,19 - 100 = -46,81$$

olarak hesaplanır. Fark negatif bir değer olduğu için, “2007 yılındaki kamyon sayısı, 2005 yılındaki kamyon sayısına göre % 46,81 azalmıştır.” şeklinde yorumlanır.

SIRA SİZDE 2

Örnek 2’deki verileri kullanarak 2010 yılı için sabit esaslı endeks değerini hesaplayınız ve bulduğunuz değeri yorumlayınız.

Değişken Esaslı Endeksler

Temel devre, her yeni endeks değeri hesaplanırken değişiyorsa hesaplanan endeks **değişken esaslı endekstir.**

Zaman serisinin herhangi bir zaman dönemine ilişkin gözlem değerindeki artış veya azalış bilgisi, *bir önceki zaman dönemi* gözlem değerine göre elde edilmesi isteniyorsa **değişken esaslı endeks** hesaplanır.

Değişken esaslı endeks hesaplamasında temel devre değeri, endeks değeri hesaplanacak devreden bir önceki devrenin değeridir.

Değişken esaslı endeks (D.E.E.), aşağıdaki eşitlik yardımıyla hesaplanır:

$$I_{i/i-1} = \frac{X_i}{X_{i-1}} \cdot 100 \qquad (4.4)$$

X_i = *i. devre değeri*

X_{i-1}: (*i* -1). *devre değeri*

$I_{i/i-1}$: *i* -1’*inci devre DEE değeri*

ÖRNEK 3

Bir banka işletmesinin 2006 - 2010 yıllarına ilişkin vergi matrahları aşağıda verilmiştir. Değişken esaslı endeks değerlerini hesaplayınız ve yorumlayınız.

Yıllar	X_i (Bin ₺)
2006	25
2007	47
2008	65
2009	52
2010	58

Tablo 4.3
Bir banka işletmesinin 2006 - 2010 yıllarına ilişkin vergi matrahları

Çözüm:

Verilen tabloda, 2006 yılı temel devre olduğu için bu yıla 0 ve izleyen yıllara da sırasıyla 1, 2, 3, 4, 5 ve 6 devre numaraları verilir.

Buna göre; 2006 yılı değişken esaslı endeks değeri, kendinden önceki dönem olmadığı için hesaplanamaz.

2007 yılı (1.devre) için değişken esaslı endeks,

$$I_{1/0} = \frac{X_1}{X_0} \cdot 100 = \frac{47}{25} \cdot 100 = 188$$

olarak hesaplanır.

$$188 - 100 = 88$$

Bu değerin yorumu, “2007 yılında banka işletmesinin vergi matrahı, 2006 yılına göre % 88 artmıştır.” şeklinde yapılır.

2008 yılı (2.devre) için değişken esaslı endeks,

$$I_{2/1} = \frac{X_2}{X_1} \cdot 100 = \frac{65}{47} \cdot 100 = 138,29$$

olarak hesaplanır.

$$138,29 - 100 = 38,29$$

Bu değerin yorumu, “2008 yılında banka işletmesinin vergi matrahı, 2007 yılına göre % 38,29 artmıştır.” şeklinde yapılır.

2009 yılı (3.devre) için değişken esaslı endeks,

$$I_{3/2} = \frac{X_3}{X_2} \cdot 100 = \frac{52}{65} \cdot 100 = 80$$

$$80 - 100 = -20$$

Bu değerin yorumu, “2009 yılında banka işletmesinin vergi matrahı, 2008 yılına göre % 20 azalmıştır.” şeklinde yapılır.

$$I_{4/3} = \frac{X_4}{X_3} \cdot 100 = \frac{58}{52} \cdot 100 = 111{,}53$$

$$111{,}53 - 100 = 11{,}53$$

Bu değerin yorumu, "2010 yılında banka işletmesinin vergi matrahı, 2009 yılına göre % 11,53 artmıştır." şeklinde yapılır.

SIRA SİZDE 3

Bir şirkette yapılan uluslararası telefon konuşmalarının (çıkan) süreleri 2004-2009 yılları itibari ile verilmiştir. 2004 yılına ait değeri temel kabul ederek 2007 yılına ait sabit esaslı endeks değerini hesaplayınız ve hesapladığınız endeks değerini yorumlayınız.

Yıllar	Uluslararası Konuşmalar (1000 dakika)
2004	715
2005	721
2006	515
2007	490
2008	520
2009	496

Sabit Esaslı Endeksler ve Değişken Esaslı Endeksler Arasında Geçiş

Eğer çalışmada, zaman serisi gerçek gözlem değerleri yerine, sabit esaslı endeks değerleri verilip dönemlere göre değişken esaslı endeks değerleri istendiğinde veya tersine, değişken esaslı endeks değerleri verilip sabit esaslı endeks değerleri isteniyorsa, bu durumda her bir endeksten bir diğerine geçiş yapılabilir.

Sabit esaslı endeksten, değişken esaslı endekse geçiş için aşağıdaki formül kullanılır:

$$I_{i/i-1} = \frac{I_{i/0}}{I_{i-1/0}} \cdot 100 \tag{4.5}$$

Burada;

$I_{i/i-1}$: *i. devre DEE değeri*

$I_{i/0}$: *i. devre SEE değeri*

$I_{i-1/0}$: *(i -1). devre SEE değeri*

ÖRNEK 4

Aşağıda tabloda verilen sabit esaslı endeks değerlerini kullanarak değişken esaslı endeks değerleri bulunuz.

Yıllar	Devre	S.E.E. (%)
2005	0	100
2006	1	188
2007	2	266
2008	3	208
2009	4	232
2010	5	272
2011	6	256

Tablo 4.4
2005 - 2011 yıllarına ilişkin sabit esaslı endeksler

2006 yılı için D.E.E. : $I_{1/0} = \frac{I_{1/0}}{I_{0/0}} \cdot 100 = \frac{188}{100} \cdot 100 = 188$

2007 yılı için D.E.E. : $I_{2/1} = \frac{I_{2/0}}{I_{1/0}} \cdot 100 = \frac{266}{188} \cdot 100 = 141,48$

2008 yılı için D.E.E. : $I_{3/2} = \frac{I_{3/0}}{I_{2/0}} \cdot 100 = \frac{208}{266} \cdot 100 = 78,19$

2009 yılı için D.E.E. : $I_{4/3} = \frac{I_{4/0}}{I_{3/0}} \cdot 100 = \frac{232}{208} \cdot 100 = 111,53$

2010 yılı için D.E.E. : $I_{5/4} = \frac{I_{5/0}}{I_{4/0}} \cdot 100 = \frac{272}{232} \cdot 100 = 117,24$

2011 yılı için D.E.E. : $I_{6/5} = \frac{I_{6/0}}{I_{5/0}} \cdot 100 = \frac{256}{272} \cdot 100 = 94,11$

Değişken esaslı endeksten sabit esaslı endekse geçiş:

$$I_{i/0} = \frac{I_{i/i-1} \cdot I_{i-1/i-2} \cdot \ldots \cdot I_{2/1} \cdot I_{1/0}}{(100)^i} \cdot 100 \qquad (4.6)$$

veya,

$$I_{i/0} = \frac{I_{i/i-1} \cdot I_{i-1/0}}{100}$$

ÖRNEK 5

Aşağıda tabloda verilen değişken esaslı endeks değerlerini kullanarak sabit esaslı endeks değerleri bulunuz.

Yıllar	Devre	D.E.E. (%)
2005	0	100
2006	1	188
2007	2	139,29
2008	3	80
2009	4	111,53
2010	5	117,24
2011	6	94,11

Tablo 4.5
2005 - 2011 yıllarına ilişkin değişken esaslı endeksler

$$2006 \text{ yılı için S.E.E.}: I_{1/0} = \frac{I_{1/0} \cdot I_{0/0}}{100} = \frac{188 \cdot 100}{100} = 188$$

$$2007 \text{ yılı için S.E.E.}: I_{2/0} = \frac{I_{2/1} \cdot I_{1/0}}{100} = \frac{(138,29) \cdot (188)}{(100)} = 259,98$$

$$2008 \text{ yılı için S.E.E.}: I_{3/0} = \frac{I_{3/2} \cdot I_{2/0}}{100} = \frac{(80) \cdot (259,98)}{(100)} = 207,98$$

$$2009 \text{ yılı için S.E.E.}: I_{4/0} = \frac{I_{4/3} \cdot I_{3/0}}{100} = \frac{(111,53) \cdot (207,98)}{(100)} = 231,96$$

$$2010 \text{ yılı için S.E.E.}: I_{5/0} = \frac{I_{5/4} \cdot I_{4/0}}{100} = \frac{(117,24) \cdot (231,96)}{(100)} = 271,94$$

$$2011 \text{ yılı için S.E.E.}: I_{6/0} = \frac{I_{6/5} \cdot I_{5/0}}{100} = \frac{(94,11) \cdot (271,94)}{(100)} = 255,92$$

SIRA SİZDE 4

Aşağıdaki tabloda, bir müzenin ziyaretçi sayısına ilişkin sabit esaslı endeks değerleri verilmiştir. Buna göre, değişken esaslı endeks değerlerini hesaplayınız. (Temel devre = ocak ayı alınmıştır.)

Aylar	S.E.E. (%)
Ocak	100
Şubat	120
Mart	90
Nisan	140
Mayıs	95

Esas Devrenin Değiştirilmesi

Sabit esaslı endeks ile çalışıldığında, temel devre olarak ele alınan devre çalışmanın amacına göre yeniden düzenlenebilir. Temel devrenin değiştirilmesi, farklı temel devreler esas alınarak aynı konuda hesaplanan endeks serilerini, tek bir temel devreye dönüştürülmesi ihtiyacı olması durumunda kullanılır.

Temel devrenin değiştirilmesi şöyle yapılır:

$$I_{y/e} = \frac{I_{y/0}}{I_{e/0}} \cdot 100$$

y: yeni temel devre

e: eski temel devre

ÖRNEK 6

Tablo 6'da sabit esaslı endeks değerleri 2005 yılı temel devre alınarak hesaplanmıştır. Temel devrenin 2008 yılı olması istendiğinde yeni oluşan sabit esaslı endeks değerlerini bulunuz.

Yıllar	Devre	S.E.E. (%)
2005	0	100
2006	1	188
2007	2	266
2008	3	208
2009	4	232
2010	5	272
2011	6	256

Tablo 4.6
2005 - 2011 yıllarına ilişkin sabit esaslı endeksler (temel devre = 2005)

2005 yılı için yeni S.E.E. : $I_{0/3} = \dfrac{I_{0/0}}{I_{3/0}} \cdot 100 = \dfrac{100}{208} \cdot 100 = 48,08$

2006 yılı için yeni S.E.E. : $I_{1/3} = \dfrac{I_{1/0}}{I_{3/0}} \cdot 100 = \dfrac{188}{208} \cdot 100 = 90,38$

2007 yılı için yeni S.E.E. : $I_{2/3} = \dfrac{I_{2/0}}{I_{3/0}} \cdot 100 = \dfrac{266}{208} \cdot 100 = 127,88$

2008 yılı için yeni S.E.E. : $I_{3/3} = \dfrac{I_{3/0}}{I_{3/0}} \cdot 100 = \dfrac{208}{208} \cdot 100 = 100$

2009 yılı için yeni S.E.E. : $I_{4/3} = \dfrac{I_{4/0}}{I_{3/0}} \cdot 100 = \dfrac{232}{208} \cdot 100 = 111,54$

2010 yılı için yeni S.E.E. : $I_{5/3} = \dfrac{I_{5/0}}{I_{3/0}} \cdot 100 = \dfrac{272}{208} \cdot 100 = 130,77$

2011 yılı için yeni S.E.E. : $I_{6/3} = \dfrac{I_{6/0}}{I_{3/0}} \cdot 100 = \dfrac{256}{208} \cdot 100 = 123,08$

Basit ve Bileşik Esaslı Endeksler

Endeksler, bir veya birden fazla değişken için hesaplanabilir. Eğer, tek bir değişkene ilişkin oransal değişimler hesaplanıyorsa bu bir **basit endeks**tir. Şu ana kadar hesaplanan endekslerin tümü basit endekslere örnektir. Endeks hesaplamada; endeksi hesaplanacak değişken fiyat olduğunda hesaplanan endeks, "basit fiyat endeksi", miktar olduğunda ise "basit miktar endeksi" adını alır.

Tek bir değişkene ilişkin oransal değişimleri gösteren endekse **basit endeks** denir.

Ancak günümüzde uygulamada yaygın olarak birçok değişkenin hesaplamaya alındığı endeksler de kullanılır. Örneğin; Türkiye İstatistik Kurumunun (TÜİK) her ay hesaplayıp duyurduğu Tüketici Fiyatları Endeksi (TÜFE), Üretici Fiyatları Endeksi (ÜFE) gibi.

Birden fazla maddenin fiyat ve/veya miktarlarında meydana gelen oransal değişimleri gösteren endekse bileşik endeks denir. Bileşik endekslerin hesaplanmasında, birçok maddenin hem fiyatı hem de miktarı veya her ikisi de değişken olarak kullanılır.

Basit endekste, sadece bir malın ya fiyatı ya da miktarının bileşik endekste, birden fazla malın fiyatı ve/veya miktarı kullanılır.

DİKKAT

Basit ve bileşik endeksin hesaplanmasında fiyatlar p_i ve miktarlar da q_i sembolü ile gösterilir. p_i ve q_i' deki i ' ler ele alınan devreyi gösterir.

p_0 *: temel devre fiyatı*

q_0 *: temel devre miktarı*

p_i *: devre fiyatı*

q_i *: devre miktarı*

ÖRNEK 7

A maddesinin 2008 - 2010 yıllarındaki fiyatları ve miktarları Tablo 7'de verilmiştir. Buna göre;
a. Basit fiyat endeks değerlerini hesaplayınız ve yorumlayınız.
b. Basit miktar endeks değerlerini hesaplayınız ve yorumlayınız.

Tablo 4.7
2008 - 2010 yıllarına ilişkin A maddesinin fiyatları ve miktarları

Yıllar	Devre	Fiyat (₺) p_i	Miktar (adet) q_i
2008	0	$p_0 = 60$	$q_0 = 110$
2009	1	$p_1 = 40$	$q_1 = 150$
2010	2	$p_2 = 70$	$q_2 = 100$

a. 2009 yılı için basit fiyat endeksi,

$$I_{1/0} = \frac{p_1}{p_0} \cdot 100 = \frac{40}{60} \cdot 100 = 66,66$$

A maddesinin fiyatı, 2009 yılında 2008 yılına göre %33,34 oranında azalmıştır.
2010 yılı için basit fiyat endeksi,

$$I_{2/0} = \frac{p_2}{p_0} \cdot 100 = \frac{70}{60} \cdot 100 = 116,66 \qquad 116,66 - 100 = 16,66$$

A maddesinin fiyatı, 2010 yılında 2008 yılına göre %16,66 oranında artmıştır.
b. 2009 yılı için basit miktar endeksi,

$$I_{1/0} = \frac{q_1}{q_0} \cdot 100 = \frac{150}{110} \cdot 100 = 136,36 \qquad 136,36 - 100 = 36,36$$

A maddesinin miktarı, 2009 yılında 2008 yılına göre %36,36 oranında artmıştır.
2010 yılı için basit miktar endeksi,

$$I_{2/0} = \frac{q_2}{q_0} \cdot 100 = \frac{100}{110} \cdot 100 = 90,90 \qquad 90,90 - 100 = -9,1$$

A maddesinin miktarı, 2010 yılında 2008 yılına göre %9,10 oranında azalmıştır.
Bileşik endeksin hesaplanmasında kullanılan teknikler izleyen başlıklarda toplanabilir:

1. Endeksler ortalaması
2. Ortalamalar endeksi
3. Laspeyres, Paasche ve Fisher endeksi

Endeksler Ortalaması

Endeksler ortalaması, endekse girecek madde sayısının birden fazla olduğu durumda bu maddelerin fiyatında veya miktarındaki oransal değişimi gösteren endekstir.

Endeksler Ortalaması şöyle hesaplanır:

1. Her bir madde için tek tek fiyat (veya miktar) endeksleri, sabit esaslı veya değişken esaslı endeks olarak hesaplanır.
2. Her devre için hesaplanan endekslerin aritmetik ortalaması alınır.

Birinci aşamada, sabit esaslı endeksler kullanıldıysa; ikinci aşamada hesaplanan endekse "sabit esaslı bileşik endeks" denir. Eğer, birinci aşamada değişken esaslı endeksler kullanıldıysa ikinci aşamada hesaplanan endekse de "değişken esaslı bileşik endeks" denir.

ÖRNEK 8

Aşağıda tabloda üç mala ilişkin fiyatlar verilmiştir. Endeksler ortalaması tekniğiyle sabit esaslı bileşik fiyat endeks değerlerini hesaplayınız ve yorumlayınız. (Temel devre = 2007)

Tablo 4.8
A, B ve C mallarına ilişkin 2007-2011 fiyatları

Yıllar	A (₺)	B (₺)	C (₺)
2007	75	50	33
2008	90	55	35
2009	100	60	40
2010	120	70	43
2011	110	75	45

Çözüm:

Öncelikle yapılması gereken, yıllara ilişkin devrelerin ve buna bağlı olarak da A, B ve C mallarının fiyatlarının devrelere göre değerlerinin (p_i' lerin : (i = 0, 1, 2, 3, 4)) oluşturulmasıdır.

Yıllar	Devre	A (₺)	B (₺)	C (₺)
2007	0	p_0=75	p_0=50	p_0=33
2008	1	p_1=90	p_1=55	p_1=35
2009	2	p_2=100	p_2=60	p_2=40
2010	3	p_3=120	p_3=70	p_3=43
2011	4	p_4=110	p_4=75	p_4=45

İkinci aşamada, her mal için ayrı ayrı sabit esaslı fiyat endeksleri bulunur.

Yıllar	Devre	A S.E.F.E. (%)	B S.E.F.E. (%)	C S.E.F.E. (%)
2007	0	100	100	100
2008	1	120	110	106.06
2009	2	133.34	120	121.21
2010	3	160	140	130.30
2011	4	146.47	150	136.36

Üçüncü aşamada, her devre (yıl) için A, B ve C mallarına ilişkin S.E.F.E.'nin aritmetik ortalamaları hesaplanarak sabit esaslı bileşik fiyat endeksi serisi bulunur.

$$I_{2007} = \frac{100+100+100}{3} = 100 \qquad (\text{-})$$

$$I_{2008} = \frac{120+110+106,06}{3} = 112,02 \qquad (\%\ 12,02\ \uparrow)$$

$$I_{2009} = \frac{133,34+120+121,21}{3} = 124,85 \qquad (\%\ 24,85\ \uparrow)$$

$$I_{2010} = \frac{160+140+130,30}{3} = 143,43 \qquad (\%\ 43,43\ \uparrow)$$

$$I_{2011} = \frac{146,67+150+136,36}{3} = 144,34 \qquad (\%\ 44,34\ \uparrow)$$

Tabloda verilen üç malın fiyatları, 2007 yılındaki fiyatlarına göre, ortalama olarak sırasıyla 2008 yılında %12,02, 2009 yılında %24,85, 2010 yılında %43,43 ve 2011 yılında %44,34 artış gösterir. Bu değişmeler okların yönüyle açıklanmıştır.

ÖRNEK 9

Tablo 8' de üç mala ilişkin verilen fiyatları kullanarak endeksler ortalaması tekniğiyle değişken esaslı bileşik fiyat endeks değerlerini hesaplayınız ve yorumlayınız.

Çözüm:

Öncelikle yapılması gereken yıllara ilişkin devrelerin ve buna bağlı olarak da A, B ve C mallarının fiyatlarının devrelere göre değerlerinin (p_i' lerin: (i = 0, 1, 2, 3, 4)) oluşturulmasıdır. İkinci aşamada, her mal için ayrı ayrı sabit esaslı fiyat endeksleri bulunur.

Yıllar	Devre	A D.E.F.E. (%)	B D.E.F.E. (%)	C D.E.F.E. (%)
2007	0	100	100	100
2008	1	120	110	106,06
2009	2	111,11	109,09	114,28
2010	3	120	116,67	107,5
2011	4	91,67	107,14	104,65

Üçüncü aşamada, her devre (yıl) için A, B ve C mallarına ilişkin D.E.F.E.'nin aritmetik ortalamaları hesaplanarak sabit esaslı bileşik fiyat endeksi serisi bulunur.

$$I_{2007} = \frac{100+100+100}{3} = 100 \qquad (\text{-})$$

$$I_{2008} = \frac{120+110+106,06}{3} = 112,02 \qquad (\%\ 12,02\ \uparrow)$$

$$I_{2009} = \frac{111,11 + 109,09 + 114,28}{3} = 111,49 \qquad (\%\ 11,49\ \uparrow)$$

$$I_{2010} = \frac{120 + 116,67 + 107,5}{3} = 114,72 \qquad (\%\ 14,72\ \uparrow)$$

$$I_{2011} = \frac{91,67 + 107,14 + 104,65}{3} = 101,15 \qquad (\%\ 1,15\ \uparrow)$$

Tabloda verilen üç malın fiyatları, bir önceki yılın fiyatlarına göre ortalama olarak 2008 yılında %12,02; 2009 yılında %11,49; 2010 yılında %14,72 ve 2011 yılında %1,15 oranında artış gösterir.

Ortalamalar Endeksi

Birden fazla maddenin fiyatının veya miktarının kullanıldığı diğer bir bileşik endeks yaklaşımıdır. Bu teknikte birden fazla seri tek bir seri hâline getirilir ve bunun üzerinden endeks değerlerine ulaşılır.

Ortalamalar endeksi şöyle hesaplanır:

1. Her devre için farklı maddelerin aldığı değerlerin aritmetik ortalamaları bulunarak tek bir değişken için seri oluşturulur.
2. Bu seri için ya sabit esaslı bileşik endeks ya da değişken esaslı bileşik endeks hesaplanır ve yorumlanır.

ÖRNEK 10

Tablo 8'de üç mala ilişkin verilen fiyatları kullanarak, ortalamalar endeksi tekniğiyle sabit esaslı bileşik fiyat endeksini hesaplayınız ve yorumlayınız.

Çözüm:

Öncelikle yapılması gereken yıllara ilişkin devrelerin ve buna bağlı olarak da A, B ve C mallarının fiyatlarının devrelere göre değerlerinin (p_i' lerin: (i = 0, 1, 2, 3, 4)) oluşturulmasıdır.

Yıllar	Devre	A (₺)	B (₺)	C (₺)
2007	0	p_0=75	p_0=50	p_0=33
2008	1	p_1=90	p_1=55	p_1=35
2009	2	p_2=100	p_2=60	p_2=40
2010	3	p_3=120	p_3=70	p_3=43

İkinci aşamada, her devre için üç malın ortalama fiyatları hesaplanır.

$$\bar{X}_{2007} = \frac{75 + 50 + 33}{3} = 52,67$$

$$\bar{X}_{2008} = \frac{90 + 55 + 35}{3} = 60$$

$$\bar{X}_{2009} = \frac{100 + 60 + 40}{3} = 66,67$$

$$\bar{X}_{2010} = \frac{120 + 70 + 43}{3} = 77,67$$

$$\bar{X}_{2011} = \frac{110 + 75 + 45}{3} = 76,67$$

Yıllar	Devre	Ortalama Fiyat (₺)
2007	0	$\bar{P}_0 = 52,67$
2008	1	$\bar{P}_1 = 60$
2009	2	$\bar{P}_2 = 66,67$
2010	3	$\bar{P}_3 = 77,67$
2011	4	$\bar{P}_4 = 76,67$

Son aşamada, oluşturulan ortalama fiyat serisi için sabit esaslı bileşik fiyat endeks değerleri bulunur.

$$SEBFE_{2007} = \frac{52,67}{52,67} \cdot 100 = \%100$$

$$SEBFE_{2008} = \frac{60}{52,67} \cdot 100 = \%113,92 \qquad (\%\ 13,92\ \uparrow)$$

$$SEBFE_{2009} = \frac{66,67}{52,67} \cdot 100 = \%126,58 \qquad (\%\ 26,58\ \uparrow)$$

$$SEBFE_{2010} = \frac{77,67}{52,67} \cdot 100 = \%147,46 \qquad (\%\ 47,46\ \uparrow)$$

$$SEBFE_{2011} = \frac{76,67}{52,67} \cdot 100 = \%145,57 \qquad (\%\ 45,57\ \uparrow)$$

3 maddenin 2008 yılındaki ortalama fiyatı, 2007 yılına göre %13,92; 2009 yılındaki ortalama fiyatı, 2007 yılına göre %26,58; 2010 yılındaki ortalama fiyatı ise 2007 yılına göre %47,46 ve 2011 yılındaki ortalama fiyatı ise 2007 yılına göre %45,57 artmıştır.

ÖRNEK 11

Tablo 8' de üç mala ilişkin verilen fiyatları kullanarak, ortalamalar endeksi tekniğiyle değişken esaslı bileşik fiyat endeksini hesaplayınız ve yorumlayınız.

Çözüm:

Öncelikle yapılması gereken yıllara ilişkin devrelerin ve buna bağlı olarak da A, B ve C mallarının fiyatlarının devrelere göre değerlerinin ('lerin : (i = 0, 1, 2, 3, 4)) oluşturulmasıdır.

Yıllar	Devre	A (₺)	B (₺)	C (₺)
2007	0	p_0=75	p_0=50	p_0=33
2008	1	p_1=90	p_1=55	p_1=35
2009	2	p_2=100	p_2=60	p_2=40
2010	3	p_3=120	p_3=70	p_3=43
2011	4	p_4=110	p_4=75	p_4=45

İkinci aşamada, her devre için üç malın ortalama fiyatları hesaplanır.

$$\bar{X}_{2007} = \frac{75+50+33}{3} = 52,67$$

$$\bar{X}_{2008} = \frac{90+55+35}{3} = 60$$

$$\bar{X}_{2009} = \frac{100+60+40}{3} = 66,67$$

$$\bar{X}_{2010} = \frac{120+70+43}{3} = 77,67$$

$$\bar{X}_{2011} = \frac{110+75+45}{3} = 76,67$$

Yıllar	Devre	Ortalama Fiyat (₺)
2007	0	$\bar{P}_0 = 52{,}67$
2008	1	$\bar{P}_1 = 60$
2009	2	$\bar{P}_2 = 66{,}67$
2010	3	$\bar{P}_3 = 77{,}67$
2011	4	$\bar{P}_4 = 76{,}67$

Son aşamada, oluşturulan ortalama fiyat serisi için değişken esaslı bileşik fiyat endeks değerleri bulunur.

$$SEBFE_{2007} = \frac{52,67}{52,67} \cdot 100 = \%100 \qquad (-)$$

$$SEBFE_{2008} = \frac{60}{52,67} \cdot 100 = \%113,92 \qquad (\%\ 13,92\ \uparrow)$$

$$DEBFE_{2009} = \frac{66,67}{60} \cdot 100 = \%112,12 \qquad (\%\ 11,12\ \uparrow)$$

$$DEBFE_{2010} = \frac{77,67}{66,67} \cdot 100 = \%116,50 \qquad (\%\ 16,50\ \uparrow)$$

$$DEBFE_{2011} = \frac{76,67}{77,67} \cdot 100 = \%98,71 \qquad (\%\ 1,29 \downarrow)$$

3 maddenin 2008 yılındaki ortalama fiyatı, 2007 yılına göre %13,92; 2009 yılındaki ortalama fiyatı, 2008 yılına göre %11,12; 2010 yılındaki ortalama fiyatı ise 2009 yılına göre %16,50 artmıştır. 2011 yılındaki ortalama fiyatı ise 2010 yılına göre %1,29 azalmıştır.

Laspeyres, Paasche ve Fisher Endeksi

Bileşik endekslerin hesaplanmasında kullanılan ilk iki teknikte, ele alınan maddenin fiyatları veya miktarları kullanılırken maddenin önemine göre herhangi bir tartı verilmemiştir. Her malın fiyatına veya miktarına aynı derecede önem (tartı!) verilmiştir. Oysa yapılan çalışmalarda, her malın fiyatı veya miktarı aynı derecede önem taşımayabilir. Bu nedenle, ele alınan malların fiyatlarına ilişkin bileşik endeks hesaplamasında miktarlar tartı olarak (veya malların miktarlarına ilişkin bileşik endeks hesaplamasında fiyatlar tartı olarak) kullanılır.

Basit endekslerin tartılı aritmetik ortalaması üç teknikle hesaplanır:

1. Laspeyres Endeksi
2. Paasche Endeksi
3. Fisher Endeksi

Laspeyres Endeksi

Bileşik endeks hesaplamasında tartı olarak temel devre değerlerini (fiyat veya miktar) alan endekse **Laspeyres** (fiyat veya miktar) **endeksi** denir.

Bileşik endeks hesaplamasında tartı olarak temel devre değerleri (fiyat veya miktar) kullanılıyorsa bu endekse **laspeyres** (fiyat veya miktar) **endeksi** denir. Laspeyres fiyat endeksinde, tartı olarak temel devre miktarı kullanılırken laspeyres miktar endeksinde, tartı olarak temel devre fiyatı kullanılır. Buna göre;

Laspeyres fiyat endeksi:

$$_{p}I_{L} = \frac{\sum(p_i \cdot q_0)}{\sum(p_0 \cdot q_0)} \cdot 100$$

Laspeyres miktar endeksi:

$$_{q}I_{L} = \frac{\sum(p_0 \cdot q_i)}{\sum(p_0 \cdot q_0)} \cdot 100$$

p_0: *temel devre fiyatı*

q_0: *temel devre miktarı*

p_i: *devre fiyatı*

q_i: *devre miktarı*

Paasche Endeksi

Bileşik endeks hesaplamasında, tartı olarak hesaplaması yapılan devre değerlerini (fiyat veya miktar) alan endekse **Paashe** (fiyat veya miktar) **endeksi** denir.

Bileşik endeks hesaplamasında, tartı olarak hesaplaması yapılan devre değerleri (fiyat veya miktar) kullanılıyorsa bu endekse **paashe** (fiyat veya miktar) **endek-**

si denir. Paasche fiyat endeksinde, tartı olarak hesaplaması yapılan devre miktarı kullanılırken paasche miktar endeksinde, tartı olarak hesaplaması yapılan fiyatı kullanılır. Buna göre;

Paasche fiyat endeksi:

$$_pI_p = \frac{\sum (p_i \cdot q_i)}{\sum (p_0 \cdot q_i)} \cdot 100$$

Paasche miktar endeksi:

$$_qI_p = \frac{\sum (p_i \cdot q_i)}{\sum (p_i \cdot q_0)} \cdot 100$$

Fisher Endeksi

Laspeyres ve paasche endeksleri, tartı olarak farklı devre değerlerini kullandıkları için farklı sonuçlar verir. Laspeyres endeksi, tartı olarak temel devre değerlerini kullandığı için fiyat ya da miktar artışlarını olduğundan fazla gösterir. Buna karşın, paasche endeksi de tartı olarak hesaplanan endeks devresi değerlerini kullandığından, fiyat veya miktar artışlarını olduğundan daha az gösterir. Bu nedenle, her iki indeksinde içerdiği hatayı azaltmak amacıyla laspeyres ve paasche indekslerinin geometrik ortalaması alınarak **fisher endeksi** önerilmiştir.

Laspeyres ve paasche indekslerinin geometrik ortalaması alınarak hesaplanan endekse **fisher endeksi** denir.

Fisher fiyat endeksi:

$$_pI_F = \sqrt{_pI_L \cdot {_pI_P}}$$

Fisher fiyat endeksi:

$$_qI_F = \sqrt{_qI_L \cdot {_qI_P}}$$

ÖRNEK 12

Aşağıdaki tabloda, belli bir yöredeki 2009 ve 2010 yıllarına ilişkin ürünlerin üretim miktarları ile bunlara karşılık gelen fiyatları verilmiştir. 2009 yılını temel yıl alarak 2010 yılı laspeyres fiyat endeksini hesaplayınız ve yorumlayınız.

	2009 Yılı		**2010 Yılı**	
Ürünler	Fiyat (₺)	Miktar (ton)	Fiyat (₺)	Miktar (ton)
A	25	50	30	60
B	20	60	25	75
C	30	40	25	50
D	15	30	20	25

Tablo 4.9
A, B, C ve D mallarına ilişkin fiyatlar ve miktarlar

Çözüm:

Laspeyres fiyat endeksi,

$$_pI_L = \frac{\sum (p_i \cdot q_0)}{\sum (p_0 \cdot q_0)} \cdot 100$$

formülü ile hesaplanmaktadır. İlgili işlemler aşağıdaki tabloda gösterilmiştir:

	2009 Yılı		2010 Yılı			
Ürünler	Fiyat (₺) p_0	Miktar (ton) q_0	Fiyat (₺) $p1$	Miktar (ton) $q1$	p_1q_0	p_0q_0
A	25	50	30	60	1500	1250
B	20	60	25	75	1500	1200
C	30	40	25	50	1000	1200
D	15	30	20	25	600	450
					4600	4100

Tablodaki değerlerden yararlanılarak laspeyres fiyat endeksi

$$_pI_L = \frac{\sum(p_i \cdot q_0)}{\sum(p_0 \cdot q_0)} \cdot 100 = \frac{4600}{4100} \cdot 100 = 112,20$$

olarak bulunur.

Söz konusu 4 ürünün fiyatları, 2010 yılında 2009 yılına göre ortalama olarak %12,20 oranında artmıştır.

ÖRNEK 13 *Tablo 9'a göre, 2009 yılını temel yıl alarak 2010 yılı paache fiyat endeksini hesaplayınız ve yorumlayınız.*

Çözüm:

Paache fiyat endeksi,

$$_pI_L = \frac{\sum(p_i \cdot q_0)}{\sum(p_0 \cdot q_0)} \cdot 100$$

formülü ile hesaplanmaktadır. İlgili işlemler aşağıdaki tabloda gösterilmiştir:

	2009 Yılı		2010 Yılı			
Ürünler	Fiyat (₺)	Miktar (ton)	Fiyat (₺)	Miktar (ton)	p_1q_1	p_0q_1
A	25	50	30	60	1800	1500
B	20	60	25	75	1875	1500
C	30	40	25	50	1250	1500
D	15	30	20	25	500	375
					5425	4875

Tablodaki değerlerden yararlanılarak Paache fiyat endeksi

$$_pI_P = \frac{\sum(p_i \cdot q_i)}{\sum(p_0 \cdot q_i)} \cdot 100 = \frac{5425}{4875} \cdot 100 = 111,28$$

olarak bulunur.

Söz konusu 4 ürünün fiyatları, 2010 yılında 2009 yılına göre ortalama olarak %11,28 oranında artmıştır.

ÖRNEK 14

Tablo 9'a göre, 2009 yılını temel yıl alarak 2010 yılı fisher fiyat endeksini hesaplayınız ve yorumlayınız.

Çözüm:
Fisher fiyat endeksi,

$$_{p}I_{F} = \sqrt{\frac{\sum(p_1 q_0)}{\sum(p_0 q_0)} \cdot \frac{\sum(p_1 q_1)}{\sum(p_0 q_1)}} \cdot 100 = \sqrt{_{p}I_{L} \cdot {_{p}I_{P}}}$$

formülü ile hesaplanmaktadır. Bilindiği gibi fischer fiyat endeksi, laspeyres ve paache fiyat endekslerinin geometrik ortalamasıdır. Laspeyres ve paache fiyat endeksleri sırasıyla örnek 12 ve örnek 13'te bulunmuştu. Buna göre, fischer fiyat endeks değeri;

$$_{p}I_{F} = \sqrt{(112,20)(111,28)} = 111,74$$

olarak bulunur.
Söz konusu 4 ürünün fiyatları, 2010 yılında 2009 yılına göre ortalama olarak %11,74 oranında artmıştır.

ÖRNEK 15

Tablo 9'a göre, belli bir yöredeki 2009 ve 2010 yıllarına ilişkin ürünlerin üretim miktarları ile bunlara karşılık gelen fiyatları verilmiştir. 2009 yılını temel yıl alarak 2010 yılı laspeyres miktar endeksini hesaplayınız ve yorumlayınız.

Çözüm:
Laspeyres miktar endeksi,

$$_{q}I_{L} = \frac{\sum(p_0 \cdot q_i)}{\sum(p_0 \cdot q_0)} \cdot 100$$

formülü ile hesaplanmaktadır. İlgili işlemler aşağıdaki tabloda gösterilmiştir:

	2009 Yılı		2010 Yılı			
Ürünler	Fiyat (₺) p_0	Miktar (ton) q_0	Fiyat (₺) p_1	Miktar (ton) q_1	p_0q_0	p_0q_1
A	25	50	30	60	1250	1500
B	20	60	25	75	1200	1500
C	30	40	25	50	1200	1500
D	15	30	20	25	450	375
					4100	4875

Tablodaki değerlerden yararlanılarak laspeyres fiyat endeksi

$$_{q}I_{L} = \frac{\sum(p_0 \cdot q_1)}{\sum(p_0 \cdot q_0)} \cdot 100 = \frac{4875}{4100} \cdot 100 = 118,90$$

olarak bulunur.

Söz konusu 4 ürünün miktarları, 2010 yılında 2009 yılına göre ortalama olarak %18,90 oranında artmıştır.

ÖRNEK 16

Tablo 9'a göre, belli bir yöredeki 2009 ve 2010 yıllarına ilişkin ürünlerin üretim miktarları ile bunlara karşılık gelen fiyatları verilmiştir. 2009 yılını temel yıl alarak 2010 yılı paasche miktar endeksini hesaplayınız ve yorumlayınız.

Çözüm:

Paasche miktar endeksi,

$$_qI_P = \frac{\sum(p_1 \cdot q_1)}{\sum(p_1 \cdot q_0)} \cdot 100$$

formülü ile hesaplanmaktadır. İlgili işlemler aşağıdaki tabloda gösterilmiştir:

	2009 Yılı		2010 Yılı			
Ürünler	Fiyat (₺) p_0	Miktar (ton) q_0	Fiyat (₺) p_1	Miktar (ton) q_1	p_1q_1	p_1q_0
A	25	50	30	60	1800	1500
B	20	60	25	75	1875	1500
C	30	40	25	50	1250	1000
D	15	30	20	25	200	600
					5425	4600

Tablodaki değerlerden yararlanılarak paache fiyat endeksi

$$_qI_P = \frac{\sum(p_1 \cdot q_1)}{\sum(p_1 \cdot q_0)} \cdot 100 = \frac{5425}{4600} \cdot 100 = 117,93$$

olarak bulunur.

Söz konusu 4 ürünün miktarları, 2010 yılında 2009 yılına göre ortalama olarak %17,93 oranında artmıştır.

ÖRNEK 17

Tablo 9'a göre, belli bir yöredeki 2009 ve 2010 yıllarına ilişkin ürünlerin üretim miktarları ile bunlara karşılık gelen fiyatları verilmiştir. 2009 yılını temel yıl alarak 2010 yılı fisher miktar endeksini hesaplayınız ve yorumlayınız.

Çözüm:

Fisher miktar endeksi,

$$_qI_F = \sqrt{\frac{\sum(p_0q_1)}{\sum(p_0q_0)} \cdot \frac{\sum(p_1q_1)}{\sum(p_1q_0)}} \cdot 100 = \sqrt{_qI_L \cdot {_qI_P}}$$

formülü ile hesaplanmaktadır. Bilindiği gibi fischer miktar endeksi, laspeyres ve paache miktar endekslerinin geometrik ortalamasıdır. laspeyres ve paache miktar endeksleri sırasıyla örnek 15 ve örnek 16 'de bulunmuştu. Buna göre, fischer miktar endeks değeri;

$$_qI_F = \sqrt{(118,90)(117,93)} = 118,41$$

olarak bulunur.

Söz konusu 4 ürünün miktarları, 2010 yılında 2009 yılına göre ortalama olarak %18,41 oranında artmıştır.

Uygulamada Önemli Bileşik Endeksler

Bileşik endekslerden en çok bilineni Türkiye İstatistik Kurumu (TÜİK) tarafından her ay hesaplanan ve duyurulan "Tüketici Fiyatları Endeksi" (TÜFE) ve "Üretici Fiyatları Endeksi" (ÜFE)' dir. Bu endeksler enflasyon oranı olarak adlandırılmaktadır.

TÜFE, hane halklarının tüketime yönelik mal ve hizmet fiyatlarının zaman içindeki değişimini ölçen bir endekstir. Amacı, piyasada tüketime konu olan 447 maddenin, ay veya yıl içerisinde fiyatlarının oransal değişmelerinin ortalamasını belirlemektir. Bu endeksin hesaplanmasında, zincirleme laspeyres tekniği kullanılmaktadır.

ÜFE, ülke ekonomisinde üretimi yapılan ve yurt içinde satışa konu olan ürünlerin üretici fiyatlarının ay veya yıl içerisindeki değişmelerini ölçen endekstir.Bu endeks bugün için 785 madde ele alınarak hesaplanmaktadır. Hesaplamada kullanılan teknik, TÜFE hesaplamasında olduğu gibi, zincirleme laspeyres tekniğidir.

TÜFE ve ÜFE hesaplamasında ele alınan maddeler için, zaman içerisinde toplumun ekonomik, kültürel ve sosyal yapısındaki değişikliklere ve gelişmelere paralel olarak bu maddelerde değişiklikler yapılabilir.

Özet

AMAÇ 1

Endeks kavramlarını tanımlamak.

Bir değişken veya değişkenler grubunun, zamana veya mekâna göre oransal değişimlerini gösteren istatistiksel bir ölçüye endeks denir. Endekslerden oluşan seriye de endeks serisi denir.

AMAÇ 2

Endeks Kavramlarını kendi içinde türlerine ayırmak.

Endeksler, çalışmanın amacına, değişken türüne ve sayısına göre farklı tekniklerle hesaplanır. Bu hesaplama, bir değişken (yer, zaman, fiyat, miktar vb.) için olabileceği gibi birden fazla değişken içinde hesaplanabilir.

Mekân endeksi: İlgilenilen değişkenin mekânlara göre ortalamasından değişiminin oransal bir göstergesidir.

Zaman endeksi: Zaman serisinde, ilgilenilen değişkenin (veya değişkenlerin) zaman içindeki değişiminin oransal bir göstergesidir. Zaman endeksi hesaplanırken temel devre değeri sabit bir dönem değeri olabileceği gibi değişebilir. Ayrıca, zaman endeksi bir tek değişken için hesaplanabileceği gibi birden fazla değişkenin birlikte değişimi olarak da hesaplanabilir.

Sabit esaslı endeksler: Zaman serisindeki dönemlere ilişkin değişken değerlerinin, belirlenen sabit bir dönem değerine (temel devre değeri) göre değişimlerini gösteren oransal bir göstergesidir.

Değişken esaslı endeksler: Zaman serisindeki her dönem için artış veya azalışların bir önceki devreye göre gösteren endekslerdir.

Basit endeks: Tek bir değişkene ilişkin oransal değişimleri gösteren endekse basit endeks denir.

Bileşik endeks: İki ve ikiden fazla değişkene ilişkin oransal değişimleri gösteren endekse bileşik endeks denir.

Laspeyres endeksi: Bileşik endeks hesaplamasında tartı olarak temel devre değerlerini (fiyat veya miktar) alan endekse laspeyres (fiyat veya miktar) endeksi denir.

Paasche endeksi: Bileşik endeks hesaplamasında, tartı olarak hesaplaması yapılan devre değerlerini (fiyat veya miktar) alan endekse paashe (fiyat veya miktar) endeksi denir.

Fisher endeksi: Laspeyres ve paasche indekslerinin geometrik ortalaması alınarak hesaplanan endekse fisher endeksi denir.

Kendimizi Sınayalım

1. Bir değişkenin ya da değişkenlerin, zaman veya mekâna göre değişimini gösteren istatistiksel ölçüye ne denir?

a. Ortalama
b. Standart Sapma
c. Varyans
d. Endeks
e. Değişim Katsayısı

2. A maddesinin 2009 yılındaki satış miktarı 100 ton, 2010 yılındaki satış miktarı 120 ton ve 2011 yılındaki satış miktarı 150 olarak verilmiştir. 2011 yılı için değişken esaslı miktar endeks değerinin yorumu aşağıdakilerden hangisidir?

a. A maddesinin 2011 yılındaki satış miktarı 2009 yılına göre %50 artmıştır.
b. A maddesinin 2011 yılındaki satış miktarı 2010 yılına göre %25 artmıştır.
c. A maddesinin 2011 yılındaki satış miktarı 2009 yılına göre %50 azalmıştır.
d. A maddesinin 2011 yılındaki satış miktarı 2009 yılına göre %25 azalmıştır.
e. A maddesinin 2011 yılındaki satış miktarı 2009 yılına göre %100 artmıştır.

3. Bir maddeye ilişkin olarak $I_{1/0}$ = %120, $I_{2/0}$=%160, $I_{3/0}$=%140 endeks değerleri verilmiştir. Buna göre, I3/2 endeks değeri kaçtır?

a. %125
b. %25
c. %12,5
d. % 87,5
e. %20 Bileşik

4. Birden fazla değişkenin ve/veya miktarının zaman içinde oransal değişimini gösteren endekse ne ad verilir?

a. Basit Endeks
b. Değişken Esaslı Endeks
c. Sabit Esaslı Endeksi
d. Mekân Endeksi
e. Bileşik Endeks

5. İzleyen tabloda 2009-2010 yılları için X ve Y mallarına ilişkin fiyatlar ve miktarlar verilmiştir. Buna göre 2009 yılı temel devre olmak üzere 2010 yılındaki laspeyres miktar endeks değeri aşağıdakilerden hangisidir?

	2009 Yılı		**2010 Yılı**	
Ürünler	Fiyat (₺) P_0	Miktar (kg) q_0	Fiyat (₺) P_1	Miktar (kg) q_1
X	400	20	440	25
Y	500	30	600	35

a. 117,35
b. 118,46
c. 119,56
d. 119,78
e. 120,69

6. 7. ve 8. soruları aşağıdaki bilgileri kullanarak cevaplayınız. 2010 yılı temel devre olmak üzere 6 maddeye ilişkin olarak;

$$\sum p_0 q_0 = 620 \qquad \sum p_0 q_1 = 590$$

$$\sum p_1 q_0 = 560 \qquad \sum p_1 q_1 = 520$$

verilmiştir. Buna göre;

6. Laspeyres fiyat endeks değeri kaçtır?

a. 88,14
b. 89,46
c. 90,32
d. 90,59
e. 91,44

7. Paasche miktar endeks değeri kaçtır?

a. 88,14
b. 89,22
c. 90,32
d. 91,67
e. 92,85

8. Fisher fiyat endeks değeri kaçtır?
a. 88,14
b. 89,22
c. 90,32
d. 91,59
e. 92,76

9. Üç maddeye ve istenilen bir yıla ilişkin laspeyres fiyat endeks değeri 155,45 ve paasche fiyat endeksi 147,42 olarak verilmiştir. Buna göre fisher fiyat endeks değeri aşağıdakilerden hangisidir?
a. 147,58
b. 148,63
c. 149,74
d. 150,21
e. 151,38

10. Beş maddeye ve istenilen bir yıla ilişkin fisher miktar endeks değeri 64 ve paasche miktar endeksi 60 olarak verilmiştir. Buna göre laspeyres miktar endeks değeri aşağıdakilerden hangisidir?
a. 47,58
b. 58,48
c. 67,88
d. 68,26
e. 69,32

Kendimizi Sınayalım Yanıt Anahtarı

1. d	Yanıtınız yanlış ise “Endeksler” konusunu gözden geçiriniz.
2. b	Yanıtınız yanlış ise “Mekân Endeksi” konusunu gözden geçiriniz.
3. d	Yanıtınız yanlış ise “Sabit Esaslı Endeksler” konusunu gözden geçiriniz.
4. e	Yanıtınız yanlış ise “Bileşik Esaslı Endekseler” konusunu gözden geçiriniz.
5. c	Yanıtınız yanlış ise “Laspeyres Endeksi” konusunu gözden geçiriniz.
6. c	Yanıtınız yanlış ise “Laspeyres Endeksi” konusunu gözden geçiriniz.
7. e	Yanıtınız yanlış ise “Paasche Endeksi” konusunu gözden geçiriniz.
8. b	Yanıtınız yanlış ise “Fisher Endeksi” konusunu gözden geçiriniz.
9. e	Yanıtınız yanlış ise “Bileşik Esaslı Endeksler” konusunu gözden geçiriniz.
10. d	Yanıtınız yanlış ise “Bileşik Esaslı Endeksler” konusunu gözden geçiriniz.

Sıra Sizde Yanıt Anahtarı

Sıra Sizde 1

$$I_{\text{Eskişehir}} = \frac{X_{\text{Eskişehir}}}{\overline{X}}.100 = \frac{17}{13}.100 = 130,76$$

144,68-100=44,68

Eskişehir'de sinema bilet fiyatı, verilen 5 ilin ortalama fiyatına göre %30,76 oranında daha yüksektir.

Sıra Sizde 2

2010 yılı için devre numarası 5 olduğuna göre,

$$I_{5/0} = \frac{X_5}{X_0} \cdot 100 = \frac{68}{47} \cdot 100 = 144,68$$

$$144,68 - 100 = 44,68$$

Fark pozitif bir değer olduğu için 2010 yılındaki kamyon sayısı 2005 yılındaki kamyon sayısına göre %44,68 artmıştır.

Sıra Sizde 3

2004 yılı temel devre olduğuna göre, 2007 yılının devre numarası 3 olur.Buna göre 2007 yılı için sabit esaslı endeks değeri;

$$I_{3/0} = \frac{X_3}{X_0} \cdot 100 = \frac{490}{715} \cdot 100 = 68,53$$

$$68,53 - 100 = -31,47$$

Fark negatif bir değer olduğu için 2007 yılında şirkette yapılan telefon konuşmalarının süresi 2004 yılına göre %31,47 azalmıştır.

Sıra Sizde 4

Temel devre ocak ayı olduğuna göre, ocak ayı devre numarası 0 olmak üzere izleyen tablo oluşturulur.

Aylar	Devre	S.E.E (%)
Ocak	0	100
Şubat	1	120
Mart	2	90
Nisan	3	140
Mayıs	4	95

Buna göre;
Şubat ayı için D.E.E. değeri:

$$I_{1/0} = \frac{I_{1/0}}{I_{0/0}} \cdot 100 = \frac{120}{100} \cdot 100 = 120$$

Mart ayı için D.E.E. değeri:

$$I_{2/1} = \frac{I_{2/0}}{I_{1/0}} \cdot 100 = \frac{90}{120} \cdot 100 = 75$$

Nisan ayı için D.E.E. değeri:

$$I_{3/2} = \frac{I_{3/0}}{I_{2/0}} \cdot 100 = \frac{140}{90} \cdot 100 = 155,55$$

Mayıs ayı için D.E.E. değeri:

$$I_{4/3} = \frac{I_{4/0}}{I_{3/0}} \cdot 100 = \frac{95}{140} \cdot 100 = 67,85$$

Yararlanılan ve Başvurulabilecek Kaynaklar

Atlas, Mahmut (2010). **İstatistik 1,** Ak Ofset Matbaacılık

Çömlekçi, Necla (1989). **İstatistik,** Bilim Teknik Yayınevi

Serper, Ö. (2004). **Uygulamalı İstatistik 1,** Ezgi Kitabevi.

www.tuik.gov.tr

5

Amaçlarımız

Bu üniteyi tamamladıktan sonra;

- Olasılık kavramını ve olasılığa ilişkin yaklaşımları tanımlayabilecek,
- Toplama kuralları, çarpma kuralları ve Bayes teoremi yardımıyla herhangi bir olayın gerçekleşme olasılığını hesaplayabilecek,
- Bir denemedeki olası sonuç sayısını belirlemek üzere saymanın temel ilkesini uygulayabilecek, permütasyon ve kombinasyon sayılarını hesaplayabilecek,
- Rassal değişken ve olasılık dağılımı kavramlarını açıklayabilecek, kesikli olasılık dağılımları için ortalama, varyans ve standart sapma hesaplayabilecek,
- Binom ve normal dağılımın özelliklerini açıklayabilecek ve bu dağılımlar yardımıyla olasılık hesaplayabileceksiniz.

Anahtar Kavramlar

- Olasılık
- Deneme
- Sonuç
- Örneklem Uzayı
- Olay
- Karşılıklı Ayrık Olaylar
- Bütüne Tamamlayan Olaylar
- Ortak Olasılık
- Bağımsızlık
- Koşullu Olasılık
- Bayes Teoremi
- Saymanın Temel İlkesi
- Permütasyon
- Kombinasyon
- Rassal Değişken
- Olasılık Dağılımı
- Binom Dağılımı
- Normal Dağılım
- Standart Normal Dağılım

İçindekiler

İstatistik → Olasılık Kuramı →

Olasılık Kuramı

GİRİŞ

Önceki ünitelerde açıklandığı üzere tanımlayıcı istatistik, geçmişte gerçekleşen olaylardan derlenen verilerin özetlenmesi ile uğraşır. İstatistiksel çıkarımda ise bir durumun ortaya çıkma şansının hesaplanmasıyla ilgilenildiği için, bu ünitede istatistiksel çıkarım için çok önemli bir kavram olan olasılık incelenecektir.

Bir olaya ya da duruma ilişkin karar verilmesi gerektiğinde, çoğu zaman tüm bilgilere sahip olamayız. İstatistikte çoğu zaman gözlenen veriler üzerinde çıkarımda bulunma ve belirsizlik durumunda karar alma sorunuyla ilgilenilir. İstatistiksel çıkarımda, bir evrenden çekilen örneklemden yararlanarak o evrene ilişkin çıkarımlarda bulunmayla ilgilenilir.

Karar vermede belirsizlik söz konusu olduğundan, konuya ilişkin tüm risklerin bilimsel olarak değerlendirilmesi önemlidir. Genellikle belirsizlik bilimi olarak adlandırılan olasılık kuramı ise bu değerlendirmeye yardımcı olur. Olasılık kuramı, karar vericinin sınırlı bilgi ile olası riskleri analiz etmesine ve bu riskleri en aza indirgemesine imkân tanır.

İstatistiksel çıkarımda olasılık kavramlarının çok önemli bir yeri olduğu için bu ünitede öncelikle olasılığın temel kavramları verilmiştir.

OLASILIK KAVRAMI

Olasılık, ihtimal ya da *şans* gibi terimler genellikle günlük yaşantılarımızda gelecekte ne ile karşılaşacağımız, başımıza nelerin gelebileceği gibi soruların cevaplanmaya çalışıldığı durumları ifade etmek üzere kullanılır. Olasılık genellikle, herhangi bir durumun ya da olayın gerçekleşme şansını ifade eden bir sayı olarak tanımlanır ve şansa bağlı olayın sonucuna ilişkin bir çıkarımda bulunma aracı olarak kullanılabilir. Sıfır ile bir kapalı aralığında (0 ve 1 dahil) değerler alan ve bir olayın ortaya çıkma şansını belirten sayıya **olasılık** adı verilir.

Olasılık, bir olayın ortaya çıkma şansını ifade eden, sıfır ile bir kapalı aralığında (0 ve 1 dahil) bir değerdir.

Olasılık değerleri, 0,50; 0,34 ya da 0,83 gibi ondalıklı sayılarla gösterilebileceği gibi, 4/7, 64/100 ya da 8/10 gibi kesirli sayılarla da ifade edilebilir. Olasılıklar; 0 ve 1 dahil olmak üzere, bu sınırlar arasında herhangi bir değer alabilir. Örneğin, bir sınıfta yirmi öğrenci bulunsun. Bu yirmi öğrencinin isimlerini *bir kağıda yazıp,* bu kağıtları bir torbaya koyduktan sonra rastgele bir çekiliş yaptığımızda, *bu sınıftaki yirmi öğrenciden birini çekme* olasılığı 1'dir. Çünkü, torbadaki yazılı kağıtların hepsinde bu yirmi öğrencinin isimleri vardır. Bu torbadan *yan sınıfta bulunan bir öğrencinin isminin çıkması* olasılığı ise 0 olur. Çünkü rastgele çekiliş yapılan torbada, yan sınıftaki öğrencilerin isimleri bulunmamaktadır. Buna göre, bir olayın gerçekleşmesi kesin ise olasılığı 1, gerçekleşmesi imkânsız ise olasılığı 0 olacaktır.

Olasılık değeri sıfıra yaklaştıkça olayın gerçekleşme şansı gittikçe azalıyor, olasılık değeri bire yaklaştıkça da o olayın gerçekleşme şansı gittikçe artıyor denir.

Olasılık incelemelerinde *deneme, sonuç, örneklem uzayı* ve *olay* olmak üzere dört temel kavram karşımıza çıkar.

Çeşitli olası gözlemlerden yalnızca birinin gerçekleşmesi ile sonuçlanan sürece deneme adı verilir. Bir denemenin iki ya da daha fazla olası sonucu bulunur ve bu sonuçlardan hangisinin gerçekleşeceği belirsizdir.

Bir denemenin sona erme biçimine sonuç adı verilir. Örneğin, tek bir bozuk paranın atılması bir denemedir. Paranın havaya atılışı gözlemlenebilir ancak "yazı" ya da "tura"dan hangisinin geleceği önceden bilinemez. Bu denemedeki olası sonuçlardan biri paranın yazı gelmesi, diğeri ise paranın tura gelmesidir. Dolayısıyla bu deneme için olası sonuç sayısı ikidir. Benzer olarak, bir zar atma denemesinde 1, 2, 3, 4, 5 ve 6 olmak üzere altı olası sonuç bulunur.

Bir denemenin tüm olası sonuçlarından oluşan kümeye örneklem uzayı adı verilir ve bu küme S harfi ile ifade edilir. Denemenin olası tüm sonuçlarına örneklem uzayının bir elemanı karşılık gelir. Bu elemana da *örneklem uzayı noktası* adı verilir.

Bir denemenin bir ya da daha fazla sonucundan oluşan kümeye ise bir **olay** adı verilir. Dolayısıyla, örneklem uzayının herhangi bir alt kümesi, bir olay olacaktır.

Tablo 5.1'de, deneme, sonuç, örneklem uzayı ve olaylara ilişkin bazı örnekler verilmektedir.

Tablo 5.1
Deneme, Sonuç, Örneklem Uzayı ve Olaylara İlişkin Bazı Örnekler

Deneme	Sonuçlar	Örneklem Uzayı	Olay
Tek bir zar atılması	1, 2, 3, 4, 5, 6	S = {1, 2, 3, 4, 5, 6}	Zarın 2'den büyük gelmesi
İki bozuk paranın birlikte atılması	YY, YT, TY, TT	S = {YY, YT, TY, TT}	Bir yazı, bir tura gelmesi
Çalışanların medeni durumlarının belirlenmesi	Evli, Bekâr	S = {Evli, Bekâr}	Çalışanın evli olması
Hisse senedinin günlük durumunun belirlenmesi	Artma, Azalma, Değişmeme	S = {Artma, Azalma, Değişmeme}	Hisse senedi değerinde artma olması

OLASILIK TANIMLARI

Olasılık tanımlarına ilişkin, nesnel (objektif) ve öznel (subjektif) olasılık olmak üzere iki yaklaşım bulunur. *Nesnel olasılık,* klasik olasılık ve deneysel olasılık olmak üzere iki kısımda incelenir. Klasik olasılık eşit olasılıklı sonuçlara dayanırken deneysel olasılık göreli sıklıklara dayanır. Öznel olasılık ise eldeki mevcut bilgilerden elde edilen olasılıktır.

Formülasyonlarda, herhangi bir A olayının gerçekleşme olasılığı P(A) ile belirtilecektir.

Klasik Olasılık

Klasik olasılık tanımı, bir denemenin sonuçlarının eşit olasılıklı olduğu varsayımına dayanır. Klasik bakış açısıyla, *bir olayın gerçekleşme olasılığı ilgilenilen sonuçların sayısının, olası tüm sonuçların sayısına bölünmesi yoluyla hesaplanır.* İlgilenilen sonuç sayısı n ile ve olası tüm sonuçların sayısı N ile belirtildiğinde, klasik olasılık yaklaşımıyla bir A olayının gerçekleşme olasılığı;

$$P(A) = \frac{n}{N} \tag{1}$$

eşitliği yardımıyla hesaplanabilir.

ÖRNEK 1

Bir zarın atılması denemesinde, "gelen zarın 4'den büyük olması" olayının gerçekleşme olasılığını bulunuz.

Çözüm:

Bu deneme için olası sonuçlar: 1, 2, 3, 4, 5 ,6 olarak belirlenir. Bu 6 eşit olasılıklı sonuç içerisinde ilgilenilen sonuçlar "5" ve "6" olup bunların sayısı 2'dir. A olayı; "Zarın 4'den büyük gelmesi" olarak tanımlandığında,

$$\mathrm{P(A)} = \frac{2}{6} = 0{,}33 \text{ olur.}$$

Önceki ünitelerde incelenen frekans dağılımlarında sınıflar oluşturulurken belirli bir olay yalnızca tek bir sınıfta yer alabiliyordu. Bu duruma göre, belli bir zamanda çeşitli olaylardan sadece bir tanesi gerçekleşebilir.

Bir olay gerçekleştiği anda diğer olayların hiçbiri gerçekleşmiyorsa bu olaylara karşılıklı ayrık olaylar adı verilir. Örneğin, "cinsiyet" değişkeninin erkek ve kadın olmak üzere karşılıklı ayrık iki sonucu bulunur. Bir şirkette rassal olarak seçilecek tek bir personel ya erkek ya da kadın olacaktır. İkisi birden olamayacaktır. Bir fabrikada üretilen bir ürün, kalite kontrol departmanında incelendikten sonra hatalı ya da hatasız ürün olarak kaydedilecektir. Aynı anda hem hatalı hem de hatasız ürün olamayacaktır. Üretilen ürünlerden oluşan bir örneklemde, hatalı bir ürün seçme olayı ile hatasız bir ürün seçme olayı karşılıklı ayrık olaylardır.

Eğer bir deneme, olası tüm sonuçları içeren bir olaylar kümesinden oluşuyorsa, bu olaylar kümesine bütüne tamamlayan olaylar adı verilir. Örneğin, zar atma denemesinde, "zarın 4'ten küçük gelmesi" olayı ile "zarın 3'ten büyük gelmesi" olayı bütüne tamamlayan olaylardır. Çünkü bu denemedeki tüm sonuçlar ya 4'ten küçük (1, 2 veya 3) ya da 3'ten büyük (4, 5 veya 6) olacaktır.

DİKKAT

Eğer bir olaylar kümesi bütüne tamamlayan olaylardan oluşuyorsa ve bu olaylar aynı zamanda karşılıklı ayrık olaylar ise olayların gerçekleşme olasılıkları toplamı 1'e eşittir.

Klasik olasılık yaklaşımı kullanıldığında, bir olayın gerçekleşmesi olasılığını belirlemek için deneme yapma zorunluluğu yoktur. Çünkü denemeden önce olası tüm sonuçların sayısı bilinir. Hilesiz bir para atılması denemesinde olası tüm sonuçların sayısı ikidir ya da zar atma denemesinde olası tüm sonuçların sayısı altıdır. Para atıldığında yazı gelmesi ya da zar atıldığında altı gelmesi olasılıkları bu yaklaşımla mantıksal olarak elde edilebilir.

Deneysel Olasılık

Olasılık tanımlarından bir diğeri, *göreli frekanslar*a dayalı olarak yapılır. Bu yaklaşımda bir olayın gerçekleşme olasılığını hesaplamak için, geçmişte benzer olayların gerçekleşme sayısının oranına bakılır. Geçmişte ortaya çıkan olayların sayısı f ile ve toplam gözlem sayısı N ile belirtildiğinde deneysel olasılık yaklaşımıyla bir A olayının gerçekleşme olasılığı;

$$P(A) = \frac{f}{N} \tag{2}$$

eşitliği ile hesaplanır.

ÖRNEK 2

Bir süt ürünleri firmasına ait şehirlerarası nakliye yapan kamyonlar, ayda 300 nakliye gerçekleştirmektedir. Bu ay yapılan 300 nakliyeden 18 tanesinde çeşitli nedenlerle teslimat gecikmiştir. Buna göre, firmaya ait kamyonlardan birinin gelecek ay yapacağı teslimatlardan herhangi birini geciktirme olasılığı kaçtır?

Çözüm:
A; "Firmaya ait kamyonlardan birinin gelecek ay yapacağı teslimatlardan herhangi birini geciktirmesi" olayı olsun.

$$\text{Teslimatın gecikme olasılığı} = \frac{\text{Geçen ay geciken teslimat sayısı}}{\text{Toplam nakliye sayısı}}$$

ya da

$$P(A) = \frac{18}{300} = 0{,}06$$

Bu değeri tahmini olasılık değeri olarak kabul edebiliriz. Diğer bir deyişle, geçen ay yaşanan tecrübeye dayanarak bu ayki yapılacak teslimatlardan herhangi birinin gecikmesi olasılığını % 6 olarak hesaplarız.

Öznel (Subjektif) Olasılık

Öznel Olasılık, belirli bir olayın gerçekleşme olasılığının, eldeki mevcut bilgilere dayalı olarak belirlendiği olasılık yaklaşımıdır.

Öznel olasılık, olasılığın belirlenmesi için herhangi bir geçmiş deneyim ya da bilginin bulunmadığı durumlarda başvurulan yaklaşımdır. Bu yaklaşımda, mevcut görüşler ve eldeki diğer bilgiler değerlendirilerek olaya ilişkin olasılık tahmin edilir ya da belirlenir. Belirlenen bu olasılığa, **öznel olasılık** adı verilir. Öznel olasılık için, İstatistik dersini alan bir öğrencinin final sınavından 80'in üzerinde puan alma olasılığının tahmin edilmesi, 2011 yılı gelirleri için Türkiye'de en fazla kurumlar vergisi ödeyecek kuruluşun tahmin edilmesi vb. gibi örnekler verilebilir.

Bir olasılık ifadesi, henüz gerçekleşmemiş bir olayın gerçekleşme şansını belirtir. Burada, esas olarak incelenen olaya ilişkin eldeki bilgilere dayalı olarak bu olasılığı içeren belirsizlik derecesinin önemli bir yeri vardır. Zar atma denemesine ilişkin pek çok bilgi mevcuttur ve bu zarın 1 gelme olasılığının altıda bir olacağı belirlenebilir. Halbuki piyasaya yeni sürülen denenmemiş bir ürünün tutulup tutulmayacağına ilişkin çok fazla bilgi mevcut değildir. Örneğin, şirketin piyasa araştırmaları yöneticisi piyasaya yeni sürülen bu ürünü 10 fabrika satış mağazasında test etmiş ve ürünün pazar payının % 40 olacağını belirlemiş olsun. Buna rağmen bu yöneticinin, ürünün yurt genelindeki satış miktarlarının nasıl olacağına yönelik çok fazla bilgisi yoktur. Her iki örnekte de ilgilenilen olaya ilişkin bir olasılık değeri belirlenmiştir. Aradaki fark ise ilk örnekte olasılığı belirleyen kişi yapacağı olasılık tahmininde gerçek değeri kesin olarak belirleyebilirken ikinci örnekte yapacağı tahminin kesinliğine olan güveni az miktardadır.

OLASILIK HESAPLAMA KURALLARI

Bu bölümde, iki veya daha fazla olayın olasılığını hesaplamada kullanılan kurallar incelenecektir.

Toplama Kuralları

Özel Toplama Kuralı

Özel toplama kuralı, yalnızca karşılıklı ayrık olaylar için uygulanabilir. Eğer A ve B olayları karşılıklı ayrık olaylar ise özel toplama kuralına göre bu olaylardan birinin veya diğerinin gerçekleşme olasılığı, bu olayların ayrı ayrı gerçekleşme olasılıklarının toplamına eşittir. Özel toplama kuralına ilişkin eşitlik:

$$P(\text{A veya B}) = P(A)+P(B) \qquad (3)$$

ile verilir. A, B ve C gibi karşılıklı ayrık üç olay için kural yazılacak olursa:

$$P(\text{A veya B veya C}) = P(A)+P(B)+P(C) \qquad (4)$$

ÖRNEK 3

Bir kargo firması göndereceği kargoları, kargonun gideceği mesafeye göre "yakın mesafe", "kısa mesafe", "orta mesafe" ve "uzun mesafe" olmak üzere dört kısma ayırmaktadır. Firmanın bir günde göndereceği 200 adet kargonun 5 tanesi yakın mesafe, 80 tanesi kısa mesafe, 100 tanesi orta mesafe ve 15 tanesi uzun mesafeye gönderilecektir. Buna göre, gönderilecek herhangi bir kargonun kısa mesafeye veya orta mesafeye gönderilme olasılığı kaçtır?

Çözüm:

Firmanın göndereceği kargonun kısa mesafeye gönderilmesi K olayı, orta mesafeye gönderilmesi ise O olayı olsun. Özel toplama kuralına göre;

$$P(K \text{ veya O}) = P(K) + P(O) = \frac{80}{200} + \frac{100}{200} = \frac{180}{200} = 0{,}90 \text{ olacaktır.}$$

Firmanın göndereceği kargonun yakın mesafeye gönderilmesine Y olayı, uzak mesafeye gönderilmesine de U olayı denirse, Y, K, O ve U olayları karşılıklı ayrık olaylardır. Diğer bir deyişle, gönderilecek herhangi bir kargo yakın, kısa, orta, ve uzun mesafeye aynı anda gönderilemez. Ayrıca, bu olaylar bütüne tamamlayan olaylar olacaktır. Çünkü gönderilecek kargo ya kısa ya yakın ya orta ya da uzun mesafeye gönderilecektir.

Tümleyen Kuralı

Bir olayın gerçekleşme olasılığının, bu olayının gerçekleşmeme olasılığının 1'den çıkarılması yoluyla belirlendiği kuraldır. A olayının gerçekleşme olasılığı P(A), gerçekleşmeme olasılığı ise $P(\bar{A})$ ile gösterilirse bu iki olasılığın toplamının 1 olacağı kolaylıkla tahmin edilebilir. Buna göre,

$$P(A) + P(\bar{A}) = 1 \qquad (5)$$

olur. Bu eşitlikten,

$$P(A) = 1 - P(\bar{A}) \tag{6}$$

olduğu sonucuna varılır. Bu son yazılan eşitlik tümleyen kuralının matematiksel gösterimidir.

DİKKAT

A ve $\bar{A}$ olayları karşılıklı ayrık ve bütüne tamamlayan olaylardır. Bu nedenle bu iki olayın olasılıkları toplamı 1'e eşittir.

ÖRNEK 4

Bir yatırımcı belli bir dönemde A, B ya da C hisselerinden birine yatırım yapmayı planlamaktadır. A hissesine yatırım yapma olasılığı 0,45; B hissesine yatırım yapma olasılığı 0,25 ise C hissesine yatırım yapma olasılığı kaçtır?

Çözüm:

A: {Yatırımcının A hissesine yatırım yapması},
B: {Yatırımcının B hissesine yatırım yapması} ve
C: {Yatırımcının C hissesine yatırım yapması}
olsun. Yatırımcının parasını C hissesine **yatırmama** olasılığı, A hissesine yatırma olasılığı ile B hissesine yatırma olasılıklarının toplamına eşittir. Dolayısıyla,

$$P(\bar{C}) = P(A \text{ veya } B) = P(A)+P(B) = 0{,}45+0{,}25 = 0{,}70$$

olacaktır. Yatırımcı A ya da B hissesine yatırım yapmayacaksa, C hissesine yatırım yapacaktır. Dolayısıyla, C hissesine yatırım yapma olasılığı;

$$P(C) = 1\text{-}P(\bar{C}) = 1\text{-}0{,}70 = 0{,}30 \quad \text{olur.}$$

Genel Toplama Kuralı

Bazı denemelerin sonuçları karşılıklı ayrık değildir. Örneğin, Eskişehir'de üretim yapan büyük çaplı bir fabrika, ürünlerini satışa sunulmak üzere Ankara ve İstanbul illerine gönderiyor olsun ve ay içerisinde bu illere gönderilen kamyonlar arasından 120 kamyonluk bir örneklem rassal olarak seçilmiş olsun. Örneklemden elde edilen verilere göre, ay içerisinde Ankara'ya 90 adet, İstanbul'a ise 48 adet kamyon görevlendirilmiştir. Bu örneklemden rassal olarak seçilen bir kamyonun ya Ankara'ya ya da İstanbul'a gitmiş olması olasılığı hesaplanmak istensin. Burada, Ankara'ya giden kamyonun seçilmiş olma olasılığı 90/120=0,75 olur. Aynı şekilde, İstanbul'a gönderilen kamyonun seçilme olasılığı ise 48/120=0,40'dır. Özel toplama kuralına göre, Bu olasılıkların toplamı 0,75+0,40=1,15 olur. Ancak bildiğimiz üzere, herhangi bir olasılık değeri 1'den büyük olamaz. Buna göre, her iki il için de görevlendirilmiş olan bazı kamyonlar bulunmaktadır ve *bunlar iki kere sayılmıştır.* Örneklem bilgileri incelendiğinde, 120 kamyondan 54 tanesinin her iki ile de gönderildiği sonucuna varılmış olsun. Buna göre seçilen kamyonun her iki ile de gönderilmiş olma olasılığı 54/120=0,45 olacaktır.

Bu durumda istenen olasılık, seçilen kamyonun Ankara'ya gönderilme olasılığı ile İstanbul'a gönderilme olasılığını toplayıp bu değerden her iki ile de gönderilme olasılığını çıkarmak yoluyla hesaplanabilir.

A: {Seçilen kamyonun Ankara'ya gönderilmesi} ve B: {Seçilen kamyonun İstanbul'a gönderilmesi} olsun. Buna göre;

$$P(A \text{ veya } B) = P(A)+P(B)\text{-}P(A \text{ ve } B) = 0{,}75+0{,}40\text{-}0{,}45 = 0{,}70 \quad \text{olur.}$$

İki olayın birlikte gerçekleşme olasılığına **ortak olasılık** adı verilir. Yukarıdaki örnekte seçilen kamyonun her ikisi ile de gönderilmiş olma olasılığı, ortak olasılıktır.

Ortak Olasılık, iki ya da daha fazla olayın aynı anda gerçekleşme şansını ölçen olasılıktır.

A ve B gibi iki olay için genel toplama kuralı;

$$P(A \text{ veya } B) = P(A)+P(B)-P(A \text{ ve } B) \qquad (7)$$

eşitliği ile verilir. P(A veya B) ifadesindeki "veya" kelimesi A'nın ya da B'nin gerçekleşebileceğini öne sürmektedir. *Bu ifade, A ve B olaylarının birlikte gerçekleşmesi olasılığını da içermektedir.*

İncelenen olaylar karşılıklı ayrık olduğunda, özel toplama ve genel toplama kuralları arasında önemli bir farklılık ortaya çıkar. Eğer olaylar karşılıklı ayrık ise P(A ve B) ortak olasılığı sıfır olur ve bu durumda A veya B'nin gerçekleşme olasılığının hesaplanmasında özel toplama kuralı uygulanabilir. Olaylar karşılıklı ayrık değil ise ortak olasılık değeri de hesaba katılmalı ve genel toplama kuralı uygulanmalıdır.

52 kartlık bir oyun destesinden rassal olarak seçilen bir kartın ikili veya karo olması olasılığı kaçtır?

SIRA SİZDE 1

Çarpma Kuralları ve Koşullu Olasılık

Toplama kuralları, olayların *birleşiminin* olasılığını bulmada kullanılır. Bu bölümde ise olayların aynı anda gerçekleşme olasılığı incelenecektir. Örneğin, A ve B gibi iki olayın birlikte gerçekleşme olasılığı hesaplanmak istendiğinde, bu iki olayın kesişimine ilişkin olasılık değerlendirilir ve bu olasılığın hesaplanmasında çarpma kurallarından yararlanılır.

Özel Çarpma Kuralı

Özel çarpma kuralının uygulanabilmesi için olayların birbirinden bağımsız olması gerekir. Eğer bir olayın gerçekleşmesi, bir diğer olayın gerçekleşme olasılığını değiştirmiyorsa bu iki olay birbirinden **bağımsız**dır.

Bağımsızlık, bir olayın gerçekleşmesinin, bir diğer olayın gerçekleşme olasılığını etkilememesi durumunda bu iki olaya bağımsız olaylar adı verilir.

A ve B olayları farklı zamanlarda gerçekleştiğinde, söz gelimi A olayının gerçekleşmesinden sonra B olayı gerçekleştiğinde, A olayı B olayının gerçekleşme olasılığını etkilemiyorsa bu A ve B olayları bağımsız olaylardır. Örneğin, bir zarın iki kez atılması denemesinde ikinci zardan elde edilen sonuç, birinci zarın kaç geldiğinden bağımsızdır.

*Özel çarpma kuralı*na göre; A ve B bağımsız olayları için A ve B'nin birlikte gerçekleşme olasılığı, bu iki olayın ayrı ayrı gerçekleşme olasılıklarının çarpımına eşittir.

$$P(A \text{ ve } B) = P(A)\cdot P(B) \qquad (8)$$

Özel çarpma kuralı ikiden fazla olay için genelleştirilebilir. Örneğin, A, B ve C bağımsız olayları için özel çarpma kuralı, bu üç olayın birlikte gerçekleşme olasılığını hesaplamada kullanılabilir. Bu olasılık:

$$P(A \text{ ve } B \text{ ve } C) = P(A)\cdot P(B)\cdot P(C) \qquad (9)$$

olacaktır.

ÖRNEK 5

Geçen yıl İstatistik dersi alan öğrencilerinin %70'i, bu dersten başarılı olmuştur. Bu öğrenciler arasından rassal olarak seçilen üç öğrencinin, her üçünün de geçen yıl İstatistik dersinden başarılı olmuş olma olasılığı kaçtır?

Çözüm:
K: {Seçilen birinci öğrencinin geçen yıl İstatistik dersinden başarılı olması},
L: {Seçilen ikinci öğrencinin geçen yıl İstatistik dersinden başarılı olması} ve
M: {Seçilen üçüncü öğrencinin geçen yıl İstatistik dersinden başarılı olması}
olayları tanımlansın. Verilen bilgilere göre, P(K) = 0,70, P(L) = 0,70 ve P(M) = 0,70 olur. Dersi alan öğrencilerin sayısı çok fazla olduğu için K, L ve M olaylarının bağımsız olduğu varsayılabilir. Buna göre soruda istenen olasılık özel çarpma kuralı yardımıyla hesaplanabilir:

$$P(K \text{ ve } L \text{ ve } M) = P(K) \cdot P(L) \cdot P(M) = (0{,}70)(0{,}70)(0{,}70) = 0{,}343$$

Eğer iki olay bağımsız değilse, bu olaylara bağımlı olaylar adı verilir. Örneğin, bir torbada 8 tanesi kırmızı, 4 tanesi sarı olmak üzere toplam 12 adet bilye bulunsun. Torbadan rassal olarak bir bilye çekilsin. Çekilen bilyenin sarı renkli bilye olma olasılığı 4/12'dir ve çekilen bilyenin kırmızı olma olasılığı 8/12 olur. Daha sonra torbadan çekilen *birinci bilye yerine koyulmadan* ikinci bir bilye çekilsin. Çekilen ikinci bilyenin sarı olma olasılığı, çekilen ilk bilyenin sarı ya da kırmızı olmasına göre değişir. Bu olasılık; çekilen ilk bilye sarı renkli ise 3/11 olur. Çünkü, kalan 11 bilyeden yalnızca 3 tanesi sarı renklidir. Çekilen ilk bilye kırmızı ise 4/11 olur. Çünkü kalan 11 bilyeden 4 tanesi sarı renklidir.

Buradaki 3/11 ve 4/11 olasılıklarına *koşullu olasılık* adı verilir. Çünkü bu değerler, torbadan çekilen ilk bilyenin sarı ya da kırmızı olmasına bağlıdır.

Koşullu Olasılık

B olayının gerçekleştiği bilindiğinde, A olayının gerçekleşme olasılığına koşullu olasılık adı verilir ve bu olasılık $P(A|B)$ ile gösterilir. $P(A|B)$ ifadesi, *"B bilindiğine göre, A'nın koşullu olasılığı"* olarak okunur.

Az önceki örnekte verilen torbadaki (8 kırmızı, 4 sarı) bilyelere ilişkin olarak aşağıdaki olaylar tanımlansın.

A: {Çekilen ikinci bilyenin sarı olması},
B: {Çekilen birinci bilyenin sarı olması} ve
C: {Çekilen birinci bilyenin kırmızı olması}

Çekilen ikinci bilyenin sarı olma olasılığı; çekilen birinci bilye sarı ise; $P(A|B)=3/11$, çekilen birinci bilye kırmızı ise $P(A|C)=4/11$ olacaktır.

Genel Çarpma Kuralı

Herhangi iki olay bağımsız olmadığında, bu iki olayın ortak olasılığını hesaplamada kullanılan kuraldır. Söz gelimi, A olayı gerçekleştikten sonra B olayı gerçekleşiyorsa ve B olayının gerçekleşmesinde A olayının etkisi varsa A ve B olayları bağımsız değildir.

Genel çarpma kuralına göre A ve B gibi iki olay için, bu olayların birlikte gerçekleşme olasılığı, A olayının gerçekleşme olasılığı ile A'nın gerçekleştiği bilindiğine göre B'nin koşullu olasılığının çarpımına eşittir.

$$P(A \text{ ve } B) = P(A) \cdot P(B|A) \qquad (10)$$

ÖRNEK 6

Yurt çapında faaliyet gösteren bir bankanın Eskişehir ili için yaptığı araştırma sonuçlarına göre, Eskişehir ilinde ikamet eden kişilerin % 16'sının, bu bankada hesabı bulunmaktadır. Bankada hesabı bulunan bu müşterilerin % 22'si de geçen yıl bankadan tüketici kredisi çekmiştir. Eskişehir ilinde ikamet eden kişiler arasından rassal olarak seçilen bir kişinin, bu bankada hesabının bulunması ve geçen yıl bu bankadan tüketici kredisi çekmiş olması olasılığı kaçtır?

Çözüm:
H = {Seçilen kişinin bu bankada hesabının olması} ve
T = {Seçilen kişinin geçen yıl bu bankadan tüketici kredisi çekmiş olması}
olayları tanımlansın. Seçilen kişinin bu bankada hesabının olması olasılığı; P(H)=0,16 ve seçilen kişinin bu bankada hesabı olduğu bilindiğinde, geçen yıl bu bankadan tüketici kredisi çekmiş olması olasılığı; P(T|H)=0,22 olacaktır. Buna göre istenen olasılık;

$$P(H \text{ ve } T) = P(H)\cdot P(T|H) = (0{,}16)\cdot(0{,}22) = 0{,}0352 \text{ olarak hesaplanır.}$$

SIRA SİZDE 2

Bir torbada 8 tanesi kırmızı, 4 tanesi sarı olmak üzere toplam 12 adet bilye bulunsun. Torbadan rassal olarak ardarda iki bilye çekilsin. Çekilen her iki bilyenin de sarı renkli bilye olma olasılığı kaçtır?

Genel çarpma kuralı ikiden fazla olay için genelleştirilebilir. Örneğin A, B ve C gibi üç olay için genel çarpma kuralı;

$$P(A \text{ ve } B \text{ ve } C) = P(A)\cdot P(B|A)\cdot P(C|A \text{ ve } B) \qquad (11)$$

olacaktır.

Torbada sarı ve kırmızı bilyelerin bulunduğu Sıra Sizde 2 sorusunda, torbadan rassal olarak art arda üç bilye çekilsin ve aşağıda verilen olaylar tanımlansın.

S_1: {Çekilen birinci bilyenin sarı olması},
S_2: {Çekilen ikinci bilyenin sarı olması} ve
S_3: {Çekilen üçüncü bilyenin sarı olması}
Çekilen bu üç bilyenin de sarı olması olasılığı;

$$P(S_1 \text{ ve } S_2 \text{ ve } S_3) = P(S_1)\cdot P(S_2|S_1)\cdot P(S_3|S_1 \text{ ve } S_2)$$

$$= \frac{4}{12}\cdot\frac{3}{11}\cdot\frac{2}{10} = \frac{24}{1320} = 0.0182$$

olarak hesaplanır.

BAYES TEOREMİ

A_1 ve A_2 olayları, karşılıklı ayrık ve bütüne tamamlayan olaylar olsun. Bayes teoremine göre, B olayının gerçekleştiği bilindiğine göre, A_i'nin gerçekleşme olasılığı;

$$P(A_i\;B) = \frac{P(A_i)P(B\;\;A_i)}{P(A_1)P(B\;\;A_1) + P(A_2)P(B\;\;A_2)} \quad , \quad i = 1,2 \text{ için} \qquad (12)$$

eşitliği ile hesaplanır.

Bayes kuralını bir örnek üzerinde açıklamaya çalışalım. Bir fast-food restoranı sahibi, şehir içine sipariş dağıtımlarını yapmak üzere iki adet motosikletli kurye çalıştırmaktadır. Verilen herhangi bir siparişi birinci kuryenin götürmesi olasılığı % 40 olsun. Herhangi bir zamanda verilen bir siparişi birinci kuryenin götürmesi K_1 olayı; ikinci kuryenin götürmesi de K_2 olayı olarak tanımlansın. Verilen siparişi birinci kuryenin götürmesi olasılığı $P(K_1)=0,40$ olacaktır. Buradaki $P(K_1)$=P(verilen siparişi birinci kuryenin götürmesi) = 0,40 değeri, önsel olasılık olarak adlandırılır. Önsel olasılık denilmesinin nedeni, bu değerin herhangi bir deneysel veri elde edilmeden önce belirlenmiş olmasındandır. Dolayısıyla önsel olasılık, eldeki mevcut bilgi düzeyine göre belirlenen başlangıç olasılığıdır.

Buna göre, verilen herhangi bir siparişi ikinci kuryenin götürmesi önsel olasılığı $P(K_2)=0,60$ olacaktır. K_1 ve K_2 olayları, karşılıklı ayrık ve bütüne tamamlayan olaylar oldukları için bu değer, $P(K_2)=1-P(K_1)=1-0,40=0,60$ biçiminde de bulunabilir.

G olayı; "Verilen herhangi bir siparişin müşteriye geç teslim edilmesi" olarak tanımlansın. Bu kuryelerin önceki teslimat verilerine göre, eğer verilen bir siparişi birinci kurye götürmüşse, bu siparişin müşteriye geç teslim edilmesi olasılığının 0,25 olduğu biliniyor olsun. Daha önce verilen koşullu olasılık tanımlarına göre bu ifade; $P(G|K_1)=0.25$ olarak yazılabilir. Bunun yanında, eğer verilen bir siparişi ikinci kurye götürmüşse bu siparişin müşteriye geç teslim edilmesi olasılığı da 0,05, dolayısıyla, $P(G|K_1)=0,05$ olarak verilsin.

Restorana dışarıdan verilen bir sipariş rassal olarak seçilsin ve bu siparişin müşteriye geç teslim edildiği tespit edilmiş olsun. Bu durumda, "Müşteriye geç teslim edilen bu siparişin birinci kurye tarafından götürülmüş olma olasılığı kaçtır?" sorusunda, $P(K_1|G)$ olasılığı, diğer bir deyişle, P(Verilen siparişi birinci kuryenin götürmesiÔVerilen siparişin müşteriye geç teslim edilmesi) olasılığı sorulmaktadır. Buradaki $P(K_1|G)$ olasılığı *sonsal olasılık* olarak adlandırılır. Buna göre sonsal olasılık, eklenen bilgiye dayalı olarak yeniden düzenlenmiş olasılıktır.

Bayes teoremi yardımıyla, soruda istenen sonsal olasılık

$$P(K_1 \mid G) = \frac{P(K_1)P(G \mid K_1)}{P(K_1)P(G \mid K_1) + P(K_2)P(G \mid K_2)} = \frac{(0,40)(0,25)}{(0,40)(0,25)+(0,60)(0,05)} = 0,77$$

olarak hesaplanır.

Buna göre, verilen siparişin müşteriye geç teslim edildiği bilindiğinde, bu geç teslimatı birinci kuryenin yapmış olma olasılığı 0,77'dir. Bu sonucu yorumlamak gerekirse verilen siparişlerden bir tanesi rassal olarak seçildiğinde, bu siparişin birinci kurye tarafından teslim edilmiş olma olasılığı 0,40'tır. Verilen siparişin geç teslim edildiği belirlenirse bu siparişin birinci kurye tarafından götürülmüş olma olasılığı 0,40'ten 0,77'ye yaklaşık olarak iki kat artmaktadır.

Bu örnekte yalnızca iki karşılıklı ayrık ve bütüne tamamlayan olay için Bayes teoremi incelenmiştir. Bayes teoremi $A_1, A_2, ..., A_n$, n adet karşılıklı ayrık ve bütüne tamamlayan olay için genellenirse;

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \ldots + P(A_n)P(B \mid A_n)} \qquad (13)$$

eşitliği ile hesaplanır.

SIRA SİZDE 3

Bir matbaada basılan kitapların %35'i birinci makinede, %25'i ikinci makinede ve %40'ı üçüncü makinede basılmaktadır. Bu makinelerde basılan herhangi bir kitabın hatalı basılması olasılıklarının sırasıyla; %2, %6 ve %3 olduğu bilinmektedir. Basılan kitaplar arasından rassal olarak seçilen bir kitabın hatalı basıldığı görülmüştür. Bu hatalı kitabın üçüncü makinede basılmış olma olasılığı kaçtır?

SAYMA KURALLARI

Bir denemedeki olası sonuçların sayısı küçük olduğunda bunların sayılması bazen işlemleri kolaylaştırır. Örneğin, bir bozuk para atma denemesinde, birincisi "Yazı" ve ikincisi "Tura" olmak üzere yalnızca iki olası sonuç bulunur.

Bozuk paranın 50 kez atılması denemesinde elde edilen yazı ve turaların sayısı gibi, olası sonuç sayısının çok fazla olduğu durumlarda ise 50 yazı 0 tura, 49 yazı 1 tura, 48 yazı 2 tura, ... vb. gibi tüm olası sonuçların sayılması oldukça yorucu bir işlem olacaktır. Bu bölümde, sayma işlemini kolaylaştırmak amacıyla birincisi *saymanın temel ilkesi, ikincisi permütasyon kuralı* ve üçüncüsü *kombinasyon kuralı* olmak üzere üç sayma kuralı incelenecektir.

Saymanın Temel İlkesi

Eğer bir işlem m farklı yolla, bir başka işlem de n farklı yolla gerçekleşebiliyorsa bu iki işlem birlikte m × n farklı yolla gerçekleşir.

$$\text{Toplam farklı düzen sayısı} = m \times n \qquad (14)$$

Bu kural ikiden fazla olay için genelleştirilebilir. Örneğin, m, n ve o gibi üç olay için saymanın temel ilkesi;

$$\text{Toplam farklı düzen sayısı} = m \times n \times o \qquad (15)$$

olarak ifade edilir.

ÖRNEK 7

Bir şirketin yönetim kurulu için 3 kurul başkanı, 6 başkan yardımcısı, 4 yazman adayı arasından, her pozisyon için birer kişi seçilmek üzere kaç farklı seçim yapılabilir?

Çözüm:

Saymanın temel ilkesine göre sorunun cevabı; $3 \times 6 \times 4 = 72$ olacaktır.

Permütasyon ve kombinasyon kurallarına geçmeden önce bu kurallarda kullanılacak olan n faktöriyel gösterimini tanımlayalım. n pozitif tamsayısından küçük ve eşit bütün pozitif tamsayıların çarpımı, n faktöriyel olarak adlandırılır ve bu ifade "n!" ile gösterilir. Bu tanıma göre;

$$n!=n(n-1)(n-2)...2.1=n(n-1)!$$

olarak yazılabilir. Örneğin, $8!=8\cdot7\cdot6\cdot5\cdot4\cdot3\cdot2\cdot1=40.320$ olacaktır.
Özel olarak, $0!=1$ ve $1!=1$'dir.

Permütasyon Kuralı

Saymanın temel ilkesi iki ya da daha fazla grup için olası düzen sayısını bulmada kullanılırken permütasyon kuralı, yalnızca bir nesne grubu için olası sıralama sayısını bulmada kullanılır. Örneğin, bir kişi mobilya üretimi yapan bir firmadan dört adet mobilya satın almış olsun. Mobilyaların eve teslimatı sırasında bu mobilyalar kamyondan herhangi bir sırayla indirilebilir. Söz gelimi, ilk olarak yemek masası, ikinci olarak televizyon sehpası, üçüncü olarak komidin ve son olarak gardrop biçiminde indirme işlemi yapılabilir. Burada yapılan bu sıralamaya bir permütasyon adı verilir. Dolayısıyla, n olası nesnenin tek bir grubundan seçilen r adet nesnenin herhangi bir sıralamasına permütasyon adı verilir.

Burada dikkat edilmesi gereken nokta, (yemek masası, televizyon sehpası, komidin, gardrop) sıralaması ile (gardrop, yemek masası, komidin, televizyon sehpası) sıralamaları, *farklı* permütasyonları ifade etmektedir.

DİKKAT

Permütasyonda, sıralanan nesnelerin dizildiği sıra önemlidir.

Toplam farklı permütasyon sayısını elde etmek için gerekli eşitlik;

$$ {}_nP_r = \frac{n!}{(n-r)!} \tag{16} $$

olarak verilir. Burada n: toplam nesne sayısını, r: seçilen nesne sayısını belirtmektedir ve nesneler sıralarda ancak bir kez kullanılabilirler.

ÖRNEK 8

Yukarıda verilen dört adet mobilyanın herhangi bir sırayla kamyondan indirilmesi örneği için, mobilyalar kamyondan kaç farklı şekilde indirilebilir?

Çözüm:
Kamyondan indirilecek dört adet mobilya bulunmaktadır. Buna göre, n=4'tür. Bu mobilyaların dördü birden kamyondan indirileceği için r=4 olur. Permütasyon kuralına göre sorunun cevabı;

$$ {}_4P_4 = \frac{4!}{(4-4)!} = \frac{4!}{0!} = \frac{4!}{1} = 24 $$

olacaktır.

Burada permütasyon kuralı uygulanarak elde edilen farklı düzen sayısının sağlaması yapılabilir. Soruda dört adet mobilya ve indirme işlemi için dört adet sıra bulunmaktadır. İlk sıra için olası dört mobilya bulunur. Bir mobilya birinci sırada indirildiğinden ikinci sıra için olası üç mobilya bulunur. İki mobilya ilk iki sırada indirildiğinden üçüncü sıra için olası iki mobilya bulunur ve üç mobilya ilk üç sırada indirildiğinden son sıra için olası tek bir mobilya bulunur. Dolayısıyla;

$4 \cdot 3 \cdot 2 \cdot 1 = 24$ permütasyon

elde edilir.

Çoğu zaman, n olası nesnenin tümünün değil de bir kısmının seçildiği ya da sıralandığı, diğer bir deyişle seçilen nesne sayısının toplam nesne sayısından küçük olduğu (r<n) durumlarla karşılaşılır.

ÖRNEK 9

Bir derneğin 15 üyesi arasından biri başkan, biri başkan yardımcısı ve biri mali işler sorumlusu olmak üzere üç kişilik bir kurul kaç farklı şekilde seçilebilir?

Çözüm:
Başkanlık için 15, başkan yardımcılığı için 14 ve mali işler sorumluluğu için 13 olası üye söz konusudur. Buna göre dernek üyeleri arasından üç kişilik kurul, 15·14·13=2730 farklı şekilde seçilebilir. Aynı sonuca permütasyon kuralıyla da ulaşılabilir. n=15 ve r=3 olmak üzere,

$$ {}_{15}P_3 = \frac{15!}{(15-3)!} = \frac{15!}{12!} = \frac{15 \cdot 14 \cdot 13 \cdot 12!}{12!} = 2730 \text{ olur.} $$

Kombinasyon Kuralı
Seçilecek nesnelerin diziliş sırası önemli değilse yapılan her bir seçime bir kombinasyon adı verilir. Kombinasyon kuralına göre, n adet nesneden oluşan bir kümeden r nesnelik kombinasyon sayısı;

$$ {}_nC_r = \frac{n!}{r!(n-r)!} \tag{17} $$

ile hesaplanır.

Örneğin, A, B, C ile gösterilen nesneler, bu nesnelerin *sırası göz önüne alınmadan* bir grup olarak seçilmek istendiğinde, yalnızca tek bir olası kombinasyon söz konusu olur. A,B,C ile B,C,A seçimleri, bu nesnelerin seçim sırası önemli olmadığından aynı seçimler olacaktır. Kombinasyon eşitliği kullanıldığında,

$$ {}_nC_r = \frac{3!}{3!(3-3)!} = \frac{3!}{3!0!} = 1 $$

olduğu görülür.

ÖRNEK 10

6 farklı meyve arasından 2 meyve kaç değişik şekilde seçilebilir?

Çözüm:
Seçilen meyvelerin seçim sırası önemli değildir. Yapılabilecek farklı seçim sayısı;

$$ {}_6C_2 = \binom{6}{2} = \frac{6!}{2!(6-2)!} = \frac{6!}{2!\,4!} = \frac{6 \cdot 5 \cdot 4!}{2!\,4!} = \frac{30}{2} = 15 \quad \text{olur.} $$

RASSAL DEĞİŞKENLER VE OLASILIK DAĞILIMLARI

Bu bölümde, öncelikle rassal değişken ve olasılık dağılımı kavramları açıklanacak, kesikli rassal değişenler için olasılık dağılımı, ortalama, varyans ve standart sapması hakkında bilgi verilecek ve önemli kesikli olasılık dağılımlarından binom dağılımı incelenecektir. Sonraki bölümde ise sürekli rassal değişkenler için olasılık dağılımı tanımlanacak ve son olarak uygulamalarda en sık kullanılan sürekli olasılık dağılımlarından normal dağılım ayrıntılı olarak incelenecektir.

Rassal Değişken

Olasılık denemelerinde, deneme sonuçlarının bir veya birkaç yönden incelenmesi gereken durumlarla karşılaşılabilir. Bu incelemeler sırasında, belirlenen bir değişken üzerinde gözlemler yapılır. Örneğin, şirket araçlarının aylık kat ettiği mesafe ölçümlerinde, her araç için kat edilen mesafeler farklı değerler alacağı için buradaki kat edilen mesafe değeri ölçülebilir türden bir rassal değişken olacaktır. Yapılan ölçümler rastgele ve öngörülemeyen bir biçimde değişim gösterdiği için, rassal değişken kavramının tanımlanması ihtiyacı ortaya çıkar. Çünkü gözlemlenen olaylarda, genellikle olasılık denemelerinin sonuçlarına ilişkin sayılarla, diğer bir deyişle rassal değişkenlerin aldığı değerlerle ilgilenilir. Bazı denemelerin sonuçları nicel iken (uzunluk, ağırlık, tura sayısı vs.), bazılarınınki de niteldir (cinsiyet, medeni hâl, saç rengi vs.).

Örneğin, bir şirkette belli bir günde mazeretleri nedeniyle işe gelmeyenlerin sayısı 0, 1, 2, ... olacaktır. Burada işe gelmeyenlerin sayısı bir rassal değişkendir. Bir sınıftaki öğrencilerin ağırlıkları ölçüldüğünde, 68,2 kg, 76,8 kg, 81,3 kg. gibi değerler elde edilebilir. Burada öğrencilerin ağırlığı bir rassal değişkendir. Hilesiz bir zar atıldığında, üst yüze gelen sayı 1, 2, 3, 4, 5 ya da 6 olabilir. Bu denemede üst yüze gelen sayının değeri, şansa bağlı bir sonuç olduğundan, bir rassal değişkendir.

Denemeden denemeye farklı değerler alan ve aldığı bu değerleri belli bir olasılıkla alan değişkenlere rassal değişken adı verilir. Rassal değişkenler genellikle X, Y, Z, ... gibi büyük harflerle, değişkenin aldığı değerler ise x, y, z, ... gibi küçük harflerle belirtilir.

Rassal değişkenler, kesikli ve sürekli rassal değişkenler olmak üzere iki şekilde ortaya çıkarlar.

Kesikli Rassal Değişken

Sonlu ya da sayılabilir sayıda farklı değeri bulunan rassal değişkenlere kesikli rassal değişken adı verilir. Eğer bir şirkette çalışanların sayısı 500 ise belli bir günde mazeretleri nedeniyle işe gelmeyenlerin sayısı yalnızca 0'dan 500'e kadar tamsayılar olabilir. Kesikli rassal değişkenler aldıkları değerleri genellikle incelenen olaya konu olan birimlerin sayılması sonucunda alırlar.

Kesikli rassal değişkenlerin, kesirli ya da ondalık değerler aldığı durumlar da bulunabilir. Ancak bu değerler birbirinden ayrılmış olmalı ya da aralarında belli uzaklıklar olmalıdır. Örneğin, 1'den 10'a kadar ondalıklı değerlerle değerlendirilen bir sınav sonucunda öğrenciler, 3,5; 8,9 ya da 6,7 vb. gibi notlar alabilir. 5,5 ve 5.6'da olduğu gibi, alınan notlar arasında 0,1 puanlık uzaklıklar olacağı için öğrencilerin aldığı notlar kesikli rassal değişkendir.

Sürekli Rassal Değişken

Sayılamayacak ya da sonsuz sayıda olası değeri bulunan ve bir sayı aralığı ya da aralık kümesi üzerinde tanımlanan rassal değişkenlere sürekli rassal değişken adı verilir. Fabrikada üretilen bir vidanın uzunluğu, bir tansiyon hastasının büyük ve küçük tansiyon değerleri, bir odanın sıcaklığı ölçüldüğünde, ölçülen bu değişkenler belli bir aralıkta sonsuz sayıda değerler alabilecekleri için sürekli rassal değişkenlerdir.

Örneğin, hassas bir derece ile ölçülen vücut sıcaklıkları 36,36 °C, 37,23 °C gibi değerler alabildiği için sürekli rassal değişkendir. Ankara-Eskişehir hattında yüksek hızlı trenin sefer tamamlama süreleri 01:32:08, 01:29:44 vb. gibi değerler alabilir. Burada, sefer tamamlama süresi sürekli bir rassal değişkendir.

Olasılık Dağılımları

İstatistiksel çıkarımda, gerçekleşmesi şansa bağlı olan olaylarla ilgilenilir. Burada amaç, evrenden seçilen ve daha az sayıda gözlemden oluşan örneklem istatistiklerinden yararlanarak evrene ilişkin çıkarımlarda bulunmaktır. Buraya kadarki bölümde, olasılığın 0 ile 1 kapalı aralığında değerler aldığını belirtmiş, toplama ve çarpma kuralları yardımıyla nasıl hesaplandığını incelemiştik. Burada ise *olasılık dağılımları* incelenecektir.

Bir denemedeki olası tüm sonuçların ve bu sonuçların her birine ilişkin olasılıkların yer aldığı listeye olasılık dağılımı adı verilir. Olasılık dağılımlarında, herhangi bir denemede ortaya çıkabilecek tüm değerlerin tanım bölgesi verilir. Olasılık dağılımları bu yönüyle göreli frekans dağılımlarına benzemektedir. Aradaki fark ise geçmişte herhangi bir zamanda gerçekleşmiş olan olayları tanımlamak yerine, bir olayın gelecekte gerçekleşme olasılığının ne kadar olduğunu tanımlamasıdır.

Olasılık dağılımlarının iki temel özelliği bulunur. Bunlar:

1. Belli bir sonucun olasılığı 0 ile 1 kapalı aralığında değerler alır.
2. Tüm karşılıklı ayrık olayların olasılıkları toplamı 1'e eşittir.

Kesikli bir rassal değişkenin olası değerler kümesi düzenlendiğinde, bu dağılım kesikli olasılık dağılımı olacaktır. Eğer rassal değişken sürekli ise olasılık dağılımı sürekli olasılık dağılımı adını alır. Bu nedenle, olasılık dağılımları, kesikli ve sürekli rassal değişkenler için ayrı ayrı incelenecektir.

Kesikli Rassal Değişkenler İçin Olasılık Dağılımları

Eğer X kesikli bir rassal değişken ise X'in olasılık dağılımı, rassal değişkenin alabileceği tüm olası x değerlerine ilişkin olasılıkların yer aldığı listedir.

Kesikli rassal değişkenler için olasılık dağılımının oluşturulmasını aşağıdaki örnek ile açıklayalım.

ÖRNEK 11

İki adet hilesiz bozuk paranın aynı anda birer kez atılması denemesinde olası sonuçlar: sıfır tura, bir tura ve iki tura olduğuna göre, gelen tura sayısı için olasılık dağılımını elde ediniz.

Çözüm:

Bu denemede, her iki paranın tura gelmesi, birinci paranın tura ikinci paranın yazı gelmesi, birinci paranın yazı ikinci paranın tura gelmesi ya da her iki paranın da yazı gelmesi olmak üzere olası dört sonuç bulunur. Burada olası sonuç sayısı, çarpma kuralı yardımıyla $2 \times 2=4$ olarak bulunabilir. Her iki paranın tura geldiği durumda elde edilen tura sayısı 2, birinci paranın tura ikinci paranın yazı geldiği durumda elde edilen tura sayısı 1, birinci paranın yazı ikinci paranın tura geldiği durumda elde edilen tura sayısı 1 ve son olarak her iki paranın da yazı geldiği durumda elde edilen tura sayısı 0 olacaktır.

Buna göre, iki tura elde edilmesi sonucu bir kez, bir tura elde edilmesi sonucu iki kez ve hiç tura elde edilememesi sonucu bir kez ortaya çıkmaktadır. Dolayısıyla, iki tura elde edilmesi sonucu dört denemede bir kez gerçekleşmektedir. Buna göre, iki tura elde etme olasılığı dörtte bir, bir tura elde etme olasılığı dörtte iki ve hiç tura elde edememe olasılığı dörtte bir olacaktır. Burada dikkat edilmesi gereken nokta, belirtilen bu sonuçlardan biri kesin olarak gerçekleşeceğinden, olası tüm olayların olasılıkları toplamı 1'dir ve bu durumda olasılık dağılımı Tablo 5.2'de verildiği gibidir.

Tablo 5.2
Örnek 11 İçin Olasılık Dağılımı

Tura Sayısı x	Elde Edilen Sonucun Olasılığı P(x)
2	1/4 = 0,25
1	2/4 = 0,50
0	1/4 = 0,25
Toplam	4/4 = 1,00

Tabloda elde edilen tura sayılarını x ile gösterdiğimizde, bu sonuçların gerçekleşme olasılıkları da P(x) ile belirtilmektedir. Örneğin, P(2 tura)=0,25, P(1 tura)=0,50 ve P(0 tura)=0,25 biçiminde ifade edilir.

Kesikli Bir Olasılık Dağılımının Ortalama, Varyans ve Standart Sapması

Daha önceki ünitelerde frekans dağılımları için konum ve değişkenlik ölçüleri incelenmişti. Ortalama, eldeki verilerin merkezî eğilimini verirken varyans ve standart sapma değeri verinin değişkenliğini tanımlamada kullanılır. Benzer biçimde, olasılık dağılımları da ortalaması ve varyansıyla birlikte verilir. *Bir olasılık dağılımının ortalaması μ ile, varyansı σ^2 ile ve standart sapması σ ile gösterilir.*

Ortalama, olasılık dağılımlarının merkezi eğilimini gösteren tipik bir değerdir ve bir olasılık dağılımının ortalamasına, o olasılık dağılımının beklenen değeri adı verilir. Beklenen değer, bir rassal değişkenin aldığı değerlerin, bu rassal değişkenlere ilişkin olasılıklarla ağırlıklandırıldığı bir ağırlıklı ortalamadır.

Kesikli olasılık dağılımları için ortalama;

$$\mu = \sum xP(x) \tag{18}$$

eşitliği yardımıyla hesaplanır. Burada P(x); belli bir x değerine karşılık gelen olasılık değeridir. Eşitliğe göre kesikli bir olasılık dağılımının beklenen değeri, tüm x değerleri kendilerine ilişkin olasılıklarla çarpıldıktan sonra bu çarpımların toplanması sonucu elde edilir.

Ortalama, kesikli bir olasılık dağılımını özetleyen tipik bir değer olmakla birlikte, dağılımın yayılmasını ya da değişkenliğini ölçen bir ölçü değildir. Değişkenliğin en önemli iki ölçüsü varyans ve standart sapmadır. Kesikli bir olasılık dağılımının varyansı;

$$\sigma^2 = \sum (x-\mu)^2 P(x) \tag{19}$$

eşitliği ile hesaplanır. Hesaplamada öncelikle, her bir x değerinden ortalama değeri çıkartılır ve bu farkın karesi alınır. Tüm karesel fark değerleri, olasılık değerleri ile çarpılır. Son olarak da, bu çarpımların toplamı alınarak varyans değerine ulaşılır. σ^2 varyans değerinin pozitif karekökü standart sapma olarak tanımlanır ve σ ile gösterilir. Dolayısıyla,

$$\sigma = \sqrt{\sigma^2} \tag{20}$$

olur.

ÖRNEK 12 *X rassal değişkenine ilişkin olasılık dağılımı;*

x	P(x)
1	0,1
2	0,3
3	0,2
4	0,4

olarak verilmektedir. Buna göre, dağılımın ortalaması, varyansı ve standart sapması kaçtır?

Çözüm:
Rassal değişkenin aldığı değerler 1'den 4'e kadar tamsayılar olduğu için, verilen olasılık dağılımı bir kesikli olasılık dağılımıdır. Kesikli olasılık dağılımları için ortalama formülüne göre dağılımın ortalaması;

$$\mu = \sum xP(x) = 1(0,1) + 2(0,3) + 3(0,2) + 4(0,4) = 2,9$$

olur. Bu örnek için X rassal değişkeninin değerleri 1, 2, 3 veya 4 tam sayılarından birisi olacağı için, değişkenin 2.9 değerini alamayacağı açıktır. Buradaki 2.9 değeri dağılımın beklenen değeri ya da ortalaması olarak yorumlanmalıdır.

Dağılımın varyansı;

$$\sigma^2 = \sum (x-\mu)^2 P(x)$$

$$= (1 - 2,9)^2(0,1) + (2 - 2,9)^2(0,3) + (3 - 2,9)^2(0,2) + (4 - 2,9)^2(0,4)$$

$$= 0,361 + 0,243 + 0,002 + 0,484 = 1,09$$

olarak hesaplanır. Dağılımın standart sapması ise varyansın pozitif karekökü olduğu için;

$$\sigma = \sqrt{\sigma^2} = \sqrt{1,09} = 1,044$$

olacaktır. Bu değer ise şu şekilde yorumlanabilir: X rassal değişkeninin başka bir olasılık dağılımının ortalaması da 2.9, ancak standart sapması bu sefer 2.106 olsun. 2,106>1,044 olduğundan ikinci olasılık dağılımının değişkenliği, örneğimizdeki olasılık dağılımının değişkenliğinden daha yüksek olacaktır.

Binom Dağılımı

Kesikli olasılık dağılımları arasında uygulamalarda en sık rastlanan olasılık dağılımıdır. Dağılımın en önemli özelliği, aynı koşullarda tekrarlanan denemelerin her tekrarında yalnızca iki olası sonuç bulunur. Örneğin, bir bozuk para atma denemesinde, para ya yazı ya da tura gelecektir. Atılan bu para aynı anda hem yazı hem de tura gelemeyeceğine göre, bu iki olası sonuç karşılıklı ayrıktır. Bir öğrenci sınavına girdiği herhangi bir dersten ya başarılı ya da başarısız olacaktır. Sınıftan rassal olarak seçilen bir öğrenci ya kız ya da erkek öğrenci olacaktır. Binom denemelerindeki olası iki sonuç genellikle "başarılı-başarısız", "var-yok", "ölü-sağ", "pozitif-negatif" vb. ikililer olarak sınıflandırılır.

Binom dağılımının bir diğer özelliği, rassal değişkenin aldığı değerlerin sayma yoluyla elde edilmesidir. Diğer bir deyişle binom denemesinde, toplam deneme sayısındaki istenilen sonuçların sayısı sayılır. Örneklerimize dönersek bozuk para 10 kez atıldığında, gelen yazı sayısı sayılarak bulunacaktır. Herhangi bir dersin sınavına giren 50 öğrenciden başarılı olanların sayısı sayılarak elde edilecektir ya da sınıftan seçilen 8 öğrenciden kaç tanesinin kız öğrenci olduğu sayılarak görülebilecektir.

Binom dağılımının üçüncü özelliği, ilgilenilen sonucun gerçekleşme olasılığının denemeden denemeye değişmeyerek sabit kalmasıdır. Bozuk paranın ilk atılışında paranın yazı gelme olasılığı 1/2'dir. İkinci atışta paranın yazı gelme olasılığı yine 1/2'dir ve bundan sonraki atışların her birinde bu olasılık değişmeyecek ve 1/2 olarak kalacaktır. Dersin sınavına giren öğrencilerden rassal olarak seçilen bir öğrencinin başarılı olma olasılığı belli bir değer ise bu değer bir sonraki rassal seçimde de değişmeyecektir ya da sınıftan rassal olarak seçilen bir öğrencinin kız olma olasılığı % 50 ise bir sonraki seçimde de bu olasılık yine %50 olacaktır.

Binom dağılımına ilişkin son özellik, yapılan denemelerin birbirinden bağımsız olarak gerçekleştirilmesidir. Diğer bir deyişle, bir denemenin sonucu başka bir denemenin sonucunu herhangi bir şekilde etkilemez.

Binom dağılımının özellikleri özetlenecek olursa;

a) Denemeler, daima aynı koşullarda tekrarlanmalıdır.
b) Yapılacak her denemenin sonunda, var olan karşılıklı ayrık iki sonuçtan yalnızca biri ortaya çıkmalıdır. Bu sonuçlardan biri ilgilenilen sonuç, diğeri ise bunun tümleyeni olan ilgilenilmeyen sonuçtur.
c) Rassal değişken, sabit sayıda denemedeki ilgilenilen durumun sayısını belirtir.
d) Tek bir denemede ilgilenilen sonucun gerçekleşme olasılığı, tüm denemelerde aynı kalmalıdır.
e) Denemeler birbirinden bağımsız yapılmalıdır.

Binom olasılık dağılımı oluşturulurken deneme sayısından ve tek bir denemedeki ilgilenilen sonucun olasılığından yararlanılır. Örneğin, bir adet bozuk paranın 10 kez atıldığı binom denemesinde, deneme sayısı 10'dur. Her denemede birincisi yazı ve ikincisi tura olmak üzere olası iki sonuç bulunur ve bu denemeler için yazı gelme olasılığı 0,50'dir. Dolayısıyla, yukarıda belirtilen binom dağılımı koşulları sağlanır.

Binom olasılık dağılımı

$$P(x)= {}_nC_x p^x (1-p)^{n-x} \tag{21}$$

eşitliği ile verilir. Burada; n: Deneme sayısı, x: n denemede ilgilenilen sonuç sayısı olarak tanımlanan rassal değişken, p: Tek bir denemede ilgilenilen sonucun gerçekleşme olasılığı ve 1-p: Tek bir denemede ilgilenilmeyen sonucun gerçekleşme olasılığıdır.

ÖRNEK 13

Bir fabrikada üretilen herhangi bir ürünün kusurlu olma olasılığının 1/20 olduğu bilinmektedir. Bu fabrikada üretilen ürünler arasından rassal olarak seçilen 4 üründen,
a) İki tanesinin kusurlu olma olasılığı kaçtır?
b) En fazla birinin kusurlu olma olasılığı kaçtır?

Çözüm:

Binom dağılımının koşullarını inceleyecek olursak üretilen ürünler arasından rassal olarak 4 ürün seçildiği için, n=4'tür. Her ürün için kusurlu olma ya da kusurlu olmama olarak tanımlanan iki sonuç bulunur. X rassal değişkeni, seçilen 4 ürün arasındaki kusurlu ürün sayısını belirtir. Ürünün kusurlu olma olasılığı p=1/20 değeri sabittir ve son olarak seçilen bir ürünün kusurlu olması, bir diğerinin kusurlu olup olmamasını etkilemez. Dolayısıyla bu örnek olay, bir binom denemesidir ve istenen olasılıklar binom olasılık fonksiyonu yardımıyla bulunabilir.

a) $$P(2) = {}_nC_x p^x (1-p)^{n-x} = {}_4C_2 \left(\frac{1}{20}\right)^2 \left(1-\frac{1}{20}\right)^{4-2}$$

$$= 6(0{,}0025)(0{,}9025)= 0{,}0135$$

b) Ürünlerden en fazla birinin kusurlu olması olasılığını bulabilmek için, özel toplama kuralından faydalanılabilir. Çünkü ürünlerden hiçbirinin kusurlu olması ve sadece bir tanesinin kusurlu olması olayları karşılıklı ayrık olaylardır. Bu nedenle soruda istenen olasılık, ürünlerden hiçbirinin kusurlu olmaması olasılığı ile sadece bir tanesinin kusurlu olması olasılıklarını hesaplayıp bu olasılıkların toplanması yoluyla bulunur.

$$P(x \le 1) = P(x=0) + P(x=1) \;=\; {}_4C_0 \left(\frac{1}{20}\right)^0 \left(1-\frac{1}{20}\right)^{4-0} + {}_4C_1 \left(\frac{1}{20}\right)^1 \left(1-\frac{1}{20}\right)^{4-1}$$

$$= 1(1)(0{,}8145) + 4(0{,}05)(0{,}8574) = 0{,}9860$$

Dolayısıyla, fabrikada üretilen ürünler arasından rassal olarak seçilen 4 üründen en fazla birinin kusurlu olma olasılığı 0,986'dır.

Örnek 13 için kusurlu ürün sayısının olasılık dağılımını oluşturunuz.

SIRA SİZDE 4

Bir kesikli rassal değişken binom dağılımına sahipse ortalaması ve varyansı aşağıdaki eşitliklerle verilir:

Binom Dağılımının Ortalaması: $\mu = np$ (22)
Binom Dağılımının Varyansı: $\sigma^2 = np(1-p)$ (23)

ÖRNEK 14

Örnek 13'te tanımlanan rassal değişkene ilişkin dağılımın ortalamasını ve varyansını hesaplayınız.

Çözüm:
Örnekte n=4 ve p=1/20 olduğundan,

Ortalama kusurlu ürün sayısı: $\mu = np = 4 \cdot \frac{1}{20} = 0{,}2$ ve

Varyansı: $\sigma^2 = np(1-p) = 4 \cdot \frac{1}{20} \cdot \frac{19}{20} = 0{,}19$ olacaktır.

Sürekli Rassal Değişkenler İçin Olasılık Dağılımları

Sürekli olasılık dağılımları, bir nesnenin uzunluğu, ağırlığı, sıcaklığı gibi aldığı değerleri genellikle bir ölçüm sonucunda alan sürekli rassal değişkenlere ilişkin dağılımlardır. Tanımı gereğince, sürekli rassal değişkenler için deneme sonuçları bir değer aralığı üzerindeki noktalarla belirtilir ve değişkenin aldığı sayısal değerler, olasılık dağılımları yardımıyla uygun noktalarla ilişkilendirilir. Sürekli rassal değişkenler, bir aralık üzerinde bulunan sonsuz sayıda noktayla ilişkilendirilebilir. Bu nedenle, sürekli rassal değişkenin olası her değeri için olasılık hesaplanması söz konusu değildir. Yapılabilecek şey, belli bir aralık için olasılığın hesaplanmasıdır. Bu nedenle, sürekli bir olasılık dağılımı için *belli bir değer aralığında* ortaya çıkan gözlem sonuçlarının oranı belirlenmeye çalışılır.

Sürekli rassal değişkenler, belli bir aralıkta sonsuz sayıda değer alırlar. Dolayısıyla, sürekli rassal değişkenin belli bir değeri alma olasılığı yerine, belli bir aralıkta yer alma olasılığı hesaplanır.

DİKKAT

Bu bölümde, belli bir aralıkta sonsuz sayıda değerler alan sürekli bir rassal değişken için olasılık bulmada kullanılan ve uygulamada en çok rastlanan sürekli olasılık dağılımı olan normal dağılım incelenecek, ortalaması ve standart sapması tanımlanacaktır.

Normal Dağılım

İstatistik teorisinde kullanılan temel olasılık dağılımlarından en önemlisi olan normal dağılımın günlük yaşamda pek çok uygulamasıyla karşılaşılır. Örneğin, bir fabrikada üretilen ürünlerin raf ömürleri, şirket çalışanlarının aylık performans puanları, bir nakliye aracına yüklenen ürünlerin toplam ağırlığı vb. gibi sürekli değişkenlerin birçoğunun dağılımı normal dağılıma uyar. Bununla birlikte, istatistiksel analiz tekniklerinin büyük bir kısmında, evren dağılımının normal dağılıma uyduğu varsayımı kullanılır.

Normal olasılık dağılımı;

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \tag{24}$$

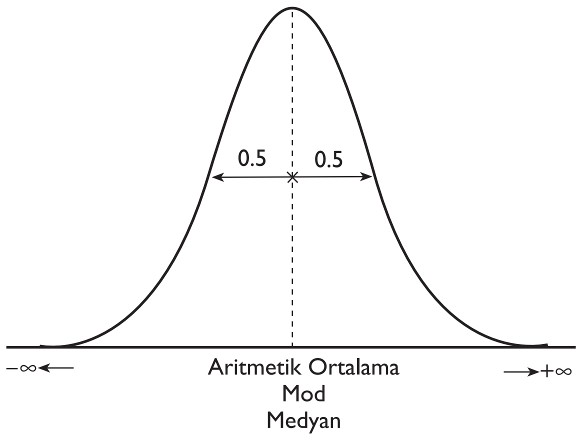
eşitliği ile verilir. Burada, X; sürekli rassal değişkenin aldığı değeri, μ; evren ortalamasını, σ; evren standart sapmasını, π; pi sayısı olan 3,14'ü ve e; doğal logaritma sisteminin tabanı olan 2,718 sayısını belirtmektedir. Dolayısıyla normal dağılımda, birincisi ortalama, ikincisi de standart sapma olmak üzere iki adet bilinmeyen parametre bulunur.

Şekil 5.1'de görüldüğü gibi, normal olasılık dağılımının şekli çan eğrisi biçimindedir ve merkezinde tek bir tepe noktasına sahiptir. Dağılımın aritmetik ortalama, mod ve medyan değerleri eşittir ve bu değerler dağılımın merkezinde yer alırlar. Dolayısıyla, normal eğrinin altında kalan alanın yarısı bu noktanın sağında, diğer yarısı da bu noktanın solunda yer alır. Buna göre, dağılım aritmetik ortalama, mod ve medyana göre simetrik bir görünümdedir. Dağılım eğrisi, merkez değerinden her iki yöne doğru düzgün bir şekilde azalır. Dolayısıyla, dağılım asimptotiktir. Bu da, eğrinin X-eksenine gitgide yaklaştığı ancak hiçbir zaman kesmediği anlamını taşır. Bir başka deyişle, eğrinin iki ucu teorik olarak X-ekseni ile sonsuzda kesişir. Fakat uygulamalarda hemen hemen tüm gözlem değerlerinin, aritmetik ortalamadan $\mp$ 3σ uzaklığa kadar değer aldığı görülür ve bu aralığın dışında kalan bölge için eğri altında kalan alanın sıfır olduğu kabul edilmektedir.

Sürekli olasılık dağılımlarında, dağılım eğrisinin altında kalan herhangi bir aralığın alanı, değişkenin bu aralıkta yer alma olasılığını verir. Ayrıca olasılık dağılımlarının özelliği gereği, olası tüm sonuçlar için normal dağılım eğrisinin altında kalan toplam alan 1'dir ve dağılımın aritmetik ortalamaya göre simetrik olması nedeniyle, aritmetik ortalamanın sağında ve solunda kalan yarı alanların büyüklüğü 0,5'e eşittir.

Şekil 5.1

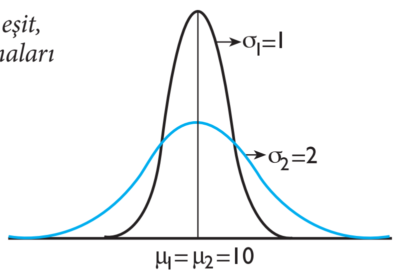
Normal Dağılım Eğrisi



Normal dağılımın konumu, μ aritmetik ortalamasına, yayılması ise s standart sapmasına göre belirlenir. Sadece tek bir eğri değil, birçok normal dağılım eğrisi çizilebilir. Şekil 5.2'de aritmetik ortalamaları eşit ancak standart sapmaları farklı iki adet normal dağılım eğrisi görülmektedir. Burada verilen eğrilerin yukarıda sözü edilen özellikleri sağlamasının yanı sıra dağılımın standart sapma değeri küçüldükçe veriler aritmetik ortalama etrafında daha fazla yoğunlaşmakta ve eğri daha sivri bir şekil almaktadır. Şekil 5.3'te, bu kez standart sapmaları aynı ancak aritmetik ortalamaları farklı, Şekil 5.4'te ise hem aritmetik ortalamaları hem de standart sapmaları farklı iki adet normal dağılım eğrisi görülmektedir.

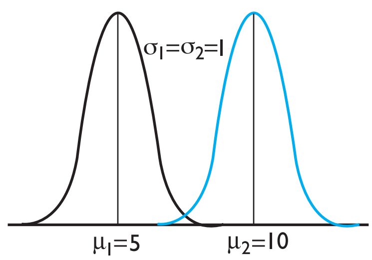
Şekil 5.2

Ortalamaları eşit, standart sapmaları farklı normal dağılımlar



Şekil 5.3

Ortalamaları farklı, standart sapmaları eşit normal dağılımlar

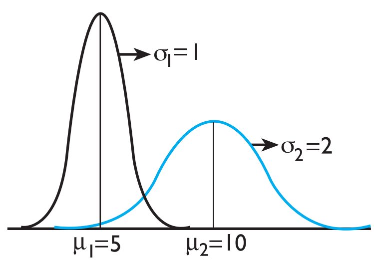


Aritmetik ortalama ve standart sapma değerlerine ilişkin olarak sonsuz sayıda farklı normal dağılım eğrisi çizilebileceğinden, tüm eğriler için olasılık hesabında kullanılacak tabloların oluşturulması mümkün değildir. Bu nedenle, olasılıkların belirlenmesinde standart normal dağılımdan yararlanılır.

Normal dağılımın özel bir durumu olan standart normal dağılım, elde edilebilecek tüm normal dağılımlar için olasılık belirlemede kullanılabilmektedir. **Aritmetik ortalaması 0 ve standart sapması 1 olan normal dağılıma standart normal dağılım adı verilir.**

Şekil 5.4

Ortalamaları ve standart sapmaları farklı normal dağılımlar



z Değeri, seçilen bir X değeri ile aritmetik ortalama (m) arasındaki farkın, standart sapma (s)'ya oranına z değeri adı verilir.

Herhangi bir normal dağılım, gözlem değerlerinden aritmetik ortalama değeri çıkartıldıktan sonra, bu farkın standart sapmaya bölünmesi yoluyla standart normal dağılıma dönüştürülebilir. Bu işlem sonucu elde edilen değerlere **z değerleri** ya da standart normal değerler adı verilir. z değerleri standart değerler olduğu için birimleri yoktur.

X gözlem değerleri aşağıdaki eşitlik yardımıyla z değerine dönüştürülür:

$$z = \frac{X - \mu}{\sigma} \qquad (25)$$

Burada, X; Herhangi bir gözlem ya da ölçüm değerini, μ; dağılımın aritmetik ortalamasını, σ ise dağılımın standart sapmasını belirtir.

μ aritmetik ortalama ve σ standart sapması ile normal dağılıma sahip X değerleri standartlaştırıldığında elde edilen z değerlerinin dağılımı, aritmetik ortalaması 0 ve standart sapması 1 olan normal dağılım, dolayısıyla standart normal dağılımdır.

DİKKAT

z değerleri, X gözlem değerlerinin aritmetik ortalamadan kaç standart sapma uzaklıkta olduğunu belirten değerlerdir.

Aritmetik ortalaması ve standart sapması belli bir normal dağılımda μ ve σ değerleri arasında değişmez ilişkiler vardır. Bu ilişki yardımıyla aritmetik ortalamadan ne kadar uzaklaşılırsa hangi büyüklükte bir alan elde edileceği belirlenebilir.

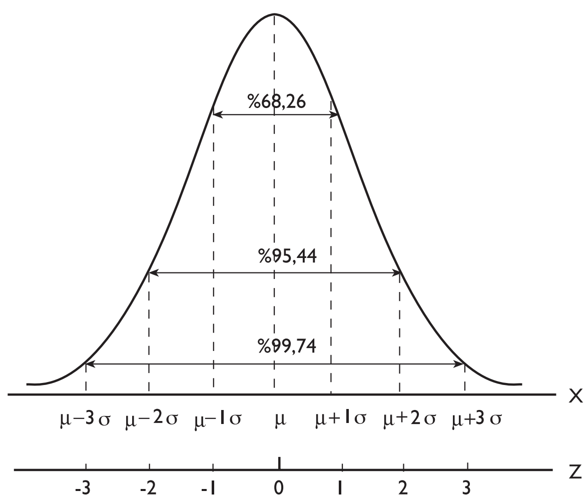
Örneğin, aritmetik ortalamadan 1μ kadar sağ ve sol tarafa ilerlenirse bu noktalarda sınırlanan yarı alanlar % 34,13 olup bu iki alanın toplamı % 68,26 olur. Bunun gibi, aritmetik ortalamadan 2σ ve 3σ kadar sağ ve sol tarafa ilerlendiğinde elde edilecek değerler Tablo 5.3'te görülmektedir.

Tablo 5.3
Normal Dağılım İçin Bazı Alan Değerleri

Aritmetik Ortalama		Standart Sapma	Kapsadığı Alan
μ	∓	1σ	0,3413+0,3413=0,6826
μ	∓	2σ	0,4772+0,4772=0,9544
μ	∓	3σ	0,4987+0,4987=0,9974

Şekil 5.5'te, aritmetik ortalamadan ∓ 1s'dan ∓ 3σ'ya kadar uzaklaşıldığında ortaya çıkan alanlar görülmektedir.

Şekil 5.5
Normal Dağılım Eğrisinin Altında Kalan Alanlar



Burada, x değerleri z değerlerine dönüştürüldüğünde, ölçek de değişmektedir. Dönüşüm sonucu elde edilecek z değerleri de grafiğin alt kısmında görülmektedir. Örneğin, μ değerine karşılık gelen z değeri 0'dır, $\mu+1\sigma$ değerine karşılık gelen z değeri ise +1'dir.

Belli bir aralıkta eğrinin altında kalan alanın hesaplanabilmesi amacıyla, standart normal dağılım tablosu kullanılır. Ek'te verilen bu tabloda, z değerleri ve bunlara karşılık gelen alanlar bulunur. Bu alanlar aynı zamanda standart normal dağılıma sahip bir rassal değişkenin 0 ile z aralığında bir değer alması olasılıklarını vermektedir. Normal dağılımın aritmetik ortalamaya göre simetrik bir dağılım olmasından dolayı, z'nin negatif değerleri için de aynı olasılıklar kullanılabilir.

Örnek 15'te, Ek-1'de verilen standart normal dağılım tablosundan yararlanarak eğri altında kalan alanların bulunuşu açıklanmaktadır.

ÖRNEK 15

X rassal değişkeni, 100 aritmetik ortalaması ve 10 standart sapması ile normal dağılıma sahiptir. X'in 100 ile 123,3 arasında olma olasılığını bulunuz.

Çözüm:

Burada, P(100<x<123,3) olasılığı sorulmaktadır. Bu olasılığın bulunabilmesi için, ilk olarak alt sınır 100 ve üst sınır 123,3 değerlerinin, 25 no'lu eşitlik kullanılarak z değerlerine dönüştürülmesi gerekir.

x=100 değerine karşılık gelen z değeri, $z=\dfrac{100-100}{10}=0$ 'dır. Yukarıda belirtildiği üzere, aritmetik ortalamaya karşılık gelen z değeri daima 0'dır. x=123,3 değerine karşılık gelen z değeri ise, $z=\dfrac{123,3-100}{10}=2,33$ olarak elde edilir. Ardından standart normal dağılım tablosuna bakılır. Tablo 5.4'te bu tablonun bir bölümü görülmektedir. z=0 ile z=2,33 arasında kalan bölgenin alanını bulabilmek için tabloda z harfi ile belirtilen ilk sütunda 2,3 değerini bulana kadar aşağıya doğru ilerlenir. Daha sonra, bu satırda sağa doğru ilerlenir ve 0,03 ile başlayan sütunun altındaki alan değeri 0,4901 olarak bulunur. Bu da, 0,00 ile 2,33 arasında eğrinin altında kalan alanın 0,4901 olduğu anlamına gelir ve bu değer X'in 100 ile 123,3 arasında yer alma olasılığıdır.

Tablo 5.4
Normal Eğri Altında Kalan Alanlar

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	...
⋮								
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	
⋮								

SIRA SİZDE 5

Standart normal eğri altındaki z=-1,16'nın sağındaki alanı bulunuz.

ÖRNEK 16

Otomobil lastiği üretimi yapan bir fabrikada, üretilen lastiklere uygulanan bir dayanıklılık testi sonucunda lastiklerin teste dayanma sürelerinin, aritmetik ortalaması 242 sn. ve standart sapması 36 sn. ile normal dağılıma sahip olduğu bilinmektedir. Bu fabrikada üretilen ve teste tabi tutulan lastikler arasından rassal olarak seçilen bir lastiğin teste dayanma süresinin,

a) 300 sn.'den fazla olma,

b) 220 ile 290 sn. arasında olma,

c) 256 ile 308 sn. arasında olma olasılıklarını bulunuz.

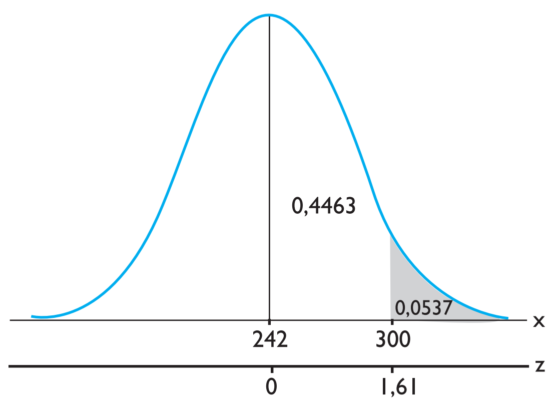
d) Toplam 150 lastiğe test uygulandığı bilindiğine göre, bu teste 300 sn.'den daha fazla dayanabilen lastik sayısı kaçtır?

Çözüm:

X rassal değişkeni, "Lastiğin dayanıklılık testine dayanma süresi" olarak tanımlandığında, soruda verilen bilgilere göre bu X rassal değişkeninin dağılımı, ortalaması 242 sn. ve standart sapması 36 sn. ile normal dağılımdır.

a) Seçilen lastiğin teste dayanma süresinin 300 sn.'den fazla olma olasılığını bulabilmek için öncelikle bu X değerine karşılık gelen z değeri hesaplanmalıdır. Diğer bir deyişle, z dönüşümü uygulanmalıdır.

x=300 noktasına karşılık gelen z değeri, $z = \frac{x-\mu}{\sigma} = \frac{300-242}{36} = 1,61$ olarak elde edilir. Buna göre soruda istenen olasılık; z=1,61 değerinin sağında kalan bölgenin alanı olacaktır. Dikkat edilirse aritmetik ortalamanın (ya da z=0 noktasının) sağında ve normal dağılım eğrisinin altında kalan bölgenin alanı, toplam alanın yarısı olan 0,5'tir. Standart normal dağılım tablosuna göre z=0 ile z=1,61 arasındaki bölgenin alanı 0,4463 olarak bulunur. Buna göre, soruda istenen olasılığı bulabilmek için 0,5 değerinden, z=1,6 değerine karşılık gelen alanı, yani 0,4463 değerini çıkartmak gerekir.



Buna göre, seçilen lastiğin teste dayanma süresinin 300 sn.'den fazla olma olasılığı;

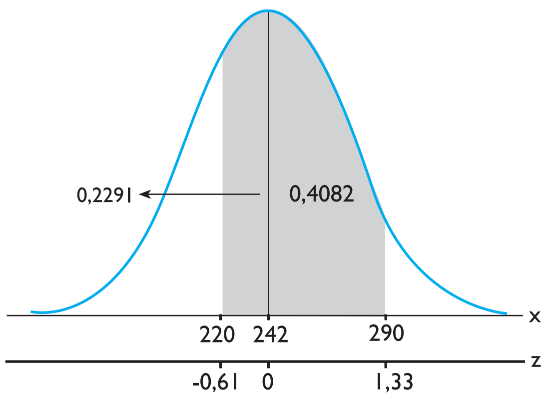
P(x>300)=P(z>1,61)=0,5-P(0<z<1,61)

=0,5-0,4463=0,0537 olarak bulunur.

b) Seçilen lastiğin teste dayanma süresinin 220 ile 290 sn. arasında olma olasılığını bulabilmek için bu değerlere karşılık gelen iki tane z değeri hesaplanmalıdır.

x_1=220 noktasına karşılık gelen z_1 değeri, $z_1 = \dfrac{220-242}{36} = -0,61$ ve x_2=290 noktasına karşılık gelen z_2 değeri, $z_2 = \dfrac{290-242}{36} = 1,33$ olarak elde edilir.

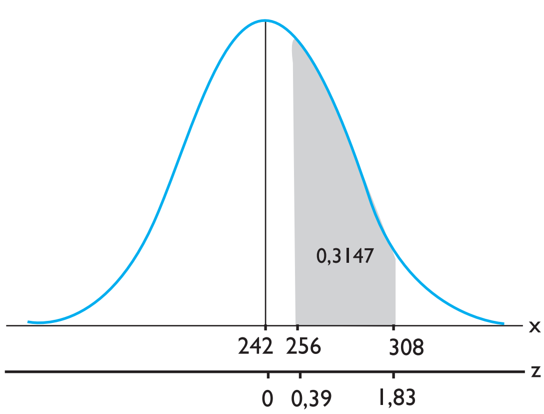
Standart normal dağılım tablosuna göre, z_1=-0,61 değeri için eğri altında kalan alan 0,2291 ve z_2=1,33 değeri için alan 0,4082 olarak bulunur. Buna göre, soruda istenen olasılığı bulabilmek için bu iki olasılığı toplamak gerekir.



Buna göre, seçilen lastiğin teste dayanma süresinin 220 ile 290 sn. arasında olma olasılığı;

$P(220<x<290)=P(-0,61<z<1,33)=P(0<z<0,61)+P(0<z<1,33)$
$=0,2291+0,4082=0,6373$ olarak bulunur.

c) x_1=256 noktasına karşılık gelen z_1 değeri, $z_1 = \dfrac{256-242}{36} = 0,39$ ve x_2=308 noktasına karşılık gelen z_2 değeri, $z_2 = \dfrac{308-242}{36} = 1,83$ olarak elde edilir. Standart normal dağılım tablosuna göre, z_1=0,39 değeri için eğri altında kalan alan 0,1517 ve z_2=1,83 değeri için alan 0,4664 olarak bulunur. Buna göre, soruda istenen olasılığı bulabilmek için büyük olasılık değerinden küçük olanı çıkartmak gerekir.



Buna göre, seçilen lastiğin teste dayanma süresinin 256 ile 308 sn. arasında olma olasılığı;

$$P(256<x<308)=P(0,39<z<1,83)=P(0<z<1,83)-P(0<z<0,39)$$
$$=0,4664-0,1517=0,3147 \text{ olarak bulunur.}$$

d) Sorunun (a) şıkkında, fabrikada üretilen ve dayanıklılık testine tabi tutulan lastikler arasından rassal olarak seçilen bir lastiğin teste dayanma süresinin 300 sn.'den fazla olma olasılığı 0,0537 olarak bulunduğuna göre, bu değer teste tabi tutulan toplam 150 lastiğin % 5,37'sinin teste dayanma süresinin 300 sn.'den fazla olduğunu gösterir. Dolayısıyla, kaç adet lastiğin teste dayanma süresinin 300 sn.'den fazla olduğunu bulabilmek için toplam lastik sayısı ile olasılık değerini çarpmak gerekir.

Sorunun cevabı; $150 \times 0,0537 \cong 8$ lastik olarak hesaplanır.

SIRA SİZDE 6

Normal dağılıma sahip bir rassal değişkenin aritmetik ortalaması 6,4 ve standart sapması 1,2 ise bu rassal değişkenin 7'den küçük olma olasılığı kaçtır?

Özet

AMAÇ 1

Olasılık kavramını ve olasılığa ilişkin yaklaşımları tanımlamak.

Olasılık, herhangi bir durumun ya da olayın gerçekleşme şansını ifade eden 0 ile 1 kapalı aralığında bir sayı olarak tanımlanır. Bir olayın gerçekleşmesi kesin ise olasılığı 1, gerçekleşmesi imkânsız ise olasılığı 0'dır. Olasılık değeri sıfıra yaklaştıkça olayın gerçekleşme şansı gittikçe azalmakta, olasılık değeri bire yaklaştıkça da o olayın gerçekleşme şansı gittikçe artmaktadır.

Klasik olasılık tanımı, bir denemede n adet eşit olasılıklı sonuç bulunduğu durumda uygulanır. Klasik bakış açısıyla, bir olayın gerçekleşme olasılığı ilgilenilen sonuçların sayısının, olası tüm sonuçların sayısına bölünmesi yoluyla hesaplanır.

Deneysel olasılık, bir olayın gerçekleşme sayısı toplam gözlem sayısına bölünerek hesaplanır.

Öznel (Subjektif) Olasılık, eldeki mevcut bilgilere dayanan olasılıktır.

AMAÇ 2

Toplama kuralları, çarpma kuralları ve Bayes teoremi yardımıyla herhangi bir olayın gerçekleşme olasılığını hesaplamak.

Eğer A ve B olayları karşılıklı ayrık olaylar ise özel toplama kuralına göre bu olaylardan birinin veya diğerinin gerçekleşme olasılığı, bu olayların ayrı ayrı gerçekleşme olasılıklarının toplamına eşittir. P(A veya B)=P(A)+P(B)

Tümleyen kuralına göre, bir olayın gerçekleşme olasılığı, bu olayının gerçekleşmeme olasılığının 1'den çıkarılması yoluyla belirlenir: P(A)=1-P($\bar{A}$)

A ve B gibi iki olay için genel toplama kuralı; P(A veya B)=P(A)+P(B)-P(A ve B) eşitliği ile verilir.

Özel çarpma kuralına göre; A ve B bağımsız olayları için A ve B'nin birlikte gerçekleşme olasılığı, bu iki olayın ayrı ayrı gerçekleşme olasılıklarının çarpımına eşittir: P(A ve B)=P(A)·P(B)

B olayının gerçekleştiği bilindiğinde, A olayının gerçekleşme olasılığına koşullu olasılık adı verilir ve bu olasılık P(A   B) ile gösterilir.

Genel çarpma kuralına göre A ve B gibi iki olay için, bu olayların birlikte gerçekleşme olasılığı, A olayının gerçekleşme olasılığı ile A'nın gerçekleştiği bilindiğine göre B'nin koşullu olasılığının çarpımına eşittir: P(A ve B)=P(A)·P(B|A)

Bayes teoremine göre, B olayının gerçekleştiği bilindiğine göre, A_i'nin gerçekleşme olasılığı:

$$P(A_i \; B)=\frac{P(A_i)P(B \; A_i)}{P(A_1)P(B \; A_1)+P(A_2)P(B \; A_2)+\ldots+P(A_n)P(B \; A_n)}$$

eşitliği ile hesaplanır.

AMAÇ 3

Bir denemedeki olası sonuç sayısını belirlemek üzere saymanın temel ilkesini uygulamak, permütasyon ve kombinasyon sayılarını hesaplamak.

Bir denemedeki olası sonuç sayısını belirlemede kullanılan üç sayma kuralı bulunur. Saymanın Temel İlkesine göre, eğer bir işlem m farklı yolla, bir başka işlem de n farklı yolla gerçekleşebiliyorsa, bu iki işlem birlikte m × n farklı yolla gerçekleşir.

n olası nesnenin tek bir grubundan seçilen r adet nesnenin herhangi bir sıralamasına permütasyon adı verilir. Toplam farklı permütasyon sayısı:

$${}_nP_r=\frac{n!}{(n-r)!}$$

Seçilecek nesnelerin diziliş sırası önemli değilse, yapılan her bir seçime bir kombinasyon adı verilir. n adet nesneden oluşan bir kümeden r nesnelik kombinasyon sayısı:

$${}_nC_r=\frac{n!}{r!(n-r)!}$$

AMAÇ 4

Rassal değişken ve olasılık dağılımı kavramlarını açıklamak, kesikli olasılık dağılımları için aritmetik ortalama, varyans ve standart sapma hesaplamak.

Denemeden denemeye farklı değerler alan ve aldığı bu değerleri belli bir olasılıkla alan değişkenlere rassal değişken adı verilir. Rassal değişkenler, kesikli ve sürekli rassal değişkenler olmak üzere iki şekilde ortaya çıkarlar. Sonlu ya da sayılabilir sayıda farklı değeri bulunan rassal değişkenlere kesikli rassal değişken adı verilir. Sayılamayacak ya da sonsuz sayıda olası değeri bulunan ve bir sayı aralığı ya da aralık kümesi üzerinde tanımlanan rassal değişkenlere sürekli rassal değişken adı verilir.

Bir denemedeki olası tüm sonuçların ve bu sonuçların her birine ilişkin olasılıkların yer aldığı listeye olasılık dağılımı adı verilir. Belli bir sonucun olasılığı 0 ile 1 kapalı aralığında değerler alır ve tüm karşılıklı ayrık olayların olasılıkları toplamı 1'e eşittir.

Eğer X kesikli bir rassal değişken ise X'in olasılık dağılımı, rassal değişkenin alabileceği tüm olası x değerlerine ilişkin olasılıkların yer aldığı listedir.

Bir olasılık dağılımının ortalamasına, o olasılık dağılımının **beklenen değeri** adı verilir. Kesikli olasılık dağılımları için ortalama; $\mu = \sum xP(x)$ ve kesikli bir olasılık dağılımının varyansı; $\sigma^2 = \sum (x-\mu)^2 P(x)$ eşitlikleri ile hesaplanır. σ^2 varyans değerinin pozitif karekökü standart sapma olarak tanımlanır ve σ ile gösterilir. Dolayısıyla, $\sigma = \sqrt{\sigma^2}$ olur.

AMAÇ 5

Binom ve normal dağılımın özelliklerini açıklamak ve bu dağılımlar yardımıyla olasılık hesaplamak.

Binom dağılımına ilişkin özellikler: Denemeler, daima aynı koşullarda tekrarlanmalıdır. Yapılacak her denemenin sonunda, var olan karşılıklı ayrık iki sonuçtan yalnızca biri ortaya çıkmalıdır. Bu sonuçlardan biri ilgilenilen sonuç, diğeri ise bunun tümleyeni olan ilgilenilmeyen sonuçtur. Rassal değişken, sabit sayıda denemedeki ilgilenilen durumun sayısını belirtir. Tek bir denemede ilgilenilen sonucun gerçekleşme olasılığı, tüm denemelerde aynı kalmalıdır ve denemeler birbirinden bağımsız yapılmalıdır.

Binom olasılık dağılımı; $P(x)={}_nC_x p^x(1-p)^{n-x}$, Binom dağılımının ortalaması; m=np ve varyansı; $\sigma^2=np(1-p)$ ile verilir.

Normalolasılıkdağılımı; $P(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2}$ eşitliği ile verilir. Normal olasılık dağılımının şekli çan eğrisi biçimindedir ve merkezinde tek bir tepe noktasına sahiptir. Dağılımın aritmetik ortalama, mod ve medyan değerleri eşittir ve bu değerler dağılımın merkezinde yer alırlar. Dağılım eğrisinin iki ucu teorik olarak X-ekseni ile sonsuzda kesişir.

Aritmetik ortalaması 0 ve standart sapması 1 olan normal dağılıma standart normal dağılım adı verilir. Herhangi bir normal dağılım, gözlem değerlerinden aritmetik ortalama değeri çıkartıldıktan sonra, bu farkın standart sapmaya bölünmesi yoluyla standart normal dağılıma dönüştürülebilir: $z = \dfrac{X-\mu}{\sigma}$. Belli bir aralıkta eğrinin altında kalan alanın hesaplanabilmesi amacıyla standart normal dağılım tablosu kullanılır. Bu tabloda, z değerleri ve bunlara karşılık gelen olasılıklar bulunur.

Kendimizi Sınayalım

1. Aşağıdakilerden hangisi, eldeki mevcut bilgilere dayalı olarak hesaplanan olasılık yaklaşımını ifade eder?

a. Deneysel Olasılık
b. Objektif Olasılık
c. Klasik Olasılık
d. Öznel olasılık
e. Nesnel Olasılık

2. A olayının gerçekleşme olasılığı 0,30, B olayının gerçekleşme olasılığı 0,45'dir. A veya B olayının gerçekleşme olasılığı 0,50 olduğuna göre, A ve B olaylarının birlikte gerçekleşme olasılığı kaçtır?

a. 0,05
b. 0,10
c. 0,15
d. 0,20
e. 0,25

3. Bir bankanın müşterilerinin %40'ının vadesiz hesabı bulunmaktadır. Vadesiz hesabı bulunan müşterilerin %30'unun yatırım hesabı bulunduğuna göre, bu banka müşterileri arasından rassal olarak seçilen bir müşterinin hem vadesiz hem de yatırım hesabı bulunması olasılığı kaçtır?

a. 0,10
b. 0,12
c. 0,38
d. 0,70
e. 0,75

4. Bir alışveriş merkezi müşterilerinin %40'ı yaptığı alışverişlerde nakit para, %60'ı ise kredi kartı kullanmaktadır. Nakit para kullananların %30'unun, kredi kartı kullananların ise %60'ının yaptığı alışveriş 50 TL'nin üzerindedir. Bu alışveriş merkezi müşterileri arasından rassal olarak seçilen bir müşterinin yaptığı alışverişin 50 TL'nin üzerinde olduğu bilindiğine göre, bu müşterinin alışverişinde nakit para kullanmış olma olasılığı kaçtır?

a. 0,12
b. 0,25
c. 0,36
d. 0,48
e. 0,75

5. Bir bilimsel araştırma projesi için altı bilim adamı arasından bir proje sorumlusu ve bir araştırmacı kaç farklı şekilde seçilebilir?

a. 4
b. 6
c. 10
d. 15
e. 30

6. Bir lojistik firmasının 8 adet kamyonu bulunmaktadır. Firmanın ilgili sorumlusu, şehir dışına nakliyat yapmak üzere bu 8 kamyondan 3 tanesini kaç farklı şekilde seçebilir?

a. 24
b. 27
c. 56
d. 336
e. 512

7. X: "İstatistik dersi final sınavında, öğrencilerin cevapladığı soru sayısı" rassal değişkeni ile ilgili olarak aşağıda verilen ifadelerden hangisi doğrudur?

a. X, bir kesikli rassal değişkendir.
b. X rassal değişkeni, $-\infty$ ile $+\infty$ aralığında sürekli değerler alır.
c. X'in olası tek bir değeri için olasılık hesaplanamaz.
d. X'in olası değerleri, bir sayı aralığı üzerinde ve sonsuz sayıdadır.
e. X, normal dağılıma sahiptir.

8. Bir rassal değişkenin alabileceği değerler; 1, 4, 7 ve bu değerleri alması olasılıkları sırasıyla 0,2, 0,7 ve 0,1 olduğuna göre, bu olasılık dağılımının aritmetik ortalaması ve varyansı aşağıdakilerden hangisinde doğru olarak verilmiştir?

a. $\mu=3,3;\ \sigma^2=2,61$
b. $\mu=3,3;\ \sigma^2=1,62$
c. $\mu=3,7;\ \sigma^2=2,61$
d. $\mu=3,7;\ \sigma^2=1,62$
e. $\mu=4;\ \ \sigma^2=2,61$

9. Bir binom denemesinde deneme sayısı 5 ve tek bir denemede ilgilenilen sonucun gerçekleşme olasılığı 0,40 ise bu 5 denemede ilgilenilen sonuç sayısının 3 olma olasılığı kaçtır?

a. 0,77
b. 0,54
c. 0,46
d. 0,23
e. 0,11

10. Normal dağılıma sahip bir rassal değişkenin aritmetik ortalaması 40 ve standart sapması 6 ise bu rassal değişkenin 37 ile 52 arasında olma olasılığı kaçtır?

a. 0,1915
b. 0,2857
c. 0,3313
d. 0,4772
e. 0,6687

Kendimizi Sınayalım Yanıt Anahtarı

1. d	Yanıtınız yanlış ise "Olasılık Tanımları" konusunu yeniden gözden geçiriniz.
2. e	Yanıtınız yanlış ise "Olasılık Hesaplama Kuralları" konusunu yeniden gözden geçiriniz.
3. b	Yanıtınız yanlış ise "Olasılık Hesaplama Kuralları" konusunu yeniden gözden geçiriniz.
4. b	Yanıtınız yanlış ise "Bayes Teoremi" konusunu yeniden gözden geçiriniz.
5. e	Yanıtınız yanlış ise "Sayma Kuralları" konusunu yeniden gözden geçiriniz.
6. c	Yanıtınız yanlış ise "Sayma Kuralları" konusunu yeniden gözden geçiriniz.
7. a	Yanıtınız yanlış ise "Rassal Değişkenler ve Olasılık Dağılımları" konusunu yeniden gözden geçiriniz.
8. c	Yanıtınız yanlış ise "Rassal Değişkenler ve Olasılık Dağılımları" konusunu yeniden gözden geçiriniz.
9. d	Yanıtınız yanlış ise "Rassal Değişkenler ve Olasılık Dağılımları" konusunu yeniden gözden geçiriniz.
10. e	Yanıtınız yanlış ise "Rassal Değişkenler ve Olasılık Dağılımları" konusunu yeniden gözden geçiriniz.

Sıra Sizde Yanıt Anahtarı

Sıra Sizde 1

A: {Seçilen kartın ikili olması} ve B: {Seçilen kartın karo olması} olayları tanımlansın. 52 kartlık oyun destesinde 4 adet ikili ve 13 adet karo, 1 tane de karo ikilisi bulunur. Buna göre;

$$P(\text{A veya B}) = P(A) + P(B) - P(\text{A ve B}) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = 0,3077$$

Sıra Sizde 2

S_1 = {Çekilen birinci bilyenin sarı renkli olması} ve S_2= {Çekilen ikinci bilyenin sarı renkli olması} olayları tanımlansın. Çekilen bilyenin sarı renkli olma olasılığı; $P(S_1) = \frac{4}{12}$ ve çekilen birinci bilyenin sarı renkli olduğu bilindiğinde, çekilen ikinci bilyenin de sarı renkli olma olasılığı; $P(S_2\ S_1) = \frac{3}{11}$ olacaktır. Buna göre istenen olasılık;

$$P(S_1 \text{ ve } S_2) = P(S_1) \cdot P(S_2\ S_1) = \frac{4}{12} \cdot \frac{3}{11} = \frac{1}{11}$$

olarak hesaplanır.

Sıra Sizde 3

M_1 = {Seçilen kitabın birinci makinede basılması}, M_2 = {Seçilen kitabın ikinci makinede basılması}, M_3 = {Seçilen kitabın üçüncü makinede basılması} ve H = {Seçilen kitabın hatalı basılması} olayları tanımlansın.

$$P(M_3\ H) = \frac{P(M_3)P(H\ M_3)}{P(M_1)P(H\ M_1) + P(M_2)P(H\ M_2) + P(M_3)P(H\ M_3)}$$

$$= \frac{(0,40)(0,03)}{(0,35)(0,02) + (0,25)(0,06) + (0,40)(0,03)}$$

$$= \frac{0,012}{0,007 + 0,015 + 0,012} = 0,3529$$

Sıra Sizde 4

Kusurlu Ürün Sayısı	Olasılık
0	$P(0) = {}_4C_0 \left(\frac{1}{20}\right)^0 \left(\frac{19}{20}\right)^4 = 0.81450625$
1	$P(1) = {}_4C_1 \left(\frac{1}{20}\right)^1 \left(\frac{19}{20}\right)^3 = 0.17147500$
2	$P(2) = {}_4C_2 \left(\frac{1}{20}\right)^2 \left(\frac{19}{20}\right)^2 = 0.01353750$
3	$P(3) = {}_4C_3 \left(\frac{1}{20}\right)^3 \left(\frac{19}{20}\right)^1 = 0.00047500$
4	$P(4) = {}_4C_4 \left(\frac{1}{20}\right)^4 \left(\frac{19}{20}\right)^0 = 0.00000625$
Toplam	**1.00000000**

Sıra Sizde 5

Normal dağılım aritmetik ortalamaya göre simetrik olduğuna göre, z=-1,16'dan z=0'a kadar olan bölgenin alanı, z=0'dan z=1,16'ya kadar olan bölgenin alanına eşittir. Standart normal dağılım tablosunda z sütununda 1,1 değerini bulunur. Daha sonra, bu satırda sağa doğru ilerlenir ve 0,06 ile başlayan sütunun altındaki alan 0,3770 olarak bulunur.

z	...	0,04	0,05	0,06	0,07	0,08	...
⋮							
0,9		0,3264	0,3289	0,3315	0,3340	0,3365	
1,0		0,3508	0,3531	0,3554	0,3577	0,3599	
1,1		0,3729	0,3749	**0,3770**	0,3790	0,3810	
1,2		0,3925	0,3944	0,3962	0,3980	0,3997	
1,3		0,4099	0,4115	0,4131	0,4147	0,4162	
⋮							

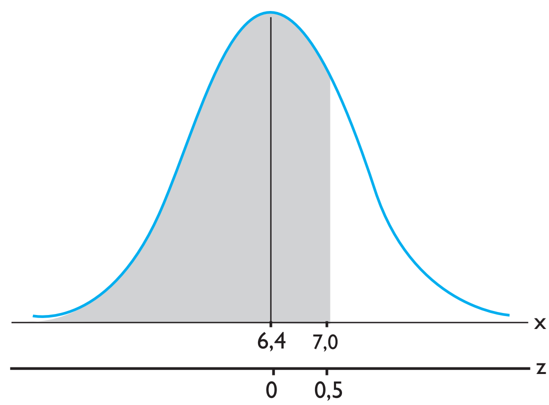
Dolayısıyla, z=-1,16'dan z=0'a kadar olan bölgenin alanı 0,3770 olur. Soruda eğrinin altında ve z=-1,16'nın sağında kalan alan sorulduğu için, 0,3770 değerine tüm alanın yarısı olan 0,5 değerini eklemek gerekir. Sonuç olarak sorunun cevabı P(z>-1,16)=0,3770+0,5=0,8770'dir.

Sıra Sizde 6

x=7 noktasına karşılık gelen z değeri,

$z = \frac{x-\mu}{\sigma} = \frac{7-6,4}{1,2} = 0,5$'dir. Soruda istenen olasılık;

z=0,5 değerinin solunda kalan bölgenin alanıdır. Standart normal dağılım tablosuna göre z=0 ile z=0,5 arasındaki bölgenin alanı 0,1915'tir. İstenen olasılığı bulabilmek için 0,5 değeri ile 0,1915 değerini toplamak gerekir.



Buna göre, rassal değişkenin 7'den küçük olma olasılığı;

P(x<7) = P(z<0,5)=0,5+P(0<z<0,5)
= 0,5+0,1915 = 0,6915'tir

Yararlanılan ve Başvurulabilecek Kaynaklar

Bluman, A.G. (2005). **Elementary Statistics: A Step by Step Approach,** McGraw-Hill.

Er, F. ve Peker, K.Ö. (2009). **Biyoistatistik,** TC. Anadolu Üniversitesi Yayını, No: 1932, Açıköğretim Fakültesi Yayını, No: 1013, Eskişehir.

Freund, J.E. (1992). **Mathematical Statistics,** Prentice-Hall, Inc., USA.

İnal, H.C. ve Günay, S. (1999). **Olasılık ve Matematiksel İstatistik,** H.Ü. Fen Fakültesi Yayınları, Ankara.

Lind, D.A., Marchal, W.G. ve Wathen, S.A. (2005). **Statistical Techniques in Business & Economics,** McGraw-Hill.

Yüzer, A.F., Ağaoğlu, E., Tatlıdil, H., Özmen, A. ve Şıklar, E. (2003). **İstatistik,** TC. Anadolu Üniversitesi Yayını, No: 1448, Açıköğretim Fakültesi Yayını, No: 771, Eskişehir.

6

Amaçlarımız

Bu üniteyi tamamladıktan sonra;

- Tam sayım ve örnekleme kavramlarını ayırt edebilecek,
- Örneklemeye başvurmanın yararlarını açıklayabilecek,
- Örnekleme planı hazırlayabilecek,
- Bir örnek araştırma için örnekleme uygulaması yapabilecek ve istenen bilgileri üretebileceksiniz.

Anahtar Kavramlar

- Tam sayım
- Çerçeve
- Parametre
- Değişken
- Örnekleme
- Örneklem
- İstatistik

İçindekiler

Örnekleme ve Bazı Örnekleme Dağılımları

TAM SAYIM VE ÖRNEKLEME

Anahtar kelimeler başlığı altında verilen kavramlar örnekleme konusunu açıklayabilmek için de bilinmesi gereken kavramlardır. Bu kavramlar birinci ünitede açıklandığı için bu üniteye sadece tam sayım ve örnekleme kavramlarıyla ilgili hatırlatıcı bilgiler verilerek başlanmıştır.

Bu ünitedeki konuları kolayca anlayabilmeniz için birinci ünitede açıklanan temel kavramları tekrar okuyunuz.

DİKKAT

Bilindiği gibi istatistiksel araştırmaların amacı tanımlanan **evren**in özellikleriyle ilgili bilgiler üretmektir. Bu bilgiler ya tam sayım sonucu elde edilen veri kümesinin (evren veri kümesinin) çözümlenmesiyle ya da örneklemden elde edilen veri kümesinin (örneklem veri kümesinin) çözümlenmesiyle üretilebilir.

Hakkında araştırma yapılacak birimler topluluğuna **evren** denir.

Tam Sayım

Planlanan bir istatistiksel araştırma için tanımlanan sonlu evrenin bütün birimleri üzerinden araştırmaya konu olan değişkenler itibariyle veri derleniyorsa yapılan işleme tam sayım denir. Tam sayım sonucu elde edilen veri kümesinin çözümlenmesiyle elde edilen bilgiler (**parametre** değerleri) veri derleme ve çözümleme hatası işlenmemiş ise kesin ve doğru bilgilerdir.

Tam sayım sonucu elde edilen veriler kullanılarak hesaplanan sayısal değerlere **parametre** denir.

Tam sayım genellikle sonlu ve küçük hacimli evrenlere uygulanır. Bununla birlikte büyük hacimli sonlu evrenlerin bütün birimlerine ulaşabilmek olanaklıysa, karşılaşılan özel problemin çözümü için mümkün bütün verilerin elde edilmesine gereksinim varsa tam sayım yapılmalıdır.

ÖRNEK 1

Yeni bir ücret sisteminin uygulandığı 50 işçisi olan bir işletmede, işçilerin bu yeni ücret sisteminden memnuniyetleri araştırılmak istenmektedir. Tam sayım yapılabilir mi?

Çözüm:

Burada evren hacmi N=50 işçiden oluşmaktadır ve küçük hacimli bir evrendir. İşçilerin her birine ulaşmak ve onlardan memnuniyetleri konusunda veri elde etmek hem kolaydır hem de çok zaman almaz. Bu nedenle istenen araştırma için tam sayım yapılabilir.

ÖRNEK 2

Bir banka şube müdürü, genel müdürlükte izleyen gün yapılacak bir toplantı için, şubesinin, toplantı öncesindeki son işgünü itibariyle mevduat durumuna ilişkin bilgiye ihtiyaç duymaktadır. Şubenin 150.000 mevduat müşterisi bulunmaktadır. İhtiyaç duyulan bilgi için tam sayım yapılabilir mi?

Çözüm:

Evren hacmi N=150.000 müşteridir. Büyük hacimli bir evrendir. Ancak şubenin bilgisayar donanımına sahip olması ve mevduat hesaplarıyla ilgili verilerin veritabanında bulunması nedeniyle, evren hacmi büyük bile olsa çok kısa zamanda gerekli olan bilgilerin elde edilmesi, tam sayım yapılması mümkündür.

Ancak tanımlanan evrenin bütün birimleri üzerinden veri derlemek veya tam sayım yapmak her zaman izleyen kısımda açıklanacak çeşitli nedenlerle mümkün olamaz, parametre değerleri hesaplanamaz. Böyle bir durumda evrenin özellikleriyle ilgili bilgiler ve genellemeler örnekleme uygulamasıyla elde edilebilir.

Örnekleme

Örneklem, evrenden belirli yöntemlerle seçilmiş olan ve seçildiği evreni temsil ettiği düşünülen birimler kümesidir.

Tanımlanan evrenden onu ilgilenilen değişkenler bakımından temsil eden sınırlı sayıda birimin belirli yöntemler kullanılarak seçilmesi işlemine örnekleme, seçilen birimlerin oluşturduğu topluluğa **örneklem** denir.

ÖRNEK 3

Bir anaokulu işletmecisinin 5 ayrı bölgedeki okullarında 1000 öğrencisi bulunmaktadır. Bu işletmeci öğrencilerine uyguladığı beslenme programıyla ilgili öğrenci ailelerinin görüşlerini almak amacıyla bir araştırma planlıyor.

Araştırmanın Amacı: Uygulanan beslenme programıyla ilgili ailelerin görüşlerinin alınması.

Araştırmanın Evreni: 5 bölgedeki okullarda okuyan öğrencilerin ailelerinin oluşturduğu topluluktur.

Örnekleme: Her bölgedeki okuldan 20şer olmak üzere toplam n=100 aile seçiliyor. Ailelerin seçim işlemine örnekleme, seçilen 100 ailenin oluşturduğu topluluğa örneklem adı verilir. Öğrencilerin annelerinin beslenme programıyla ilgili görüşleri ise değişkendir.

Seçilen ailelerin oluşturduğu topluluk örneklem, öğrencilerin anneleri gözlem birimi, aileler örnekleme birimidir.

Evreni temsil eden, onun bir modeli olan örneklemden elde edilen veri kümesi kullanılarak yapılacak çözümlemenin sonucu olan bilgi (örneklem istatistiği) evren bilgisi anlamında kullanılabilir. Başka bir deyişle örneklem istatistiğinin değeri bilinmeyen evren parametre değeri hakkında genelleme yapmak amacıyla kullanılabilir. Örneklemede önemli olan, eğer evren doğru, net tanımlanmış ise örneklemin araştırmaya konu olan değişkenler itibariyle evreni temsil edip etmemesi konusudur. İyi, temsili bir örneklem evrenin sadece birim sayısı bakımından küçük, özellikleri bakımından benzeri ve modeli olan örneklemdir. Eğer ilgilenilen değişkenler bakımından evrendeki ve örneklemdeki birimler benzer dağılım gösteriyorsa oluşturulan örneklem temsili bir örneklemdir.

ÖRNEK 4

Bir fakültede kayıtlı olan ve hacmi N=2000 öğrenci olan bir evren tanımlayalım. Bunların %40'ı erkek ve erkeklerin de %80'inin yaşları 20-24 arasında değişiyor olsun. Diyelim ki n=200 öğrenciden oluşan bir örneklem seçilsin ve bu öğrencilerin %40'ı (80 öğrenci) erkek ve bunların %80'inin (64 öğrencinin) yaşları 20-24 arasında değişiyorsa bu örneklem cinsiyet ve yaş aralığı değişkenleri bakımından temsili örneklemdir.

SIRA SİZDE 1

- **Bir kitapçıya gelen müşterilerin kitap okuma alışkanlığı araştırılmak isteniyor. Tam sayım yapılabilir mi? Tartışınız.**
- **Örneklem istatistiğinin değeri her zaman parametre değeri için bilgi niteliği taşır mı? Tartışınız.**
- **Örneklemenin temel amacı nedir? Açıklayınız.**

ÖRNEKLEMEYİ GEREKLİ KILAN NEDENLER

KİTAP

Bu ünitenin izleyen kısımlarındaki konular için, A. F. Yüzer, E. Şıklar, E. Ağaoğlu, H. Tatlıdil, A. Özmen, Editör: A. F. Yüzer (2011). İstatistik, Ünite 7 ve 8, Eskişehir Anadolu Üniversitesi Yayını İsimli kitaptan, editör ve ünite yazarlarının izni alınarak yararlanılmıştır.

- Evrenin sonsuz evren olması.

Tanımlanan evrenin sonsuz evren olması durumunda tam sayım mümkün olmaz. Çünkü incelenecek birimler X rassal değişkeninin teorik olasılık dağılımının türettiği sonuçlardır. Dolayısıyla incelenecek birim sayısında ve gözlem değeri sayısında bir sınır yoktur. Bu nedenle örnekleme uygulamasına başvurmak kaçınılmazdır.

ÖRNEK 5

Bir bisküvi fabrikasında üretilen paketlerin planlanan ağırlıkta üretilip üretilmediği, sistemin doğru çalışıp çalışmadığı araştırılmak isteniyor. Bu amaçla üretilen paketler arasından 250 paket seçiliyor. Tam sayım yapılıp yapılmayacağını açıklayınız.

Çözüm:

Bisküvi paketleri bir üretim sürecinin çıktıları niteliğindedir. Üretim sürdükçe bisküvi paketleri evrenine yeni paketler dahil olacaktır. Bu nedenle bisküvi paketleri evreni sonsuz bir evrendir. İlgili araştırma için örnekleme uygulaması zorunludur.

- Evrenin sonlu evren olması.

Daha önce de açıklanmış olduğu gibi tanımlanan evren sonlu evren olduğunda evrenle ilgili bilgiler hem tam sayım uygulayarak hem de örneklemeye başvurularak elde edilebilir. Bu durumda tam sayım mı, örnekleme mi uygulanacağına karar verebilmek için aşağıdaki kriterler değerlendirilir.

KİTAP

Neter J., Wasserman W., Whitmore G. A. (1993). Applied Statistics 4. Edition, Allyn and Bacon

- **Maliyet:** Örnekleme bütçesi, örneklemeyi tam sayıma tercih etmede en önemli belirleyicidir. Örnekleme tam sayıma göre daha az maliyetle bilgi üretme imkanı sağlar. Öte yandan eğer evren hacmi küçükse veya tam sayım yapmak bütçe olanaklarıyla da mümkünse tam sayım tercih edilmelidir. Burada dikkat edilmesi gereken nokta tam sayım yapma maliyetinin, elde edilecek bilginin değerinden küçük olması gerekir. Aksi durumda örneklemeye başvurmak uygun olacaktır.

ÖRNEK 6

Belirli bir bölge için planlanan bir siyasi araştırmada, partilerin bugünkü oy dağılımı hakkında bilgi edinmek amaçlanmış olsun. Tam sayıma mı örneklemeye mi başvurursunuz?

Çözüm:

Bu araştırmadaki evren ilgili bölgedeki seçmen sayısıdır, sonlu evrendir. İstenen bilgilerin üretilmesi amacıyla tam sayım yapmaya karar verilirse üstlenilmesi gereken maliyet bölgedeki genel seçim maliyetine eşdeğer olacaktır. Bölgede yapılmış genel seçim harcamalarına bakıldığında bunun araştırma yapacak kişi ya da kuruluş tarafından karşılanması oldukça zor görünebilir. Kaldı ki genel seçimlerde bütün seçmenler oy kullanmamaktadırlar. Tam sayım yapmanın imkansız ve maliyetli olduğu bu gibi durumlarda en akılcı yol adı geçen araştırma için örneklemeye başvurmak olmalıdır.

- **Zaman:** Bir araştırma sonunda ulaşılacak bilgiye duyulan ihtiyacın zaman sınırları, araştırmanın tam sayımla mı yoksa örneklemeyle mi yapılacağına karar verirken değerlendirilecek diğer önemli bir etkendir. Örnekleme, tam sayıma göre daha kısa zamanda ve yeterli ayrıntıda bilgi elde etme olanağı verir. Örnekleme bilgiye çok hızlı gereksinim olduğu durumlarda bilhassa önemlidir.

Hem tam sayımdan hem de bir örneklemden elde edilecek bilgi için gerekli olan zaman, bir alternatif maliyet üstlenmeyi de gerektirir. Çünkü bilgi elde etme süresine bağlı olarak verilecek kararın erken ya da geç oluşu kazanç kadar kayıplara da neden olabilir.

ÖRNEK 7

Seçmenlerin oy verme tercihleri üzerinde pek çok faktörün etkisi bulunmaktadır. Partileri, bugünkü oy dağılımını belirlemeye yönelik bir araştırma uzun bir zamana yayılırsa, seçmenlerin araştırmaya başlandığı günkü görüşleriyle araştırmanın sonuçlandığı zamandaki görüşleri arasında önemli farklılıklar oluşabileceğinden üretilen bilginin değeri ve geçerliliği giderek azalacaktır. Bu nedenle örneklemeye başvurmak önem arz eder.

- **Doğru veri elde etme:** Her ne kadar tam sayım yapılınca kesin, doğru bilgiye ulaşılır denilse de tam sayımın yapılabilmesi için gerekli olan sayıda veri derleme aracı ve istenen özelliklere sahip, veri derleme hatası yapmayacak gözlemci ya da görüşmeci bulmak ya da yetiştirmek oldukça zor hatta olanaksızdır. Bu nedenlerle örnekleme uygulamaları tam sayıma göre daha doğru veri derleme ve daha doğru bilgi üretme imkânı verir.
- **İncelenecek birimlerin fiziksel zarara uğraması:** Tanımlanan evrende yer alan birimler, veri derlemek ya da ölçüm yapmak amacıyla fiziksel zarara uğratılıyorsa örneklemeye başvurmak zorunludur. Örneğin bir savunma sanayi kuruluşunda belirlenen bir gün içinde üretilen mermilerin içerisindeki hatalı mermi oranının belirlenmesi için yapılacak bir araştırmada tam sayım benimsenirse gerekli verilerin derlenmesi amacıyla üretilen tüm mermilerin patlatma testine tabi tutulması gerekir. Bu durumda üretimin amacı gerçekleşmemiş olur. Bu anlamsız bir testtir. Bu gibi durumlarda örneklemeden yararlanmak kaçınılmaz olur.
- **Evreni oluşturan birimlerin değişkenliği:** Evreni oluşturan birimler araştırmaya konu olan değişkenler bakımından heterojen olduğunda mümkün ise tam sayım yapmak, değil ise büyük hacimli örneklem seçmek gerekir.

- **Evren hacmi küçük, parasal imkânların yeterli olduğu bir araştırma için tam sayım mı yoksa örnekleme mi tercih edersiniz? Açıklayınız.**
- **1,5 milyon öğrencinin olduğu Açıköğretim Sisteminin değerlendirileceği bir araştırma için gerekli olan zaman ve parasal imkânlar yeterli ise tam sayım mı yoksa örnekleme mi uygularsınız? Tartışınız.**

SIRA SİZDE 2

Örneklem İçin Birim Seçme Yöntemleri

Örnekleme girecek birimlerin seçiminde kullanılan yöntemler keyfi seçim yöntemi ve rassal seçim yöntemi şeklinde sınıflandırılmaktadır.

Keyfi Seçim

Örneklem oluşturulurken tanımlanan evreni oluşturan birimler arasında fark gözetilir, yani bütün birimlere bilinen bir olasılıkla seçilme şansı verilmez ise bu türden birim seçimine keyfi seçim adı verilir. Bu seçim yönteminde araştırmacı, hangi birimlerin örnekleme seçileceğini bilerek ve isteyerek belirler. Örneğin yaşadığınız ildeki öğretmenlerin sorunlarını belirlemek amacıyla yapılacak bir araştırma için her bir öğretmen evi lokalinde birlikte oturduğunuz öğretmenler arasından tanıdığınız öğretmenleri seçmek suretiyle bir örneklem oluşturursanız yapmış olduğunuz seçim keyfi seçimdir.

Rassal Seçim

Sonlu Evrenlerde Rassal Örneklem Seçimi

Sonlu evrenlerde rassal birim seçim imkânı veren iki seçim uygulaması bulunmaktadır. Bunlar kura seçimi ve sistematik seçimdir.

Serper Ö. (2004). Uygulamalı İstatistik 2, Genişletilmiş 5. Baskı, Bursa: Ezgi Kitabevi

K İ T A P

Kura Seçimi

Rassal birim seçimi için kura usulü uygulanacak ise aşağıdaki adımlar izlenir:

- Tanımlanan evrenle ilgili oluşturulacak güncel çerçevedeki bütün birimlere birden N e kadar numara verilir. Bu numaralar fişlere yazılır ve bir torbaya veya bir kaba atılır.
- Fişler iyice karıştırıldıktan sonra n tane fişin çekilmesi işlemine başlanır. Çekilen fiş her çekilişten sonra torbaya iade edilir veya edilmez. Çekilen fiş torbaya iade ediliyorsa birim seçimine iadeli seçim, iade edilmiyorsa iadesiz seçim adı verilir.
- Seçilen n sayıdaki birim örneklemi oluşturur.

Bu birim seçim uygulamasıyla evreni oluşturan çerçevede yer alan birimler arasında örneklemde yer almaları bakımından ayrıcalık yapılmamış ve eşit seçilme olasılığı tanınmış olur.

Birim seçimi iadeli yapıldığında aynı birim tekrar tekrar örnekleme seçilmiş olabilir. Bu durumda örnekleme kuramının önemli koşullarından biri olan bağımsızlık koşulu sağlanmış olur ve herhangi bir birimin seçilmesi bir başka birimin seçilmesi olasılığını etkilememiş olur.

Gerçekte uygulanan örnekleme planlarında iadeli seçim genellikle uygulanmaz. Birim seçimi iadesiz yapıldığında, seçilen bir birimin tekrar seçilmesi engellenmiş olur. Ancak evren hacmi N çok büyük, örneklem hacmi n küçük olduğunda her birimin seçiminin diğerinden bağımsız olduğu ve iadeli seçimdeki bağımsızlık koşulunun sağlanabilecek olduğu varsayılır.

Belirlenen sonlu evrende yer alması gereken birim oluşturulacak çerçevede yer almıyorsa veya birden fazla yer alıyorsa kura seçiminin rassallığı etkilenir. N hacimli sonlu bir evrende rassal iadeli seçimle N^n tane farklı örneklem seçmek mümkün iken, aynı evrende aynı hacimli iadesiz seçim uygulandığında $C_N^n = \frac{N!}{(N-n)!.n!}$ tane farklı örneklem seçmek mümkün olur.

Sistematik Seçim

Eğer rassal seçim için sistematik seçim uygun görülürse aşağıdaki aşamalar izlenir:

- Güncel çerçevedeki birimler birden N ye kadar numaralandırılır.
- Örneklem hacmi belirlenir.
- k = N / n oranı hesaplanır. Bu oran “büyütme faktörü” olarak isimlendirilir.
- 1, 2, ….. , k adet sayı arasından rassal olarak bir sayı çekilir. Çekilen sayı a ile gösterilsin. a, örnekleme girecek birinci birimin sıra numarasını gösterir.
- a’ıncı, a + k’ıncı, ….. , a + (n − 1)k’ıncı sıra nolu birimlerin seçilmesiyle n hacimli örneklem oluşturulur.

Hem kura usulü seçimde hem de sistematik seçimde seçilecek bir birimin belirlenen n hacimli örneklemde yer alması olasılıkları aynı (n / N) olmasına rağmen olası örneklemlerden birinin incelenen örneklem olma olasılıkları farklıdır. Bu olasılıklar kura usulü (iadesiz) seçimde 1 / C_N^n olduğu hâlde, sistematik seçimde örneklemi oluşturabilme şansına sahip kombinasyonların her biri için eşit 1 / k , diğerlerininkinde ise 0 (sıfır) dır.

Olasılıklı örneklemenin üç önemli üstünlüğü vardır:

- Örneklemden elde edilen verilerden hesaplanan istatistikler evren parametreleri hakkında genelleme yapmak üzere kullanılabilir.
- Örneklem hatasının büyüklüğü hakkında bilgi elde edilebilir.
- Keyfi seçimde söz konusu olabilecek yanlılık (sistematik hata) giderilmiş olur.

Olasılıklı örneklem oluşturma prensibi esas olmak üzere, uygulamada ya birim seçim işlemini kolaylaştırmak ya da evreni temsil edecek daha iyi bir örneklemin oluşturulmasını sağlamak üzere çeşitli rassal örnekleme yöntemleri geliştirilmiştir. Bu yöntemler izleyen başlıklar altında açıklanmıştır.

Bu örnekleme yöntemlerinden herhangi birini bir örnekleme uygulaması için seçerken yöntemlerin etkinlik ve doğruluk kriterlerine göre değerlendirilmesi gerekir. Örnekleme yöntemlerinin etkinlikleri farklılık gösterir. Etkinlik, örnekleme maliyeti ve doğruluğu arasındaki dengeyi yansıtan bir kavramdır. Doğruluk ise ölçülecek özelliğin belirsizliği ile ilgili düzeyi gösterir. Doğruluk ile örnekleme hataları arasında ters ilişki varken maliyetle aynı yönde ilişki vardır. Yani daha çok maliyet daha doğru bilgi, daha doğru bilgi daha az hatalı karar ve tahmin demektir.

ÖRNEKLEME SÜRECİNİN AŞAMALARI

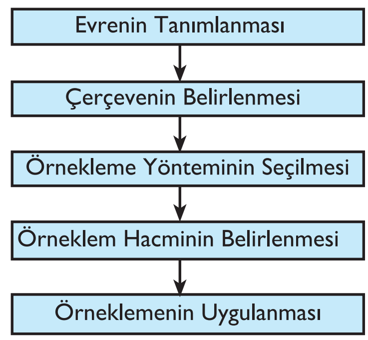
Genel olarak örnekleme süreci 5 aşamadan oluşmaktadır. Birbirleriyle ve bir araştırma sürecinin diğer aşamalarıyla sıkı sıkıya ilişki içinde olan bu aşamalar Şekil 6.2 de gösterilmiştir.

Evrenin Tanımlanması

Örnekleme süreci evrenin tanımlanmasıyla başlar ve bir araştırma sürecinde araştırmacının ilk yapacağı işlerden biridir.

Evren, araştırmacı tarafından belirlenen bir tanıma uyan ve hakkında bilgilerin üretileceği, çıkarımların yapılacağı birimlerden oluşan topluluktur. Evrenin ayrıntılı bir biçimde tanımlanmasıyla, hangi birimlerin araştırma kapsamına alınacağı, hangilerinin alınmayacağı belirlenmiş olur..

Şekil 6.1



Örnekleme sürecinin aşamaları.

***Kaynak:** N. K. Malhotra, Marketing Research An Applied Orientation, Prentice Hall International Inc., 1996*

Evrenin tanımlanması genel olarak örnekleme birimi, gözlem birimi, yer ve zaman kavramlarıyla yapılmaktadır. Araştırmanın konusunu tanımlayan değişkenlerle ilgili verilerin derlendiği birimlere gözlem birimi adı verilir. Örneklem birimi ise örnekleme seçilecek birimlerdir.

ÖRNEK 8

Araştırmanın konusu: Eskişehir Merkez İlçesinde ikamet eden 250000 ailede hangi marka bulaşık deterjanı kullanıldığını araştırmak.
Evren: Eskişehir Merkez İlçesinde araştırma yapıldığı tarihte ikamet etmekte olan ailelerin oluşturduğu topluluktur. Sonlu bir evrendir. Evren hacmi 250.000 ailedir.
Örnekleme Birimi: Her bir aile.
Gözlem Birimi: Ailede çamaşır yıkama görevini üstlenen veya deterjan satın alma tercihinde bulunan kişidir.
Yer: Eskişehir Merkez İlçesi
Zaman: Araştırmanın yapıldığı tarih.

Örnekten anlaşılabileceği gibi önce örnekleme aile seçilecek sonra ailedeki çamaşır yıkayan, deterjan tercihini yapan kişiden veri derlenecektir.

Bu örnekte olduğu gibi her araştırmada gözlem birimi ve örnekleme birimi ayrımı olmayabilir. Örnekleme seçilen ve veri derlenen birim aynı olabilir. Bu durumda sadece birim kavramı kullanılır.

ÖRNEK 9

Araştırmanın Konusu: 2010-2011 öğretim yılında Anadolu Üniversitesi İktisat Fakültesi Kamu Yönetimi Bölümünde öğrenim gören öğrencilerin kitap okuma alışkanlığını araştırmak.
Evren: 2010-2011 öğretim yılında Kamu Yönetimi bölümünde kayıtlı olan öğrenciler topluluğudur.
Birim: Kamu Yönetimi bölümünde kayıtlı olan her bir öğrenci hem gözlem birimi hem de örnekleme birimidir. Çünkü örnekleme seçilecek birim de, veri derlenecek birim de öğrencilerdir.

DİKKAT

Örnekleme ve gözlem birimi aynı olduğu zaman araştırmalarda sadece birim kavramı kullanılmaktadır.

Gözlem birimi ve örnekleme birimi ayrımına gidilmesinin nedeni gözlem birimleri ile ilgili bir çerçevenin temin edilmesinin veya hazırlanmasının zor, maliyetli ve çok zaman alacak olmasıdır. Örnekleme birimi birden çok gözlem birimini kapsayacak şekilde de tanımlanabilir. Örnek 3 üzerinden açıklamak gerekirse, beslenme programıyla ilgili öğrencinin hem annesinin hem de babasının görüşlerine başvurulabilir. Bu durumda gözlem birimi öğrencinin hem annesi hem de babası olur.

Evrenin tanımlanması yukarıda yapılan açıklamalarda olduğu gibi her zaman kolay olmayabilir. Örnek 8 üzerinden açıklama yapılacak olursa evren tanımı yapılırken örneğin öğrenci evleri, yabancı uyruklu olup Eskişehir Merkez İlçesinde ikamet edenler, tek kişilik yaşamın yaşandığı konutlar, araştırmada ifade edilen aile tanımı içine alınacak mı yoksa alınmayacak mı, karar verilmelidir. Gözlem birimi olarak ailedeki anne mi, kız mı, yoksa annenin yardımcısı olarak çalışan kişi mi seçilecektir?

Bir araştırmanın evrenini tanımlarken açıklık, kesinlik, amaca uygunluk ve örnekleme uygulaması için güçlük yaratmaması gibi ilkelerin de göz önünde bulundurulması gerekir.

Çerçevenin Belirlenmesi

Çerçeve sonlu bir evrenin bütün birimlerinin kayıtlı olduğu bir listedir, tablodur veya cetveldir. Nüfus kayıtları, seçmen kütükleri, tapu ve sicil kayıtları, ticaret ve sanayi odaları üye listeleri, ekonomik büyüklüklerine göre sanayi kuruluşlarının listesi, telefon rehberi, öğrenci kayıt listeleri, su, elektrik abonelik listeleri vb. çerçeve olarak kullanılabilecek araçlardır.

DİKKAT

Sonsuz evrenler için yapılaak örnekleme uygulamalarında çerçeve söz konusu olmaz.

Örneklemeye başlamadan önce amaca uygun bir çerçevenin var olup olmadığı, yoksa sağlanıp sağlanamayacağı öncelikle araştırılmalıdır. Araştırmaya uygun bir çerçevenin var olması durumunda bu çerçevenin güncel olup olmadığının araştırılması da önemli bir konudur. Dikkat edilmelidir ki çerçeve olmadan ne tam sayım ne de örnekleme yapılabilir.

Bir çerçeve yoksa yeni bir çerçevenin hazırlanması problemiyle karşılaşılır. Yeni bir çerçevenin hazırlanmasında çerçeve maliyeti ve kapsam hatası özellikle göz önünde tutulmalıdır. Bazen tanımlanan evrenin bazı birimleri çerçevede yer almadığı gibi tanımlanan evrenin dışında kalması gereken birimler de çerçevede yer alabilir ya da bazı birimler tekrar tekrar çerçevede yer alabilir. Bu özellikteki çerçevelerde kapsam hatası işlenmiş olur. Güncel çerçeve bulmak zordur. Kapsam hatası işlenen mevcut çerçevelerin de güncelleştirilmesi uzun zaman alır ve maliyetli olur. Bu nedenlerle uygulamada güncel olmayan bir çerçevenin kabul edilebilirliği onun güncelleştirilmesinin maliyeti ve sağladığı zaman tasarrufu ile ilişkilendirilerek belirlenir. Çerçeve, kabul edilebilir bir çerçeve hatası düzeyinde evren birimlerinin çok büyük kısmını kapsamalıdır. Şüphesiz amaç evren tanımında yer alan bütün birimleri kapsayan bir çerçeve elde etmek veya oluşturmaktır.

Örnekleme Yönteminin Seçimi

Örneklemeye girecek birimlerin belirlenmesine imkan veren yöntemlere örnekleme yöntemleri denir. Bu yöntemler örneklem için birim seçiminde uygulanan usulün keyfi ya da rassal oluşuna göre iki sınıfa ayrılır. Birinci durumda olasılıklı olmayan örnekleme, ikinci durumdaysa olasılıklı örnekleme söz konusu olur. Örnekleme yönteminin seçimiyle ilgili en önemli karar bir örnekleme planında ne tür bir örnekleme yöntemi uygulanacağıdır. Bu konu örnekleme yöntemleri başlığı altında ayrıntılı bir biçimde ele alınacaktır.

Örneklem Hacminin Belirlenmesi

Örneklem hacmi, örnekleme girecek birimlerin sayısını gösterir ve "n" simgesiyle gösterilir. Bu sayının ne olacağına ilişkin kesin yanıt vermek mümkün değildir. Ancak, bu sorunun yanıtlanabilmesi için aşağıda açıklanan faktörlere ilişkin yapılacak nitel değerlendirmelere ve örnekleme dağılımı başlığı altında açıklanacak olan nicel yöntemlere başvurulur.

Malhotra N. K. (1996). Marketing Research An Applied Orientation, Prentice Hall International Inc.

KİTAP

Nitel Değerlendirmede Esas Olan Faktörler

- **Evrenin homojenliği:** Ele alınan evrenin ilgilenilen değişken bakımından homojen ya da heterojen olması örneklem hacminin belirlenmesine etki eder. Eğer evrenin bütün birimleri ilgilenilen değişken itibariyle aynı değere sahipse bir birimin incelenmesi amaca ulaşmak için yeterlidir. Ancak birimlerin özellikler bakımından farklılığı artıkça evreni temsil edebilecek bir örneklem oluşturabilmek için örneklem hacminin de giderek büyümesi gerekir.
- **Araştırmada verilecek kararın önemi:** Önemli kararlar için olabildiğince çok veriye ve ayrıntılı bilgiye gereksinim vardır. Bu gibi durumlar büyük hacimli bir örneklem üzerinde araştırma yapmayı gerekli kılar. Ancak örneklem hacmi arttıkça maliyet ve gereksinim duyulan zaman ve nitelikli personel sayısı da artar. Burada dikkat edilmesi gereken husus, bir yandan küçük hacimli örneklem oluşturmak suretiyle bu örneklemin evreni temsil etmesi bakımından yetersiz kalmasını engellemek, diğer taraftan da gereksiz yere çok büyük hacimli örneklem seçerek zaman ve maliyet yönünden kayba uğramamak için uygun büyüklükte bir örneklem hacmini belirlemektir. Örneklem hacmi arttıkça örnekleme seçilecek her yeni birimin alınacak kararın, yapılacak tahminin doğruluğuna katkısının azalabileceğini dikkate almak gerekir.
- **Araştırmanın yapısı:** Araştırmanın doğası da örneklem hacmi üzerinde etkilidir. Uygulamada genellikle nitel araştırmalarda küçük hacimli örneklemlerde, nicel araştırmalarda ise örneğin betimsel araştırmalarda daha büyük hacimli örneklemlerle çalışılır. Ayrıca araştırmalarda değişken sayısı arttıkça örneklem hacminin artırılması bilginin niteliği açısından ihtiyaç olur. Örneğin çok değişkenli analiz teknikleri ve yöntemlerinin kullanıldığı araştırmalarda örneklem hacmi büyük olmalıdır.

Örneklemin Seçimi

Örnekleme sürecinin bu son aşamasında örnekleme girecek birimler keyfi veya rassal seçim uygulamalarıyla seçilirler. Seçilen birimlerden gerekli veriler derlenir. Bu aşamada yapılacak işlemleri, uygun özellikte büro ve çalışma ortamı ile nitelikli işgörenlerin temini gibi sayabiliriz. Önceki aşamalarda yapılan yanlış uygulamalar ve dikkatsizlikler bu aşamada büyük sorunların yaşanmasına neden olur.

ÖRNEKLEME YÖNTEMLERİ

Örnekleme yöntemleri evrenden örnekleme birim seçiminde uygulanan usule göre Şekil 6.2'de gösterildiği gibi olasılıklı olmayan örnekleme yöntemleri ve olasılıklı örnekleme yöntemleri gibi iki sınıfa ayrılır.

KİTAP **N. K. Malhotra, Marketing Research An Applied Orientation, Prentice Hall International Inc., 1996**

Şekil 6.2

Örnekleme Yöntemleri.

Örnekleme Yöntemleri

Olasılık Olmayan Örnekleme Yöntemleri

- -Kolayda Örnekleme
- -Yargısal Örnekleme
- -Kota Örneklemesi
- -Kartopu Örneklemesi

Olasılıklı Örnekleme Yöntemleri

- -Basit Rassal Örnekleme
- -Tabakalı Örnekleme
- -Sistematik Örnekleme
- -Tek ve Çok Anlamlı Örnekleme

Olasılıklı Olmayan Örnekleme Yöntemleri

Olasılıklı olmayan örnekleme, birim seçiminin keyfi olarak yapıldığı örneklemedir.

Araştırmayı planlayan ya da örnekleme uygulamasını yapan kişi ya da grubun istekleri ve değer yargıları örnekleme seçilecek birimlerin ve örneklem hacminin belirlenmesinde etkili oluyorsa yapılan örnekleme **olasılıklı olmayan örnekleme**dir.

Bu örnekleme yöntemleri, örneklem için birim seçiminde keyfi seçim usulünün uygulandığı örnekleme yöntemleridir. Örneklem oluşturulurken, tanımlanan evreni oluşturan birimler arasında fark gözetilir ve bütün birimlere, bilinen bir olasılıkla seçilme şansı verilmezse yapılan seçim keyfi seçimdir. Keyfi seçimle oluşturulan olasılıklı olmayan örneklemin evreni temsil etmeyeceği anlamına gelmez. Ancak olasılık kuramının uygulanamayacağı anlamına gelir. Olasılıklı olmayan örnekleme uygulandığında örneklemin evreni temsil etme olasılığı bilinemez. Oysa bu bilgi araştırmacılar için çok önemlidir.

Temsili örneklem oluşturma bakımından olasılıklı olmayan örnekleme yönteminin başarısı, örnekleme uygulamasını yürüten kişi ya da grubun araştırma konusuyla ilgili deneyimine, tanımlanan evrenin özellikleri hakkındaki öncül bilgilerine ve bu evrenin ilgilenilen özelliklerinin homojenliğine bağlıdır.

Temsili örneklem oluşturma ve uygulama kolaylığı sağlaması amacıyla çeşitli olasılıklı olmayan örnekleme yöntemleri geliştirilmiştir. Uygulamada sıkça kullanılan ve aşağıda incelenen bu yöntemlerin ortak özellikleri:

- Örneklem için birim seçimi keyfidir.
- Örneklem hacmi keyfi olarak belirlenir.
- Örneklemden hesaplanan istatistikler evren parametreleri hakkında genelleme amacıyla kullanılamaz.

Kolayda Örnekleme

Burada amaç, araştırma konusu ile ilgili ve kolayca ulaşılabilir olan birimlerden bir örneklemin oluşturulmasıdır. Araştırma konusu ile ilgili olan ve doğru yerde, doğru zamanda bulunan birimler arasından keyfi olarak birimler seçiliyorsa yapılan örneklemeye kolayda örnekleme denir. Kolayda örnekleme gönüllülük esasına göre katılan birimlerden oluşur.

ÖRNEK 10

Eskişehir'de yaşayan insanların hayat pahalılığı ile ilgili görüşlerini öğrenmek amacıyla bir araştırma planlanıyor. Hayat pahalılığının yaşandığı yer olarak haftanın herhangi bir günü Eskişehir'de kurulan semt pazarlarına alışverişe gelenler arasından keyfi olarak belirlenen ve araştırmanın amacıyla ilgili mülakata katılmayı kabul eden kişilere görüşleri soruluyor. Mülakata katılanların oluşturduğu topluluk kolayda örneklemedir. Bu örnekte doğru yer Eskişehir'de kurulan semt pazarlarıdır. Doğru zaman ise semt pazarının kurulduğu gündür. Bu araştırmada bir gün içinde veriler derlenebilir ve bu veriler çözümlenerek istenen bilgi üretilebilir.

Uygun görülen sokaktan, uygun görülen zamanda gelip geçen bireylerle görüşme yapılması ya da bir konferansa katılan belirli sayıdaki katılımcıdan araştırma konusuyla ilgili görüşlerinin alınması, birer kolayda örnekleme uygulamasıdır. Bu örnekleme uygulamasında örnekleme birimlerine kolayca ulaşılabilir, ilgilenilen değişkenlerle ilgili veriler kolayca derlenebilir ve birimlerle iş birliği sağlanabilir.

En kısa zamanda ve en az maliyetle bilgi üretilmesine ihtiyaç duyulduğu durumlarda kolayda örnekleme yöntemi bir seçenektir.

Bu örnekleme yönteminde en önemli sorun, seçilen örneklemin seçildiği evreni ne kadar temsil edebildiğidir. Kolayda örnekleme uygulaması ile oluşturulan örneklem, birim seçimindeki yanlılık nedeniyle tanımlanan bir evreni temsil etmeyebilir. Betimleyici ve ilişki araştırıcı araştırmalarda kolayda örnekleme uygun bir yöntem değildir. Kolayda örnekleme fokus gruplar, soru kağıtlarının (anket formlarının) ön testi veya pilot çalışmalar için kullanılabilir.

Yargısal Örnekleme

Bu örnekleme de bir tür kolayda örneklemedir. **Yargısal örnekleme,** örneklemin araştırmacının ya da örneklemeyi yapan kişinin kişisel arzu, düşünce ve deneyimlerine göre seçilmiş olduğu örneklemedir.

Evrenden temsili örneklem oluşturacağına inanılan kriterlere dayalı olarak birimlerin seçilmesi işlemine **yargısal örnekleme** denir.

Bu yöntemin kolayda örneklemeden farkı örnekleme birim seçimi için araştırmacının uzman fikirleriyle belirlediği ölçütler kullanması ve bu ölçütlerin temsili bir örneklem oluşturacak ölçütler olduğuna inanıyor olmasıdır.

ÖRNEK 11

A Üniversitesinin sorunlarını araştırmak amacıyla bu üniversitenin üst düzey yöneticilerinden seçim yapılması yargısal örnekleme için bir örnektir. Çünkü üniversitenin üst düzey yöneticileri üniversite sorunlarını en iyi bilen kişilerdir. Bu düşünceyle seçimin bu kişiler arasından yapılması temsili bir örneklem oluşturabilir.

Kolayda örneklemede olduğu gibi yargısal örneklemede de örnekleme birimlerine kolayca ulaşılabilir ve verilerin çok hızlı biçimde derlenmesi mümkün olur. Yargısal örnekleme pazarlama araştırmasında, kamuoyu araştırmalarında ve biyolojik araştırmalarda yaygın olarak kullanılmaktadır.

Eğer evreni oluşturan birimler araştırmaya konu olan değişkenler bakımından homojen ise kolayda ve yargısal örnekleme uygulamaları temsili örneklem oluşturma imkânı verir. Bu örnekleme uygulamalarının maliyeti, uzman çalıştırılacağı için kolayda örneklemeye göre daha yüksektir.

Kota Örneklemesi

Örneklem için birim seçiminin keyfi olarak yapıldığı yöntemlerden biri de kota örneklemesidir. Tanımlanan sonlu evren heterojen özelliklere sahip birimlerden oluşuyorsa olasılıklı olmayan örnekleme yöntemleri grubundan kota örneklemesi temsili örneklem oluşturma amacıyla tercih edilmelidir. Bu yöntemin başarıyla uygulanabilmesi için tanımlanan sonlu evrenle ilgili bir çerçevenin var olması, ilgili evrenin homojen veya heterojen özelliğe sahip olup olmadığının sorgulanabilmesi için evren hakkında öncül bilgilere sahip olunması, evrenin heterojen olduğuna karar verilmiş ise hangi kritere göre heterojen birimlerden oluşan bu evrenin homojen birimlerden oluşacak tabakalara ayırmada kullanılacak kriterin belirlenmesi ve tabaka hacimlerinin bilinmesi gerekir.

Kota örneklemesi sürecindeki adımlar aşağıdaki gibidir:

- Evren hacmi N ve tabaka hacimleri N_h , (Tabaka sayısı h=1, 2, ...) belirlenir.
- Örneklem hacmi n keyfi olarak belirlenir.
- Her tabakanın, evren hacmi içindeki oranı N_h / N belirlenir.
- Her tabakada keyfi seçimle $n_h = (N_h / N) . n$ sayıda birim seçilir ve bu seçilen birimler örneklemi oluşturur.

ÖRNEK 12

Anadolu Üniversitesi yönetimi verdiği açıköğretim hizmetlerinden memnun olan öğrencilerin oranını belirlemek amacıyla bir araştırma planlamıştır. Araştırmayı gerçekleştirecek grup kota örneklemesi uygulamayı düşünmektedir.

Çözüm:

Cinsiyet tabakalama kriterine göre öğrencilere ilişkin bilgiler;
$N = 600.000$ Evren hacmi.
$N_E = 400.000$ Erkek öğrenci hacmi.
$N_K = 200.000$ Kız öğrenci hacmi.
$n = 3.000$ Örneklem hacmi.
Erkek öğrenci tabakasından (N_E) seçilecek öğrenci hacmi:
$n_E = (N_E / N) . n = (400.000/600.000) . 3000 = 2000$ olarak bulunur.
Benzer hesaplama kız öğrenciler tabakası için de yapılırsa:
$n_K = 1000$ öğrenci bulunur.

Erkek ve kız öğrenci tabakalarından sırasıyla 2000 ve 1000 öğrenci keyfi seçimle seçilmek suretiyle n = 2000 + 1000 = 3000 hacimli örneklem seçilmiş olur. Bu örneklemdeki birimler üzerinden gerekli veriler derlenir ve istenilen bilgi üretilir.

Kota örneklemesi kolayda ve yargısal örneklemeye göre daha temsili örneklem oluşturma çalışmasıdır. Ancak bu örnekleme yönteminin uygulanması sonucu oluşturulan örneklemin evreni temsil etmesinin garantisi yoktur. Çünkü kota örneklemesi uygulamasında belirlenen tabakalardaki birimlerin homojen özellikli birimlerden oluştuğunun, tabaka oranlarının doğruluğunun garantisi yoktur. Ayrıca tabakalardan birimler keyfi olarak seçildiği için yanlılık söz konusu olabilir. Bu örnekleme uygulaması sonucu oluşturulan örneklemden elde edilen bilgiler evren bilgisi için genelleme amacıyla kullanılamaz.

Kartopu Örneklemesi

Kartopu örneklemesi, özellikle bir çerçevenin mevcut olmaması ya da oluşturulmasının imkânsız olduğu durumlarda faydalı bir örneklemedir. Bu yöntemde örnekleme süreci tanımlanan evrende yer alan bir bireyin genellikle rassal olarak seçilmesiyle başlar. Belirlenen bu birey örnekleme giren birinci birimdir. Bu bireyden aynı evren tanımında yer alan tanıdığı bir bireyin olup olmadığı öğrenilir. Varsa bu bireye ulaşılır. Böylece örneklemde yer alacak ikinci birime ulaşılmış olur. Benzer şekilde bu süreç, referanslarla keyfi olarak belirlenen hacimde örnekleme ulaşılıncaya kadar sürdürülür.

ÖRNEK 13

Bir bölgedeki uyuşturucu madde kullananlar üzerinde bir araştırma yapılacak olsun. Bu bölgede uyuşturucu kullananlarla ilgili bir liste bulmak mümkün değildir. Bölgede bir ya da iki uyuşturucu kullanan tanımlanabilirse kartopu örnekleme süreci başlar. Örnekleme seçilmiş olan bu kişi ya da kişilere uyuşturucu kullanan arkadaşları ya da tanıdıklarının olup olmadığı sorulur. Varsa adresleri öğrenilir, bu kişilere ulaşılır ve bunlar da bu örnekleme seçilirler. Bu süreç keyfi olarak belirlenen n hacimli örneklem oluşturuluncaya kadar sürdürülür.

Çete üyeleri ve bir ülkeye yasal olmayan yollarla girmiş kişilerle ilgili araştırmalarda, bir kentte İnternet üzerinden alışveriş yapanlarla ilgili araştırmalarda kartopu örneklemesi uygulanır. Bu örnekleme, endüstriyel ürün alan ve satanlar hakkında yapılacak araştırmalarda da kullanılabilir. Bu yöntem uygulandığında temsili örneklem oluşturmak olanaklıdır. Kartopu örneklemesinin maliyeti ve örneklem değişkenliği düşüktür.

Tüm olasılıklı olmayan örnekleme yöntemlerinde örnekleme girecek birimlerin seçiminin keyfi olması tek yönlü hatalara neden olur. Bu tür hatalardan kaçınmak için izleyen kısımlarda ele alınacak olan olasılıklı örnekleme yöntemleri tercih edilmelidir.

SIRA SİZDE 3

- **Kolayda örnekleme mi yargısal örnekleme mi daha temsili örneklem oluşturur?**
- **Evrenin birimleri ilgilenilen özellik bakımından heterojen ise hangi olasılıklı olmayan örnekleme yöntemi kullanılır? Açıklayınız.**

Olasılıklı Örnekleme Yöntemleri

Olasılıklı örnekleme, ilgilenilen evrendeki her örnekleme birimine hesaplanabilir ve sıfırdan farklı bir olasılıkla seçilme imkânı veren örneklemedir.

Örneklem için birim seçiminde rassal seçimin uygulandığı yöntemlere **olasılıklı örnekleme** yöntemleri denir.

Rassal örnekleme yöntemleri olarak da alınan bu örnekleme yöntemleri örnekleme planlarında yaygın olarak uygulanır. Bu tür örneklemede örnekleme girecek birimlerin seçimi rassal olarak yapılır. Rassal seçim evrenden örnekleme girecek birimleri seçerken herhangi bir ayrıcalığın uygulanmadığı seçimdir.

Basit Rassal Örnekleme

Sonlu Evrenlerde Basit Rassal Örnekleme

KİTAP

Serper Ö. (2004). Uygulamalı İstatistik 2, 5. Baskı, Bursa.
Trochim W. M. (2001). Research Methods Knowledge Base, Cornell University.

Örnekleme planlarında uygulanan en temel olasılıklı örnekleme basit rassal örneklemedir. Basit rassal örnekleme hacmi N olan sonlu bir evrenden birbirinden farklı ve n hacimli oluşturulabilecek C_N^n sayıdaki olası örneklemlerin her birine incelenecek örneklem olması bakımından eşit şans tanıyan örnekleme yöntemidir. Bu tanımda belirtilen özellikleri taşıyan C_N^n sayıdaki mümkün örneklemlerin her birine basit rassal örneklem denir. Bu örnekleme yöntemi evrendeki bütün birimlere hacmi n olarak belirlenen örnekleme girmeleri bakımından bilinen ve birbirine eşit seçilme olasılığı sağlar.

ÖRNEK 14

Hizmet içi eğitim programına katılan bir firmanın 4 personelinin yapılan sınav sonuçları incelenecektir. Sınav sonuçları 60, 40, 50 ve 70 puandır. Sonlu evreni oluşturan bu 4 personel, A, B, C ve D olarak simgelendirelim ve 2 personellik

$$C_N^n = \frac{N!}{(N-n)!.n!} = C_4^2 = \frac{4!}{(4-2)!\,.\,2!} = 6$$

tane mümkün farklı örneklem aşağıdaki gibi oluşturulabilir.

Tablo 6.1 *4 birimlik evrenden iadesiz seçimle oluşturulabilecek 2 hacimli mümkün örneklemler.*

BİRİMLER	MÜMKÜN ÖRNEKLEMLER
A	A, B
B	A, C
C	A, D
D	B, C
	B, D
	C, D

Bu mümkün farklı 6 rassal örneklemden birinin incelenen örneklem olması olasılığı $\frac{1}{C_N^n} = \frac{1}{C_4^2} = \frac{1}{6}$ olur. Burada herhangi bir birimin yapılacak bir rassal seçimde seçilmesi olasılığı $\frac{1}{N} = \frac{1}{4} = 0,25$; herhangi bir birimin n = 2 hacimlik örneklemde yer alması olasılığı da $\frac{n}{N} = \frac{2}{4} = 0,5$ olacaktır.

DİKKAT

Örnekleme uygulamalarında mümkün örneklemler oluşturulmaz, oluşturulabileceği varsayılır. Sadece bu mümkün örneklemlerin biri oluşturulur ve araştırma bu örneklem üzerinden yapılır.

Sonlu bir evrenden iadesiz seçimle n hacimli bir rassal örneklem oluşturmak için aşağıdaki adımlar izlenir:

- Güncel çerçeve temin edilir ya da hazırlanır.
- Örneklem hacmi belirlenir.
- Çerçevede yer alan n sayıdaki birime tanımlayıcı numara ya da işaret verilir.
- Evrendeki her birime eşit seçilme şansı vermek suretiyle örnekleme girecek birinci birim rassal seçim araçları kullanılarak belirlenir.
- Geriye kalan (N - 1) birimin her birine yine eşit şans vermek suretiyle ikinci birim seçilir.
- Bu birim seçim süreci n hacimli örneklem seçilinceye kadar tekrarlanır. Bu birimin seçilmesi olasılığı 1 / (N-1) olur.

Açıklanan seçim sürecinde her çekilişte seçilen birim incelendikten sonra evrene iade edilmediği için bu seçim sürecine iadesiz rassal seçim süreci adı verilir. Eğer basit rassal örnekleme planlarında önceki çekilişte seçilen birim incelendikten sonra evrene iade ediliyorsa başka bir ifadeyle birimler tekrar tekrar seçilme şansına sahipse bu seçim sürecine iadeli rassal seçim süreci adı verilir. Bu seçim sürecinde evren hacmi çekilişten çekilişe değişmez. Sonlu bir evrenden iadeli seçimle bir rassal örneklem seçilirse sonsuz evrenden basit rassal örnekleme yapılıyormuş gibi bir anlam ifade eder. Bu çekiliş sürecinde evrendeki her birimin yapılacak çekilişlerin her birinde seçilme şansı birbirine eşittir, $\frac{1}{N}$ olan seçilme olasılığına sahiptir ve birbirini izleyen çekilişler bağımsızdır. Sonlu evrenlerde evren hacminin büyük ya da küçük oluşu iadeli ya da iadesiz seçimler için önemli farklılıklar gösterir. Evren hacmi büyük, örnekleme oranı $\frac{n}{N}$ küçük olduğu zaman iadeli ve iadesiz örneklemler benzer özellikler gösterirler. Çünkü iadeli çekiliş uygulandığında önceki çekilişlerde seçilmiş olan bir birimin yeniden örnekleme seçilmiş olma olasılığı çok küçüktür. Ancak evren hacmi küçükse iadeli ve iadesiz rassal seçimlerle oluşturulan aynı hacimli örneklemler için hesaplanan örneklem istatistikleri ile evren parametreleri karşılaştırılırsa iadesiz basit rassal seçimle oluşturulan örneklem iadeli olana göre daha az hatayla tahminleme imkânı sağlar. Bu özellik nedeniyle de iadesiz rassal seçim uygulamada genellikle başvurulan yöntem olmaktadır.

İlgilenilen özellik bakımından evrenin homojen olması durumunda basit rassal örnekleme tercih edilmesi gereken bir yöntemdir. Örnekleme planlarında basit rassal örnekleme yöntemlerinin tercihlerini etkileyen önemli sınırlayıcılar vardır. Bunlardan birincisi güncel bir çerçeve oluşturma ya da hazırlama güçlüğüdür. İkincisi evrenin birimleri geniş bir coğrafi alana yayılmışsa basit rassal örnekleme uygulaması çok zaman alır ve veri derleme maliyeti giderek artar. Üçüncüsü evren homojen değilse basit rassal örneklem sonuçlarının başarısı diğer olasılıklı örnekleme yöntemleri sonuçlarının başarısından düşüktür.

Sonsuz Evrenlerde Basit Rassal Örnekleme

Sonsuz evrenlerde basit rassal örnekleme ile ilgili bilgiler için Neter J., Wasserman W., Whitmore G. A. (1993). Applied Statistics, (Boston: Fourth Edition, Allyn and Bacon) adlı kitaptan yararlanmıştır.

KİTAP

Birinci ünitede sonsuz evren, “Aynı koşullar altında işleyen bir sürecin sonuçlarının oluşturduğu topluluktur.” şeklinde tanımlanmıştı ve bir bisküvi fabrikasındaki bisküvi paketleme süreci örnek verilmişti. Bisküvi paketleme süreci sonucu olan her bir paket bir birim, paketleme süreci devam ettikçe yeni bisküvi paketleri evrene dahil olduğu için “Bu sürecin sonuçları olan paketler sonsuz evreni oluşturur.” denmişti. Örnekten de anlaşılabileceği gibi sonsuz evrendeki bir başka ifadeyle aynı koşullar altında işleyen bir sürecin bütün sonuçları listelenemez. Bütün birimlerle ilgili bir çerçeve hazırlanamaz, bunun yerine bu birimlerin ilgilenilen bir değişken X için bir teorik olasılık dağılımı tarif edilebilir. Tarif edilen bu teorik olasılık dağılımına sonsuz evren adı verilir.

Sonsuz evrenin birimlerinin X değişkeni için ölçümlenen değerleri (gözlem değerleri) bu birimler için X değişkeninnin gerçekleşen değerleridir. Eğer sonsuz evrenin birimleri kararlı (aynı koşullar altında meydana gelen) bir sürecin sonuçları, birimleri ise n sayıdaki birimin ilgilenilen X değişkeni bakımından aldığı x_1 x_2, ..., x_n değerleri (gözlem değerleri) X_1, X_2, ... X_n rassal değişkenlerinin birer gerçekleşmesi olduğu düşünülür. Buna göre, bir süreç tarafından türetilen birbirinden bağımsız ve benzer olasılık dağılımına sahip n sayıdaki X_1, X_2, ... X_n rassal değişkenlerinin oluşturduğu topluluğa sonsuz evrenden seçilmiş basit rassal örneklem denir. Bu rassal değişkenlerin teorik olasılık dağılımına sonsuz evren adı verilir.

ÖRNEK 15

Bir bisküvi fabrikasının paketleme sürecinin planlanan ağırlıkta üretimi gerçekleştirip gerçekleştirmediğini incelemek olsun. Bisküvi paketleme süreci aynı koşullar altında işleyen kararlı bir süreçtir. Sürecin sonuçları olan paketler sonsuz evrenin birimleridir. İncelenen değişken üretilen ve üretilecek olan paketlerin ağırlığıdır. Bu üretim sürecinin sonucu olan paketlerin seçilen n tanesinin ölçüm ağırlıkları olan x_1 x_2, ... , x_n sırasıyla X_1, X_2, ... X_n rassal değişkenlerinin gerçekleşmeleridir. Buna göre paketleme sürecinden rassal olarak seçilen n tane paketin ölçülen ağırlıkları kullanılarak hesaplanan olasılık dağılımı X_1, X_2, ... X_n teorik olasılık dağılımını ifade eden sonsuz evrenden seçilmiş basit rassal örneklemi oluşturduğu söylenebilir.

Tabakalı Örnekleme

Tabakalı örnekleme evren birimlerinin tabakalara ayrıldığı ve her tabakadan rassal seçimle örneklemin oluşturulduğu örneklemedir.

Tanımlanan evrenin birimleri araştırmaya konu olan değişkenler bakımından heterojen ise, önemli farklılıklar gösteriyorsa temsili örneklem oluşturabilmek için **tabakalı örnekleme** tercih edilmelidir. Tabakalı örnekleme evren birimlerinin tabakalara ayrıldığı ve her tabakadan rassal seçimle örneklemin oluşturulduğu örneklemedir.

Tabakalı örnekleme üzerinde araştırma yapılacak evren ilgilenilen değişkenler yönünden heterojen olduğunda evren parametre tahminine ilişkin varyansın olabildiğince küçük olmasını sağlayan örneklemedir.

Tabakalı örnekleme 4 aşamalı bir süreçtir:

- Tabakalama kriterinin belirlenmesi.
  Tabakalı örnekleme uygulaması yapacak araştırmacı önce incelenecek değişkenler açısından önemli farklılıklar gösteren N hacimli evrenin birimlerini homojen birimlerden oluşacak tabakalara ayırmada kullanılacak kriter belirler. Burada önemli olan belirlenecek kriterin tabakalar içindeki birimleri olabildiğince homojen, tabakalar arasında ise birimlerin heterojen olmasını sağlayacak kriter olmasıdır. Aynı zamanda bu kriterin uygulama ve ölçme kolaylığı da sağlamak suretiyle maliyet artırmadan tahminleme hatasını azaltması gerekir. Tabakalama kriteri ilgilenilen parametre ile sıkı sıkıya ilişki içindedir. Belirlenen tabakalama kriterinin uygunluğu örneklem değişkenliğinin etkinliği üzerinde olumlu yönde etkilidir. Bu nedenle asimetrik bölünmeye sahip evrenlerde tabakalı örnekleme uygulamasını tercih etmek bir zorunluluktur. Tabakalama amacıyla kullanılabilecek kriterlere demografik özellik, tüketici türü, sosyoekonomik sınıf, meslek grubu, firma büyüklüğü, coğrafi yerleşim yeri, fakülte türü vb. örnek olarak gösterilebilir.
- Tabakaların oluşturulması.
  Belirlenen tabakalama kriteri itibariyle N hacimli bir evren daha homojen, L sayıda ve hacimleri N_1, N_2, ... , N_L olan tabakalara ayrılır. Bu aşamada önemli olan tanımlanan evrendeki her bir birimin yalnız bir tabakaya ait olması ve hiçbir birimin açıkta kalmamasının sağlanmasıdır. Başka bir ifadeyle

  $$N_1 + N_2 + \ldots + N_L = \sum_{h}^{L}=1 \qquad N_h = N$$

  olmalıdır. Tabaka sayısı L arttıkça tabakaların homojenliği de artacağından tabaka varyansları giderek küçülecek ve buna bağlı olarak da tahminlerin güvenilirliği giderek artacaktır. Tabaka sayısının artması maliyetleri yükseltir ve uygulama zorluğu yaratır. Bu nedenlerle tabaka sayısı L belirlenirken tabaka sayısının yaratacağı maliyet, uygulama zorluğu ve elde edilecek tahminlerin güvenilirliği birlikte değerlendirilmelidir. Deneyimler ve uygulamalar tabaka sayısının 6 'dan fazla olmamasını önermektedir.

- Tabakalardan birimlerin seçilmesi.
  Her tabakadan basit rassal seçimle sırasıyla n_1, n_2, ... , n_L hacimli alt örneklemler oluşturulur. Alt örneklem hacimleri toplamı örneklem hacmine eşittir. Başka bir ifadeyle n örneklem hacmini göstermek üzere

  $$n_1 + n_2 + \ldots + n_L = \sum_{h=1}^{L} n_h = n$$

  olmalıdır.
- Verilerin derlenmesi.
  Oluşturulan alt örneklem birimleri üzerinden veriler derlenir, bu veriler kullanılarak araştırma amaçları için gerekli olan istatistikler hesaplanır ve bu istatstiklere dayanarak istatistiksel çıkarımlar yapılır.

Daha önce de vurgulandığı gibi tabakalar içi homojenlik arttıkça tabakalar içi varyanslar küçülür. Bu da ilgili evren parametre tahminleyicisinin varyansını küçültür. Bu sonuca göre heterojen evrenlerde aynı örneklem hacmi için basit rassal örnekleme uygulamasının örnekleme hatası, tabakalı örneklemenin örnekleme hatasından büyük olur. Heterojen evrenler için tabakalı örnekleme yöntemi daha etkindir.

Tabakalı örneklemenin diğer bir üstünlüğü ilgilenilen evrenin yanısıra her tabaka içinde ayrı bilgi elde etme olanağı sağlamasıdır. Uygulamada evrene göre tabakalar için çerçeve oluşturmak daha kolay olabilir. Ancak sağladığı kolaylıklara rağmen tabakalı örneklemenin bazı güçlükleri de vardır. Örneğin tabaka hacimleri ve bunların toplamı olan evren hacminin bilinmesi gerekir. Bu her zaman mümkün olamamaktadır. Ayrıca ilgilenilen evrenin homojenliğinin sorgulanması için de bu evren hakkında pek çok öncül bilgiye gereksinim vardır. Bu öncül bilgilerin yetersizliği ve geçersizliği oluşturulacak örneklemin temsil niteliğini olumsuz yönde etkiler. Tabakalı örnekleme sürecinin adımlarıyla ilgili yukarıdaki kuramsal bilgileri örnek araştırma üzerinde uygulayalım.

ÖRNEK 16

Amaç: Bir üniversitede görev yapan 1000 öğretim elemanının bilgisayar kullanım alışkanlığını araştırmak.

Evren: 1000 öğretim elemanının oluşturduğu topluluktur. Sonlu evrendir, hacmi N= 1000 kişidir.

Örnekleme Yöntemi: Evren, sonlu evren olduğu için araştırmacı tam sayım da uygulayabilir, örneklemeye de başvurabilir. Araştırmacı örnekleme yapmayı gerekli kılan nedenleri değerlendirmiş ve örnekleme yapmaya karar vermiştir. Araştırmacının gözlemlerine ve öncül bilgilerine göre öğretim elemanlarının bilgisayar kullanım alışkanlıkları bakımından heterojendir. Çünkü yaşları genç olan araştırma görevlileri, öğretim görevlileri daha sık ve daha farklı amaçlarla bilgisayar kullanırken profesörler daha az süreli ve daha sınırlı amaçlarla bilgisayar kullanmaktadırlar. Bu değerlendirmeye göre araştırmacı tabakalı örnekleme yöntemini örnekleme amacıyla tercih etmiş ve uygulama adımlarını aşağıdaki gibi izlemiştir:

- Öğretim elemanlarının ünvan türü kriterine göre bilgisayar kullanma alışkanlığı bakımından homojen tabakalara ayrılabileceği düşünülmüştür.
- Tabakalar öğretim üyesi dışındakiler (araştırma görevlisi, öğretim görevlisi), yardımcı doçent doktor (Yrd. Doç. Dr.) , doçent doktor (Doç. Dr.) ve

profesör doktor (Prof. Dr.) şeklinde belirlenmiştir. Yönetimden elde edilen bilgilere göre tabaka hacimleri öğretim üyesi dışındakiler $N_1 = 400$, Yrd. Doç. Dr. $N_2 = 300$, Doç. Dr. $N_3 = 200$ ve Prof. Dr. $N_4 = 100$ birim olduğu görülmüştür.

- n = 100 olarak belirlenen örneklem hacmi tabaka hacimlerinin (N_h) evren hacmi N içindeki paylarıyla orantılı olarak dağıtılmıştır:

$$n_h = (\frac{N_h}{N}) \;.\; n$$

$$n_h = (\frac{N_1}{N}) \;.\; n = (400 / 1000) \;.\; 100 = 40$$

Benzer şekilde h = 2, 3, 4 için hesaplama yapılırsa $n_2 = 30$, $n_3 = 20$, $n_4 = 10$ hacimlik örneklemler rassal olarak seçildi ve

$$n_1 = 40 + n_2 = 30 + n_3 = 20 + n_4 = 10 = n = 100$$

hacimlik tabakalı örneklem oluşturulmuştur.

- Oluşturulan örneklemdeki öğretim elemanlarından birinci elden veri derleme yöntemiyle veriler derlenip çözümleme yapılır ve gerekli bilgiler üretilebilir.

Sistematik Örnekleme

Örneklem için birim seçimi aşağıda ele alınan bir sistematiğe uygun olarak yapılan örnekleme sürecine sistematik örnekleme adı verilir. Bu yöntemin sınırlayıcıları ilgili evrene ilişkin bir çerçevenin var olup olmaması veya birimlerin doğal bir sıraya sahip olup olmamasıdır.

Bir sistematik örneklem oluşturmak için aşağıdaki adımlar izlenir:

- Evrendeki birimler 1'den N'ye kadar numaralandırılır.
- Araştırma için yeterli olacak örneklem hacmi n belirlenir.
- $k = \frac{N}{n}$ büyütme faktörü hesaplanır. Bu oran örnekleme aralığını gösterir.
- 1 ile k arasında bir tamsayı rassal olarak seçilir. Bu sayı a ile gösterilirse a örnekleme girecek birinci birimin sıra numarası olur.
- a'ıncı birimi k aralıklarıyla izleyen a + k'ıncı , a + 2k'ıncı , ... , a + (n - 1) k'ıncı sıra nolu birimler örnekleme seçilir ve n hacimli sistematik örneklem oluşturulur.
- Oluşturulan örneklemden elde edilen veriler kullanılarak gerekli istatistikler hesaplanır.

Bütün bu adımları daha açık bir şekilde bir örnek üzerinden gösterelim:

ÖRNEK 17

- *Evrendeki birim sayısı n = 1000 öğretim elemanıdır. Bu öğretim elemanlarının ünvan türü, soyadı sırası itibariyle 1'den N'ye kadar listeler üzerinde numaralandırma yapıldı.*
- *Diyelim ki n = 100 birimlik bir örneklem seçileceği tasarlandı. k = 1000 / 100 = 10, hesaplandı. Bunun anlamı, her 10'uncu birim örnekleme alınacak demektir.*
- *1, 2, ... , 10 arasından rassal olarak bir sayı seçildi. Seçilen sayı 4 olsun. A = 4. Sıradaki öğretim elemanı örnekleme girecek birinci birimdir.*
- *4. Sıra nolu öğretim elemanından başlayarak her 10. (4, 14, 24, 34,...) sıra nolu öğretim elemanı örnekleme alınarak n = 100 birimlik rassal örneklem oluşturulmuş olur.*

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1000

Tablo 6.2
Örnek problem için sistematik örnekleme uygulama tablosu.

Sistematik örneklemede uygulanacak sistematiğin belirlenmesi her zaman yukarıdaki hesaplamalarda olduğu gibi yapılmayabilir. Evreni oluşturan birimlere sıra numarası verilemiyorsa veya numaralandırma çok zaman alıyor ve masraflı oluyorsa ancak sonlu veya sonsuz evren birimleri rastgele dizilişlere veya gelişlere sahip ise sistematik örnekleme uygulanabilir. Örneğin bir üniversitenin 20.000 öğrencisi üzerinde yapılacak bir araştırma için haftanın seçilen bir öğretim günü saat 8:30 da başlayıp her yarım saat aralıklarla üniversite kampüs kapısından giren öğrencinin örnekleme alınması uygulaması da bir sistematik seçimdir. Bir süpermarketten ayrılan her k'ıncı müşteriyle görüşme yapılarak yürütülen araştırmalar da bu örneklemin uygulandığı araştırmalardır. Sistematik örnekleme evrendeki her birimin örneklemde yer alma olasılığı bakımından basit rassal örneklemeye benzer ve bu olasılık n / N dir. Ancak mümkün örneklemlerden herhangi birinin incelenen örneklem olması olasılığı basit rassal örneklemede 1 / C_N^n olduğu hâlde sistematik örneklemede belirlenen sistematiğe göre örneklem olma şansına sahip kombinasyonların her biri için eşit 1 / k diğerleri için 0'dır.

Sistematik örnekleme uygulaması sonucu oluşturulacak örneklemin temsil niteliği evren birimlerinin sıralandırılması ve k aralığı ile ilişkilidir. Bu sıralandırma araştırmaya konu olan değişkenlerle ilişkilendirilerek yapılırsa sistematik örnekleme basit rassal örneklemeye göre daha temsili örneklem oluşturma imkânı verir. Örneğin firmaların düşük ciroya sahip olandan büyük ciroya sahip olana doğru sıralanması, otellerin yıldız sayıları bakımından büyükten küçüğe doğru sıralanması, öğrencilerin küçük boyludan büyük boyluya doğru sıralandırılması durumunda sistematik örnekleme her gruptan birimin örnekleme girmesini temin edebilir. Ancak örnekleme aralığının uygun şekilde belirlenememesi hâlinde sistematik örneklemenin basit rassal örneklemeye olan yukarıda açıklanan üstünlüğü ortadan kalkar. Çünkü belirli özelliğe sahip birimlerin gereksiz oranda örnekleme girmesi söz konusu olabilir. Bu durum örneklemin temsil niteliğini olumsuz yönde etkiler.

Sistematik örnekleme önceki rassal örnekleme yöntemlerine göre daha az maliyetli ve uygulaması kolaydır.

Çerçevenin doğal yapısında tekrarlamalar varsa sistematik örnekleme kullanılmamalıdır. Örneğin, veriler aylık olarak düzenlenmiş ve k = 12 alınmışsa her yılın aynı ayı örnekleme gireceğinden bu tür bir uygulama tek yönlü hatalara neden olabilir.

Tek Aşamalı ve Çok Aşamalı Küme Örneklemesi

Bu örnekleme türünde tabakalı örneklemede olduğu gibi evrenin birimleri küme adı verilen gruplara ayrılır. Bu gruplar genellikle doğal olarak vardır. Her küme bir örnekleme birimi olarak tanımlanır. Kümeler arasından rassal olarak belirli sayıda küme seçilir ve seçilen kümelerdeki gözlem birimlerinin tamamı örneklemi oluşturur. Örneğin bir organize sanayi bölgesinde faaliyette bulunan işyerlerinde çalışan işgörenler hakkında bir araştırma planlandığında bu organize sanayi bölgesinde her bir işyerinde çalışan işgörenler bir küme olarak tanımlanabilir. Hanehalkı geliriyle ilgili bir araştırmada her mahalledeki hanehalkı topluluğu bir küme olarak tanımlanır. Örneklerden de anlaşılabileceği gibi kümeler genellikle coğrafi bir kritere göre tanımlanmaktadır.

Küme örneklemesi bir ve daha fazla kümeleme aşamaları ile de uygulanabilir. Bir kümeleme aşaması ile gözlem birimlerine ulaşılıyorsa *tek aşamalı kümeleme,* iki veya daha fazla kümeleme aşaması ile gözlem birimlerine ulaşılıyorsa *çok aşamalı kümeleme* adı verilir.

Bu örnekleme yöntemleri evrendeki birimlerin homojen, hacimlerinin çok büyük ve geniş bir coğrafi alana yayılmış olmaları ya da örnekleme girecek birimlere ilişkin bir çerçeve oluşturmanın mümkün olmadığı durumlarda tercih edilmesi gereken yöntemlerdir.

Tek aşamalı (küme) örneklemesi sürecinde aşağıdaki adımlar izlenir.

- İlgilenilen evrendeki birimler genellikle coğrafi kritere göre kümelere ayrılır. Bu, birinci düzey kümelemedir. Kümeler doğal olarak belirli bir mekânda var olan birimlerden oluşur. Küme sayısı “M” simgesiyle gösterilir. Üniversiteler, banka şubeleri, lojistik firmaları, kamu kurumları, ortaöğretim okulları birer kümedir. Örneğin ortaöğretim kurumlarını ele alalım. Bu okulların öğrencileri, sınıfları, öğretmenleri kümeleri oluşturur.
- Kümeler arasından rassal seçimle “m” sayıda küme seçilir.
- Seçilen kümelerdeki birimlerin toplamı tek aşamalı küme hacmini gösterir.

Tanımlanan birinci aşama kümelerine, benzer kritere göre ikinci, üçüncü ve n’inci aşama kümelere ayrılır ve son kümeleme aşamasındaki kümeler arasından rassal seçimle m sayılı küme seçilir ve seçilen kümelerdeki birimlerden örneklem oluşturulursa yapılan örneklemeye çok aşamalı küme örneklemesi denir.

Tek ve çok aşamalı küme örneklemesi ile ilgili kuramsal açıklamaları bir araştırma örneği üzerinden açıklayalım.

ÖRNEK 18

Anadolu Üniversitesinde 2010-2011 öğretim yılında örgün öğretim yapan lisans programlarına kayıtlı olan öğrencilerin kendilerine sunulan eğitim-öğretim ortamlarından memnun olup olmadıkları araştırılmak isteniyor.

Evren: Bu araştırmanın evreni N=26000 olan öğrencilerin oluşturduğu topluluktur. Sonlu ve büyük hacimli bir evrendir. Evren hacminin büyük olması, bütün öğrencilere ulaşmanın güçlükleri gibi nedenlerle araştırma için örneklemeye başvurulması düşünülmüştür.

Örnekleme Yöntemi: Öğrenciler iki büyük yerleşkedeki öğretim birimlerinde öğrenim görmektedirler. Araştırmacı, öğrencilerin eğitim-öğretim ortamlarından memnuniyetlerini değerlendirmeleri bakımından homojen olduğu öncül bilgisine sahiptir. Bu değerlendirmelere göre araştırma için uygun örnekleme yöntemi olarak önce tek aşamalı sonra iki aşamalı örnekleme yöntemi seçilmiştir.

Tek aşamalı örnekleme uygulamasının adımları:

- Öğrenciler lisans öğrenimi yapan program türü kriterine göre kümelere ayrılmıştır. Kümeler İktisadi ve İdari Bilimler Fakültesi, Fen Fakültesi, Güzel Sanatlar Fakültesi, Mühendislik Fakültesi, Edebiyat Fakültesi, Turizm Yüksekokulu vb. öğrencileri topluluğu şeklinde tanımlanmıştır. Bu tanıma göre küme sayısı M=6 ‘dır.
- M=6 lisans programı arasından m=2 program rassal olarak seçilir. Varsayalım ki seçilen 2 program İktisadi ve İdari Bilimler Fakültesi ve Mühendislik Fakültesi olsun. Seçilen birinci fakültedeki öğrencilerin tamamı N_1=5000 ve ikinci fakültedeki öğrencilerin tamamı N_4=4000 ise örneklem hacmi $n=N_1+N_4=5000+4000=9000$ öğrenci örnekleme seçilmiş olur.

İki aşamalı örnekleme uygulaması benimsendiğinde birinci aşamada seçilen İktisadi ve İdari Bilimler ve Mühendislik Fakülteleri bölüm türü kriterine göre tekrar kümelere ayrılırsa ikinci aşama kümeleme yapılmış olur. İkinci aşamada oluşturulan bölümlerden (kümelerden) rassal seçimle m_2 sayıda bölüm seçilir ve bu bölümlerdeki bütün öğrenciler örneklemi oluşturur.

Tek ve çok aşamalı örnekleme uygulamasında tanımlanan kümeler örnekleme birimi olarak benimsendiğinden basit rassal örneklemede olduğu gibi evrenle ilgili bir çerçeveye gerek yoktur. Sadece seçilen kümelerle ilgili çerçeveye gereksinim vardır. Bu durum örnekleme uygulamasında zaman, maliyet tasarrufu yanında uygulama kolaylığı sağlamaktadır.

Eğer birimler kümeler arasında homojen değilse seçilen kümelerdeki birimlerden oluşacak örneklemin evreni temsil niteliği tartışılır çünkü evreni oluşturan her türden birim örnekleme girmemiş olur.

Tek aşamalı ve çok aşamalı örnekleme yöntemleri, örnekleme maliyetini azaltıp etkinliği artırırken, tabakalı örnekleme doğruluğu artırmaktadır. Küme örneklemesinde kümelerdeki birimlerin mümkün oldukça heterojen olması, tabakalı örneklemede tanımlanan tabakaların ise mümkün oldukça homojen olması istenir.

Olasılık örnekleme ile olasılıklı olmayan örneklem yöntemleri arasındaki temel fark nedir? Açıklayınız.

SIRA SİZDE 4

ÖRNEKLEME DAĞILIMLARI

Daha önce de değinildiği gibi tanımlanan evrene ilişkin sayısal karakteristiklere parametre adı verilir ve parametre genel olarak θ (theta) simgesiyle gösterilir. Tam sayım yapılamadığı durumlarda araştırmacılar istatistiksel tahminleme ve karar verme (istatistiksel çıkarım) problemleri ile karşılaşırlar. Bu çıkarımlar örneklem istatistiklerine dayanır. Örneklem istatistiklerinin genel gösterimi daha önce ifade edildiği gibi $\hat{\theta}$ simgesiyle yapılır. Örneklem istatistiği bilindiği gibi rassal olarak seçilen n hacimli örneklemden elde edilen $x_1, x_2, \ldots, x_n$ gözlem değerlerinin kullanılmasıyla hesaplanan karakteristiklerin genel adıdır.

Örnekleme sürecinde rassal olarak seçilen n hacimli bir örneklem için hesaplanan istatistikler sadece ait oldukları örneklem için bilgi niteliğindedirler. Çünkü incelenen n hacimli bir örneklem aynı hacimli ve fakat farklı birimlerden oluşabilecek mümkün örneklemlerden sadece biridir ve mümkün örneklemlerin her biri için hesaplanacak istatistikler birbirinden farklı ve evren parametre değerlerine eşit $(\theta = \hat{\theta})$, büyük $(\theta > \hat{\theta})$ veya küçük $(\theta < \hat{\theta})$ olabilir. Bu nedenle örneklem istatistiklerinden yararlanarak evren parametreleri hakkında tahminleme ve karar verme sürecinde rassal olarak seçilen n hacimli bir örneklemin hesaplanan istatistiğinden değil, o istatistiğin mümkün örneklemlerde alacağı değerlerin dağılımından ve bu dağılımın özelliklerinden yararlanılır.

N hacimli sonlu bir evrenden iadesiz seçimle rassal olarak seçilebilecek n hacimli C_N^n sayıdaki mümkün bütün örneklemlerin seçildiğini ve her örneklem için $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_{C_N^n}$ istatistiklerin hesaplandığını varsayalım. Hesaplandığını varsaydığımız bu istatistiklerin dağılımına **örnekleme dağılımı** denir.

Rassal örneklemden hesaplanan i ̂ istatistiğinin olasılık dağılımına bu istatistiğin **örnekleme dağılımı** denir.

Eğer N hacimli evrenden n hacimli örneklemler iadeli seçimle oluşturulursa mümkün örneklem sayısı ve bunlar için hesaplanacak istatistik sayısı N^n olacaktır.

Uygulamada, n hacimli bir tek örneklem seçilir ve bu örneklem için tahminlenecek veya karar verilecek parametre hakkında bilgi üreten istatistik hesaplanır. n hacimli mümkün örneklemler seçilmez, örneklem istatistikleri hesaplanmaz ve bu istatistiğin dağılımı oluşturulmaz. Mümkün örneklemler seçilmiş gibi düşünülerek bu istatistiğin varsayımsal dağılımından yararlanmak suretiyle evren parametreleri hakkında çıkarımlar yapılabilir. Bu durum, örneklem istatistiğinin aynı hacimli örneklemden örnekleme farklı değerler alan rassal değişken olduğu esasına dayanır. Bir evrenden örneklem seçilmeden önce örneklem gözlem değerleri x_1, x_2 , ... , x_n rassal değişkenlerdir ve bu gözlem değerlerinden hesaplanan istatistikleri de bir rassal değişkendir. Yapılan açıklama bağlamında örnekleme dağılımı bir rassal değişken olan örneklem istatistiğinin olasılık dağılımı şeklinde tanımlanabilir.

Çeşitli amaçlar için örnekleme karar verildiği zaman dikkatlerin en çok odaklandığı parametreler evren aritmetik ortalaması (μ) ve evren oranı (Π) olmaktadır. Bu nedenle bu ünitenin izleyen kısımlarında bu parametreler hakkında bilgi üreten örneklem istatistiklerinin sırasıyla örneklem aritmetik ortalaması $\bar{X}$ ve örneklem oranı p'nin örnekleme dağılımları ve özellikleri incelenecektir.

Ortalamanın ($\bar{X}$ 'nın) Örnekleme Dağılımı

Bir örneklem istatistiği olan örneklem aritmetik ortalaması ($\bar{X}$) rassal bir değişkendir. $\bar{X}$ rassal değişkeninin olasılık dağılımına ortalamanın örnekleme dağılımı adı verilir. Bir başka anlatımla tanımlanan evrenden n hacimli bir rassal örneklem değil de aynı hacimli (veya N^n) sayıdaki mümkün rassal örneklemlerin seçildiğini ve her mümkün örneklem için $\bar{X}$ hesaplandığını varsaydığımızda { $\bar{X}_i$ } dan oluşan bir frekans dağılımı elde edilebilir. Bu dağılıma ortalamanın örnekleme dağılımı adı verilir. Bir örnek üzerinde $\bar{X}$ nın örnekleme dağılımını oluşturalım.

ÖRNEK 19

Bir banka şubesinde çalışan 4 işgören bir üst ünvana yükselme (terfi) için açılan hizmetiçi eğitim programına katılmışlardır. Bu kişilerin simgesel isimleri ve eğitim sonunda açılan sınavda aldıkları puanlar aşağıda verilmiştir. Evren ortalamasını hesaplayınız ve $\bar{X}$ nın örnekleme dağılımını oluşturunuz.

Eğitime Katılanlar (Birimler)	Başarı Puanları X_i
A	90
B	80
C	60
D	70
	300

Çözüm:

Evren ortalaması μ, tam sayım yapıldığında, gözlem değerleri x_1, x_2, ... , x_N olarak gösterildiğinde ve yukarıdaki veriler kullanıldığında

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N} = \frac{300}{4} = 75 \ puan$$

şeklinde hesaplanır. Hesaplama hatası yapılmamış ise 75 puan kesin, doğru olan ortalama başarıyı gösterir.

Tam sayım yapılamadığı durumlarda açıktır ki yukarıdaki başarı puanları (x_i değerleri) derlenememiş ve μ=75 puan bilgisi hesaplanamamış olur. Bu durumda μ hakkında tahminleme ve karar verme (istatistiksel çıkarım) problemleriyle karşılaşılır. Bu türden problemlerin çözümlenebilmesi için μ hakkında bilgi üreten örneklem istatistiği $\overline{X}$ nın örnekleme dağılımının özellikleriyle ilgili bilgilere gereksinim vardır.

$\overline{X}$ ların örnekleme dağılımının oluşturulması için aşağıdaki adımlar izlenir.

- N=4 birimlik evrenden iadeli veya iadesiz seçimle belirlenen n hacimli mümkün örneklemler seçilir. Örneğin n=2 için iadesiz seçimle oluşturulabilecek mümkün örneklem sayısı;

$$C_N^n = \frac{N!}{(N-n)!\,.\,n!} \qquad C_4^2 = \frac{4!}{(4-2)!\,.\,2!} = 6 \text{ adet}$$

olup aşağıdaki tabloda verilmiştir.

Örneklem No	Mümkün Örneklemler	Örneklem Gözlem Değerleri	$\overline{X}_i$
1	A, B	90, 80	85
2	A, C	90, 60	75
3	A, D	90, 70	80
4	B, C	80, 60	70
5	B, D	80, 70	75
6	C, D	60, 70	65
			450

Tablo 6.3
N=4 hacimli evrenden n=2 hacimli mümkün örneklemlerin $\overline{X}_i$ larının örnekleme dağılımı.

$\overline{X}_i$ lar serisinin dağılımına, $\overline{X}$ nın örnekleme dağılımı adı verilir. Bu dağılımın ortalaması $\mu_{\overline{x}}$ simgesiyle gösterilir ve aşağıdaki gibi hesaplanır:

$$\mu_{\overline{x}} = \frac{\sum_{i=1}^{C_N^n} X_i}{C_N^n} = \frac{450}{6} = 75 \ \textit{puan}$$

Görüldüğü gibi evren ortalaması μ, $\overline{X}$ nın örnekleme dağılımının ortalaması $\mu_{\overline{x}}$ ya eşittir ve

$\mu = \mu_{\overline{x}} = 75$ puan

şeklinde yazılır. Tablo 6.3'teki $\overline{X}_i$ lar frekans dağılımı olarak düzenlendiğinde n=2 hacimli farklı örneklem ortalamalarının olasılığı Tablo 6.4'teki gibi gösterilmiş olur.

$\overline{X}_i$	n_i	Örneklem Ortalamalarının Elde Edilmesi Olasılığı
65	1	1/6 = 0,167
70	1	1/6 = 0,167
75	2	2/6 = 0,333
80	1	1/6 = 0,167
85	1	1/6 = 0,166*
	6	1,000

Not: *: Olasılıklar toplamını 1'e eşitlemek için düzeltme yapılmıştır.

Tablo 6.4
N=4 hacimli evrenden n=2 hacimli mümkün örneklemlerin $\overline{X}_i$ larının frekans dağılımı.

Tablodaki bilgilere göre, örneğin $\overline{X}$ = 75 puan değerini elde etme olasılığı %33,3'tür bilgisi üretilebilir.

Örnekleme girecek birimlerin seçimi iadeli yapılmış olsaydı, N=4 birimden, n=2 birimlik Tablo 6.5 teki $N^n = 4^2 = 16$ farklı örneklem oluşurdu. 16 farklı örneklem için yukarıdaki işlemler yapılırsa $\overline{X}$ ların örnekleme dağılımı oluşturulabilir. Bu durumda da

$$\mu = \mu_{\overline{x}} = 75 \text{ puan}$$

olduğu görülebilir.

Tablo 6.5
N=4 birimlik evrenden iadeli seçimle oluşturulabilecek mümkün örneklemler.

	A	B	C	D
A	A, A	A, B	A, C	A, D
B	B, A	B, B	B, C	B, D
C	C, A	C, B	C, C	C, D
D	D, A	D, B	D, C	D, D

Özetle

- Her örneklem hacmi için bir istatistiğe ilişkin örnekleme dağılımı olduğu düşünülür.
- Örnekleme birim seçimi iadeli de yapılsa, iadesiz de yapılsa hesaplanan örneklem ortalamalarının dağılımı evren ortalamasına eşit olur.

Ancak hiçbir araştırmada istatistiksel çıkarsama amacıyla yukarıda açıkladığımız işlemler yapılmaz. Bunun yerine, n hacimli tek bir örneklem seçilir, bunun $\overline{X}$ ortalaması hesaplanır ve örnekleme dağılımı ile ilgili yukarıda açıklanan bilgilerden yararlanılarak μ hakkında çıkarım yapılır.

($\overline{X}$ 'nınDağılımının Özellikleri

$\overline{X}$ rassal değişkeninin örnekleme dağılımının özellikleri, bu dağılımın ortalaması μ (evren ortalaması, $\overline{X}$ nın örnekleme dağılımının ortalaması $\mu_{\overline{x}}$ ya eşit olduğu için $\mu_{\overline{x}}$ yerine μ kullanılmıştır) ve standart sapması $\sigma_{\overline{x}}$ (standart hata) ile açıklanır. Standart hatanın karesi ise varyans olarak isimlendirilir ve $\sigma^2_{\overline{x}}$ simgesi ile gösterilir.

$\overline{X}$ 'nın Dağılımının Ortalaması

$\overline{X}$ nın örnekleme dağılımının ortalaması veya aynı anlama gelen, $\overline{X}$ nın beklenen değeri E($\overline{X}$) şeklinde gösterilirse

$$E(\overline{X}) = \mu$$

yazılabilir. Bu sonuca göre $\overline{X}$ nın örnekleme dağılımının ortalaması ile ilgili aşağıdaki değerlendirmeler yapılabilir:

- Örneklem hacmi n arttıkça $\overline{X}$ nın örnekleme dağılımının ortalaması evren ortalamasına yaklaşır. Örneklem hacmi yeterli büyüklüğe ulaştığında $\overline{X}$ nın örnekleme dağılımı, normal olur.
- Evrenin dağılım şekli çarpık bir dağılım gösterse bile, örneklem hacmi arttıkça $\overline{X}$ nın dağılımı normal dağılıma yaklaşır.

Örneğin, bankadaki işgörenlerin girdiği sınavla ilgili örnek ele alındığında ve rassal olarak seçilen n=2 hacimli örneklemdeki birimler A ve D birimleri olduğunda örneklem ortalaması;

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\sum_{i=1}^{2} X_i}{2} = \frac{90+70}{2} = 80 \ \ puan$$

olarak hesaplanır ve μ nün tahmini

E($\bar{X}$) = μ = 80 puan şeklinde yazılabilir.

$\bar{X}$ 'nın Dağılımının Standart Hatası

$\bar{X}$ nın standart sapması veya aynı anlama gelecek şekilde, $\bar{X}$ nın örnekleme dağılımının standart hatası $\sigma_{\bar{x}}$ simgesiyle gösterilir ve standart hata olarak da isimlendirilir.

Standart hata $\sigma_{\bar{x}}$ ortalamanın örnekleme dağılımının değişkenliğini gösterir. Yani, mümkün örneklem ortalamalarının ($\bar{X}_i$ ların) evren ortalamasından farklarının ($\bar{X}_i$ – μ = hata) ortalama ölçüsüdür. Standart hatanın karesi ($\sigma^2_{\bar{x}}$), $\bar{X}$ nın dağılımının varyansını ifade eder.

Evren, sonsuz bir evren ise basit rassal örneklemede birim seçimi iadeli seçimle yapılıyorsa örneklem hacmi n ≥ 30 birim veya örnekleme oranı $\frac{n}{N} \leq 0{,}05$ ise $\sigma_{\bar{x}}$ ve $\sigma^2_{\bar{x}}$ aşağıdaki eşitlikler yardımıyla hesaplanır:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad , \qquad \sigma^2_{\bar{x}} = \frac{\sigma^2}{n}$$

Bu eşitliklerden $\sigma_{\bar{x}}$'nın özellikleri ile ilgili aşağıdaki değerlendirmeler yapılabilir.

- Standart hata evren standart sapması σ ya ve örneklem hacmi n'e bağlıdır. Bir başka ifadeyle evren değişkenliği σ büyük ise herhangi bir örneklem hacmi için $\sigma_{\bar{x}}$ da büyük olur.
- Örneklem hacmi arttıkça örneklem istatistiğinden yararlanarak μ hakkında daha az hatalı, daha güvenilir bilgi üretmek mümkün olur.
- Örneklem hacminin kare kökü ile $\sigma_{\bar{x}}$ arasında ters yönde ilişki vardır. Yani örneklem hacmini artırdıkça $\sigma_{\bar{x}}$ küçülür. Ancak örneklem hacmini artırarak $\sigma_{\bar{x}}$ yı düşürmeye çalışmak örneklemeye başvurmayı gerekli kılan nedenlerden dolayı bazı güçlüklere yol açar. Örneğin n=100 birim iken $\bar{X}$ standart sapmasını yarıya indirebilmek için örneklem hacmi 4 kat artırılmalıdır.

Evren standart sapması genellikle bilinmediğinden $\sigma_{\bar{x}}$ hesaplanırken σ yerine onun yansız bir taminleyicisi olan örneklem standart sapması s kullanılır. Bu durumda standart hata $S_{\bar{x}}$ simgesiyle gösterilir ve $S_{\bar{x}} = \frac{s}{\sqrt{n}}$ eşitliğiyle hesaplanır.

Örneklem standart sapması,

$$s = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}}$$

şeklinde hesaplanır.

Eğer ilgilenilen evren sonlu bir evren ve örnekleme oranı n/N ≥ 0,05 ise standart hata hesaplanırken $\sqrt{\frac{N-n}{N-1}}$ şeklindeki bir çarpan, düzeltme faktörü olarak kullanılır ve standart hata hesaplanması

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

veya

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

eşitlikleriyle yapılır.

Basit rassal örneklemede örneklem hacmi arttıkça nın örnekleme dağılımı normal dağılıma yaklaşır. Bu sonuca, istatistikte önemli bir yeri olan aşağıdaki teorem yardımıyla ulaşılır:

Merkezi Limit Teoremi

Evrenin dağılım şekli ne olursa olsun, basit rassal örneklem hacmi büyüdükçe, $\bar{X}$ nın örneklem dağılımı normal dağılıma yaklaşır. Bu dağılımın ortalaması μ, varyansı σ^2/n dir. Örneklem hacmi n için yeterli büyüklük, kesin olmamakla birlikte uygulamada $n \geq 30$ birim olarak kabul edilmektedir.

Eğer $\bar{X}$ ortalaması μ ve varyansı σ^2 olan normal dağılımlı bir evrenden seçilmiş n hacimlik basit bir rassal örneklemin ortalaması ise $\bar{X}$ nın örnekleme dağılımı ortalaması μ, varyansı σ^2 / n olan bir normal dağılımdır.

$\bar{X}$ rassal değişkeninin dağılımı normal olduğunda,

$$z_i = \frac{\bar{X}_i - \mu}{\sigma / \sqrt{n}}$$

Eşitliğiyle standart değişkene dönüştürülür. Böylece, normal dağılımın özellikleri kullanılarak örneklem aritmetik ortalamasından evren aritmetik ortalaması hakkında bilgi üretmek kolaylaşır.

Normal dağılan bir evrenden, rassal olarak seçilebilecek birbirinden farklı $n<30$ birimlik mümkün bütün örneklemlerin seçildiğini, her örneklem için $\{\bar{X}_i\}$ ları ve onların $\frac{\bar{x}-\mu}{S_x}$ standart değerlerini hesaplandığını düşünelim. Değerler aralığı $-\infty < \frac{\bar{x}-\mu}{S_x} < \infty$ olan istatistiğin dağılımı (n-1) serbestlik derecesi (sd = n-1) ile t dağılımı adı verilen sürekli bir dağılım gösterir ve bu istatistik

$$t = \frac{\bar{x}-\mu}{S_{\bar{x}}}$$

Burada, $S_{\bar{x}} = \dfrac{s}{\sqrt{n-1}}$

şeklinde hesaplanır.

t dağılımı ortalaması sıfır olan tek modlu ve simetrik bir dağılımdır. Dağılımın şekli standart normal dağılıma benzer fakat değişkenliği daha büyüktür. Bu değişkenlik serbestlik derecesi ile ters orantılıdır. Örneklem hacmi artarken (sd = n-1) büyür, t değerinin hesaplanmasında $S_{\bar{x}}$ nın kullanılması nedeniyle ortaya çıkan değişkenlik küçülür ve t dağılımı standart normal dağılıma (z dağılımına) yaklaşır. t örnekleme dağılımının özelliklerinden yararlanarak evren ortalaması μ ile ilgili bilgilerin nasıl üretileceği de izleyen ünitede örneklerle açıklanacaktır.

ÖRNEK 20

Otomobil lastiği üreticisi bir fabrikanın yöneticisi ürettikleri lastiklerin ortalama ömrünü lastiklerin katettiği km olarak tahmin etmek istiyor. Bu amaçla rassal olarak 100 lastik seçilmiş ve bu lastiklerin ortalama ömrünün $\bar{X} = 40000$ km ve standart sapmasının s=1500 km olduğu tespit edilmiştir. Yönetim, ürettikleri lastiklerin 35000 Km ömürlü olmasını planlamıştır.
Bu bilgileri kullanarak;

- *$\bar{X}$ nın örnekleme dağılımının ortalaması nedir? Hesaplayınız.*
- *İstenen tahminleme yapılırken işlenebilecek hata nedir? Hesaplayınız.*
- *$\bar{X}$ nın standart z değerini hesaplayınız.*

Çözüm:

- $E(\bar{X}) = \mu = 40000$ km
- n= 100 lastik olduğu için standart hata ($n \geq 30$ birim) evren standart sapması bilinmediği için)

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{1500}{\sqrt{100}} = 150 \text{ km}$$

hesaplanır. Üretilen lastiklerin tümünün ömrünü yukarıdaki verilere göre tahminlerken işlenebilecek hata düzeyi 150 km'dir bilgisi elde edilebilir.

Örneklem Oranı p'nın Örnekleme Dağılımı

Örnekleme planlarında ele alınan evrenin araştırılmak istenen özelliklerinin bazıları iki sonuçlu olmaktadır. Örneğin bir fabrikada üretilen ürünler, hatalı ya da hatasız ürün, bir fakültedeki öğrenciler, başarılı ya da başarısız öğrenci olmak üzere iki grupta toplanabilir. Bu iki sonuçtan birinde örneğin A sonucunda yer alan birimlerin oranıyla ilgilenilebilir. Bu durumda evren oranı evrenin birimleri içindeki ilgilenilen türden özelliğe sahip olanların oranı biçiminde tanımlanır.

ÖRNEK 21

Y sınıfındaki öğrencilerin genel başarı durumu aşağıda verilmiştir. Bu sınıfın başarılı öğrenci oranı nedir?

ÖĞRENCİ ADI	BAŞARI DURUMU
A	Başarılı
B	Başarısız
C	Başarılı
D	Başarılı

Örnekte sınıftaki başarılı öğrenci oranı, evren oranıdır ve Π ile gösterilir. Bu evrendeki ilgilenilen türden özelliğe sahip (başarılı) birim (öğrenci) sayısı R ile gösterilirse evren oranı Π,

$$\Pi = \frac{R}{N}$$

Eşitliği ile hesaplanır. Burada R = 0, 1, 2, ... , N değerlerini alabileceği için Π nin değer aralığı $0 \leq \Pi \leq 1$ olur. Sınıftaki başarısız öğrenci sayısı (ilgilenilmeyen türden özelliğe sahip birim sayısı) N-R olduğu için, başarısız öğrenci oranı Q,

$$Q = \frac{N - R}{N} = 1 - \Pi$$

olur.

Yukarıdaki örnekte başarılı öğrenci sayısı, R=3 olduğu için

$$\Pi = \frac{3}{4} = 0,75$$

Olarak bulunur. Bu sonuca göre sınıftaki öğrencilerin %75 i başarılıdır. Bu kesin bir sonuçtur.

Tam sayım yapılamadığı zaman R bilinemez ve Π hesaplanamaz. Örnekleme planlarında, Π parametresi hakkında bilgi, örneklem istatistiklerinden yararlanılarak üretilebilir.

Hacmi n olan bir basit rassal örneklemden, bu örneklemin seçildiği evrenin Π parametresi hakkında bilgi üretebilmek için iki örneklem istatistiği söz konusudur. Birincisi, hacmi n olan bir basit rassal örneklemdeki ilgilenilen türden özelliğe sahip olan birimlerin sayısıdır ve r ile gösterilir. İkincisi, ilgilenilen türden özelliğe sahip olan örneklemdeki birimlerin oranıdır. Örneklem oranı p simgesiyle gösterilir ve

$$p = \frac{r}{n}$$

eşitliği ile hesaplanır. Burada r = 0, 1, 2, ... , n değerlerini alabilir. r'nin değerlerine bağlı olarak p de $0 \leq p \leq 1$ aralığında bir değer alır. Örneklemi oluşturan birimler arasında ilgilenilen türden sonuca sahip olmayan birimlerin oranıysa q ile gösterilir. Bu sonuca sahip birimlerin sayısı n- r olduğu için

$$q = \frac{n - r}{n} = 1 - p$$

olur.

Yukarıda verilen örnekte ele alınan N=4 birimlik bir evrenden basit rassal örneklemeyle hacmi n=2 olan bir örneklem seçildiğinde ve örneklemdeki birimler (öğrenciler) B ve C olduğunda başarılı öğrenci oranı,

$$p = \frac{r}{n} = \frac{1}{2} = 0,5$$

Olarak hesaplanmış olur.

Öte yandan, evren oranına ilişkin varyans,

$$\sigma^2 = \Pi\left(1 - \Pi\right)$$

Şeklinde ifade edilir. Varyansın karekökü de standart sapmayı verdiğinden,

$$\sigma = \sqrt{\Pi\left(1 - \Pi\right)}$$

şeklinde yazılır.

Örneklem varyansı ve standart sapması da benzer şekilde sırasıyla

$$s^2 = p\left(1 - p\right)$$

ve

$$s = \sqrt{p\left(1 - p\right)}$$

olarak gösterilir.

İki sonuçlu bir evrenden, mümkün bütün n hacimli basit rassal örneklemlerin seçildiğini ve her örneklem için p oranının hesaplandığı varsayıldığında p_i oranlarından oluşan bir dağılım elde edilir. Bir evrenden seçilebilecek aynı hacimli mümkün bütün örneklemler için hesaplanan örneklem oranlarının oluşturduğu dağılıma *oranların örnekleme dağılımı* adı verilir.

Örnekleme planlarında, tanımlanan evrenden rassal olarak n hacimli sadece tek bir örneklem oluşturulur ve bu örneklem için p oranı hesaplanır. P rassal değişkeninin çekilmesi mümkün bütün n hacimli örneklemlerde aldığı değerlerin dağılımına "oranların örnekleme dağılımı" adı verilir.

Ortalama ve Varyans

Evren oranı hakkında araştırılmak istenin bilgi n hacimli tek bir örneklem için hesaplanan p istatistiğine değil, bir rassal değişken olan p istatistiğinin örnekleme dağılımının özelliklerinden yararlanılarak üretilir. Bu dağılımın özellikleri dağılımın aritmetik ortalaması ve varyansıyla belirlenebilir.

Sonsuz bir evrenden seçilen n hacimli basit rassal örneklem için hesaplanan p oranının örnekleme dağılımının aritmetik ortalaması μ_p evren oranı Π ye eşittir. Bu durum örneklem oranı p'nin evren oranı Π nin yansız (sistematik hata içermeyen) tahminleyicisi olduğunu gösterir. Bu sonuca göre,

$$E(p) = \Pi$$

Yazılır.

Sonsuz evrenlere ya da örnekleme oranı $n/N \leq 0{,}05$olan bütün sonlu evrenlere uygulanan basit rassal örnekleme planlarında örneklem oranı p nin dağılımının varyansı σ_p^2 ve standart hatası da σ_p ile gösterilir.

Eğer evren varyansı σ^2 biliniyorsa

$$\sigma_p^2 = \frac{\Pi\left(1-\Pi\right)}{n}, \qquad \sigma_p = \sqrt{\frac{\Pi\left(1-\Pi\right)}{n}}$$

Eşitlikleriyle, evren varyansı σ^2 bilinmiyorsa

$$s_p^2 = \frac{p\left(1-p\right)}{n}, \quad s_p = \sqrt{\frac{p\left(1-p\right)}{n}}$$

Eşitlikleriyle hesaplanır.

Uygulanan bir basit rassal örnekleme planında ilgilenilen evren sonlu ve örnekleme oranı $n/N \leq 0{,}05$ ise örneklem oranı p nin örnekleme dağılımının varyansı σ_p^2 ve standart hatası σ_p;

Evren varyansı biliniyorsa

$$\sigma_p^2 = \frac{\Pi\left(1-\Pi\right)}{n} \cdot \frac{N-n}{N-1}, \quad \sigma_p = \sqrt{\frac{\Pi\left(1-\Pi\right)}{n}} \cdot \sqrt{\frac{N-n}{N-1}}$$

Evren varyansı bilinmiyorsa

$$s_p^2 = \frac{p\left(1-p\right)}{n} \cdot \frac{N-n}{n}, \quad s_p = \sqrt{\frac{p\left(1-p\right)}{n}} \cdot \sqrt{\frac{N-n}{N}}$$

Eşitlikleri kullanılarak hesaplanır.

Dağılım Şekli ve Merkezi Limit Teoremi

Oranların örnekleme dağılımının şekli eğer E(p) = Π < 0,5 ise sağa çarpık, E(p) = Π > 0,5 ise sola çarpık ve E(p) = Π = 0,5 ise simetrik bir dağılım gösterir. Kolaylıkla görülebileceği gibi, E(p) = Π nin değeri 0 veya 1 e yaklaşırken dağılımın çarpıklığı artar.

Merkezi limit teoremine göre bir örnekleme planında seçilen basit rassal örneklemin hacmi n büyürken örneklem oranı p'nin örnekleme dağılımı normal dağılıma yaklaşır. Uygulamada nΠ ≥ 5 ve n(1-Π) ≥ 5 koşullarını birlikte sağlayan örneklem büyüklüğü, yeterli örneklem büyüklüğü olarak kabul edilir. Aynı teoreme göre rassal örneklem hacmi $n \geq 30$ birim olması ve evren oranı Π nin 0 ya da 1 e yakın değerler almaması koşuluyla oranların örnekleme dağılımı normal dağılıma yaklaşır. Bu koşulları sağlayan oranların örnekleme dağılımıyla ilgili problemlerin çözümlerinde normal dağılımın özelliklerinden yararlanılır. Bu amaçla p rassal değişkeni, standartlaştırılmış Z değişkeni

$$Z = \frac{p - \Pi}{\sigma_p}$$

Şeklinde yazılır. Bu standart değişken kullanılarak evren oranı hakkında bilgi üretmek mümkün olur.

SIRA SİZDE 5

- **Örnekleme dağılımı kavramını açıklayınız.**
- **Merkezi limit teoremi istatistiğe ne tür kolaylıklar getirmiştir. Açıklayınız.**

ÖRNEK 22

Bir bankada ayda 8000 işlem yapılıp fiş düzenlenmektedir. Banka yönetimi ilgili ayda düzenlenen fişler içindeki hatalı işlem oranını tahminlemek amacıyla 100 fişi rassal olarak seçiyor. Seçilen fişlerin 15 tanesinin hatalı düzenlendiği belirlenmiştir. Bu verileri kullanarak istenen tahminleme yapılırken işlenecek hata düzeyi nedir?

Çözüm:

Bilindiği gibi, tahminleme yaparken işlenebilecek hata düzeyini belirleme imkanı veren istatistik standart hatadır. Bu

$$s_p = \sqrt{\frac{P(1-P)}{n}}$$

eşitliği ile hesaplanır. Burada

$$P = \frac{r}{n} = \frac{15}{100} = 0{,}15 \quad ve \quad q = 1 - p = 1 - 0{,}15 = 0{,}85$$

$$s_p = \sqrt{\frac{0{,}15 \,.\, 0{,}85}{100}} = \sqrt{\frac{0{,}127}{100}} = 0{,}032$$

Bu bilgiye göre, bu bankada ilgili ayda düzenlenen fişler arasındaki hatalı fiş oranını tahminlerken işlenecek hata düzeyi s_p = 0,032 fiş olacaktır.

Örneklem Hacminin Belirlenmesinde Nicel Yöntemler

Karşılanabilecek Maliyeti Esas Alan Yöntem: Örneklem hacmi n, araştırma bütçesine bağlı olarak,

$$n = \frac{c - c_0}{c_t}$$

eşitliği ile hesaplanır. Burada,

C = Araştırma bütçesini,

C_0 = Araştırmanın sabit maliyetini,

C_t = Örnekleme birimi için değişken maliyeti gösterir.

ÖRNEK 23

Araştırma bütçesinin ₺2.200 ile sınırlı olduğu bir araştırmada, sabit maliyet ₺800 ve örnekleme seçilecek her örnekleme birimi için maliyet ise ₺5'dır. Bu bütçeyle oluşturulabilecek örneklem hacmi en fazla ne olabilir?

Çözüm:

$$n = \frac{c - c_0}{c_t} = \frac{2.200 - 800}{5} = 280$$

Örneklem hacmi en az 280 birim olmalıdır.

Kabul Edilebilir Hata Düzeyini Esas Alan Yöntem: Örneklem istatistiğinin dağılımının normal olduğu varsayımı altında bu yöntemle örneklem hacminin belirlenmesi için aşağıdaki eşitlikten yararlanılır.

Kabul Edilebilir Hata Düzeyi $(\bar{X} - \mu) = d$ Olduğunda;

$$n = \frac{z^2 \sigma^2}{d^2}$$

Kabul Edilebilir Hata Düzeyi $(p - \Pi)$=d Olduğunda;

$$n = \frac{z^2\ \Pi(1 - \Pi)}{d^2}$$

Bu eşitliklerde

n = Örneklem hacmini

$d = (\bar{X} - \mu)$ veya $(p - \Pi)$ araştırmacının belirlediği kabul edilebilir değeri

z = Belirlenen $1 - \alpha$ güven düzeyinde standart normal dağılım tablo değerini

s = Standart sapmayı

P = Evren oranını

gösterir.

Örneğin, kabul edilebilir hata düzeyi $d = (\bar{X} - \mu)$ esas alındığında örneklem hacminin

$$n = \frac{z^2 \sigma^2}{d^2}$$

eşitliği ile hesaplanabilmesi için araştırmacının σ anlamlılık düzeyini ve d^2 değerini belirlemesi ve evren varyansı hakkında σ^2 bilgiye sahip olması gerekir. Evren varyansı σ^2 genellikle bilinmez. Bu durumda, σ^2 ile ilgili bilgi geçmiş yıllarda yapılmış olan aynı ya da benzer konudaki çalışmalardan elde edilebileceği gibi, bir pilot çalışmadan ya da en büyük değerli gözlem değeri x_{max} ve en küçük değerli gözlem değeri x_{min} biliniyorsa ve x_i rassal değişkeni normal dağılıyorsa $\alpha = 0.01$ için

$$s = \frac{x_{\max} - x_{\min}}{6}$$

tahmincisi kullanılarak da hesaplanabilir.

ÖRNEK 24

Bir araştırmacı X ilinin merkez ilçesinde ikamet eden ailelerin ortalama aylık mutfak harcama tutarını tahminlemek istiyor. Ayrıca bu tahminlemede 0.05 anlamlılık düzeyinde ₺10 'lik bir yanılgı payı amaçlıyor. Örneklem hacmi ne olmalıdır? Benzer amaçla bu il merkezinde yapılan araştırmalardan ailelerin aylık mutfak giderleriyle ilgili standart sapmanın ₺150 olduğu öğrenilmiştir.

Çözüm:
d = ₺10.
z = 0.95, $\sigma = 0.05$
s = ₺150.

$$n = \frac{(1.96)^2\, 150^2}{10^2} = \frac{86436}{100} = 864{,}36 \rightarrow 865 \text{ birim}$$

en az 865 aile rassal olarak seçilmelidir.

Özet

AMAÇ 1

Tam sayım ve örnekleme kavramlarını ayırt etmek.

Birinci ünitede de açıklandığı gibi bir araştırmada tanımlanan evren sonlu evren ise gerekli bilgilerin üretilebilmesi için tam sayım yapılabilir veya örneklemeye başvurulur. Araştırma için gerekli zamana, ekonomik imkânlara ve araçlara sahip olunduğunda tam sayım yapılmalıdır. Çünkü tam sayım sonucu elde edilen verileri kullanarak hesaplanan bilgiler kesin bilgilerdir. Tanımlanan evren sonsuz evren ise örnekleme zorunludur. Tam sayım uygulamasının imkânsız olduğu durumlarda örneklemeye başvurmak kaçınılmazdır.

AMAÇ 2

Örneklemeye başvurmanın yararları.

Örneklemeye başvurulduğunda araştırmacı zaman ve ekonomik tasarruf sağlar. Ayrıca tam sayım yapmayı engelleyen diğer nedenlerin varlığında örnekleme araştırma yapmaya imkan verir.

AMAÇ 3

Örnekleme planı hazırlamak.

Örnekleme planı 5 aşamalı bir süreçtir. Bu aşamalarda araştırmacı araştırma yapacağı evreni tanımlar. Bu evren sonlu evren olduğunda evrenle ilgili güncel bir çerçeve hazırlar veya belirli bir kaynaktan nasıl temin edilebileceğini belirler. Daha sonra olasılıklı ve olasılıklı olmayan örnekleme yöntemlerinden hangisini temsili bir örneklem oluşturma amacıyla araştırmasında kullanacağına karar verir. Seçilecek örneklem hacminin ne olacağını belirler ve örnekleme uygulaması sonucu oluşturduğu örneklemdeki birimler üzerinden araştırmaya konu olan değişkenler itibariyle veriler derlenir.

AMAÇ 4

Bir örnek araştırma için örnekleme uygulaması yapmak ve istenen bilgileri üretmek.

Örneklemden derlenen veriler için araştırmada istenen bilgileri üreten istatistikler, örneklem aritmetik ortalaması, örneklem oranı vb. gibi istatistikler hesaplanır. Bu istatistikler ve bu istatistiklerle ilgili dağılımın özellikleri kullanılarak bu istatistiklerin bilgi ürettiği parametreler, evren aritmetik ortalaması μ, evren oranı Π için gerekli çıkarım bilgileri üretilebilir.

Kendimizi Sınayalım

1. Bir örneklemin özelliklerine ilişkin değerlere ne ad verilir?

a. Örnek istatistiği
b. Evren
c. Anlamlı fark
d. Örnekleme
e. Parametre

2. Bir evrenden rastgele seçilen birden fazla örneğin sonuçlarının birbirinden farklı olduğu gözlenmiştir. Bu farklılığın nedeni aşağıdakilerden hangisidir?

a. Beklenen frekans
b. Yöntem farklılığı
c. Örnek değişkenliği
d. Örnek istatistiği
e. Parametre

3. İki sonuçlu bir evrenden, mümkün n hacimli basit rassal örneklemlerin seçildiğini ve her iki örneklem için p oranı hesaplandığı varsayımı altında, *pi* lerin dağılımına (i=1, 2, … , n) ne ad verilir?

a. Oranların örnekleme dağılımı
b. Örneklem istatistiği
c. Binom dağılımı
d. Normal dağılım
e. Ortalamaların örnekleme dağılımı

4. Aşağıdakilerden hangisi olasılıklı örnekleme yöntemlerinden biri **değildir?**

a. Tabakalı örnekleme
b. Basit rassal örnekleme
c. Kartopu örneklemesi
d. Küme örneklemesi
e. Sistematik örnekleme

5. X rassal değişkeni, ortalaması (μ) 50 ve standart sapması (σ) 10 olmak üzere normal dağılmıştır. Buna göre, hacmi n=100 olan örneklemin ortalaması $\bar{X}$=55 değerinin standart normal değeri (z) kaçtır?

a. 0,5
b. 1
c. 1,5
d. 2
e. 5

6. Evren oranının değer aralığı aşağıdakilerden hangisidir?

a. $0 \leq \Pi \leq 1$
b. $-1 < \Pi < 0$
c. $-1 < \Pi < 1$
d. $\Pi < -1$
e. $\Pi > 1$

7. Bir örneklemin gözlem değerleri için hesaplanan karakteristik değerlere ne ad verilir?

a. Ortalama
b. İstatistik
c. Frekans
d. Anlamlı fark
e. Parametre

8. Aşağıdakilerden hangisi, tam sayım yapmayı engelleyen nedenlerden biri **değildir?**

a. Maliyet
b. Ölçüm için birimlerin tahrip edilmesi olasılığı
c. Evren hacminin küçük olması
d. Zaman
e. Evren hacminin sonsuz sayıda olması

9. Tabaka hacimleri sırasıyla 50, 250 ve 200 birimden oluşan bir evrenden kota örneklemesiyle 50 birimlik örneklem oluşturulmak istenmektedir. Bu örneklem hacmi tabaka hacimlerinin evren içindeki paylarıyla orantılı olarak dağıtılırsa üçüncü sıradaki tabakadan seçilecek birim sayısı kaç olmalıdır?

a. 4
b. 15
c. 20
d. 35
e. 40

10. Evren hacmi küçük olduğunda, örnekleme seçilen bir birimin diğerlerinin seçilme şansını etkilememesi için aşağıdaki seçim yöntemlerinden hangisi kullanılır?

a. Rassal iadeli
b. Rassal iadesiz
c. Sistematik
d. Keyfi ve sistematik birlikte
e. Keyfi

Kendimizi Sınayalım Yanıt Anahtarı

1. a	Yanıtınız yanlış ise "Tam Sayım ve Örnekleme" bölümünü yeniden gözden geçiriniz.
2. c	Yanıtınız yanlış ise "Olasılıklı Örnekleme ve Örnekleme Dağılımı" bölümünü yeniden gözden geçiriniz.
3. a	Yanıtınız yanlış ise "Oranların Örnekleme Dağılımı" bölümünü yeniden gözden geçiriniz.
4. c	Yanıtınız yanlış ise "Olasılıklı ve Olasılıklı Olmayan Örnekleme" bölümünü yeniden gözden geçiriniz.
5. e	Yanıtınız yanlış ise "Ortalamanın Örnekleme Dağılımının Özellikleri" bölümünü yeniden gözden geçiriniz.
6. a	Yanıtınız yanlış ise "Basit Doğrusal Regresyon Analizi" bölümünü yeniden gözden geçiriniz.
7. b	Yanıtınız yanlış ise "Oranların Örnekleme Dağılımının Özellikleri" bölümünü yeniden gözden geçiriniz.
8. c	Yanıtınız yanlış ise "Örnekleme Başvurma Nedenleri" bölümünü yeniden gözden geçiriniz.
9. c	Yanıtınız yanlış ise "Kota Örneklemesi Uygulaması" bölümünü yeniden gözden geçiriniz.
10. a	Yanıtınız yanlış ise "Keyfi ve Rassal Seçim Uygulamaları" bölümünü yeniden gözden geçiriniz.

Sıra Sizde Yanıt Anahtarı

Sıra Sizde 1

- Tam sayım yapılamaz. Çünkü kitapçıya gelen müşteriler evreni sonsuz evrendir.
- Örnekleme giren birimlerin rassal olarak seçilmesi ve örneklemin temsil örneklem olması durumunda örneklem istatistikleri evren parametreleri için bilgi niteliğindedir.
- Örneklemin temel amacı temsili örneklem oluşturmaktır.

Sıra Sizde 2

- Tam sayım yapmayı imkânsız kılan nedenler söz konusu olmadıkça ve kesin bilgi istediği sürece tam sayım yapılır.
- Her ne kadar tam sayım için gerekli koşullar mümkün ise de örneklemeye başvurulur. Çünkü, öğrencilerin bazılarına ulaşmak imkânsız olabilir, ulaşılsa bile bazıları bilgi vermeyebilir.

Sıra Sizde 3

- Yargısal örnekleme daha temsili örneklem oluşturur. Çünkü bu örnekleme uygulamasında örnekleme birim seçimini yapacak kişi incelenecek evrenle ilgili temsili örneklem oluşturma bakımından öncül bilgilere sahiptir.
- Kota örneklemesi seçilir.

Sıra Sizde 4

- Temel fark, olasılıklı olmayan örnekleme yöntemlerinde birim seçimi keyfi yapılırken bu örneklem uygulamaları ile oluşturulan örneklem için hesaplanan istatistiklerin evren parametreleri hakkında genelleme yapmak amacıyla kullanılmamasıdır. Olasılıklı örnekleme uygulandığında birim seçimi rassal yapılır ve örneklem istatistikleri parametreler hakkında genelleme amacıyla kullanılır.

Sıra Sizde 5

- N sonlu bir evrenden rassal olarak n hacimli bir örneklem değil de mümkün aynı hacimli bütün örneklemleri seçtiğimiz ve her örneklem için bir istatistik hesapladığımızda meydana gelen dağılıma örnekleme dağılımı denir.
- Bu teorem herhangi bir örnekleme uygulamasında hesaplanan n hacimli bir örneklem için hesaplanan istatistiğin dağılım şekli ve özellikleri ile ilgili ispatlanmış bilgileri verdiği için istatistiksel çıkarımlar bu teoreme dayandırılmaktadır.

Yararlanılan ve Başvurulabilecek Kaynaklar

Fink, A. (1995). **How To Sampling in Surveys,** London: Sage Publication.

Gürsakal, N. (1997). **Bilgisayar Uygulamalı İstatistik 1,** Bursa: Marmara Kitabevi.

Malhotra, N. K. (1996). **Marketing Research An Applied Orientation,** New Jersey: Prentice Hall International.

Neter, J., Wasserman, W., Whitmore, G. A. (1993). **Applied Statistics,** Boston: Simon and Schuster.

Serper, Ö., Aytaç, M. (2000). **Örnekleme,** Bursa: Ezgi Kitabevi.

Serper, Ö. (1986). **Uygulamalı İstatistik 2,** İstanbul: Filiz Kitabevi.

Tryfos, P. (1996). **Sampling Methods for Applied Research,** New York: John Wiley and Sons Inc.

Trochim, W. M. (2001). **Research Methods Knowledge Base,** Cornell University.

Hirsch, W. Z. (1963). Introduction to Modern Statistics, New York: The Macmillan Company.

7

Amaçlarımız

Bu üniteyi tamamladıktan sonra;

- İstatistiksel tahminleme tanımı yapabilecek ve bir evren parametresi için tahminleme sürecinin aşamalarını açıklayabilecek,
- Evren aritmetik ortalamasına ve evren oranına ilişkin nokta ve aralık tahminlemesi yapabilecek,
- Hipotez, istatistiksel hipotez ayırımını yapabilecek ve istatistiksel hipotezlerin test aşamalarını açıklayabilecek,
- Evren aritmetik ortalamasına ve evren oranına ilişkin hipotez testi uygulamalarını yorumlayabileceksiniz.

Anahtar Kavramlar

- Evren
- Parametre
- Örneklem
- Örneklem İstatistiği
- Tahminleme
- İstatistiksel Hipotez

İçindekiler

İstatistiksel Tahminleme ve Karar Alma

GİRİŞ

Tam sayım yapılamayıp, örneklemeye başvurulunca araştırmacılar iki problemle karşılaşırlar. Bunlardan birincisi istatistiksel **tahminleme**, ikincisi ise istatistiksel *karar vermedir.*

Bir rassal örneklemden hesaplanan istatistikler yardımıyla ilgili evrenin parametre değerinin araştırılmasına **tahminleme** denir.

Bu ünitenin birinci kısmında istatistiksel tahminleme ve istatistiksel tahminleme türleriyle ilgili teorik bilgiler açıklanacak, sonra evren aritmetik ortalaması ve evren oranına ilişkin tahminleme uygulamalarına yer verilecektir. İkinci kısımda ise istatistiksel hipotez testi ve test sürecinin aşamalarıyla ilgili teorik bilgiler açıklanacak, evren aritmetik ortalaması ve oranına ilişkin hipotez testi uygulamalarına yer verilecektir.

K İ T A P

Bu ünitenin izleyen kısımlarındaki konular için A. F. Yüzer, E. Şıklar, E. Ağaoğlu, H. Tatlıdil, A. Özmen, Editör: A. F. Yüzer (2011). İstatistik, Ünite 8 ve 9, Eskişehir Anadolu Üniversitesi Yayını İsimli kitaptan, editör ve ünite yazarlarının izni alınarak yararlanılmıştır.

İSTATİSTİKSEL TAHMİNLEME

Günlük yaşamın hemen hemen her alanında parametre tahminlemesiyle karşılaşılır. Bir araştırma sürecinin en önemli aşaması olan örneklemeyle tahminleme birbirinin ayrılmaz parçasıdır.

D İ K K A T

Örneklem, örnekleme, tamsayım kavramları iyice özümsenmelidir.

Bir tanım vermek gerekirse, *tahminleme, tanımlanan evrenden seçilen rassal örneklemden hesaplanan istatistikler yardımıyla bu evrenin uyduğu dağılımın parametre değerlerini araştırmaktır.*

Bir tahminleme sürecinde aşağıdaki aşamalar izlenir:

- Tanımlanan evrenden önceden belirlenen n hacimli rassal bir örneklem seçilir. Örneklemdeki birimler üzerinden veriler (gözlem değerleri) derlenir.
- Derlenen bu gözlem değerleri kullanılarak tahminlenecek parametre için bilgi üretecek istatistikler hesaplanır.
- Parametre için bilgi üreten istatistiğin örnekleme dağılımının özelliklerinden yararlanılarak parametre değeri tahminlenir.

Hem örneklem için hem de evren için bilgi üreten istatistiğe ilişkin formülasyona “tahminleyici”, örneklem gözlem değerinin bir tahminleyiciye uygulanmasıyla hesaplanan değere ise "tahmin" adı verilir. Tahminleyici, tahminin nasıl yapılacağını gösterir. Tahminse sayısal bir değerdir. Örneğin,

$$p = \frac{r}{n} = \frac{56}{80} = 0.70$$

formülü ile tanımlanan örneklemin aritmetik ortalaması, evren aritmetik ortalaması için bir tahminleyicidir. Bu formül yardımıyla hesaplanan değerse evren aritmetik ortalamasının bir tahminidir.

Parametreleri, istatistiklerden hareketle, bir sayı ve bu sayıyı kapsayan bir aralığın sınırlarıyla tahminlemek mümkündür. Birinci durumda *nokta tahminlemesi,* ikinci durumdaysa *aralık tahminlemesi* söz konusu olur. Nokta ve aralık tahminlemesi izleyen bölümlerde ayrıntılı olarak ele alınacaktır.

SIRA SİZDE 1

- **İstatistiksel tahminleme nedir?**
- **Tahmin ve tahminleyici kavramları arasındaki farkı açıklayınız.**

İSTATİSTİKSEL TAHMİNLEME TÜRLERİ

İstatistiksel tahminleme, **nokta** ve **aralık tahminlemesi** olarak sınıflandırılır.

Nokta Tahminlemesi

Bir rassal örneklemden hesaplanan istatistiğin değerini, ilgili evren parametre değerine eşit kabul eden tahminleme sürecine **nokta tahminlemesi** denilir.

Bir rassal örneklemden hesaplanan istatistiğinin değerini ilgili evren parametresi $\hat{\theta}$ değerine eşit kabul eden tahminleme sürecine, **"nokta tahminlemesi"** denir. Bu süreçte, örneklem için hesaplanan $\hat{\theta}$ istatistiğinin rassal bir değişken olduğu gerçeği benimsenir. $\hat{\theta}$ ilgili evren parametresinin bir nokta tahminleyicisi durumundadır.

Bu tahminleyici yardımıyla hesaplanan istatistik, bir sayıdır ve ilgili parametrenin nokta (tek değer) tahminidir.

Tahminleme sürecinde aşağıdaki aşamalar izlenir:

- Tanımlanan evrenden, belirlenen n birimlik (hacimli) bir basit rassal örneklem seçilir.
- Bu örneklemdeki birimler üzerinden ilgilenilen değişken itibariyle veriler derlenir.
- Bu veriler kullanılarak tahminlenecek parametre θ için tahminleyici olan istatistik hesaplanır.
- Son olarak bu istatistiğin değerinden yararlanarak θ parametresi için tahminleme yapılır.

Bu ünitenin izleyen kısmında tek evren ortalaması ve oranıyla ilgili nokta ve aralık tahminlemesine yer verilmektedir.

Evren Aritmetik Ortalaması μ'nün Nokta Tahminlemesi

n hacimli rassal bir örneklemin aritmetik ortalamasının ($\bar{X}$) hesaplanabilmesi için $x_1, x_2, \ldots\ldots x_n$ gözlem değerlerine gereksinim vardır. Örnekleme uygulamalarında bu gözlem değerleri sırasıyla $X_1, X_2, \ldots\ldots X_n$ rassal değişkenlerinin gerçekleşen değerleridir. Bu düşünceyle aynı zamanda evrenden seçilebilecek aynı hacimli farklı örneklemlerde farklı değerler alabilecek olan, $\bar{X}$ bir rassal değişkendir ve μ'nün nokta tahminleyicisi olarak isimlendirilir.

Bir rassal değişken olan $\bar{X}$'nin değerinin bu istatistiğin bilgi ürettiği μ'nün değerine eşit olan tahminlemeye μ'nün nokta tahminlemesi, hesaplanan $\bar{X}$ değerine μ'nün nokta tahmini denir.

Yapılan açıklamaları bir örnek üzerinde görelim:

ÖRNEK 1

Bir araştırmacı ilaç dağıtım şirketi çalışanlarının yıllık ortalama seyahat etme sürelerini tahmin etmek istemektedir. Bu amaçla rassal olarak 100 ilaç dağıtıcısı seçilmiş ve yıllık ortalama seyahat etme sürelerinin 386 saat ve standart sapmanın 72 saat olduğu belirlenmiştir. Yıllık seyahat etme süresini tahmin ediniz.

Çözüm:

$\bar{X}$ = 386 saat

s = 72 saat

n = 100 dağıtım şirketi çalışanı

μ = ?

100 birimlik rassal örneklemin aritmetik ortalaması, evren ortalamasının bir nokta tahmini olduğundan, aylık ortalama seyahat etme süresine ilişkin tahmin:

E ($\bar{X}$) = μ = 386 saat olacaktır.

Ancak bu nokta tahminlemesiyle gerçekte evren ortalaması μ'nün 386 saate yakın bir değer olduğu yorumu yapılabilir.

Evren Oranı Π'nin Nokta Tahminlemesi

Örnekleme uygulamalarının pek çoğunda evren oranı Π hakkında bilgi elde edilmesi istenir. Bu durumda Π için bilgi üreten istatistik örneklem oranı p olur.

Evren oranı Π'ye ilişkin nokta tahminlemesi bir rassal örnekleme planında oluşturulan n hacimli örneklem için, r bir binom rassal değişkeni olmak üzere, hesaplanan $p = \frac{r}{n}$ oranının değeri Π'ye eşit olan bir tahminleme sürecidir. Burada,

$p = \frac{r}{n}$ **eşitliği tahminleyici, hesaplanacak p değeri de tahmindir**.

ÖRNEK 2

Bir elektrikli ısıtıcı üreticisi yeni ürettikleri modelden memnun olan müşterilerinin oranını tahmin etmek istiyor. Bu amaçla yeni elektrikli ısıtıcıyı kullananlar arasından 80 kişi seçiliyor ve bunların 56'sının yeni ısıtıcı modelinden memnun oldukları belirleniyor. İstenen tahmini nokta tahmini olarak yapınız.

Çözüm:

n = 80 kişi

r = 56 kişi

$$p = \frac{r}{n} = \frac{56}{80} = 0.70$$

olur. Örneklem oranı E(p) = Π = 0.70 değeri evren oranı Π'nin yansız tahminidir. Bu, örneklemdeki 80 kişinin %70'inin yeni ısıtıcı modelinden memnun olduğu anlamına gelir.

Aralık Tahminlemesi

Bilinmeyen evren parametresini, istenen bir olasılıkla, simetrik bir aralıkta tahminleme sürecine, **aralık tahminlemesi** denilir.

Daha önce açıklandığı gibi, bir tahminleme sürecinde, küçük varyansa (ya da standart hataya) sahip olan tahminleyicinin tercih edilmesi, önemli bir kriterdir. Standart hatanın küçüklüğü tahminin güvenilirliğiyle ilgilidir. Güvenilir tahmin, tanımlanan evrenden seçilen aynı hacimli farklı örneklemlerde büyük ölçüde farklılık göstermeyen tahmindir. Nokta tahminlemesi *tahminin güvenilirliği hakkında bilgi veremediği için* sınırlı bir tahminlemedir.

Güvenilirliği somut bir şekilde ortaya koymak için güven aralığı kavramı geliştirilmiştir. Bu kavram, tahminlenecek parametre değerini kapsayacak alanı belirleyen bir çift sınır değeri kullanmaktadır. Güven sınırları adı verilen bu sınırlardan biri alt sınır (A), diğeri üst sınır (Ü) olarak tanımlanır.

Bir parametrenin örneklem istatistiğine göre, örneklemenin planlama aşamasında belirlenen bir olasılığa (güven düzeyine) göre simetrik bir aralıkta belirlenmesi çalışmasına, **aralık tahminlemesi** denir.

Bu tanıma göre θ parametresinin aralık tahminlemesinin gösterimi genel olarak;

$A \leq \theta \leq Ü$ şeklinde yapılır.

Örneklem istatistiklerinin değerleri ve standart hataları örneklemden örnekleme değiştiğinden, güven aralığının sınır değerleri değişir, güven aralığı genişler ya da daralır. Bu nedenle bazı güven aralıkları parametre değerini kapsar, bazıları kapsamaz. Bu sebeple tahminleme sürecinde alt ve üst sınırlar birer rassal değişken, bu sınırların belirlediği aralık da rassal aralık niteliğindedir. Aralık tahminlemesi sürecinde, araştırmacı tarafından önceden belirlenen olasılık düzeyi ya da **güven düzeyi** (G.D.) doğru aralık tahminlemesinin yapıldığı parametre değerini kapsayan güven aralığının tahminlendiği olasılığı ifade eder ve 1-α ile gösterilir.

Güven aralığının tahmininde parametre değerini kapsayan olasılığa **güven düzeyi** denilir.

G.D.= 1-α değeri büyük seçilirse tahminlenen aralığın θ'yı kapsayan bir aralık olma olasılığı artmış, fakat tahminlerin güvenilirliği azalmış olur.

Bir aralık tahminlemesi sürecinde, güven sınırları A ve Ü'nin tahminlenebilmesi için şu adımlar izlenir.

- Güven düzeyi G.D.=1-α belirlenir. Uygulamada genellikle güven düzeyi G.D.=1-α, %95 ya da % 99 olarak seçilmektedir. Güven düzeyi tanımlanan evrenden, n hacimli, mümkün pek çok farklı örneklemler çekilirse bu, örneklemler için hesaplanan güven aralıklarının, parametreyi kapsama olasılığını gösterir. Bu nedenle, parametreyi kapsayan güven aralıkları, doğru güven aralığı olarak isimlendirilir. Burada a anlamlılık düzeyini gösterir. Güven düzeyi örneğin % 95 seçildiğinde, anlamlılık düzeyi $\alpha = 1 - 0.95 = 0.05$ olur.
- Hacmi n olan bir basit rassal örneklem seçilir. Tahminlenecek parametre için, bilgi üretecek $\hat{\theta}$ istatistiği hesaplanır.
- Bu istatistiğin dağılımıyla ilgili bilgilerden yararlanarak güven aralığı oluşturulur.

Hesaplanan istatistiğinin dağılımı normal ise güven sınırlarının belirlenebilmesi için aşağıdaki eşitliklerden yararlanılır.

Alt Sınır,

$$A = \hat{\theta} - Z_{\frac{a}{2}} \cdot s_{\hat{\theta}}$$

ve

$$Ü = \hat{\theta} + z_{\frac{a}{2}} \cdot s_{\hat{\theta}}$$

Bu eşitliklerde:

$\hat{\theta}$: Örneklem istatistiğinin değerini

$z_{\frac{\alpha}{2}}$: Belirlenen güven düzeyi için, standart normal dağılım tablo değerini,

$\sigma_{\hat{\theta}}$: σ bilindiğinde tahminin standart hatasını gösterir.

$s_{\hat{\theta}}$: σ bilinmediğinde (s yerine onun yansız tahmini s kullanıldığında) standart hata tahminini gösterir.

Belirlenen güven düzeyi için ve $\hat{\theta}$ istatistiğinin örnekleme dağılımının normal olduğu varsayımı altında θ parametresi için aralık tahminlemesinin sınırları

$$\hat{\theta} - z_{\frac{\alpha}{2}} .\sigma_{\hat{\theta}} < \theta < \hat{\theta} + z_{\frac{\alpha}{2}} .\sigma_{\hat{\theta}} \text{ (σ bilindiğinde)}$$

ve

$$\hat{\theta} - z_{\frac{\alpha}{2}} .s_{\hat{\theta}} < \theta < \hat{\theta} + z_{\frac{\alpha}{2}} .s_{\hat{\theta}} \text{ (σ bilinmediğinde)}$$

şeklinde ifade edilir.

Aralık tahminlemesinde güven aralığının mümkün olduğu ölçüde dar tutulması arzu edilir. Çünkü dar aralığın sınırları parametre değerine daha yakındır. Bu güven düzeyine ve örneklem hacmine bağlıdır. Güven düzeyi % 99'dan %95'e düştüğünde daha dar güven aralığı elde edilir. Belirlenen örneklem hacmi için hesaplanan standart hatanın küçük olması durumunda da güven aralığı daralır.

Evren Aritmetik Ortalamasının Aralık Tahminlemesi

Evren aritmetik ortalaması μ için 1-α güven sınırlarının ya da güven aralığının belirlenmesi işlemlerine μ'nün aralık tahminlemesi denir. Bu tahminleme sürecinde şu aşamalar izlenir:

- G.D.=1-α belirlenir.
- Hacmi n olan bir rassal örneklem seçilir.
- Bu örneklem için $\bar{X}$ ve s istatistikleri hesaplanır.
- $\bar{X}$'nın standart hatası $s_{\bar{x}}$ hesaplanır ya da $s_{\bar{x}}$ tahminleyicisi yardımıyla tahmin edilir.
- $\bar{X}$'nın dağılım şekliyle ilgili bilgilerden yararlanarak.

$\bar{X}_A < \mu < \bar{X}_Ü$ güven aralığı tahminlenir. Burada, $\bar{X}_A$ ve $\bar{X}_Ü$ aralığın sırasıyla alt ve üst sınır değerlerini gösterirler.

Örneklem hacmi n, $\bar{X}$'nin dağılım şeklini belirleyen önemli bir belirleyici olduğu için μ'nün aralık tahminlemesiyle ilgili açıklamalar örneklem hacminin yeterli büyüklükte olup olmaması durumuna göre iki alt başlıkta açıklanmıştır.

Büyük Örneklemlerde μ'nün Aralık Tahminlemesi

Bilindiği gibi büyük örneklem hacmi için yeterli büyüklük, n ≥ 30 birim kabul edilmektedir.

Rassal örneklem hacmi yeterli büyüklükteyse, **evren dağılım şekli ne olursa olsun,** $\bar{X}$'nın örnekleme dağılımı, ortalaması μ ve varyansı $\frac{\sigma^2}{n}$ olan normal dağılıma uyar. $\bar{X}$ rassal değişkeninin dağılımı normal dağılım özelliğine sahip olduğundan, z standart rassal değişkeni,

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

ile dönüştürülür.

İstatistiksel yorumlama amacıyla yapılacak çözümlemelerde büyük kolaylıklar sağlayan bu dönüştürmeden yararlanarak birer rassal değişken olan $\bar{X}_A$ ve $\bar{X}_Ü$ güven sınırları

$$z_A = \frac{\bar{X}_A - \mu}{\sigma_{\bar{X}}} \qquad z_Ü = \frac{\bar{X}_Ü - \mu}{\sigma_{\bar{X}}}$$

eşitlikleri yardımıyla hesaplanabilir. Burada $z_A = -z_{\frac{a}{2}}$ ve $z_A = +z_{\frac{a}{2}}$ olmak üzere belirlenecek güven aralığının alt sınır değeri,

$$\bar{X}_A = \mu - z_{\frac{a}{2}} \sigma_{\bar{x}}$$

ve üst sınır değeri de

$$\bar{X}_Ü = \mu + \sigma_{\bar{x}}$$

olur.

Daha önce de açıklandığı gibi $\bar{X}$, μ'nün yansız tahminleyicisi olduğuna göre, bilinmeyen evren parametresi μ için güven aralığı

$$\bar{X} - z_{\frac{a}{2}} \sigma_{\bar{x}} < \mu < \bar{X} + z_{\frac{a}{2}} \sigma_{\bar{x}}$$

şeklinde oluşturulur.

Tablo 7.1
Bazı (1-a) değerleri için $z_{\frac{a}{2}}$ tablo değerleri.

1-α	$Z_{\alpha/2}$	1-α	$Z_{\alpha/2}$
0.99	2.576	0.80	1.282
0.95	1.960	0.60	0.842
0.90	1.645	0.50	0.674

Yukarıdaki formüllerde z değeri, araştırmacı tarafından başlangıçta belirlenen 1-α güven düzeyine bağlı olarak standart normal eğri alanları tablosundan bulunur. Bu tablo kitabımızın sonunda yer almaktadır. Bu tablodan alınan örnek bazı 1-α güven düzeyleri için z değerleri aşağıdaki tabloda verilmiştir:

$\bar{X}$'nın standart hatası, $\sigma_{\bar{x}}$ örnekleme oranı n/N < 0.05 ise ve örneklemin seçiminde iadeli seçim uygulanırsa ya da tanımlanan evren sonsuz hacimliyse

$$s_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

formülüne göre, buna karşılık örneklemin seçiminde iadesiz seçim uygulanıyor ya da örnekleme oranı n/N ≥ 0.05 olduğunda, $\sigma_{\overline{x}}$, yukarıdaki eşitliğe $\sqrt{\frac{N-n}{N-1}}$ çarpanı ilave edilerek hesaplanır.

Evren aritmetik ortalaması μ'nün aralık tahminlemesiyle ilgili yukarıda yapılan açıklamalar evren standart sapması σ'nın bilindiği durum için geçerlidir. Ne var ki gerçek yaşamda karşılaştığımız problem çözümlemelerinde σ genellikle bilinmez. Bu durumda standart hata, σ yerine onun tahmini olan örneklem standart sapması s kullanılarak tahminlenir ve $s_{\overline{x}}$ simgesiyle gösterilir. $\overline{X}$'nın örnekleme dağılımı n ≥ 30 birim olduğunda normal dağılıma uyduğu için, bu yaklaşım önemli bir hataya neden olmaz. Rassal örneklem hacmi yeterli büyüklükte olduğunda, 1-a güven düzeyi için, µ'nün güven aralığı

$$\overline{X} - z_{\frac{a}{2}} \,.\, s_{\overline{x}} < \mu < \overline{X} + z_{\frac{a}{2}} \,.\, s_{\overline{x}}$$

şeklinde bulunur.

Ortalamanın standart hatasının tahmini; örnekleme birim seçiminde iadeli seçim uygulanır, örnekleme oranı n/N <0.05 olur, tanımlanan evren sonsuz olursa

$$s_{\overline{x}} = \frac{s}{\sqrt{n}}$$

formülüne göre hesaplanır. Buna karşılık birim seçimi iadesiz yapılır ya da örnekleme oranı n/N > 0.05 olursa

$$s_{\overline{x}} = \frac{s}{\sqrt{n}} \,.\, \sqrt{\frac{N-n}{N-1}}$$ formülüyle hesaplanır.

Örnekler ve örnek çözümleri dikkatle incelenmeli, sorunlarla karşılaşılırsa kavramsal açıklamalara geri dönülmelidir.

DİKKAT

ÖRNEK 3

Konserve bezelye üreten bir fabrikadan, üretim sırasında 64 konserve kutusundan oluşan rassal bir örneklem seçilmiştir. Bu konserve kutularının ortalama ağırlığı 492 gr. ve standart sapma 12 gr. olarak belirlenmiştir. Üretilen konservelerin ortalama ağırlığını %95 güven düzeyinde tahmin ediniz.

Çözüm:

n = 64 konserve

$\overline{X}$ = 492 gr.

s = 12 gr.

G.D. = 1 - α = 0.95 dolayısıyla α = 0.05 olur.

Örnekte evrenin değişkenliği ve bölünme şekli hakkında bilgi bulunmamaktadır.

Örneklem hacmi n = 64 birim (n >30 birimdir, ayrıca evren hacmi sonsuzdur) olduğu için $\overline{X}$'nın örnekleme bölünmesi normaldir (Merkezi Limit Teoremi). Bu bilgilere göre $z_{\frac{a}{2}}$ = ± 1.96 alınır. Buna göre

$$s_{\bar{x}} = \frac{12}{\sqrt{64}} = \frac{12}{8} = 1.5 \text{ gr}$$

ve güven aralığı

$$\bar{X} - z_{\frac{a}{2}} \, . \, s_{\bar{x}} < \mu < \bar{X} + z_{\frac{a}{2}} \, . \, s_{\bar{x}}$$

$$492\text{-}1.96(1.5) < \mu < 492 + 1.96(1.5)$$

$$489.06 < \mu < 494.94$$

olarak hesaplanır.

Bu bilginin anlamı, üretilen konserve kutularının ortalama ağırlığı %95 güven ile 489.06 gr. ile 494.94 gr. arasındadır. Başka bir ifadeyle 489.06 - 494.94 gr. aralığının üretilen konservelerin gerçek ortalama ağırlığını kapsayacağına %95 güvenilebilir.

Küçük Örneklemlerde μ'nün Aralık Tahminlemesi

Örneklem hacmi küçük (n < 30 birim) olduğu zaman, örneklem ortalamalarının standart değerleri yukarıda açıklandığı gibi normal dağılıma sahip olmaz. Bu durumda μ için aralık tahminlemesi, tanımlanan evrenin normal dağılıma sahip olup olmadığının bilinmesine bağlıdır.

Normal dağılan bir evrenden, rassal olarak seçilebilecek birbirinden farklı n < 30 birim hacimli, mümkün bütün örneklemlerin seçildiğini, her örneklem için $\bar{X}$ ve onların $(\bar{X}\text{-}\mu)$ / $s_{\bar{x}}$ standart değerlerinin hesaplandığını düşünelim. Değer aralığı $-\infty < (\bar{X} - \mu) / s_{\bar{x}} < +\infty$ olan istatistiği s.d = n - 1 serbestlik derecesinde t dağılımı adı verilen sürekli bir dağılıma uyar ve

$$t = \frac{\bar{X} - \mu}{s_{\bar{X}}}$$

şeklinde gösterilir.

t dağılımı ortalaması sıfır olan tek modlu ve simetrik bir dağılımdır. Bu dağılımın şekli normal dağılımın şekline benzer fakat değişkenliği daha fazladır. Bilindiği gibi örneklenen evrenin dağılımı normal olduğunda, $(\bar{X} - \mu) / \sigma_{\bar{x}}$ istatistiği standart normal dağılıma sahip olur. Evren değişkenliği σ'nın bilinmediği durumlarda $\sigma_{\bar{x}}$'nin yerine $s_{\bar{x}}$'nin ikame edilmesi; bu standart değişkenin üzerinde ilave bir değişkenliği tanımlar ve t istatistiği bu değişkenliği doğru bir şekilde dikkate alır. Örneklem hacmi artarken serbestlik derecesi s.d = n-1 büyür, $s_{\bar{x}}$'nin kullanılması nedeniyle ortaya çıkan değişkenlik küçülür ve t dağılımı standart normal dağılıma yaklaşır.

Yukarıda verilen bilgilerin ışığında ilgilenilen evren normal dağılıma sahip olduğunda, μ için güven aralığı büyük örneklerdekine benzer şekilde aşağıdaki gibi ifade edilir. Tek fark z yerine t istatistiğinin kullanılmasıdır.

Tanımlanan evren normal dağılıyorsa, μ için 1-α güven aralığı

$$\bar{X}\text{-t } s_{\bar{x}} < \mu < \bar{X} + \text{t } s_{\bar{x}}$$

şeklinde belirlenir.

Burada;

t: 1-α güven düzeyinde ve n-1 serbestlik derecesine bağlı olarak t değerleri tablosundan bulunur. t tablosu örneği ekte verilmiştir.

$s_{\overline{x}}$: Ortalamanın standart hata tahminidir ve

$$s_{\overline{x}} = \frac{s}{\sqrt{n-1}}$$

Tahminleyiciyle tahminlenir.

ÖRNEK 4

Bir cep telefonu üreticisi, üretilen cep telefonlarındaki bataryaların bir kez şarj edildikten sonra ortalama ne kadar süre kullanıldığını tahmin etmek istemektedir. Bu amaçla rassal olarak 17 telefon bataryası seçilmiş ve bunların ortalama 135 saat kullanılabildiği ve standart sapmanın 28 saat olduğu belirlenmiştir. %99 güven düzeyi için istenilen tahminlemeyi yapınız.

Çözüm:

n = 17 batarya

$\bar{X}$ = 135 saat

s = 28 saat

G.D. = 0.99

α = 0.01

Üretilen cep telefonu bataryalarının ortalama kullanım süresinin normal bölünmeye sahip olabileceğini, çok fazla çarpık olamayacağını biliyoruz. Bu nedenle

$$\bar{X} - t\, s_{\overline{x}} < \mu < \bar{X} + t\, s_{\overline{x}}$$

güven aralığı uygulanabilir. Belirlenen G.D = 1 -α=0.99 ve serbestlik derecesi

s.d = n - 1 = 17 - 1 = 16 dır.

% 99 güven düzeyi (v = 0.01) ve s.d. = 16 serbestlik derecesi için t Tablo Değeri t0.01;16 = 2.921 elde edilir.

Ortalamanın standart hatası

$$s_{\overline{x}} = \frac{s}{\sqrt{n-1}} = \frac{28}{\sqrt{16}} = 7 \text{ saat}$$

bulunur.

Örneklem ortalaması, $s_{\overline{x}}$ tahmininden yararlanılarak %99 güven aralığı,

$$135 - 2.921\ (7) < \mu < 135 + 2.921\ (7)$$

$$114.553 < \mu < 155.447$$

biçiminde oluşturulur.

Üretilen cep telefonlarındaki bataryaların bir kez şarj edildikten sonra ortalama kullanım süresi % 99 güvenle 114.553 saat ile 155.447 saat arasında bir değerdir.

Evren dağılımı normal değilse verilerin bir matematiksel dönüşümle normale yaklaştırabileceği düşüncesinden hareketle μ'nün güven aralığının tahminlenmesinde, (yukarıda açıklanan uygulama) dönüştürülmüş verilere uygulanır.

Örneğin logaritmik dönüşüm bu amaçla sıkça kullanılır. Çünkü dönüştürülmüş logaritmik veriler, dönüştürülmemiş verilere göre daha az çarpıktır.

Evren Oranının Aralık Tahminlemesi

Bu parametre için aralık tahmini prosedürü, bu ünitede örneklem hacmi büyük (n ≥ 30 birim) ya da *n/N* £ 0.05 olduğu durum için uygulanmıştır. Daha önce de açıklandığı gibi, büyük örneklemler için örneklem oranı, p'nin örnekleme dağılımı, yaklaşık normal dağılım gösterir. Aynı zamanda bu dağılımın ortalaması p, standart hatası

$$s_p = \sqrt{\frac{p(1-p)}{n}}$$

Buradan, rassal örneklem hacmi yeterli büyüklükteyse, örneklem oranlarının standart değerlerinin,

$$z = \frac{p - \Pi}{s_p}$$

standart normal dağılıma sahip olduğu söylenebilir ve evren ortalaması aralık tahmininde olduğu gibi 1-α güven düzeyinde evren oranı Π'nin aralık tahmini evren aritmetik ortalamasının tahminlenmesinde izlenen aşamalara benzer şekilde,

$$p - z_{\frac{a}{2}} \cdot s_p < \Pi < p + z_{\frac{a}{2}} \cdot s_p$$

hesaplanır.

ÖRNEK 5

Bir bölgede yaşayan kişiler arasından A gazetesi okuru olanların oranı belirlenmek isteniyor. Bu amaçla rassal olarak 200 kişiden oluşan bir örneklem seçiliyor ve seçilen 200 kişi arasından 58 kişinin A gazetesi okuru olduğu belirleniyor. Bu bölgedeki A gazetesi okurlarının oranını %95 güven düzeyinde tahmin ediniz

Çözüm:

n = 200 kişi

r = 58 kişi

1 - α = 0.95

α = 0.05

Π = ?

$Z_{\alpha/2} = \pm 1.96$

$$p = \frac{r}{n} = \frac{58}{200} = 0.29$$

1 - p = 0,71

$$s_p = \sqrt{\frac{(0.29)(0.71)}{200}} = 0.032$$

$$0.29 - 1.96(0.032) < \Pi < 0.29 + 1.96(0.032)$$

$$0.29 - 0.06272 < \Pi < 0.29 + 0.06272$$

$$0.22728 < \Pi < 0.35272$$

olarak elde edilir.

Bu bölgedeki A gazetesi okurlarının oranı %95 güvenle, %22.73 ile %35.27 arasında bir değer olarak belirlenir.

İSTATİSTİKSEL HİPOTEZ VE İSTATİSTİKSEL HİPOTEZ TESTİ

Hipotez, karşılaşılan özel duruma ilişkin bir önermedir. **İstatistiksel hipotez** ise bir dağılımın evren parametresine ilişkin bir önermedir.

Genel olarak **hipotez,** karşılaşılan özel duruma ilişkin bir önermedir. İstatistiksel **hipotez,** bir araştırmada ilgilenilen bir ya da daha fazla parametrenin değeri hakkında ileri sürülen ve doğruluğu, geçerliliği bu parametre hakkında bilgi üreten istatistikten ve bu istatistiğin örnekleme dağılımıyla ilgili bilgilerden yararlanarak araştırılabilen önermelerdir. İstatistiksel hipotezler bir ya da daha fazla evren parametre değeriyle ilgili olabilirler. İstatistiksel hipotezleri diğer hipotezlerden ayıran özellik, bu hipotezlerin bir frekans dağılımına ait olmasıdır. Bazı istatistiksel hipotez örnekleri aşağıda verilmiştir.

Örnekler:

1. Günlük ortalama üretimi 750 kg. olan bir ilaç fabrikasında, uygulanan yeni üretim tekniği, ortalama üretimi arttırmıştır.
2. Bir üretim sürecinde üretilen tereyağı paketleri ortalama 500 gr. ağırlığındadır.
3. Bir yerleşim yerinde ikamet eden ailelerin %10'u alışverişlerini süper marketlerden yapmaktadır.

Evren parametreleri hakkındaki hipotezler (önermeler) parametre değerleri hakkında, daha önceden bilinen, belirlenen standart bir değer ya da varsayımsal bir değer olabilir. Birinci örnekte, birinci ilaç üretim yönteminin ortalama üretim düzeyi olan 750 kg. bilinen bir değerdir. İkinci örnekteki tereyağı paketlerinin planlanan ağırlığı olan 500 gr. standart bir değerdir. Son örnekteki süper marketlerden alışveriş yapan ailelerin oranı olan %10 varsayımsal bir değerdir ya da daha önce aynı konuda yapılan araştırmalarda elde edilen bir bilgidir.

Bir istatistiksel hipotez, doğru ya da yanlış olabilir. Çünkü bu bir önermedir. Gerçeği öğrenebilmek için evren parametresi θ'nın değerini hesaplamak gerekir. Bu da tamsayım yapmayı gerektirir. Ancak örnekleme yapmayı gerektiren nedenlerden dolayı bu, her zaman mümkün değildir. Bu durumda istatistiksel hipotezlerin geçerliliği ya da doğruluğu konusunda karar verebilmek için bu hipotezlerin, tanımlanan evrenden seçilen örneklemin gözlem değerinden hesaplanan örneklem istatistiğinden ($\hat{\theta}$'dan), bu istatistiğin ($\hat{\theta}$'nın) örnekleme dağılımının özelliklerinden yararlanarak test edilmesi gerekir. **İstatistiksel hipotez testi, örneklem istatistiklerini kullanarak bir hipotezin doğru olup olmadığını ortaya koymaya yönelik olarak yapılan çalışmadır.** Yorumsal istatistikte hipotez testi, örneklem gözlem değerlerinden yararlanarak bu örneklemin seçildiği evrenin durumu hakkında yorum yapmaktır.

Daha önce de belirtildiği gibi, evrenden rassal örneklem alınmış olsa bile, örneklem gözlemlerinden hesaplanan bir istatistiğin, bu istatistiğin bilgi ürettiği parametre hakkında ileri sürülen değere ($\hat{\theta}_0$) eşit olması beklenemez. Yani örneklem istatistikleri aynı hacimli farklı örneklemlerde farklı değerler alabildiği için

$\hat{\theta} - \theta_0 > 0$, $\hat{\theta} - \theta_0 = 0$ ya da $\hat{\theta} - \theta_0 < 0$ gibi farklar olabilir. Bu nedenle, istatistiksel test sonucu verilecek kararın, güvenilir olduğu konusunda, kesin karar verilemez. Fakat, olasılık kuramından yararlanarak bir istatistiksel hipotezin yapılacak bir testle ne derece güvenle (ya da ne derece hatayla) kabul ya da reddedileceğini belirlemek olanaklı olmaktadır. Burada önemli olan $\hat{\theta} - \theta_0$ farkının istatistiksel olarak anlamlı olup olmadığını belirlemektir. Başka bir anlatımla, farkların gerçek değişmeyi mi açıkladığı, yoksa rassal olarak mı meydana geldiğini belirlemektir. Anlamlı farklılık belirlenmişse hipotez, belirli bir hata payıyla reddedilir. Tersi durumda kabul edilir.

SIRA SİZDE 2

- **İstatistiksel hipotez nedir?**
- **İstatistiksel hipotez testinin konusu nedir?**

HİPOTEZ TESTİ TÜRLERİ

Hipotez testleri, ilgilenilen değişkenin ölçülmesinde benimsenen ölçeğe bağlı olarak, parametrik hipotez testleri ve parametrik olmayan hipotez testleri şeklinde sınıflandırılırlar. Parametrik testler değişkenlerinin ölçülmesinde eşit aralıklı ya da oranlı ölçeğin kullanıldığı hipotez testleridir. Çünkü bu iki ölçekte de elde edilen veriler üzerinden aritmetik işlemler yapmak mümkündür. Parametrik hipotez testlerinde, hipotezde bilinen bir olasılık fonksiyonundaki θ parametresinin önceden bilinen θ_0 değerine eşit ya da bundan büyük, küçük ya da farklı olduğu ileri sürülebilir.

Parametrik testler örneklem sayısının tek ya da iki oluşuna ve iki örneklemin varlığında, bu örneklemlerin bağımsız ya da bağımlı oluşuna bağlı olarak sınıflandırılırlar. En önemli parametrik testler z ve t testleridir. Bu ünitede tek evren ya da tek örneklem ortalamasına ilişkin z ve t testleriyle tek evren ya da tek örneklem oranına ilişkin z testi ayrıntılı olarak incelenmiştir.

Bu kitapta parametrik olmayan testler inceleme konusu yapılmamıştır.

KİTAP

Parametrik testlerle ilgili ihtiyaç duyulan bilgi için Özer, Serper'in Uygulamalı İstatistik II (Bursa: 5. Baskı, Ezgi Kitabevi, 2004) adlı kitaptan yararlanabilirsiniz.

HİPOTEZ TESTİ SÜRECİNİN ADIMLARI

Evren parametre değerleri hakkında ileri sürülen iddiaların test edilmesinde aşağıdaki adımlar izlenir:

Sıfır hipotezi (H_0), ilgili evren parametresinin bilinen değerinde herhangi bir farklılığın beklenmediğinin ifade edildiği hipotezdir.

Karşıt hipotez (H_1), ilgili evren parametresinin bilinen değerinde istatistiksel olarak anlamlı farkların beklendiğini ifade eden hipotezdir.

Hipotezlerin İfade Edilmesi

İstatistiksel hipotezlerin testinde, iki hipotez söz konusudur. Bunlar, **"sıfır hipotezi"** ve **"karşıt hipotez"** olarak isimlendirilirler. Bu aşamada, sıfır hipotezinin ve karşıt hipotezin nasıl ifade edileceğine karar verilir.

Sıfır hipotezi H_0 simgesiyle gösterilir ve hangi hipotezin test edileceğini ifade eder. H_0 hipotezinde test süreci tamamlanıncaya kadar $\theta = \theta_0$ olarak kabul edilir. $\hat{\theta}$ ile θ parametrelerinin değerleri arasındaki farkın örnekleme hatasından kaynaklanabileceği, bu iki değer arasında gerçekte anlamlı bir farklılık olmadığı ifade edilir. Sıfır hipotezi; parametrenin önceden belirlenmiş, bilinen değerinde, hiçbir farklılığın (etkinin) beklenmediğinin ifade edildiği hipotezdir.

Bu açıklamaların ışığında H_0 hipotezi,

$$H_0: \theta = \theta_0$$

şeklinde yazılır.

H_0 hipotezinin test edilebilmesinde, bu hipotezden farklı bir hipotezin de ifade edilmesi gerekir. H_1 simgesiyle gösterilen bu hipoteze "karşıt hipotez" adı verilir. H_1 hipotezi H_0 hipotezinin belirli bir olasılıkla reddedilmesi durumunda kabul edilen ve genellikle araştırma hipotezinin ifade edildiği hipotezdir. Karşıt hipotez, parametrenin önceden belirlenmiş, bilinen değerinde anlamlı farklılığın ya da etkinin beklendiğinin ifade edildiği hipotezdir. Bir başka ifadeyle sıfır hipotezini çürüten bir hipotezdir. Bu hipotez araştırmanın amacına bağlı olarak, aşağıdaki üç farklı şekilden biriyle ifade edilmiş olur:

$H_1: \theta \neq \theta_0$

$H_1: \theta > \theta_0$

$H_1: \theta < \theta_0$

$H_1: \theta \neq \theta_0$ ifadesi, evren parametresinin belirlenen (ya da bilinen) θ_0 değerinden, her iki yöndeki (hem küçük hem de büyük yöndeki) anlamlı farklılıkların, test sonucu verilecek kararı etkileyeceği anlamına gelir. İkinci ifade, verilecek kararın, evren parametre değerinde, sadece büyük yöndeki anlamlı sapmadan etkileneceği, son ifade ise sadece küçük yöndeki anlamlı farklılığın verilecek kararı etkileyeceği anlamına gelir.

Hipotez testlerinde H_1 hipotezi, örnekleme dağılımındaki testin yönünü ya da H_0 hipotezinin reddedileceği bölgenin yerini belirleyen hipotezdir. Red bölgesi, H_0 hipotezinin reddedilmesine (H_1 hipotezinin kabul edilmesine) neden olan örneklem istatistiği $\hat{\theta}$ (ya da test istatistiği) ile ilgili değerler aralığıdır. Kabul bölgesi ise, H_0 hipotezinin kabul edilmesine (H_1 hipotezinin reddedilmesine) neden olan örneklem istatistiği $\hat{\theta}$ (test istatistiği) ile ilgili değerler aralığıdır. Hipotez testleri, H_1 hipotezinin ifade ediliş şekline göre: "iki yönlü test", "tek yönlü üst kuyruk testi" ve "tek yönlü alt kuyruk testi" olarak isimlendirilirler. Bu testlere ilişkin hipotezlerin ifade edilmiş biçimi aşağıda verilmiştir.

İki Yönlü Testlerde Hipotezler:

$H_0: \theta = \theta_0$

$H_1: \theta \neq \theta_0$

Tek Yönlü Üst Kuyruk Testlerinde Hipotezler:

$H_0: \theta = \theta_0$

$H_1: \theta > \theta_0$

Tek Yönlü Alt Kuyruk Testlerinde Hipotezler:

$H_0: \theta = \theta_0$

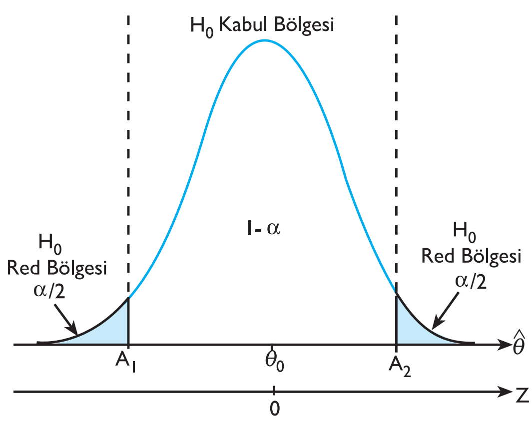
$H_1: \theta < \theta_0$

şeklinde belirlenir.

Yukarıdaki hipotez takımlarında kullanılan isimler, H_1 hipotezinde θ için verilen

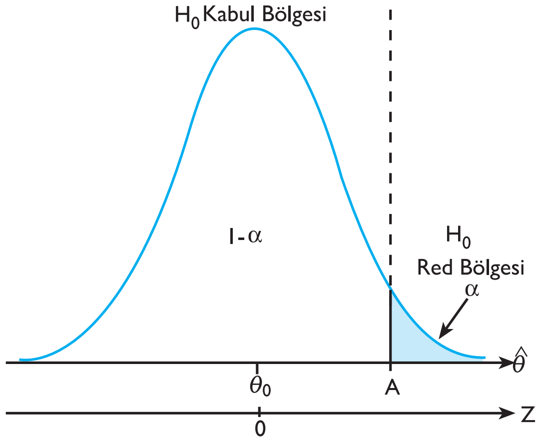
Şekil 7.1

İki Yönlü Testlerde Red Bölgeleri.



Şekil 7.2

Tek Yönlü Üst Kuyruk Testlerinde Red Bölgesi.



değer aralığını açıklamaktadır. Bu durum, örneklem istatistiği $\hat{\theta}$'nın normal dağılıma sahip olduğu kabul edilerek aşağıdaki şekillerle açıklanmıştır. Örneğin iki yönü hipotezlerde H_1 hipotezi Şekil 7.1'de görüldüğü gibi θ_0'ın her iki tarafındaki θ ile ilgili değerleri kapsamaktadır. Başka bir ifadeyle örneklem istatistiği $\hat{\theta}$'nın belirli bir A_1 değerinden küçük ya da belirli bir A_2 değerinden büyük olan değerleri H_1 hipotezi yönünde, H_0 hipotezinin red bölgesinde yer alan değerlerdir.

Tek yönlü üst kuyruk testlerinde, H_1 hipotezi, A'dan büyük olan θ_0 ile ilgili değerleri içerdiği için, bu isim verilmiştir. Tek yönlü üst kuyruk testlerinde, H_1 hipotezi, Şekil 7.2'de görüldüğü gibi $\hat{\theta}$'nın θ_0'dan büyük olmak üzere, belirli bir A değerinden büyük değerleri H_1 hipotezi yönünde, H_0 hipotezinin red bölgesinde yer almaktadır.

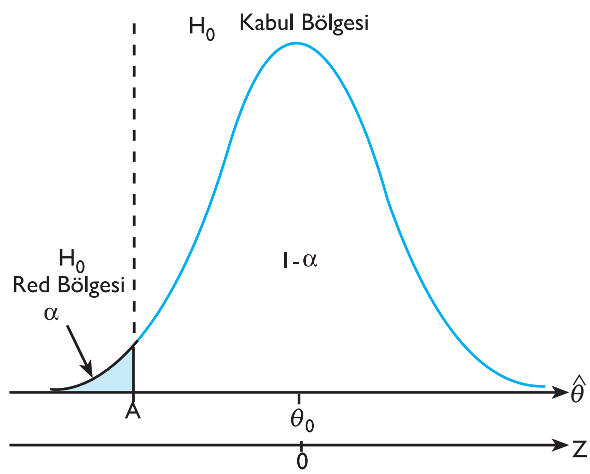
Tek yönlü alt kuyruk testlerindeyse *tek yönlü üst kuyruk testinin tam tersine,* (Şekil 7.3'te görüldüğü gibi) θ_0'ın solunda ve $\hat{\theta}$'nin A'dan küçük olan değerleri, H_0 hipotezinin red bölgesinde yer alan değerlerdir.

DİKKAT

Hipotez testlerinde kabul ya da red edilen hipotez H_0'dır.

Şekil 7.3

Tek Yönlü Alt Kuyruk Testlerinde Red Bölgesi.



Anlamlılık Düzeyinin Belirlenmesi

Bir istatistiksel hipotez testinde ya sıfır hipotezinin reddedilmesi ya da kabul edilmesi şeklinde karar verilir. Bu iki karar arasında seçim yaparken örneklem istatistiğinden yararlanıldığı için hatalı karar verme riski vardır. Çünkü aynı evrenden rassal olarak seçilen, aynı hacimli farklı örneklemler için hesaplanan istatistikler, örneklemden örnekleme değişen değerler aldığından, evren parametre değerinden farklılık göstermektedirler.

Hipotez testlerinde, sıfır hipotezinin yanlışlıkla reddedilmesi ya da kabul edilmesi sonucu işlenen hataya "yorumlama (çıkarsama) hatası" adı verilir. İki tür yorumlama hatası vardır: Bunlar, gerçekte doğru olan sıfır hipotezinin reddedilmesi durumunda işlenen hatayla gerçekte yanlış olan sıfır hipotezinin kabul edilmesi durumunda işlenen hatadır. Gerçekte *doğru olan sıfır hipotezinin reddedilmesi* durumunda işlenen hataya I. Tip hata ya da a tipi hata adı verilir. Araştırmalarda a tipi hata işlemenin maksimum olasılığına **"testin anlamlılık düzeyi"** denir. Anlamlılık düzeyinin belirlenmesi, doğru olan sıfır hipotezinin, örneklemden elde dilen bilgilere dayanarak reddedilmesi olasılığını belirleyen α'nın seçilmesidir. α anlamlılık düzeyi, araştırmacı tarafından, hipotezler ifade edilip veri derlemeye başlamadan önce seçilmelidir. Sosyal bilim araştırmalarında α için genellikle %5 veya %1 değerleri seçilmektedir. Yapılan bu seçimle birlikte, doğru olan H_0 hipotezinin reddedilme olasılığı, belirlenmiş olur. Bu olasılık örnekleme dağılımıyla ilişkilendirilerek kullanılır. Bu durumda, α anlamlılık düzeyi, doğru olan sıfır hipotezinin reddedilmesi olasılığına eşit olan, örnekleme dağılımındaki oransal alanı göstermiş olur. Örnekleme dağılımında, doğru olan sıfır hipotezinin, reddedilmesi olasılığına eşit olan oransal alana "red bölgesi" denir. Örnekleme dağılımının bu bölgesi, sıfır hipotezi doğru olduğunda, beklenmeyen örneklem istatistiği değerlerini temsil eder. Örnekleme dağılımında, red bölgesini tanımlamadan önce, örnekleme dağılımını tanımlamak gerekir. Örneklem istatistiğinin normal dağılımlı olması durumu için red ve kabul bölgeleri Şekil 1,2 ve 3'te gösterilmiştir. Şekillerdeki A, A_1 ve A_2 noktaları red bölgelerinin başlangıç noktalarıdır.

α tipi hata yapmanın maksimum olasılığına **testin anlam düzeyi** adı verilir.

Diğer taraftan, sıfır hipotezi gerçekte yanlış olabilir ve araştırmacı *yanlış olan bu hipotezi kabul ederse* yine hatalı karar vermiş olur; bu tür hataya II. Tip hata ya da β tipi hata denir. Bu türden hata yapmanın maksimum olasılığı da β ile gösterilir.

İstatistiksel uygulamalarda α tipi hatadan daha çok sakınılır ve genellikle sadece α tipi hata kontrol edilir.

Araştırmada H_0 hipotezinin doğru olduğuna inanan araştırmacı, α anlamlılık düzeyini çok küçük bir değer olarak seçebilir. Yanlış olan H_0 hipotezinin kabul edilmesi riskli ise büyük kayıplara neden oluyorsa, α olasılığı büyük tutulmalıdır.

H_0 doğruyken test sonucunda reddedilirse α **(I. tip) tipi hata**, H_0 doğru değilken test sonucunda kabul edilirse **b (II. tip) tipi hata** gerçekleşmiş olur.

Örneklem hacmi sabit olduğunda, α **tipi hata** işlemenin azalması (ya da artması), β **tipi hata** işleme olasılığının artmasına (ya da azalmasına) neden olur.

Verilerin Derlenmesi

Bir araştırma planında, hipotezlerin ifade edilmesiyle araştırmanın genel çerçevesi ortaya konur, problem ve değişkenler tanımlanmış olur. İfade edilen hipotezlerin test edilmesi için, gerekli uygun α anlamlılık düzeyi belirlendikten sonra, belirlenen evrenden, hangi hacimde bir örneklem seçileceği kararlaştırılır. Daha sonra da ilgili evrenden belirlenen hacimde rassal bir örneklem seçilerek veriler derlenir. Bu veriler kullanılarak, test edilecek parametre hakkında bilgi üreten örneklem istatistikleri hesaplanır.

Test İstatistiğinin Seçilmesi

Daha önce de belirtildiği gibi, evrenden rassal örneklem alınmış olsa bile, hesaplanan örneklem istatistiğinin evren parametresi hakkında, önceden bilinen, belirlenen değere esşit olması beklenmez. Bu durumda şu soru akla gelebilir: Örneklem istatistiğinin değeriyle bu istatistiğin bilgi ürettiği parametrenin sıfır hipotezinde ifade edilen değeri arasında nasıl bir farklılık vardır? Başka bir ifadeyle, sıfır hipotezi doğruysa anlamsız bir farklılığı veren bir örneklem istatistiği elde etmek mümkün müdür?

Bu sorunun yanıtlanabilmesi için sıfır hipotezinin test edilebilmesinde, örneklem istatistiğinin dağılımının bilinmesine ve uygun test istatistiğine gereksinim vardır.

Test istatistiği, örneklem istatistiğinin değeriyle evrenin, sıfır hipotezinde ifade edilen değeri arasındaki farkın, standartlaştırılmış değeri olarak tanımlanır. Başka bir ifadeyle test istatistiği, örneklem istatistiği $\hat{\theta}$ ile θ_0 arasındaki farkı standart hata birimiyle ifade eden ölçüdür. Test istatistiği örneklemin sıfır hipotezine ne kadar uyduğunu gösterir. Bu nedenle de test istatistiği test sonunda verilecek kararın dayandırıldığı bir örneklem istatistiğidir.

DİKKAT **Hipotez testlerinde, örneklem istatistiğinin dağılımının bilinmesi zorunludur.**

Bir örneklem istatistiğinin herhangi bir değeri, bu örneklem istatistiğinin *dağılımının* bir değeridir. Mümkün her örneklem istatistiğinin değeri için, bir test istatistiği değeri hesaplanabileceğine göre, test istatistiği örnekleme dağılımından söz edilebilir. Test istatistikleri genellikle normal dağılım (z dağılımı), t dağılımı v.b. gibi bilinen dağılımlara uyar.

Hipotez testi türleriyle ilgili bilgiler verilirken açıklandığı gibi, hipotez testleri için de uygun test istatistiğinin seçilmesiyle ilgilenilen değişkenlerin ölçülmesinde kullanılan ölçek türü, örneklem hacmi, örneklem sayısı (örneklem sayısı iki olduğunda örneklemlerin bağımsız ya da bağımlı olması) gibi hususların bilinmesi gerekir.

İzleyen bölümde, bazı parametrelere ilişkin hipotezlerin testinde z ve t test istatistiklerinin seçilme gerekçeleri ve uygulamalarına yer verilmektedir.

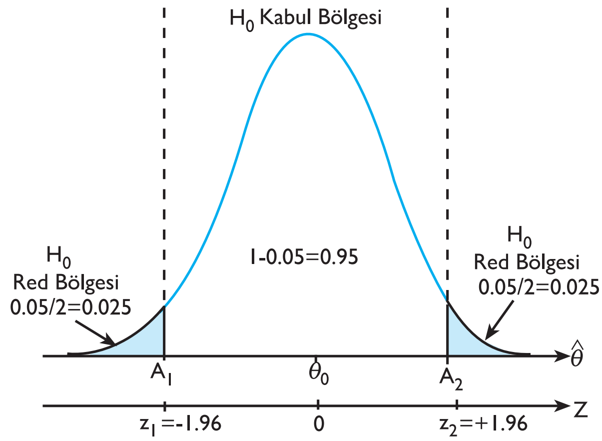
İstatistiksel Kararın Verilmesi

İstatistiksel karar vermekle eş anlamlı olan hipotez testi, aslında a anlamlılık düzeyinde H_0 hipotezinin kabul edilmesi ya da reddedilmesi kararıdır. Bu kararın verilebilmesi için bir ölçütün belirlenmesi gerekir. Test istatistiğinin, kritik değeri olarak isimlendirilen bu ölçüt, $\hat{\theta}$ istatistiğinin örnekleme dağılımında, red ve kabul bölgelerini birbirinden ayıran bir değerdir. Test istatistiğinin kritik değeri, bir örnekleme dağılımında, red bölgesinin başlama noktasını gösteren değerdir. Kritik değer, seçilen α anlamlılık düzeyinde, H_1 hipotezinin ifade ediliş biçimine ve örneklem istatistiğinin dağılım şekline bağlıdır. İzleyen açıklamalar $\hat{\theta}$ örneklem istatistiğinin ve bu istatistiğin standart değeri olan

$$z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

test istatistiğinin normal dağılıma sahip olduğu kabul edilerek yapılmıştır. Açıklamalarda $\alpha = 0.05$ seçilmiştir.

Eğer karşıt hipotez $H_1 : \theta \neq \theta_0$ şeklinde ifade edilmişse red bölgesi Şekil 7.4'te gösterildiği gibi $\hat{\theta}$ istatistiğine ilişkin normal dağılımın her iki ucunda simetrik olarak tanımlanmış olur ve her bölgenin alanı oransal olarak $\alpha/2= 0.05/2=0.025$'dir. Buna bağlı olarak kritik değerler $\hat{\theta}$ istatistiğine ilişkin normal dağılımın her iki kuyruğundaki, θ_0'a göre simetrik, A_1 ve A_2 değerleri olmaktadır.



Şekil 7.4

İki Yönlü Testlerde Red Bölgeleri ve Kritik Değerler.

Ancak; istatistiksel hipotez testlerinde, $\hat{\theta}$ örneklem istatistiği yerine, standartlaştırılmış değer kullanılmaktadır. Bu durumda kritik değerler A_1 ve A_2'nin

$$z_1 = \frac{A_1 - \theta_0}{\sigma_{\hat{\theta}}} \qquad \text{ve} \qquad z_2 = \frac{A_2 - \theta_0}{\sigma_{\hat{\theta}}}$$

standart değerleri olur. A_1 örneklem değeri θ_0'ın solunda (θ_0'dan küçük) olduğu için z_2 pozitif değer olarak ifade edilir. z_1 ya da z_2 kritik değerleri α = 0.05 anlamlılık düzeyi için Ek-1'de verilen Standart Normal Eğri Alanları Tablosundan yararlanılarak belirlenir. z tablo değeri (z_{tab}),

$z_{tab} = z_{0.5\ -\ 0.025} = z_{0.4750} = 1.96$ dır.

$z_{tab} = \pm$ 1.96 değerleri, standart normal dağılımda % 47.5'lik oransal alana karşı gelen örneklem istatistiğinin standart değerleridir. İki yönlü testte H_0 hipotezinin reddedilmesi için

$$z_{hes} = \left| \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}} \right| > z_{tab} = 1.96$$

koşulunun sağlanması gerekir. Tersi durumda H_0 hipotezi kabul edilir.

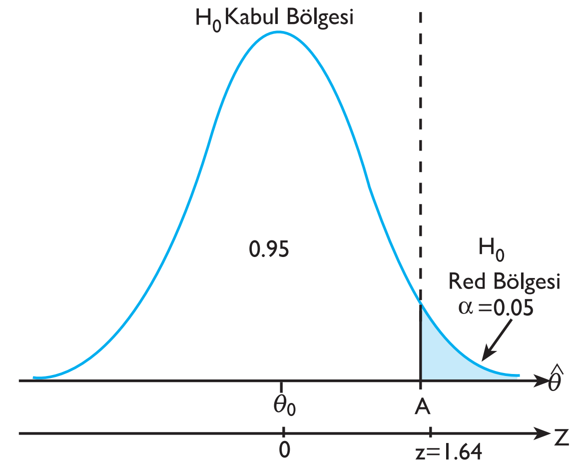
Eğer $H_1 : \theta > \theta_0$ ya da $H_1 : \theta < \theta_0$ şeklinde ifade edilmişse, H_0 hipotezinin red bölgesi, birinci durumda, istatistiğine ilişkin normal dağılımın üst kuyruğunda, ikinci durumda alt kuyruğunda tanımlanmışsa alan α =0.05 olur. Buna bağlı olarak kritik değerler sırasıyla (Şekil 7.5'te gösterildiği gibi) $\hat{\theta}$ istatistiğine ilişkin normal dağılımın üst kuyruğundaki A değeri ya da bunun standart değeri

$$z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$$

Şekil 7.6 gösterildiği gibi olur.

Şekil 7.5

Tek Yönlü Üst Kuyruk Testlerinde Red Bölgesi ve Kritik Değer.



Bu z değerleri birbirinin simetriğidir. Üst kuyruk testinde, z pozitif, alt kuyruk testinde, z negatif işaretlidir. z değerleri α =0.05 için Ek-1'de verilen Standart Normal Eğri alanları tablosundan yararlanılarak belirlenirler.

$z_{tab} = z_{0.50\ -\ 0.05} = z_{0.4500} = 1.64$ tür.

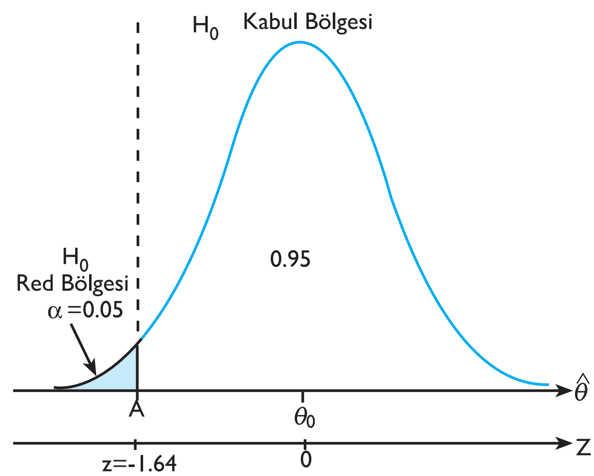
z_{tab}= 1.64 değeri, standart normal dağılımda, %45'lik oransal alana karşı gelen, örneklem istatistiğinin standart değeridir. Tek yönlü üst kuyruk testi sözkonusu olduğunda z_{tab}= +1.64, tek yönlü alt kuyruk sözkonusu olduğunda z_{tab}= -1.64 alınır. Bu bilgilere göre H_0 hipotezinin reddedilmesi için tek yönlü üst kuyruk testinde;

$$z_{hes} = \frac{\left(\hat{\theta} - \theta_0\right)}{\left(\sigma_{\hat{\theta}}\right)} > z_{tab} = 1.64$$

olmalıdır.

Şekil 7.6

Tek Yönlü Alt KuyrukTestlerinde Red Bölgesi ve Kritik Değer.



Tek yönlü alt kuyruk testinde H_0'ın reddedilebilmesi için

$$z_{hes} = \frac{(\hat{\theta} - \theta_0)}{(\sigma_{\hat{\theta}})} < z_{tab} = -1.64$$

koşulunun sağlanması gerekir. Tersi durumda H_0 hipotezi kabul edilir.

H_0 hipotezinin reddedilmesi yönündeki kararlar, örneklem değeri $\hat{\theta}$ ile evren parametresi arasında, α anlamlılık düzeyinde anlamlı bir farklılığın var olduğu, H_0 hipotezinin kabul edilmesi durumundaysa varolan farklılığın örnekleme hatasından kaynaklandığı anlamına gelir.

Probleme İlişkin Kararın Verilmesi

Hipotez testlerinde önemli olan, istatistiksel kararın, araştırma problemine ilişkin karara dönüştürülmesidir. Bu konu örnek problemler üzerinde açıklanmıştır.

SIRA SİZDE 3

- **Hangi hipotez test edilecek hipotezdir?**
- **I. Tip hata ne demektir?**

TEK EVREN PARAMETRESİYLE İLGİLİ HİPOTEZ TESTLERİ

Pek çok araştırmada, tek bir evrenin bir parametresinin değerine ilişkin, hipotezlerin ileri sürüldüğü görülmektedir. Aşağıdaki hipotezler tek evren parametresiyle ilgili hipotez testine örnek verilebilir:

A ürününün reklamını beğenenlerin oranı en az % 45'tir.

Günlük ortalama üretimi 100 Ton olan bir üretim sürecinde yapılan değişiklik, günlük ortalama üretim miktarını arttırmıştır.

Bu tür hipotezlerin testinde, tanımlanan bir evrenin ilgilenilen bir değişkenine ilişkin önceden belirlenen (ya da bilinen) bir parametre değerinin ($\hat{\theta}$'ın) değişmediği şeklindeki sıfır hipotezi test edilir. Böylece, verilen karara göre, karşıt hipotezde (araştırma hipotezinde) ileri sürülen iddianın kabul edilip edilmeyeceği ortaya çıkar.

Bu testlerde karar verilirken örneklem istatistiğinin değeriyle bu istatistiğin bilgi ürettiği parametrenin bilinen ya da belirlenen değeri θ_0 karşılaştırılır.

İzleyen bölümlerde, tek evren parametresiyle ilgili olarak, uygulamada sıkça karşılaşılan, evren ortalaması ve evren oranına ilişkin testler ele alınmıştır.

Evren Ortalamasına İlişkin Hipotez Testi

Bu testlerde, tanımlanan evren rassal olarak seçilen bir örneklem için hesaplanan $\bar{X}$ değeriyle, bu örneklemin seçildiği evrenin aritmetik ortalaması μ ile ilgili, önceden belirlenen (ya da bilinen) μ_0 gibi bir değer arasındaki farklılığın istatistiksel olarak anlamlı olup olmadığı araştırılır. Belirlenen farklılığın, sıfır hipotezini reddetmek için yeterli olup olmadığına karar verilir.

Evren ortalamasına ilişkin hipotez testleri uygulamada, sıkça kullanılan önemli parametrik testlerdir.

Bu hipotezlerin test edilmesine ilişkin açıklamalar, örneklem hacminin büyük olması ($n \geq 30$ birim) ve örneklem hacminin küçük olması ($n < 30$ birim) durumları için, alt başlıklar altında, aşağıdaki örnek problemler üzerinde ayrıntılı olarak ele alınmıştır:

Evren Ortalamasına İlişkin Büyük Örneklem Testi
Bu test türünde:
- Örneklem rassal olarak seçilir.
- Örneklem hacminin yeterli büyüklükte ($n \geq 30$) birimden oluştuğu ya da evren normal dağılımlı ve değişkenliğinin biliniyor olması gereklidir.
- $H_0 : \mu = \mu_0$ hipotezi, seçilecek bir a anlamlılık düzeyi için test edilir.

ÖRNEK 6

Margarin üreten bir fabrikada 250 g'lık paketler halinde üretim yapılması öngörülmektedir. Margarin paketlerinin ağırlığını kontrol amacıyla rassal olarak 100 paket seçilmiş ve seçilen bu paketler için ortalama ağırlık 244 g, standart sapma 18 g olarak saptanmıştır. a = 0.05 anlam düzeyinde, paketlerin ağırlığının öngörüldüğü gibi olduğu söylenebilir mi? Karar veriniz.

Çözüm:

1. Adım: Hipotezlerin ifade edilmesi

Margarin paketlerinin öngörülen ağırlığı 250 gr.dır. Bu nedenle sıfır hipotezi, üretilen margarin paketlerinin ağırlığının 250 gr. olduğu yönündedir. Her iki yöndeki **(hem küçük hem de büyük yöndeki)** anlamlı ağırlık farklılıkları bu iddiayı çürütecektir. Diğer bir ifadeyle bu anlamlı farklılıklar üretimin planlandığı gibi gerçekleşmediğini gösterecektir. Buna göre yapılacak test, iki yönlü test olacaktır.

Hipotezler:

$H_0 : \mu = 250$ gr.

$H_1 : \mu \neq 250$ gr.

şeklinde ifade edilecektir.

2. Adım: İstatistiksel test

Bu örnekte tanımlanan evren sonsuzdur. Evrenin dağılımı ve değişkenliği hakkında bilgi yoktur. Örneklem hacmi $n = 100$ paket olup $n \geq 30$ olduğundan, örneklem aritmetik ortalamasının örnekleme dağılımı, normal dağılımdır. Kullanılması gereken test istatistiği, örneklem aritmetik ortalamasının standart değeri olan z istatistiğidir. Bu nedenle z testi uygulanmalıdır.

$$z = \frac{\bar{X} - \mu_0}{s_{\bar{X}}}$$

Evren standart sapması bilinmediği için $\sigma_{\bar{x}}$'nın tahmini olan $s_{\bar{x}}$ kullanılmalıdır.

$$s_{\bar{X}} = \frac{s}{\sqrt{n}}$$

3. Adım: Anlamlılık düzeyinin belirlenmesi

Problemde doğru olan H_0 hipotezinin reddedilmesi olasılığı $\alpha = 0.05$ olarak belirlenmiştir. Red bölgeleri, ortalamanın örnekleme dağılımının her iki kuyruğunda tanımlandığı için, red bölgelerinin her birinin oransal büyüklüğü,

$\alpha /2 = 0.05 /2 = 0.025$ dir.

4. Adım: H_0'ın red bölgesinin belirlenmesi

H_1 hipotezi, testin red bölgesinin yönünü belirlemektedir. Bu testte red bölgesi, örnekleme dağılımının simetrik olması nedeniyle hem alt kuyrukta hem de üst kuyrukta tanımlanmıştır. Bu durum Şekil 7.7'de gösterilmiştir.

5. Adım: Test istatistiğinin hesaplanması

n = 100 paket

$\bar{X}$ = 244 gr.

s = 18 gr.

μ_0 = 250 gr.

α = 0.05

$$s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{18}{\sqrt{100}} = 1.8$$

ve

$$z = \frac{\bar{X} - \mu_0}{s_{\bar{X}}} = \frac{244 - 250}{1.8} = -3.33$$

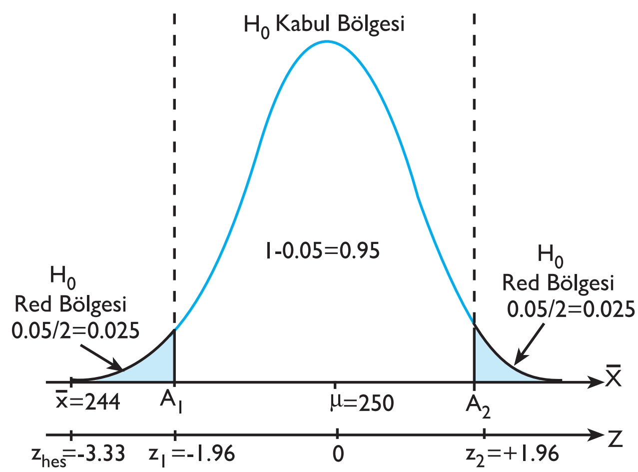
olarak hesaplanır.

$\bar{X}$'nın örnekleme dağılımı normal olduğu için bu dağılımı oluşturan değerlerin standart değerleri olan z test istatistiğinin örnekleme dağılımı, standart normaldir. İki yönlü bir test olduğu için red bölgesi dağılımın her iki kuyruğunda tanımlanmıştır ve oransal büyüklükleri $\alpha /2 = 0.05 /2 = 0.025$ dir. Buna göre, dağılımın sınır değerleri alt kuyruk bölgesi için z_1 = -1.96 ve üst kuyruk bölgesi için z_2 = +1.96 olacaktır.

Hesaplanan test istatistiği $z_{hes} = |3.33| > z_{tab} = 1.96$ olduğundan H_0 hipotezi reddedilir, dolayısıyla H_1 kabul edilir. Ayrıca bu karar Şekil 7.7'de z_{hes} = -3.33 standart değerinin Z_i eksenindeki *red bölgesinde yer aldığı* belirtilerek de açıklanabilir.

H_0 hipotezinin reddedilmesi, üretilen margarin paketlerinin ortalama ağırlığının 250 gr. olmadığını, bu değerden anlamlı bir biçimde farklı olduğunu gösterir.

Şekil 7.7



Evren Ortalaması μ İçin İki Yönlü Test Sonuçları

ÖRNEK 7

Bir firmanın geliştirdiği yeni sistemin ortalama paketleme süresini ürün başına 15 dakikanın altına indirdiği iddia edilmektedir. Bu iddiayı araştırmak amacıyla paketleme esnasında rassal olarak seçilen 225 ürünün yeni sistemde ortalama paketlenme süresi 13 dakika ve standart sapması 4.2 dakika olarak belirlenmiştir. Yeni sistemle ilgili iddia hakkında α = 0.01 anlam düzeyinde karar veriniz.

Çözüm:

1. Adım: Hipotezlerin ifade edilmesi

Burada verilecek karar, ortalama paketleme süresinin 15 dakikadan az olup olmadığıdır. Araştırma hipoteziyse 15 dakikadan az olduğudur. Buna göre hipotezler:

$H_0 : \mu = 15$ dk.

$H_1 : \mu < 15$ dk.

şeklinde ifade edilecektir.

2. Adım: İstatistiksel test

Örneklem hacmi n = 225 (n ≥ 30) olduğu için $\bar{X}$'nın örnekleme dağılımı normaldir. $\bar{X}$'nın standart sapması ya da standart hata birimi $s_{\bar{x}}$ cinsinden $\bar{X}$= 13 değerinin bilinen μ = 15 değerinden ne kadar farklılık gösterdiğini ölçmek için standartlaştırılmış z değişkeni kullanılır. z standartlaştırılmış test istatistiği olarak ifade edilir.

$$z = \frac{\bar{X} - \mu_0}{s_{\bar{X}}}$$

$\bar{X}$'nın örnekleme dağılımı normal olduğu için $\mu = \mu_0$= 15 olduğu zaman, z standart normal dağılıma sahip rassal değişkendir. Bu nedenle, bu hipotezlerin testi için z testi uygulanmalıdır.

3. Adım: Anlamlılık düzeyinin belirlenmesi

Doğru olan H_0 hipotezinin reddedilmesi olasılığı α = 0.01 olarak belirlenmiştir. H_0 hipotezinin kabul bölgesinin oransal büyüklüğü 1 - α = 0.99 iken red bölgesinin büyüklüğü α = 0.01'dir.

4. Adım: H_0'ın red bölgesinin belirlenmesi

H_1 hipotezi, testin red bölgesinin yönünü belirlemektedir. Bu testte $H_1 : \mu < 15$ olarak ifade edildiği için H_0 hipotezinin red bölgesi dağılımın alt kuyruğunda tanımlanmıştır. Hipotez, tek yönlü alt kuyruk testiyle test edilecektir. Bu durum Şekil 7.8'de gösterilmiştir.

5. Adım: Test istatistiğinin hesaplanması

μ_0 = 15 dk

$\bar{X}$= 13 dk

n = 225 ürün

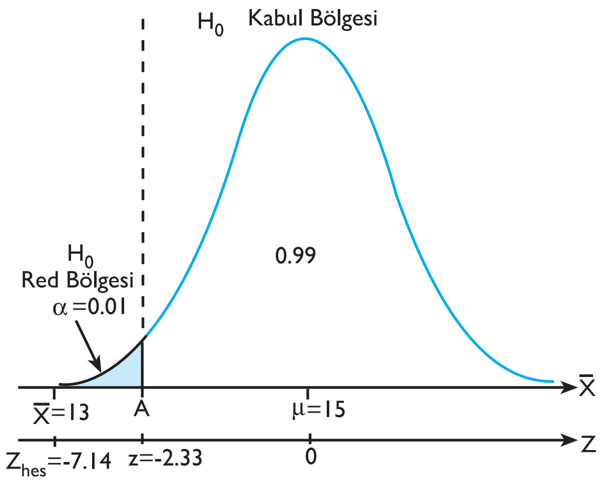
s = 4.2 dk.

α = 0.01

$$s_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{4.2}{\sqrt{225}} = 0.28$$

$$z = \frac{\bar{X} - \mu_0}{s_{\bar{X}}} = \frac{13 - 15}{0.28} = -7.142$$

olarak hesaplanan $z_{hes} = |7.142| > z_{tab} = 2.33$ olduğundan H_0 hipotezi red (dolayısıyla H_1 hipotezi kabul) edilir.



Şekil 7.8

Evren Ortalaması İçin Tek Yönlü Alt KuyrukTesti Sonuçları

Bu firmanın geliştirdiği yeni sistem, ortalama paketleme süresini ürün başına 15 dakikanın altına indirmektedir.

Evren Ortalamasına İlişkin Küçük Örneklem Testi

Çok nadir görülen bir hastalıkla ilgili araştırmada vaka sayısını, uzun süren deneylere dayanan araştırmalarla ve maliyeti yüksek olan laboratuar çalışmaları nedeniyle örneklem hacmini arttırmak çok güçtür. Örneklem hacminin az ($n < 30$ birim) olduğu bu gibi durumlarda, küçük örneklemler için geliştirilmiş test yöntemlerine başvurulur.

Örneklem hacminin küçük olması durumunda, σ yerine s'nin kullanılması istatistiksel test üzerinde etkili olur. Çünkü σ yerine s'nin kullanılması durumunda tahmin edilen istatistik $\bar{X}$-μ / $s_{\bar{x}}$ standart normal dağılım göstermemekte, dolayısıyla büyük örneklemlerde izlenen yöntem geçerli olmamaktadır. Normal dağılıma sahip ve değişkenliği bilinmeyen bir evrenden seçilen 30'dan daha az birim içeren bir örneklemin aritmetik ortalaması, n-1 serbestlik derecesiyle t dağılımına sahiptir. t istatistiği,

$$t = \frac{\bar{X} - \mu_0}{s_{\bar{X}}}$$

şeklindedir. Burada $s_{\bar{x}}$, örneklem ortalamasının standart hata tahminini gösterir ve

$$s_{\bar{X}} = \frac{s}{\sqrt{n-1}}$$

eşitliği ile hesaplanır.

t dağılımı da normal dağılım gibi simetrik bir dağılımdır ve örneklem hacmi büyüdükçe normal dağılıma yaklaşır.

Küçük örneklem kullanılarak yapılan tek evren ortalamasına ilişkin hipotez testleri, kullanılan test istatistiği dışında tek evren ortalamasına ilişkin büyük örneklem testlerine benzemektedir. Aşağıdaki örnek problem üzerinde bu testin uygulanış biçimi test sürecinin adımları itibariyle açıklanmıştır.

Tek evren ortalamasına ilişkin büyük örneklem testinde olduğu gibi, küçük örneklem testinde de örneklem aritmetik ortalaması $\bar{X}$ ile evrenin ortalaması hakkında daha önceden bilinen ya da belirlenen bir değer μ_0 arasındaki farkın istatistiksel olarak anlamlı olup olmadığı araştırılır.

ÖRNEK 8

Bir su dağıtım şirketinde doldurulan damacanaların ortalama ağırlığının 22 kg. olması gerekmektedir. Damacanaların bu ağırlığa uyup uymadığını kontrol etmek amacıyla rassal olarak 17 damacana seçilerek, bunların ağılıklarının ının ortalaması 21.7 kg ve standart sapması 0.8 kg olarak hesaplanmıştır. Su dağıtım şirketinin sattığı damacanaların olması gerektiği gibi doldurulup doldurulmadığı hakkında 0.01 anlamlılık düzeyinde karar veriniz.

Çözüm:

1. Adım: Hipotezlerin ifade edilmesi

Dolu damacanaların belirlenen ağırlığı 22 kg olup; sıfır hipotezi, damacanaların ortalama ağırlığının 22 kg'ye eşit olduğu yönündedir. Bu iddiayı her iki yöndeki anlamlı ağırlık farklılıkları çürütecektir. Buna göre yapılacak test, *iki yönlü test* olacak ve hipotezler;

$H_0 : \mu = 22$ kg.

$H_1 : \mu \neq 22$ kg.

şeklinde yazılacaktır.

2. Adım: İstatistiksel test

Örneklem hacmi n = 17 damacanadır. Kullanılması gereken test istatistiği küçük örneklem t testidir. t test istatistiği

$$t = \frac{\bar{X} - \mu_0}{s_{\bar{X}}}$$

biçiminde hesaplanır.

3. Adım: Anlamlılık düzeyinin belirlenmesi

Problemde, doğru olan H_0 hipotezinin reddedilmesi olasılığı $\alpha = 0.01$ olarak belirlenmiştir. H_0 hipotezinin kabul bölgesinin oransal büyüklüğü $1\text{-}\alpha = 0.99$ iken red bölgesinin büyüklüğü $\alpha = 0.01$'dir.

4. Adım: H_0'ın red bölgesinin belirlenmesi

H_0 hipotezi iki yönlü test edilecektir. Red bölgeleri, t dağılımının hem alt kuyruğunda hem de üst kuyruğunda tanımlanmıştır. Bu durum Şekil 7.9' da gösterilmiştir.

5. Adım: Test istatistiğinin hesaplanması

n = 17 damacana

$\bar{X}$= 21.7 kg

s = 0.8 kg

μ_0= 22 kg

α = 0.01

$$s_{\bar{X}} = \frac{s}{\sqrt{n-1}} = \frac{0.8}{\sqrt{16}} = 0.2$$

ve

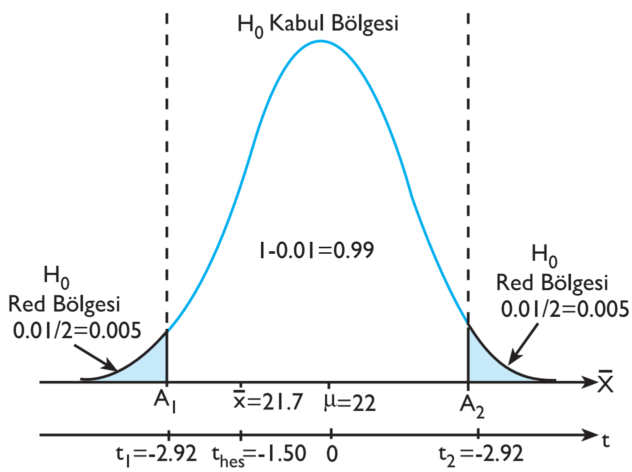
$$t_{hes} = \frac{\bar{X} - \mu_0}{s_{\bar{X}}} = \frac{21.7 - 22}{0.2} = -1.5 \; (t_{hes} : \text{Hesaplanan t değeri})$$

olarak bulunur.

Test edilecek hipotez iki yönlüdür. α = 0.01 ve n = 17 için t tablo değeri (t_{tablo})

t_{tablo} = t (1- a/2; s.d = n - 1) = 2.921 olarak belirlenir.

t_{hes} = |1.5|>t_{tab} = 2.921 olduğundan H_0 hipotezi kabul edilir, dolayısıyla H_1 hipotezi red edilir. Şekil 7.9'da görüldüğü gibi hesaplanan test istatistiği değeri, kabul bölgesinde yer almaktadır.



Şekil 7.9

Küçük Örneklemlerde μ İçin İki Yönlü Test Sonuçları.

Bu karara göre, su dağıtım şirketi su damacanalarını olması gerektiği gibi 22 kg olarak satmaktadır. Örneklem aritmetik ortalaması ile evren aritmetik ortalaması arasında anlamlı bir farklılık bulunmamıştır.

Evren Oranına İlişkin Test

Daha önce de açıklandığı gibi, bir evrenin, ilgilenilen iki şıklı bir değişkeninin, herhangi bir şıkkına sahip birimlerinin oranına "evren oranı" denir ve Π simgesiyle gösterilir. Ünitenin bu kesiminde, evren oranı Π'nin değeri hakkında ileri sürülen bir önermenin, nasıl test edileceği konusu ele alınacaktır. Tek evren oranına ilişkin test olarak isimlendirilen bu testin, örneklem oranı p ile evren oranı P'nin iddia edilen değeri Π_0 arasındaki farkın istatistiksel olarak anlamlı olup olmadığı araştırılır. Örneklem hacmi yeterli büyüklükte olduğunda ($n \geq 30$ birim), evren oranı Π hakkındaki testler daha önce açıklanan evren ortalaması μ için büyük örneklem testlerindekilere benzer şekilde yapılır. Ancak test için örneklem istatistiği olarak örneklem oranı p ve bu istatistiğin örnekleme dağılımı kullanılır. $n \geq 30$ olduğunda, örneklem oranı p'nin örnekleme dağılımı, yaklaşık normal dağılıma sahip olur. Bu durumda p örneklem istatistiği dağılımına ilişkin standart değerlerin dağılımının da normal olacağı açıktır.

Evren oranı Π'ye ilişkin testlerde örneklem hacmi büyük olduğunda, standartlaştırılmış z istatistiği kullanılır.

$$z = \frac{p - \Pi_0}{\sigma_p}$$

Burada, σ_p örneklem oranı, p'nin örnekleme dağılımının standart sapmasını gösterir ve

$$z = \sqrt{\frac{\Pi_0(1 - \Pi_0)}{n}}$$

eşitliği ile hesaplanır.

DİKKAT

Eğer örneklem oranı p ile evren oranı Π'nin ileri sürülen değeri arasındaki farkın, istatistiksel olarak anlamlı olup olmadığı büyük hacimli örneklemlerle araştırılıyorsa evren oranına ilişkin test uygulanır ve test istatistiği $z = \frac{p - \Pi_0}{\sigma_p}$ 'dir.

ÖRNEK 9

Özel bir dersane yetkilisi, üniversiteye giriş sınavına hazırlık için kendi dersanelerine gelen öğrencilerden, üniversitede istediği bölümü kazananların oranının %80'den fazla olduğunu iddia etmektedir. Bu iddiayı araştırmak amacıyla söz konusu dersaneye giden ve üniversiteyi kazanan öğrenciler arasından rassal olarak seçilen 120 öğrenciye istedikleri bölümü kazanıp kazanmadıkları sorulmuş ve 102 öğrencinin üniversitede istediği bölümü kazandığı öğrenilmiştir. Yetkilinin iddiası hakkında 0.05 anlam düzeyinde karar veriniz.

Çözüm:

1. Adım: Hipotezlerin ifade edilmesi

Bu problemde sıfır hipotezi, söz konusu dersaneye giderek üniversitede istediği bölümü kazananların oranının 0.80'den farklı olmadığı şeklindedir. Karşıt (araştırma) hipotezi ise söz konusu dersaneye giderek üniversitede istediği bölümü kazananların oranının 0.80'den fazla olduğu yönündeki tek taraflı hipotezdir. Bu hipotezler;

$$H_0 : \Pi = 0.80$$

$$H_1 : \Pi > 0.80$$

şeklinde ifade edilir.

2. Adım: İstatistiksel test

Örneklem hacmi n = 120 (n ≥ 30) olduğu için p'nin örnekleme dağılımı normaldir. Buna göre kullanılması gereken test istatistiği

$$z = \frac{p - \Pi_0}{\sigma_p}$$

olur. Evren oranına ilişkin tek yönlü üst kuyruk z testi uygulanmalıdır.

Evren oranı bilindiği için

$$\sigma_p = \sqrt{\frac{\Pi_0(1-\Pi_0)}{n}}$$

şeklinde hesaplanır.

3. Adım: Anlamlılık düzeyinin belirlenmesi

Problemde doğru olan H_0 hipotezinin reddedilmesi olasılığı $\alpha = 0.05$ olarak belirlenmiştir. Red bölgesinin oransal büyüklüğü, 0.05'dir.

4. Adım: H_0'ın red bölgesinin belirlenmesi

H_1 hipotezi, testin red bölgesinin yönünü belirlemektedir. Red bölgesi, oranlar örnekleme dağılımının üst kuyruğunda tanımlanmıştır. Bu durum Şekil 7.10'da gösterilmiştir.

5. Adım: Test istatistiğinin hesaplanması

Seçilen 120 öğrenci arasından üniversitede istediği bölümü kazananlar 102 kişidir. Bu durumda örneklem oranı;

$$p = \frac{r}{n} = \frac{102}{120} = 0.85$$

olur. Örnek için standart hata,

$$\sigma_p = \sqrt{\frac{\Pi_0(1-\Pi_0)}{n}} = \sqrt{\frac{(0.80)(0.20)}{120}} = 0.036$$

ve

$$z = \frac{0.85 - 0.80}{0.036} = 1.38$$

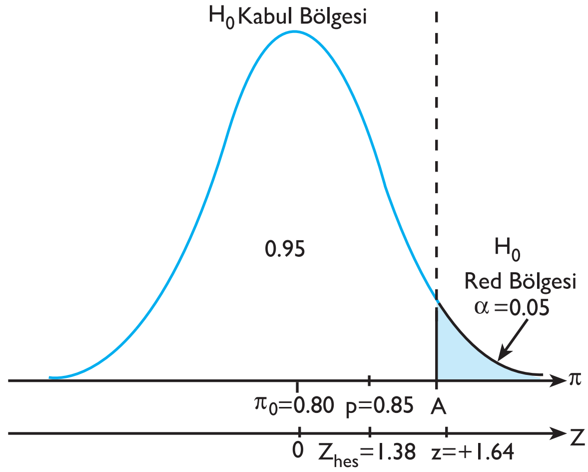
olarak elde edilir.

Hesaplanan test istatistiği $z_{hes} = 1.38 < z_{tab} = 1.645$ olduğundan H_0 hipotezi kabul edilir (Dolayısıyla H_1 hipotezi reddedilir.).

Bu kararın anlamı; söz konusu özel dersane yetkilisinin iddiası yanlıştır, üniversiteye giriş sınavına hazırlık için kendi dersanelerine gelen öğrencilerden, üniversitede istediği bölümü kazananların oranının %80'den fazla değildir.

Şekil 7.10

Evren Oranı Π İçin Test Sonuçları.



ÖRNEK 10

100 Bir yemek firması, yapmış oldukları yemeklerden memnun olmayan müşteri oranının %10'dan az olduğunu ileri sürmektedir. Konuyu araştırmak amacıyla, bu yemek firmasını tercih eden müşteriler arasından rassal olarak 150 kişi seçilmiş ve 12 kişinin yemeklerden memnun olmadığını öğrenilmiştir. 0.01 anlam düzeyinde karar veriniz.

Çözüm:

1. Adım: Hipotezlerin ifade edilmesi

Bu örnekte sıfır hipotezi, söz konusu yemek firmasını tercih eden müşteriler arasından yemeklerden memnun olmayanların oranının 0.10'dan farklı olmadığı şeklindedir. Karşıt (araştırma) hipotezi ise yemek firmasını tercih eden müşteriler arasından yemeklerden memnun olmayanların oranının 0.10'dan az olduğu yönündeki tek taraflı hipotezdir. Bu hipotezler;

$$H_0 : \Pi = 0.10$$

$$H_1 : \Pi < 0.10$$

şeklinde ifade edilir.

2. Adım: İstatistiksel test

Örneklem hacmi n = 150 (n ≥ 30) olduğu için p'nin örnekleme dağılımı normaldir. Buna göre kullanılması gereken test istatistiği

$$z = \frac{p - \Pi_0}{\sigma_p}$$

olur. Evren oranına ilişkin tek yönlü alt kuyruk z testi uygulanmalıdır.

Evren oranı bilindiği için

$$\sigma_p = \sqrt{\frac{\Pi_0(1-\Pi_0)}{n}}$$

şeklinde hesaplanır.

3. Adım: Anlamlılık düzeyinin belirlenmesi

Problemde doğru olan H_0 hipotezinin reddedilmesi olasılığı $\alpha = 0.05$ olarak belirlenmiştir. Red bölgesinin oransal büyüklüğü, 0.05'dir.

4. Adım: H_0'ın red bölgesinin belirlenmesi

Red bölgesi, oranlar örnekleme dağılımının alt kuyruğunda tanımlanmıştır. Bu durum Şekil 7.11'de gösterilmiştir.

5. Adım: Test istatistiğinin hesaplanması

Seçilen 150 müşteri arasından yemeklerden memnun olmayan müşteri sayısı 12 dir. Bu durumda örneklem oranı;

$$p = \frac{r}{n} = \frac{12}{150} = 0.08$$

olur. Örnek için standart hata,

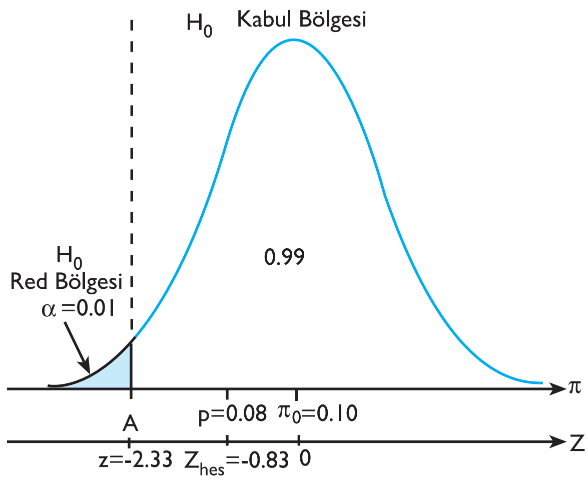
$$\sigma_p = \sqrt{\frac{\Pi_0(1-\Pi_0)}{n}} = \sqrt{\frac{(0.10)(0.90)}{150}} = 0.024 \text{ ve}$$

$$z = \frac{0.08 - 0.10}{0.024} = -0.83$$

olarak elde edilir.

Hesaplanan test istatistiği $z_{hes} = |0.83| < z_{tab} = 2.33$ olduğundan H_0 hipotezi kabul edilir (Dolayısıyla H_1 hipotezi reddedilir.).

Bu kararın anlamı; yemek firmasının yapmış olduğu yemeklerden memnun olmayan müşteri oranının %10'dan az olmadığıdır.



Şekil 7.11

Evren Oranı P İçin Test Sonuçları.

SIRA SİZDE 4

- **Örneklem hacmi $n \geq 30$ birim olduğunda, tek evren ortalamasına ilişkin bir testte, hangi test istatistiği kullanılır? Nedenini açıklayınız.**
- **Evren oranına ilişkin bir testte, test istatistiği nasıl hesaplanır?**

Özet

AMAÇ 1

İstatistiksel tahminleme tanımı yapmak ve bir evren parametresi için tahminleme sürecinin aşamalarını açıklamak.

Örnekleme ünitesinde değinildiği gibi tamsayım gerçekleştirilemediği durumlarda örneklemeye başvurulmaktadır ve örneklem istatistiklerinden yararlanarak evren parametresi hakkında tahminleme yapılmaktadır. Tahminleme süreci ise örneklem seçimi, tahminlenecek parametre için bilgi üretecek istatistiklerin hesaplanması ve parametre değerinin tahmin edilmesidir.

AMAÇ 2

Evren aritmetik ortalamasına ve evren oranına ilişkin nokta ve aralık tahminlemesi yapmak.

İstatistiksel tahminleme nokta ve aralık tahminlemesi olarak iki başlık altında sınıflandırılır. Bir rassal örneklemden hesaplanan $\bar{X}$ istatistiğinin değerini ilgili evren parametresi μ değerine eşit kabul eden tahminleme sürecine nokta tahminlemesi denir.

Evren oranı Π'ye ilişkin nokta tahminlemesi bir rassal örnekleme planında oluşturulan n hacimli örneklem için, r bir binom rassal değişkeni olmak üzere, hesaplanan $p=\frac{r}{n}$ oranının değeri Π'ye eşit olan bir tahminleme sürecidir.

Evren aritmetik ortalaması μ için 1-α güven sınırlarının ya da güven aralığının belirlenmesi işlemlerine μ'nün aralık tahminlemesi denir.

Evren oranı Π'nin aralık tahmini, 1-α güven düzeyinde evren ortalaması aralık tahmininde izlenen aşamalara benzer şekilde gerçekleştirilir.

AMAÇ 3

Hipotez, istatiksel hipotez ayrımını yapmak ve istatistiksel hipotezlerin aşamalarını açıklamak.

Genel olarak hipotez, karşılaşılan özel duruma ilişkin bir önermedir. İstatistiksel hipotez, bir araştırmada ilgilenilen bir ya da daha fazla parametrenin değeri hakkında ileri sürülür Bu önermenin doğruluğu, geçerliliği bu parametre hakkında bilgi üreten istatistikden ve bu istatistiğin örnekleme dağılımıyla ilgili bilgilerden yararlanarak araştırıabilen önermelerdir.

Sıfır hipotezinin ve karşıt hipotezin ifade edilmesi, anlam düzeyinin belirlenmesi, verilerin derlenmesi, test istatistiğinin seçilmesi ve hesaplanması, son olarak da karar verilmesi istatistiksel hipotez testinin aşamaları olarak sıralanır.

AMAÇ 4

Evren aritmetik ortalamasına ve evren oranına ilişkin hipotez testi uygulamalarını yorumlamak.

Tek evren ortalamasına ilişkin hipotez testi uygulamalarında evrenin dağılım şekline, rassal olarak seçilen örneklemin hacmine bağlı olarak z ve t testleri uygulanır. z testi evrenin dağılımının normal olması, evren değişkenliğinin biliniyor olması veya örneklem hacminin yeterli büyüklükte olması durumunda uygulanır. t testi ise örneklem hacmi küçük ve evrenin dağılım şekli normal ise uygulanır.

Örneklem hacmi yeterli büyüklükte olduğunda, örneklem oranı p rassal değişkeninin dağılım şekli normal olacağı için evren oranına ilişkin hipotez sınamalarında z test istatistiği kullanılır.

Kendimizi Sınayalım

1. Bir evrenin aritmetik ortalamasının güven sınırlarını % 99.34 güvenle tahmin edebilmek için kullanılacak standart hata katsayısı $z_{\frac{\alpha}{2}}$ kaçtır?

a. 2.66
b. 2.68
c. 2.70
d. 2.72
e. 2.74

2. ve 3. sorular aşağıdaki bilgilere göre cevaplandırılacaktır.
Pil üreten bir işletmede, ürünlerin ömrünü artırmak amacıyla bir üretim yöntemi uygulanacaktır. Bu amaçla uygulanacak yeni yöntem sınanmak istenmektedir.

2. Bu sınamada kurulacak sıfır hipotezi nedir?

a. Yeni yöntem pillerin ömrünü artırır.
b. Yeni yöntem pil üretimini artırır.
c. Yeni yöntem pillerin ömrünü değiştirmez.
d. Yeni yöntem pillerin ömrünü değiştirir.
e. Yeni yöntem pil üretimini azaltır.

3. Bu sınamada kurulacak alternatif hipotez nedir?

a. Yeni yöntem pil üretimini artırır.
b. Yeni yöntem pil üretimini azaltır.
c. Yeni yöntem pil üretimini değiştirir.
d. Yeni yöntem pillerin ömrünü artırır.
e. Yeni yöntem pillerin ömrünü değiştirir.

4. Anlam düzeyi 0.05 olan bir hipotez sınamasında, I. tip hata yapma olasılığı kaçtır?

a. 0.99
b. 0.90
c. 0.10
d. 0.05
e. 0.01

5. Aşağıdaki alternatif hipotezlerden hangisi çift yönlü bir sınama gerektirir?

a. A firmasının günlük üretim miktarı B firmasının günlük üretim miktarından fazladır.
b. A marka malın dayanıklılığı B marka malın dayanıklılığından azdır.
c. A makinesinin verimliliği B makinesinin verimliliğinden farklıdır.
d. A marka ampülün ömrü B marka ampülün ömründen kısadır.
e. A marka malın satış oranı B marka malın satış oranından fazladır.

6. İstatistik dersine ait notların ortalamasının, 80'den büyük olup olmadığı, sınanmak istenmektedir. Bu sınamada kurulacak sıfır hipotezi nedir?

a. $\mu > 80$
b. $\mu = 80$
c. $\mu \neq 80$
d. $\mu < 81$
e. $\mu > 81$

7. 0.02 anlam düzeyinde sınanan bir hipotez için, doğru olan sıfır hipotezini reddederek hatalı karar verme olasılığı kaçtır?

a. 0.01
b. 0.02
c. 0.05
d. 0.98
e. 0.99

8. Bir hipotez sınamasında red bölgesinin yönünü aşağıdakilerden hangisi belirler?

a. H_0
b. H_1
c. H_0 ve H_1
d. α hatası
e. β hatası

9. ve 10. sorular aşağıdaki bilgilere göre cevaplandırılacaktır.
Bir işletmede, üretilen ampüllerin 450 saat olan dayanma süresini artırmak için yeni bir hammaddenin kullanımı düşünülmektedir. Bu ham madde kullanılarak 1000 ürün üretilmiş ve ortalama dayanma süresi 462 saat olarak hesaplanmıştır. Ham maddenin olumlu sonuç verip vermediği %95 güvenle sınanacaktır.

9. Bu sınamada örnekleme dağılımının red bölgesi aşağıdakilerden hangisidir?

a. Sağ uçta, %2.5'lik alan
b. Sol uçta, %2.5'lik alan
c. Sağ uçta, %5'lik alan
d. Sol uçta, %5'lik alan
e. Sağ uçta, %10'luk alan

10. Bu sınamadaki alternatif hipotez nedir?

a. $\mu \neq 450$
b. $\mu > 462$
c. $\mu = 450$
d. $\mu \neq 462$
e. $\mu > 450$

Kendimizi Sınayalım Yanıt Anahtarı

1. d	Yanıtınız yanlış ise, "Evren Aritmetik Ortalamasının Aralık Tahminlemesi" bölümünü yeniden gözden geçiriniz.
2. c	Yanıtınız yanlış ise, "Hipotezlerin İfade Edilmesi" bölümünü yeniden gözden geçiriniz.
3. d	Yanıtınız yanlış ise, "Hipotezlerin İfade Edilmesi" bölümünü yeniden gözden geçiriniz.
4. d	Yanıtınız yanlış ise, "Anlamlılık Düzeyinin Belirlenmesi" bölümünü yeniden gözden geçiriniz.
5. c	Yanıtınız yanlış ise, "Hipotezlerin İfade Edilmesi" bölümünü yeniden gözden geçiriniz.
6. b	Yanıtınız yanlış ise, "Evren Ortalamasına İlişkin Hipotez Testi" bölümünü yeniden gözden geçiriniz.
7. b	Yanıtınız yanlış ise, "Anlamlılık Düzeyinin Belirlenmesi" bölümünü yeniden gözden geçiriniz.
8. b	Yanıtınız yanlış ise, "Hipotezlerin İfade Edilmesi" bölümünü yeniden gözden geçiriniz.
9. c	Yanıtınız yanlış ise, "Anlamlılık Düzeyinin Belirlenmesi" bölümünü yeniden gözden geçiriniz.
10. e	Yanıtınız yanlış ise, "Hipotezlerin İfade Edilmesi" bölümünü yeniden gözden geçiriniz.

Sıra Sizde Yanıt Anahtarı

Sıra Sizde 1

Bir rassal örneklemden hesaplanan istatistikler yardımıyla ilgili evrenin parametre değerinin araştırılmasına tahminleme denir.

Hem örneklem için hem de evren için bilgi üreten istatistiğe ilişkin formülasyona "tahminleyici", örneklem gözlem değerinin bir tahminleyiciye uygulanmasıyla hesaplanan değere ise "tahmin" adı verilir.

Sıra Sizde 2

İstatistiksel hipotez, bir dağılımın evren parametresine ilişkin bir önermedir.

İstatistiksel hipotez testi, örneklem istatistiklerini kullanarak bir hipotezin doğru olup olmadığını ortaya koymaya yönelik yapılan çalışmalardır.

Sıra Sizde 3

Test edilecek hipotez sıfır hipotezidir.

Hipotez testinde gerçekte doğru olan sıfır hipotezinin yanlış diye reddedilmesi durumunda ortaya çıkan hatadır.

Sıra Sizde 4

z test istatistiği, $z = \frac{\bar{X} - \mu_0}{s_{\bar{X}}}$ kullanılır.

Evren oranı Π 'ye ilişkin testlerde standartlaştırılmış z istatistiği, $z = \frac{p - \Pi_0}{\sigma_p}$ kullanılır.

Yararlanılan ve Başvurulabilecek Kaynaklar

Canküyer, E., Aşan, Z. (2001). **Parametrik Olmayan İstatistiksel Teknikler**, Eskişehir: Anadolu Üniversitesi Yayınları, No 1266.

Çömlekçi, N. (2001). **Bilimsel Araştırma Yöntemi ve İstatistiksel Anlamlılık Sınamaları,** İstanbul: Bilim Teknik Yayınevi.

Fink, A. (1995). **How To Sampling in Surveys,** London: Sage Publication.

Gürsakal, N. (1997). **Bilgisayar Uygulamalı İstatistik 1**, Bursa: Marmara Kitabevi.

Hinkle, D. E., Wiersma, W., Jurs S. G. (1998). **Applied Statistics For The Behavioral Sciences,** Boston.

Malhotra, N. K. (1996). **Marketing Research An Applied Orientation,** New Jersey: Prentice Hall International.

Neter, J., Wasserman, W., Whitmore, G. A. (1993). **Applied Statistics,** Boston: Simon and Schuster.

Özmen, A., Özdamar, K., Odabaşı, Y., Hoşcan, Y., Bir, A. A., Kırcaaliiftar, G., Uzuner, Y. (1999). **Sosyal Bilimlerde Araştırma Yöntemleri,** Eskişehir: TC. Anadolu Üniversitesi Yayınları No:1081; Açıköğretim Fakültesi Yayınları No:601.

Püskülcü, H., İkiz, F. (1986). İstatistiğe Giriş, (2. Baskı), İzmir: E. Ü. Mühendislik Fakültesi Yayın No:601, ege Üniversitesi Basımevi.

Serper, Ö., Aytaç, M. (2000). **Örnekleme**, Bursa: Ezgi Kitabevi.

Serper, Ö. (1986). **Uygulamalı İstatistik 2,** İstanbul: Filiz Kitabevi.

Tryfos, P. (1996). **Sampling Methods for Applied Research,** New York: John Wiley and Sons Inc.

Tull, D. D.,Hawkins, D. I. (1993). **Marketing Research Measurement and Method,** 6th Edition, New York: MacMillan Publishing Company.

8

Amaçlarımız

Bu üniteyi tamamladıktan sonra;

- Değişkenler arasındaki ilişkinin derecesini tespit edebilecek,
- Bağımlı değişkendeki değişimin yüzde kaçının bağımsız değişken tarafından açıklandığını tespit edebilecek,
- Bir bağımlı ve bir bağımsız değişken ile model kurabileceksiniz.

Anahtar Kavramlar

- Korelasyon
- Regresyon
- Korelasyon Analizi
- Pearson Korelasyon Katsayısı
- Belirlilik Katsayısı
- Korelasyon Katsayısı Anlamlılık Testi
- Basit Doğrusal Regresyon Analizi
- Tahminin Standart Hatası
- Örneklem Regresyon Doğrusu Anlamlılık Testi

İçindekiler

Regresyon ve Korelasyon Çözümlemesi

GİRİŞ

İki değişken arasındaki ilişkiyi tanımlamaya ve ölçmeye *korelasyon analizi* adı verilir. Bir veya birden çok değişkenin başka bir değişken üzerindeki ilişkisini açıklamaya *regresyon analizi* adı verilir. Yukarıdaki tanımlar her ne kadar birbirine benzer gözükse de korelasyon analizi ve regresyon analizi bir bozuk paranın iki farklı yüzü gibidirler. Regresyon terimi 19. yüzyılda İngiliz istatistikçisi Francis Galton tarafından bir biyolojik inceleme için ortaya atılmıştır. Galton babaların boyları ile oğullarının boyları arasındaki ilişkiyi araştırmıştır. Araştırmaları sonucunda ortalamaya doğru bir eğilimin varlığını fark etmiştir. Çok kısa boylu babaların oğullarının boylarının ortalama değerler etrafında (babalarından daha uzun gibi) toplandığını gözlerken bu durumun tersinin de doğru olduğunu (uzun boylu babaların oğulları da ortalama boyda bulunmaktadır) fark etmiştir. Galton oğullarının boylarının ortalamaya doğru yönlendiğini (İngilizce karşılığı "regressed" olmak üzere) belirterek regresyon kelimesinin temelini de atmıştır. Günümüzde regresyon kelimesi iki ya da daha fazla değişken arasındaki ilişkinin modellenmesi işlemlerinin tümünü içeren geniş bir anlama sahiptir.

Regresyon analizinde iki farklı değişken tanımlamasına ihtiyaç duyulmaktadır: Bunlar sırasıyla bağımlı ve bağımsız değişken kavramlarıdır. *Bağımlı değişken* araştırmacının üzerinde çalıştığı değişken olup bu değişken üzerinde meydana gelen değişimlerin ya da bu değişkenin toplam değişkenliğinin açıklanılmasına çalışılmaktadır. *Bağımsız değişken* ya da değişkenler ise ilgilenilen bağımlı değişkende meydana gelen değişim ya da toplam değişkenliğinin üzerinde etkisi olabileceği düşünülen değişken ya da değişkenlerdir. Örneğin bağımlı değişken trafik akım hızı ise bağımsız değişkenler birim otomobil cinsinden trafik hacminin kapasiteye oranı, birim otomobil cinsinden trafik hacmi içindeki bisiklet sayısı, yol üzerindeki ticari yoğunluk oranı, güzergah üzerinde bulunan sinyalli ve sinyalsiz önemli kavşak sayısı olabilir. Regresyon analizinde bir ya da daha fazla bağımsız değişken olabilir. Bu ünite içerisinde tek bağımsız değişken olması durumu *basit doğrusal regresyon analizi* olarak ele alınacak, birden fazla bağımsız değişken olması durumu *çoklu doğrusal regresyon analizi* ele alınmayacaktır.

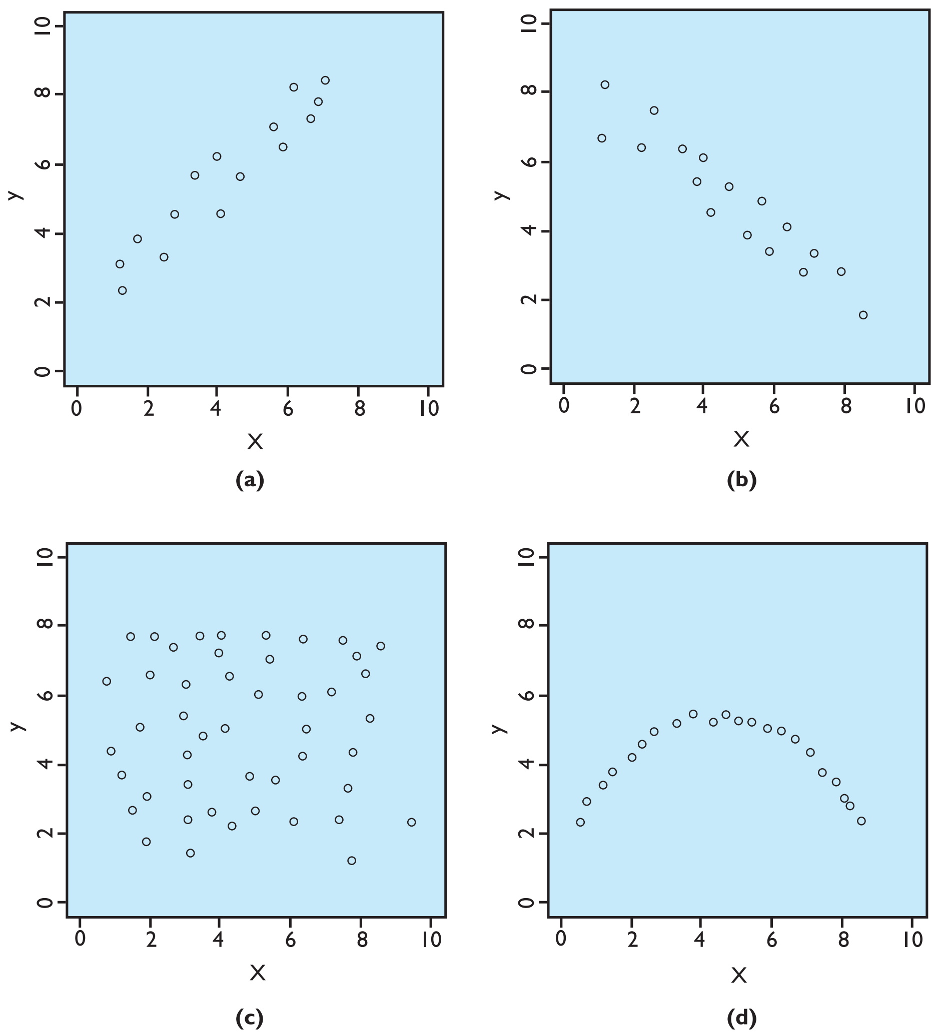
İlgilenilen iki değişken arasındaki ilişkinin derecesi için korelasyon analizi kullanılır. Korelasyon analizinin regresyon analizinden farklılık gösterdiği nokta, korelasyon analizinin değişkenler arasındaki ilişkinin yanlızca derecesini göstermesidir. İki değişken arasında yüksek korelasyon olması bu iki değişkenden birinin diğerinin nedeni olabileceğini göstermez. Korelasyon analizi iki değişken arasındaki nedensellik için kullanılmaz. Nedensellik araştırması için farklı istatistik tekniklerinin kullanılması gerekir.

KORELASYON ANALİZİ

Korelasyon analizi en genel anlamı ile iki değişken arasındaki ilişkinin tanımlanması ve ilişkinin derecesinin belirlenmesidir. Değişkenler arasında var olabilecek ilişkinin derecesinin tespit edilebilmesi amacı ile çeşitli teknikler kullanılabilir. En basit şekli ile iki değişken arasındaki ilişkiyi gözlemlemek için bu değişkenlerin dağılım grafikleri çizilebilir. Şekil 8.1'de iki değişken arasında gözlemlenebilecek dört farklı durum örneklenmiştir.

Şekil 8.1

İki Değişken İçin Farklı Dağılım Grafikleri



Şekil 8.1 (a) ve (b) grafiklerinde yer alan iki değişkenin arasındaki ilişkinin derecesi birbirine eşittir, fakat iki değişken arasındaki ilişkinin yönü farklıdır. Dikkat edilirse Şekil 8.1 (a)'da x değişkenin değeri artarken y değişkenin değeri de artmaktadır. Ancak Şekil 8.2 (b)'de x değişkenin değeri artarken y değişkenin değeri azalmaktadır.

Şekil 8.1 (c) ve (d) grafiklerinde yer alan iki değişkenin dağılım grafikleri incelendiğinde 8.1. (c) grafiğinde yer alan değişkenler arasındaki ilişkinin rassal olduğu gözlemlenmektedir. Bu nedenle korelasyon analizi yapıldığında aralarında bir ilişki çıkmayacaktır. 8.1. (d) dağılım grafiğinde ise iki değişken arasındaki ilişkinin eğrisel olduğu görülmektedir.

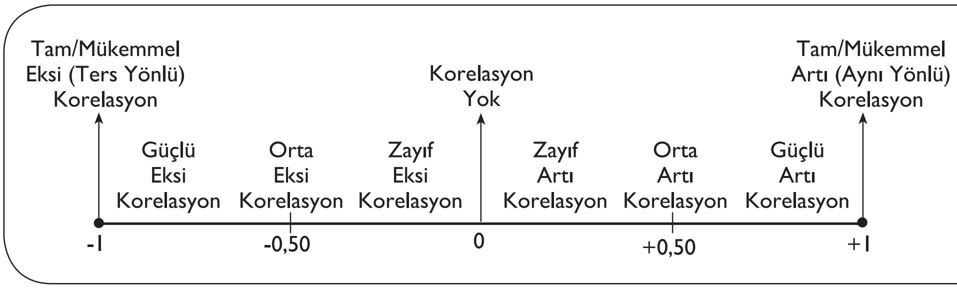
İki değişken arasındaki ilişkinin yanlızca grafikler ile incelenmesi yeterli olmayacaktır. İlişkinin derecesini gösteren istatistiklere ihtiyaç duyulmaktadır. İzleyen bölümde ilişkinin derecesi için hesaplanacak Pearson korelasyon katsayısı incelenmiştir.

Pearson Korelasyon Katsayısı

Korelasyon katsayısı iki değişken arasındaki doğrusal ilişkinin derecesini ifade eder.

İki ya da daha fazla, oranlı ve eşit aralıklı ölçeğe uygun şekilde ölçümlenmiş değişkenler arasındaki ilişkinin derecesini belirlemek için **Pearson korelasyon katsayısı** kullanılır. Evrendeki x ve y değişkenlerinin korelasyon değeri ρ ile sembolize edilir ve bu sembol "ro" şeklinde okunur. Örneklemdeki x ve y değişkenlerinin korelasyon değeri r simgesi ile gösterilir. Yapılan çalışmalar genellikle örneklemler üzerinden gerçekleştirildiğinden bu ünite içinde Pearson korelasyon katsayısı sembolü olarak r kullanılmıştır. Pearson korelasyon katsayısı r, -1 ile +1 arasında değişen değerler almaktadır. Pearson korelasyon katsayısı r'nin -1 ve +1 değerlerine eşit sonuçlar, mükemmel/tam ilişkinin varlığını gösterir. Pearson korelasyon katsayısının eksi r değerleri değişkenler arasındaki ters yönlü ilişkiyi gösterirken (biri artarken diğerinin azalması gibi), artı r değerleri değişkenler arasındaki aynı yönlü ilişkinin (biri artarken diğeri de artmaktadır gibi) var olduğunu gösterir. Eğer iki değişken arasında hiç ilişki yok ise bir başka ifade ile değişkenler bağımsız ise Pearson korelasyon katsayısı 0 (sıfır) değerini alır. Şekil 8.1 (c) değişkenler arasında ilişkinin olmadığını bir başka ifade ile değişkenler arasındaki korelasyonun sıfır değerine eşit olduğunu göstermektedir. Perason korelasyon katsayısı r, -1 ve +1 değerlerine yaklaştıkça ilişkinin derecesinin arttığı ifade edilirken sıfır değerine yaklaştıkça ilişkinin derecesinin azaldığı/zayıfladığı ifade edilir. r değeri -0,50 ya da +0,50 etrafında bir değer ise değişkenler arasında orta düzeyli bir ilişkinin varlığı ifade edilir. Şekil 8.2'de Pearson korelasyon katsayısı r için ilişkinin derecesi ve yönü özetlenmektedir.

Şekil 8.2



Pearson Korelasyon Katsayısı İlişki Yön ve Dereceleri

x ve y değişkenleri arasındaki Pearson korelasyon katsayısı, x'in standart sapması s_x ve y'nin standart sapması s_y olmak üzere,

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y}$$

eşitliği yardımıyla hesaplanır. Bu eşitliğin hesaplanabilmesi için değişkenlerin aritmetik ortalamalarının ve standart sapmalarının hesaplanması gerekmektedir.

Bu hesaplamaların yürütülmesi zaman alacağından Pearson korelasyon katsayısı ortalama ve standart sapmalara ihtiyaç duyulmaksızın,

$$r=\frac{n\sum_{i=1}^{n}x_iy_i-(\sum_{i=1}^{n}x_i)(\sum_{i=1}^{n}y_i)}{\sqrt{n\sum_{i=1}^{n}x_i^2-(\sum_{i=1}^{n}x_i)^2}\sqrt{n\sum_{i=1}^{n}y_i^2-(\sum_{i=1}^{n}y_i)^2}}$$

eşitliği yardımıyla da hesaplanabilir. Her iki eşitlikte aynı sonucu verecektir. Orijinal verinin var olması durumunda verilen ikinci eşitlik hesaplama kolaylığı sağlamaktadır.

ÖRNEK 1

Şemsiye satışı yapan bir firma bir yıl içindeki yağmurlu gün sayısı ile şemsiye satışları arasında bir ilişki olup olmadığını araştırmak istemiştir. Buna göre; bir yıl içindeki aylara göre yağmurlu gün sayısını bir değişken ve ilgili aylara göre şemsiye satış miktarını diğer bir değişken olarak tanımlamıştır. Bu amaçla, bir yıl içindeki aylara göre yağmurlu gün sayısı ve şemsiye satış miktarı değişkenlerine ilişkin derlenen veriler Tablo 8.1'de verilmiştir.

Tablo 8.1
Bir Aylık Ziyaret ve Satış Sayıları.

***Kaynak:** Prof.Dr. Haluk Cillov, (1993). İstatistik Metotları, İ.Ü. Basımevi ve Film Merkezi, İstanbul.*

Aylar	1	2	3	4	5	6	7	8	9	10	11	12
Yağmurlu Gün Sayısı	15	8	18	10	20	10	3	5	6	11	13	13
Şemsiye Satış Miktarı	150	100	200	120	175	70	30	40	70	95	110	290

Derlenen verileri kullanarak yağmurlu gün sayısı ve şemsiye satış miktarları değişkenleri arasındaki Pearson korelasyon katsayısını hesaplayınız?

Aritmetik ortalama ve standart sapma değerlerini hesaplamadan Pearson korelasyon katsayısı hesaplanabilir. Öncelikle eşitlik içerisinde gerekli olacak değerleri hesaplayalım. Bu amaçla Tablo 8.2 oluşturulmuştur.

Tablo 8.2
Pearson Korelasyon Katsayısı İçin Gerekli İşlemler Tablosu.

Aylar	Yağmurlu Gün Sayısı (x)	Şemsiye Satış Miktarı (y)	x_i^2	y_i^2	$x_i y_i$
1	15	150	225	22.500	2.250
2	8	100	64	10.000	800
3	18	200	324	40.000	3.600
4	10	120	100	14.400	1.200
5	20	175	400	30.625	3.500
6	10	70	100	4.900	700
7	3	30	9	900	90
8	5	40	25	1.600	200
9	6	70	36	4.900	420
10	11	95	121	9.025	1.045
11	13	110	169	12.100	1.430
12	13	290	169	84.100	3.770
Toplam	132	1.450	1.742	235.050	19.005

Bu bilgilere göre Pearson Korelasyon katsayısı değeri,

$$r = \frac{12(19005) - (132)(1450)}{\sqrt{12(1742) - (132)^2}\sqrt{12(235050) - (1450)^2}} = \frac{36660}{\sqrt{11400}\sqrt{718100}} = 0,733$$

olacaktır. Pearson korelasyon değerine göre; iki değişken arasında güçlü aynı yönlü korelasyon olduğu söylenebilir.

SIRA SİZDE 1

Tablo 8.2'deki veri seti için Pearson korelasyon katsayısını ortalama ve standart sapmalar yardımıyla hesaplayınız.

Belirlilik Katsayısı

Bağımlı değişken y'de ortaya çıkan toplam değişkenliğin, x değişkenin değişkenliği tarafından açıklanan kısmını/oranını tespit etmek için **belirlilik katsayısı** hesaplanır.

Pearson korelasyon katsayısı yardımıyla değişkenler arasındaki ilişkinin derecesi belirlenmiş ve bu bilgiyi kullanarak ilişkinin zayıf, orta ilişki gibi nitelemesini yapmıştık. Araştırmacılar, bağımlı değişkende meydana gelen değişim içerisinde bağımsız değişkenin payının ne olduğunu bilmek isteyebilir. Bu bilgiyi gösteren istatistiğe **belirlilik katsayısı** adı verilir. Evrendeki x ve y değişkenlerinin belirlilik katsayısı ρ^2 ile sembolize edilir. Örneklemdeki x ve y değişkenlerinin belirlilik katsayısı r^2 simgesi ile gösterilir. Araştırmalar genellikle örneklemler üzerinden yapıldığından bu ünite içinde belirlilik katsayısının simgesi olarak r^2 kullanılacaktır. Belirlilik katsayısı, Pearson korelasyon katsayısı r'nin karesinin alınması ile hesaplanır ve r^2 ile gösterilir. 0 ile 1 arasında değerler alır ve oran olarak ifade edilir.

ÖRNEK 2

Örnek 8.1'in verilerinden faydalanarak şemsiye satış miktarı değişkeninin ne kadarlık kısmının aylara göre yağmurlu gün sayısı ile açıklanabileceğini hesaplayınız?

Bu hesaplamanın yapılabilmesi için öncelikle Pearson korelasyon katsayısına ihtiyaç duyulur. Örnek 8.1'de Pearson korelasyon katsayısı r =0,733 olarak hesaplanmıştır. Belirlilik katsayısı bu değerin karesi olan r^2= (0,733)2=0,537 olacaktır. Bu değere göre; aylara göre yağmurlu gün sayısı değişkeni, şemsiye satış miktarı değişkeninin %53,7'sini açıkladığı söylenebilir. Diğer bir deyişle, şemsiye satış miktarı değişkeninin üzerinde, şemsiye satış miktarı değişkeni dışında kalan farklı faktörlerin de etkisinin olabileceği sonucuna ulaşılabilir.

Korelasyon Katsayısı Anlamlılık Testi

Araştırmalarda çoğunlukla örneklemler üzerinden çalışılmaktadır. Bir araştırmada örneklem sonuçlarına göre belirlenen korelasyon değerinin evren değeri için test edilmesi istenebilir. Örneklem sonuçlarına göre "Korelasyon yoktur." yönünde karar verildiyse bu kararın evren içinde geçerli olup olmadığı korelasyon katsayısı anlamlılık testi yardımıyla yapılır. Korelasyon katsayısının anlamlılık testi, aşağıdaki adımsal süreçte gerçekleştirilebilir.

Adım 1: Hipotezlerin ifade edilmesi

Korelasyon katsayısına ilişkin hipotezler aşağıdaki gibi ifade edilir:

H_0: $\rho = 0$ (x ve y değişkenleri arasında korelasyon yoktur.)

H_1: $\rho \neq 0$ (x ve y değişkenleri arasında korelasyon vardır.)

Adım 2: Anlamlılık düzeyinin belirlenmesi

Anlamlılık düzeyinin belirlenmesi, doğru olan sıfır hipotezinin, örneklemden elde edilen bilgilere dayanarak reddedilmesi olasılığını belirleyen α'nın seçilmesidir. α anlamlılık düzeyi, araştırmacı tarafından, hipotezler ifade edilip veri derlenmeye başlanmadan önce seçilmelidir. Sosyal bilim araştırmalarında α için genellikle %5 ve %1 değerleri seçilmektedir. Yapılan bu seçimle birlikte, doğru olan H_0 hipotezinin reddedilme olasılığı belirlenmiş olur. Bu olasılık örnekleme dağılımıyla ilişkilendirilerek kullanılır. Bu durumda, α anlamlılık düzeyi, doğru olan sıfır hipotezinin reddedilmesi olasılığına eşit olan örnekleme dağılımındaki oransal alanı göstermiş olur.

Adım 3: İstatistiksel test

Testin gerçekleştirilmesinde n-2 serbestlik derecesi ile t dağılımı tablosundan elde edilen kritik değerler kullanılır. Daha sonra belirlenen anlam düzeyine göre t tablosu yardımıyla kritik değer tespit edilir. Örneklem korelasyon değeri yardımıyla,

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

istatistiği hesaplanılarak tablodan elde edilen kritik değer ile karşılaştırılır.

Adım 4: İstatistiksel kararın verilmesi

Eğer hesaplanan t değeri, t tablosu yardımıyla belirlenen kritik değer (-kritik değer, + kritik değer) aralığında yer alıyor ise H_0 (x ve y değişkenleri arasında korelasyon yoktur.) hipotezi kabul edilir. Bu aralık dışında yer alıyor ise H_0 (x ve y değişkenleri arasında korelasyon yoktur.) hipotezi reddedilir, H_1 (x ve y değişkenleri arasında korelasyon vardır.) hipotezi kabul edilir.

ÖRNEK 3

Örnek 8.1'de tanımlanan problem için evren korelasyon değeri sıfıra eşit midir? %5 anlam düzeyine göre test ediniz.

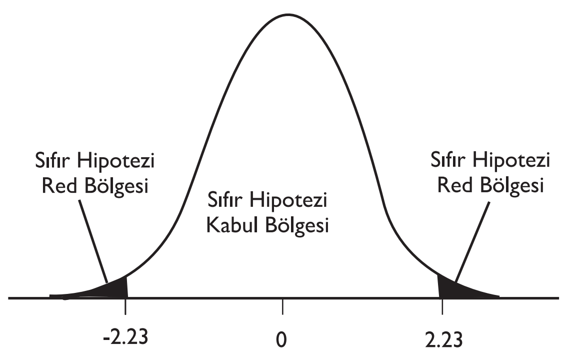
İlk olarak hipotezleri yazalım.

H_0: $\rho = 0$ (x ve y değişkenleri arasında korelasyon yoktur.)
H_1: $\rho \neq 0$ (x ve y değişkenleri arasında korelasyon vardır.)

Daha sonra tablodan gerekli kritik değerler elde edilir. t dağılımı tablosu yardımıyla %5 anlam düzeyi ve n-2=12-2=10 serbestlik derecesi için elde edilen kritik değerler ile H_0 hipotezi (Evrendeki x ve y değişkenleri arasında korelasyon yoktur.) red bölgeleri Şekil 8.3'teki gibidir.

Şekil 8.3

Korelasyon Katsayısı Anlamlılık Testi Kabul ve Red Bölgeleri



Bu aşamadan sonra test için gerekli olan gözlemlenen t istatistiği hesaplanır. Bu değer

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,733\sqrt{12-2}}{\sqrt{1-(0,733)^2}} = \frac{2,3179}{0,6802} = 3,4075$$

olacaktır. Hesaplanan 3,4075 değeri Şekil 8.3'de gösterilen kritik değerler aralığı dışında yer aldığından H_0 hipotezi (Evrendeki x ve y değişkenleri arasında korelasyon yoktur.) hipotezi ret edilir, H_1 (Evrendeki x ve y değişkenleri arasında korelasyon vardır.) hipotezi kabul edilir.

BASİT DOĞRUSAL REGRESYON ANALİZİ

Önceki kısımlarda Pearson korelasyon katsayısı yardımıyla iki değişken arasındaki doğrusal ilişkinin derecesi tespit edilmiştir. İki değişken arasındaki ilişki bir modelle açıklanmak istenebilir. İstatistiksel anlamda iki değişken arasındaki ilişki denildiğinde, bu değişkenlerin değerlerinin karşılıklı değişimleri arasında bir neden sonuç ilişkisi veya bir bağlılık ilişkisi anlaşılır. Regresyon kelimesi, bir değişkenle bir başka değişken arasında ilişki kurma işini ve ilişkinin biçimini anlatır. Bağımsız değişken x'in değerleri değişirken, buna bağlı olarak bağımlı değişken y'nin değerleri de değişiyorsa bu iki değişken arasında ilişki olduğu söylenebilir. Değişkenler arasındaki bu ilişkinin matematiksel bir fonksiyonla ifade edilmesi, regresyon analizinin konusunu oluşturur. İki değişken arasındaki ilişkinin gösteriminde kullanılan fonksiyona veya eşitliğe *regresyon denklemi* adı verilir. Değişkenler arasındaki ilişki için gerekli fonksiyonun veya eşitliğin hesaplanması ve tahminlerin oluşturulması işlemi *regresyon analizi* olarak adlandırılır. Regresyon analizinde değişkenler arasındaki ilişki doğrusal olabileceği gibi eğrisel de olabilir. Eğer değişkenler arasındaki ilişki doğrusal ise doğrusal regresyon analizi, değilse doğrusal olmayan regresyon analizi adını alır. Regresyon denklemi, bir bağımlı ve bir bağımsız değişkenden meydana geliyorsa ve değişkenler arasındaki ilişki doğrusal ise yapılan regresyon analizi basit doğrusal regresyon analizi adını alır. Bu ünitede sadece değişkenler arasındaki doğrusal ilişki incelenecektir ve doğrusal modelin nasıl tahminleneceği konusu işlenecektir. Doğrusal regresyon denkleminin tahmini için bir çok teknik kullanılmaktadır. En yaygın kullanılan teknik olan en küçük kareler tekniği bu ünite içerisinde ele alınacaktır.

n hacimli örneklemden elde edilen veri setinde x ve y değişkenleriyle ilgili gözlem değerleri yer alır. x ve y değişkenleri arasında var olan kuramsal ilişki, izleyen doğrusal model yardımıyla araştırılabilir. Basit doğrusal regresyon modeli,

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

olarak yazılır. Model'de

y_i ; bağımlı değişken y'nin i'inci gözlem değerini,

x_i ; bağımsız değişken x'in i'inci gözlem değerini,

ε_i ; i'inci gözlem için ortalaması sıfır ve tüm gözlemler için sabit σ standart sapmalı normal dağılıma sahip olduğu varsayılan rassal hatayı,

α ve β ; tahminlenecek parametre değerlerini gösterir. α, x=0 olduğunda y'nin alacağı değeri, β, x'te meydana gelecek birim değişikliğin y'deki oransal değişmeyi göstermektedir. Ancak uygulamada her zaman evren değerlerinin tamamına ulaşmak mümkün olmadığından örneklemeye başvurulur, örneklem için x ve y değişkenleri arasındaki kuramsal ilişki aşağıdaki doğrusal tahmin denklemi ile araştırılabilir:

$$y_i = a + bx_i + e_i$$

Bu modelde:

y_i ; bağımlı değişken y'nin i'inci gözlem değerini,

x_i ; bağımsız değişken x'in i'inci gözlem değerini,

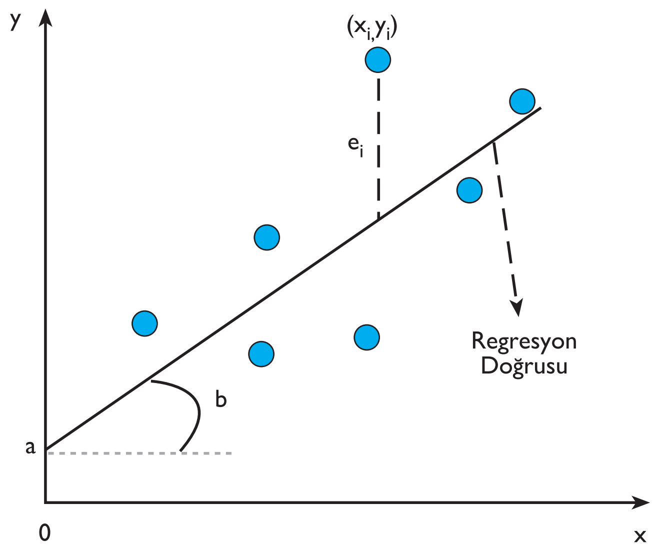
e_i ; i'inci gözlem için ortalaması sıfır ve tüm gözlemler için sabit σ standart sapmalı normal dağılıma sahip olduğu varsayılan rassal hatayı,

a ve b; sırasıyla α ve β'nın tahmin değerlerini gösterir. Benzer şekilde a, x=0 olduğunda y'nin değerini, b ise x'te meydana gelecek birim değişikliğin y'deki oransal etkisini gösterir.

Şekil 8.4 tahmin edilen regresyon modelinin grafiksel gösterimidir. Burada modelin parametre tahminleri için en küçük kareler analizi kullanılmıştır. En küçük kareler analizi hataları en küçükleme tekniğidir. Bu teknikte gözlemlenen y_i değerleri ile tahminlenen modelin sonuçları $\hat{y}_i$ olan değerleri arasındaki düşey farkların kareleri toplamının en küçüklenmesi sağlanır. En küçük kareler tekniği yardımıyla, basit doğrusal tahmin regresyon modelinde yer alan a ve b tahminleri yapılır. Bu modelde b, regresyon doğrusunun eğimini gösterirken a'da doğrunun y eksenini kesim noktasını gösterir. e_i ise ($e_i = y_i - \hat{y}_i$) gerçek y_i değerleri ile tahmin sonucu elde edilen $\hat{y}_i$ değerleri arasındaki düşey farklardır.

Şekil 8.4

Tahmin Edilen Regresyon Modelinin Grafiksel Gösterimi



En küçük kareler tekniği yardımıyla α ve β parametrelerinin tahminleri olan a ve b'nin hesaplanması için,

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

veya araştırmada ilgilenilen x ve y değişkenlerinin standart sapmalarının bilinmesi durumunda ise

$$b = r\frac{s_y}{s_x}$$

ve

$$a = \bar{y} - b\bar{x}$$

eşitlikleri kullanılır. Tahminlenen doğrusal regresyon denklemi,

$$\hat{y}_i = a + bx_i$$

olarak yazılır. Parametre değerleri tahminlenen regresyon doğrusu Şekil 8.4'teki gibi gösterilir. Araştırmacı tahminlenen regresyon denklemi yardımıyla herhangi bir x değeri için y'nin alacağı değeri tahmin edebilir. Denklemde x yerine ilgilendiği değeri yazan araştırmacı, y'nin modele göre beklenen değerini hesaplamış olur. Benzer şekilde y'nin bir değeri için x'in modele göre beklenen değeri de hesaplanabilir.

a parametresi regresyon doğrusunun y eksenini kesim noktasını gösterirken b parametresi doğrunun eğimini göstermektedir.

Bir regresyon analizi tercih edildiğinde, regresyon çözümlemesinden elde edilen sonuçların güvenilirliliği aşağıdaki beş varsayımın geçerliliğine bağlıdır. Eğer varsayımlar geçerli değilse diğer regresyon modelleri ve çözümlemesi yaklaşımlarına başvurulmalıdır. Bu varsayımlar:

b için bir başka eşitlikte olarak yazılabilir.

$$b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}$$

olarak yazılabilir.

- ε rassal değişkeni x'in değerlerinden istatistiksel olarak bağımsızdır.
- ε rassal değişkeni normal dağılıma sahiptir.
- ε rassal değişkeninin sıfır aritmetik ortalamaya sahiptir.
- ε_i ve ε_j gibi birbirinden farklı iki hatanın istatistiksel olarak bağımsız olmalıdır.
- ε_i rassal değişkenleri x_i'lerin tüm değerleri için sabit bir varyansa sahiptir,

şeklinde sıralanırlar.

ÖRNEK 4

Tablo 8.3, Ulaştırma Bakanlığı raporundan elde edilen, yıllara göre beton asfalt onarım uzunluğu (km) (BAKOU) x_i bağımsız değişken ve onarım maliyeti (Milyon ₺) (OM) y_i bağımlı değişken olmak üzere, gözlem değerlerini göstermektedir. Basit doğrusal regresyon denklemini tahminleyiniz ve grafiğini çiziniz.

Yıllar	2005	2006	2007	2008	2009	2010
BAKOU (x_i)	1603	711	1404	654	1597	742
OM (y_i)	513	170	371	144	455	165

Tablo 8.3
Regresyon Doğrusu İçin Gözlemler

Öncelikle ilgilenilen her iki değişken için Tablo 8.3'teki veriler kullanılarak aritmetik ortalama değerlerini hesaplayalım.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{1603+711+...+1597+742}{6} = \frac{6711}{6} = 1118{,}50$$

ve

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{513+170+...+455+165}{6} = \frac{1818}{6} = 303$$

Şimdi bu ortalamaları kullanarak en küçük kareler tekniğini uygulayalım. Tablo 8.4'te görüldüğü gibi x_i bağımsız değişkenine ait gözlem değerleri, bu değişkene ait aritmetik ortalama değerinden çıkartılmıştır. Benzer şekilde y_i bağımlı değişkenine ait gözlem değerleri, bu değişkene ait aritmetik ortalama değerinden çıkartılmıştır. Bulunan değerler bir sonraki aşamada birbirleri ile çarpılmıştır. Daha sonra x_i bağımsız değişkenine ait gözlem değerleri, bu değişkene ait aritmetik ortalama değerlerinin kareleri alınmıştır.

Tablo 8.4
Regresyon Denklemi Hesaplamaları Tablosu.

Yıllar	x_i	y_i	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(x_i-\bar{x})\ (y_i-\bar{y})$	$(x_i-\bar{x})^2$	$\hat{y}_i$	$e_i = y_i-\hat{y}_i$
2005	1.603	513	484,50	210	101.745,00	234.740,25	472,091	40,909
2006	711	170	-406,50	-133	54.064,50	165.242,25	160,783	9,217
2007	1.404	371	286,50	68	19.482,00	82.082,25	402,64	-31,64
2008	654	144	-463,50	-159	73.696,50	214.832,25	140,89	3,11
2009	1.597	455	479,50	152	72.884,00	229.920,25	469,997	-14,997
2010	742	165	-376,50	-138	51.957,00	141.752,25	171,602	-6,602
Toplam	6.711	1.818			373.829,00	1.068.569,50		

Tablo yardımıyla regresyon denklemi katsayıları sırasıyla,

$$b=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}=\frac{373.829,00}{1.068.569,50}=0,349$$

ve

$a = \bar{y}-b\bar{x} = 303 - (0,349)\ (1118,50) = -87,356$

değerlerine eşittir. Bu değerler yardımıyla basit doğrusal regresyon denklemi tahmini,

$\hat{y}_i = -87,356 + 0,349x_i$

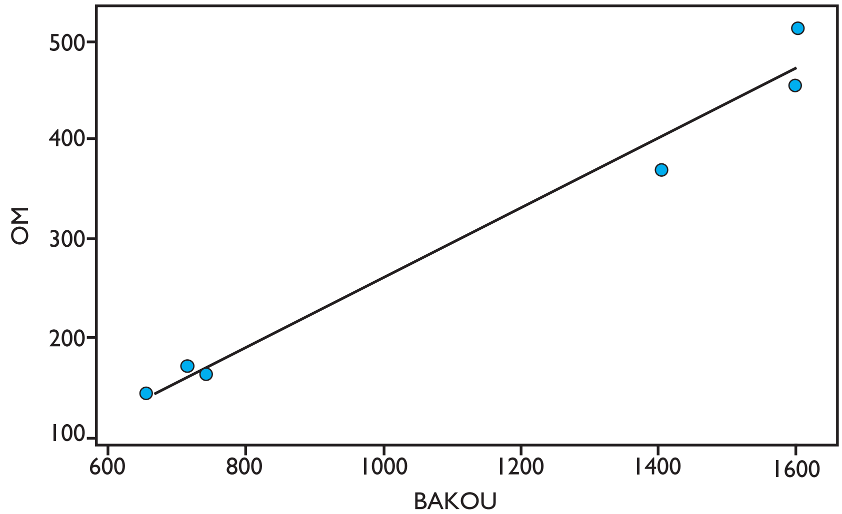
şeklinde yazılır. Tahmin edilen basit doğrusal regresyon denkleminde, x_i'nin sıfır değerini almasıyla $\hat{y}_i$ değeri -87,356 değerine eşit olur. Diğer bir ifadeyle tahmin edilen basit doğrusal regresyon denklemi y ekseniyle -87,356 noktasında birleşir. x_i'deki bir birimlik artış, $\hat{y}_i$ değerini 0,349 oranında artışla etkileyecektir.

Regresyon denklemi yardımıyla araştırmacı gözlemlediği herhangi bir x değeri için y'nin alacağı değeri tahmin edebilir. Örneğin x=1500 olarak gözlemlenmiş ise y'nin tahmini,

$\hat{y}_i = -\ 87,356 + 0,349x = -87,356 + (0,349)\ (1500) = 436,144$

olur. Hesaplanan 436,144 değeri; 1500 km için yaklaşık ₺436,144 milyon onarım maliyeti olduğunun tahminidir. Benzer şekilde y'nin bir değeri için x değeri de tahmin edilebilir. Şekil 8.5, problemin dağılım grafiği ve tahmin edilen basit doğrusal regresyon doğrusu göstermektedir. Regresyon doğrusunun çizilebilmesi için gözlemlenen y değerlerine karşılık gelen regresyon denklemi tahmin değerlerinin tespit edilmesi gereklidir.

Şekil 8.5



Dağılım Grafiği ve Regresyon Doğrusu

Tahminin Standart Hatası

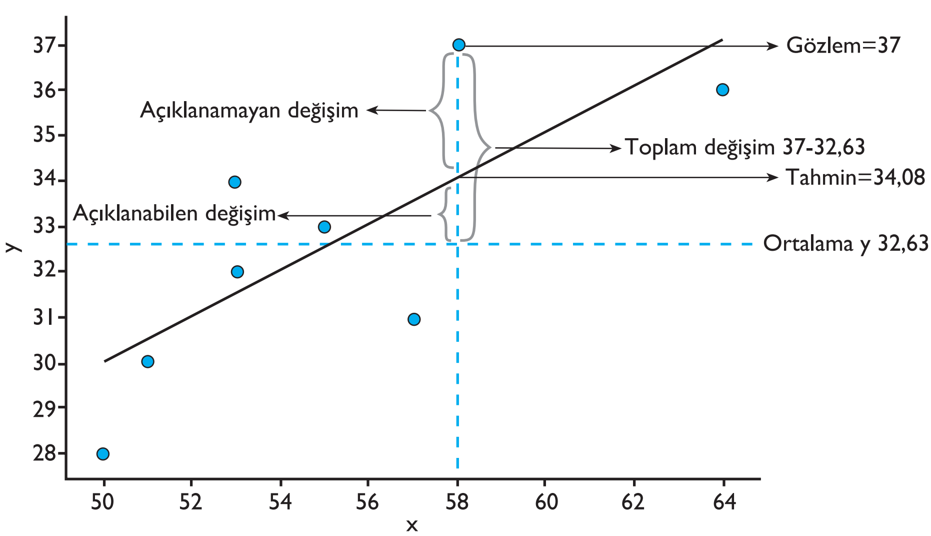
Tahminin standart hatası kavramını inceleyebilmek için öncelikle regresyon analizinde ele alınan toplam değişkenliğin bileşenlerini incelemekte fayda vardır. Regresyon analizinde y_i ile y'lerin ortalaması $\bar{y}$ arasındaki fark toplam değişim olarak adlandırılır. Bu değişim miktarı iki bileşene kolaylıkla ayrılabilir. Bu bileşenler açıklanabilen değişim ve açıklanamayan değişim bileşenleridir. Açıklanamayan değişim $e_i = y_i - \hat{y}_i$ olarak ifade edilir ve i'inci gözlemin hata terimi olarak adlandırılır. Açıklanabilen değişim $\hat{y}_i - \bar{y}$ olacaktır ve i'inci gözlemin regresyon denklemi tarafından açıklanan kısmını temsil edecektir. Genel bir ifade ile toplam değişim için,

Toplam değişim = Açıklanamayan değişim + Açıklanan değişim

$$(y_i - \bar{y}) = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$$

yazılabilir. Toplam değişim için oluşturulan bu yapı Şekil 8.6'da sekiz birimlik bir regresyon problemi kullanılarak grafik üzerinde gösterilmiştir.

Şekil 8.6



Açıklanan ve Açıklanamayan Değişim

Bileşenlerine ayrılan toplam değişim kavramından faydalanarak sapmaların kareleri toplamı,

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

eşitlikleri yardımıyla oluşturulur. Bu eşitliğin sol tarafı toplam değişkenlik ya da genel kareler toplamı (GKT) olarak ifade edilir. Eşitliğin sağ tarafında yer alan ilk toplam değeri açıklanamayan değişkenliktir ve hata kareler toplamı (HKT) olarak adlandırılır. Eşitliğin son bileşeni ise açıklanan değişkenliktir ve regresyon kareler toplamı (RKT) olarak adlandırılır. Bu ifadeleri kullanarak toplam değişkenlik,

GKT = HKT + RKT

şeklinde ifade edilebilir. Toplam değişkenliğin bileşenlerine ayrılmasının en büyük faydası hata kareler toplamı büyüklüğü bakımından gözlemlenen ve model yardımıyla hesaplanan değerler arasındaki uyumun iyiliğine bakabilmesidir. Eğer mükemmel uyum var ise HKT=0 olacaktır.

Regresyon doğrusu için hesaplanacak olan regresyon doğrusu değişkenliği hata kareler toplamı değerinin serbestlik derecesine bölünmesi ile elde edilecektir. y için tahmin değerleri belirlenmeden önce modelde yer alan iki parametrenin (α ve β'nın kestiricileri olan a ve b) tahmin edilmesi gerektiğinden burada serbestlik derecesi n-2 olacaktır. Tahminin standart hatası,

$$s_e = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} = \sqrt{\frac{\sum_{i=1}^{n}(e_i)^2}{n-2}}$$

eşitliği yardımıyla hesaplanır. Tahminin standart hatası yorumlanırken hataların Normal dağıldığı varsayımından faydalanılabilir. Hatırlanırsa normal dağılımda terimlerin %68,30'u aritmetik ortalamadan bir standart sapma uzaklıkta yer alacaklardır. Bu bilgi kullanılarak veri ile hesaplanan regresyon doğrusu hakkında çıkarsama yapılabilir. Eğer aynı veri seti için iki adet regresyon doğrusu hesaplanmış ise bu doğrulardan daha küçük standart hataya sahip olan kullanılmalıdır.

Örneklem Regresyon Doğrusunun Anlamlılık Testi

İlgilenilen iki değişken arasındaki doğrusal ilişki için en küçük kareler tekniği kullanılarak bir regresyon doğrusu tahmini işlemlerini buraya kadar yürüttükten sonra dikkatin "y_i değerlerini tahmin ederken x bağımsız değişkeninin değerlerini bilmenin gerçekten faydası var mıdır?" sorusunun cevabının araştırılmasına verilmesi gerekmektedir. Örneğin doğrunun eğimini veren β katsayısı sıfıra eşit ise ya da istatistiksel olarak test edilerek sıfıra eşit olarak bulunur ise modelden parametrenin çıkartılması gerekecektir. Dolayısıyla, x değişkenine ihtiyaç kalmayacaktır. Bu durumda evren regresyon doğrusu $\hat{y} = \bar{y}$ olacak şekilde düz bir doğru olacaktır. Eğer β değeri sıfıra eşit değilse y'nin değerlerinin tahmininde x değişkeni kullanılabilecektir. Bundan dolayı y'nin değerlerinin tahmininde regresyon doğrusu kullanımının faydası olup olmadığını görmek için β =0, sıfır hipotezinin test edilmesi gerekir. Alternatif hipotez ise β'nın sıfırdan büyük ya da küçük olmasına göre kurulabileceği gibi genellikle $\beta \neq 0$ olacak şekilde çift yönlü olarak kurulur. Hipotezler,

$H_0 : \beta = 0$ (Regresyon doğrusu anlamlı değildir.)
$H_1 : \beta \neq 0$ (Regresyon doğrusu anlamlıdır.)

şeklinde yazılırlar. Testin yürütülmesinde t testi kullanılır. Evren β değerini tahmin etmek için kullandığımız b regresyon katsayısının standart hatası,

$$s_b = s_e\sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

eşitliği yardımıyla hesaplanır. Sıfır hipotezini test etmek için hesaplanacak olan t istatistiği,

$$t = \frac{b}{s_b}$$

n-2 serbestlik derecesi ile belirli anlam düzeyine göre tablodan tespit edilen kritik değer ile karşılaştırılacaktır. Sıfır hipotezini test etmek için hesaplanacak olan t istatistiği n-2 serbestlik derecesi ile belirli anlam düzeyine göre t dağılımı tablosundan tespit edilen kritik değerden büyükse sıfır hipotezi reddedilir. Diğer bir ifadeyle evren β değerini tahmin etmek için kullandığımız b regresyon katsayısı istatistiksel olarak anlamlıdır, yani elde edilen regresyon doğrusu amaca uygun olarak kullanılabilir. Eğer sıfır hipotezini test etmek için hesaplanacak olan t istatistiği n-2 serbestlik derecesi ile belirli anlam düzeyine göre t dağılımı tablosundan tespit edilen kritik değerden küçükse sıfır hipotezi kabul edilir. Diğer bir ifadeyle evren β değerini tahmin etmek için kullandığımız b regresyon katsayısı anlamlı değildir, yani elde edilen regresyon doğrusu kullanılamaz denilebilir.

β'nın sıfıra eşit olup olmadığının test edilmesine ek olarak β için güven aralığı da tespit edilebilir. Regresyon katsayısı b için n-2 serbestlik derecesi ve s_b standart sapması ile t dağılımı uyumu bilindiğine göre evren regresyon doğrusu eğimi β'nın güven aralığı,

$$b - t_{\frac{a}{2}, n-2}\, s_b \leq \beta \leq b + t_{\frac{a}{2}, n-2}\, s_b = 1 - a$$

şeklinde verilir.

ÖRNEK 5

Bir şirkette çalışan araştırmacı, şirketinin vermiş olduğu reklam harcamaları (bin ₺) (x_i) ile şirketin satışları (bin ₺) (y_i) arasındaki ilişkiyi araştırmaktadır. Bu amaçla seçilen 4 ay için Tablo 8.5'teki gözlem değerlerini alarak regresyon modeli oluşturmuştur. Basit doğrusal regresyon denklemini tahminleyiniz. Tahminlenen basit doğrusal regresyon doğrusundaki b katsayısının anlamlılığını sınayınız. Evren β değeri için güven aralığını hesaplayınız. Anlamlılık düzeyi olarak %5 alınız.

Tablo 8.5
Regresyon Doğrusu İçin Gözlemler

Aylar	Haziran	Temmuz	Ağustos	Eylül
Reklam Harcamaları (bin ₺) (x_i)	2	1	3	4
Şirketin Satışları (bin ₺) (y_i)	7	3	8	10

Öncelikle ilgilenilen her iki değişken için Tablo 8.5'teki veriler kullanılarak aritmetik ortalama değerlerini hesaplayalım.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2+1+3+4}{4} = \frac{10}{4} = 2,50$$

ve

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{7+3+8+10}{4} = \frac{28}{4} = 7$$

Şimdi bu ortalamaları kullanarak en küçük kareler tekniğini uygulayalım. Tablo 8.6'da görüldüğü gibi x_i bağımsız değişkenine ait gözlem değerleri, bu değişkene ait aritmetik ortalama değerinden çıkartılmıştır. Benzer şekilde y_i bağımlı değişkenine ait gözlem değerleri, bu değişkene ait aritmetik ortalama değerinden çıkartılmıştır. Bulunan değerler bir sonraki aşamada birbirleri ile çarpılmıştır. Daha sonra x_i bağımsız değişkenine ait gözlem değerleri, bu değişkene ait aritmetik ortalama değerlerinin kareleri alınmıştır.

Tablo 8.6
Regresyon Doğrusu Hesaplamaları Tablosu.

Aylar	x_i	y_i	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$	$\hat{y}_i$	$e_i = y_i-\hat{y}_i$	$(e_i)^2$
Haziran	2	7	-0,5	0	0	0,25	5,9	1,1	1,21
Temmuz	1	3	-1,5	-4	6	2,25	3,7	-0,7	0,49
Ağustos	3	8	0,5	1	0,5	0,25	8,1	-0,1	0,01
Eylül	4	10	1,5	3	4,5	2,25	10,3	-0,3	0,09
Toplam	10	28			11	5			1,8

Tablo yardımıyla regresyon denklemi katsayıları sırasıyla,

$$b = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{11}{5} = 2,2$$

ve

$a = \bar{y} - b\bar{x} = 7 - (2,2)\ (2,5) = 1,5$

değerlerine eşittir. Bu değerler yardımıyla basit doğrusal regresyon denklemi tahmini,

$\hat{y}_i = 1,5 + 2,2x_i$

şeklinde yazılır. Tahmin edilen basit doğrusal regresyon denkleminde, x_i'nin sıfır değerini almasıyla $\hat{y}$ değeri 1,5 değerine eşit olur. Diğer bir ifadeyle tahmin edilen basit doğrusal regresyon denklemi y ekseniyle 1,5 noktasında birleşir. x_i'deki bir birimlik artış, $\hat{y}$ değerini 2,2 oranında artışla etkileyecektir.

Tahminlenen regresyon modelinde x_i değerleri yerine konulduğunda Tablo 8.6'daki $\hat{y}_i$ değerleri ve gözlenen yi değerlerinden $\hat{y}_i$ değerleri çıkartıldığında $e_i = y_i - \hat{y}_i$ değerleri elde edilmiş olur.

Tahminlenen regresyon denklemine dayanarak yapılacak tahminlerin hata düzeyi standart hata ölçüsü ile belirlenir. Standart hata, hata kareler toplamının $((e_i)^2)$ değeri serbestlik derecesine bölünmesi ile elde edilir.

$$s_e = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} = \sqrt{\frac{\sum_{i=1}^{n}(e_i)^2}{n-2}} = \sqrt{\frac{1,8}{4-2}} = 0,95$$

Tahminlenen regresyon modelinin standart hatası ₺0,95 olarak hesaplanmıştır. Diğer bir deyişle tahminlenen regresyon denklemine dayanarak yapılacak satış tahminleri ₺0,95'lik ortalama bir hataya sahip olacaktır.

Modelin uygunluğunun araştırılması, bir başka ifade ile regresyon katsayısı b'nin anlamlılığının test edilmesi gerekir. Yapılacak test sürecinin aşamaları aşağıdaki gibidir. Elde edilen örneklem regresyon doğrusunun sınanması için hipotezler aşağıdaki gibi yazılabilir.

$H_0 : \beta = 0$ (Regresyon doğrusu anlamlı değildir.)
$H_1 : \beta \neq 0$ (Regresyon doğrusu anlamlıdır.)

Evren β değerini tahmin etmek için kullandığımız b regresyon katsayısının standart hatası ve sıfır hipotezini test etmek için hesaplanacak olan t istatistiği sırasıyla aşağıdaki gibi hesaplanır.

$$s_b = s_e \sqrt{\frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = 0,95\sqrt{\frac{1}{5}} = 0,42$$

$$t = \frac{b}{s_b} = \frac{2,2}{0,42} = 5,24$$

α =0,05 ve n-2=4-2=2 serbestlik derecesi ile $t_{0,05;2}$=4,303 hesaplanan t istatistiği 5,24'den küçük olduğundan sıfır hipotezi reddedilecektir. Diğer bir ifadeyle evren β değerini tahmin etmek için kullandığımız b regresyon katsayısı istatistiksel olarak anlamlıdır, yani x değişkeni y değişkenini açıklamada önemli bir açıklayıcı değişkendir. Satışların kestiriminde reklam harcamaları değişkeni açıklayıcı değişken olarak modelde yer almalıdır.

Evren β değeri için güven aralığı aşağıdaki gibi aşağıdaki gibi hesaplanabilir.

$$b - t_{\frac{a}{2};n-2}\, s_b \leq \beta \leq b + t_{\frac{a}{2};n-2}\, s_b = 1 - a$$

$$(2,2 - (4,303)\,(0,42) \leq \beta \leq 2,2 + (4,303)\,(0,42)) = 0,95$$

$$(0,39 \leq \beta \leq 4,007) = 0,95$$

Buna göre evren regresyon doğrusu eğimi β'nin 0,95 olasılıkla alabileceği değerler 0,39 ile 4,007 arasında olacaktır.

SIRA SİZDE 2

Tablo 8.7'de bağımlı değişken y_i ve bağımsız değişken x_i gözlem değerleri verilmiştir. Basit doğrusal regresyon denklemini tahminleyiniz. Tahminlenen basit doğrusal regresyon doğrusundaki b katsayısının anlamlılığını sınayınız. Evren β değeri için güven aralığını hesaplayınız. Anlamlılık düzeyi olarak %5 alınız.

Değişkenler	1	2	3	4	5
Bağımsız Değişken (x_i)	14	7	3	15	11
Bağımlı Değişken (y_i)	6	5	3	9	7

Tablo 8.7
Regresyon Doğrusu için Gözlemler

Özet

AMAÇ 1

Değişkenler arasındaki ilişkinin derecesini tespit etmek.

Çoğunlukla araştırmacılar ele aldıkları iki değişken artasındaki ilişkinin nasıl olduğunu araştırmak istemektedir. Örneğin "İlgilenilen a değişkenindeki değişim b değişkeninde de aynı etkiyi gösterecek mi?", "a değişkeninin değeri artarsa b değişkeninin de değeri artar mı?" ve benzeri sorular üzerinde durulabilir. Böyle durumlarda iki değişken arasındaki ilişkinin yönünü ve derecesini belirtebilmek için Pearson korelasyon katsayısını hesaplayabilirsiniz.

AMAÇ 2

Bağımlı değişkendeki değişimin yüzde kaçının bağımsız değişken tarafından açıklandığını tespit etmek.

Araştırmacılar, bağımlı değişkende meydana gelen değişimin ne kadarının bağımsız değişken tarafından açıklandığını bilmek isteyebilir. Bu ilişkiyi gösteren değere belirlilik katsayısı adı verilir. Belirlilik katsayısını Pearson korelasyon katsayısı r'nin karesini alarak hesaplayabilirsiniz. 0 ile 1 arasında değerler alan belirlilik katsayısını yorumlayabilirsiniz.

AMAÇ 3

Bir bağımlı ve bir bağımsız değişken ile model kurmak.

Bir bağımlı değişkende meydana gelen değişkenlik bir bağımsız değişken ile açıklanmak istenebilir. Örneğin bir fabrikadaki işçilerin kıdemleri ile aynı işçilerin haftalık üretimleri üzerindeki etkisi araştırılabilir. Basit doğrusal regresyon analizi yardımıyla bir model oluşturularak işçilerin haftalık üertimleri modellenir.

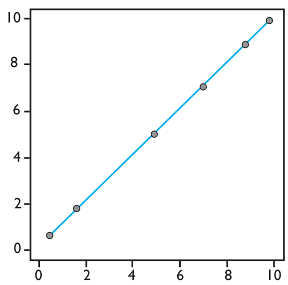
Kendimizi Sınayalım

1. Bir araştırma da ilgilenilen değişkenlerden birisinin değeri artarken diğerinin değeri azalmaktadır. Bu çalışma da Pearson korelasyon katsayısı hesaplanırsa sonuç aşağıdaki seçeneklerden hangisi olabilir?

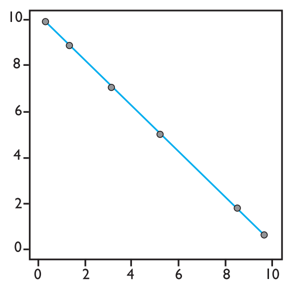
a. 0
b. +1
c. -0,90
d. +0,50
e. Korelasyon katsayısı hesaplanamaz.

2. Aşağıda verilen grafiklerden hangisinde korelasyon katsayısı -1 olacaktır?

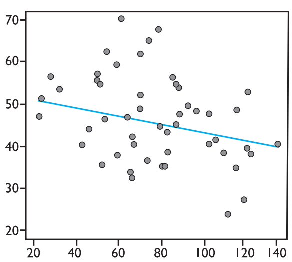
a.



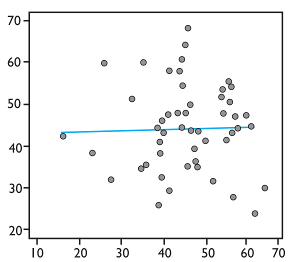
b.



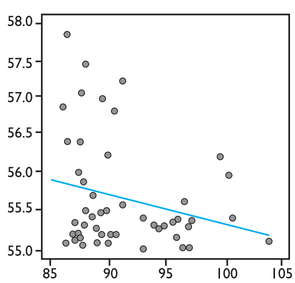
c.



d.



e.



3. İki değişken arasındaki Pearson korelasyon katsayısı r=0,68 olarak hesaplanmıştır. Bu sonuca göre aşağıdaki ifadelerden hangisi doğrudur?

a. İki değişken arasında korelasyon yoktur.
b. İki değişken arasında zayıf artı korelasyon vardır.
c. İki değişken arasında orta artı korelasyon vardır.
d. İki değişken arasında güçlü artı korelasyon vardır.
e. İki değişken arasında tam/mükemmel artı korelasyon vardır.

4. y=23+(7,23)x olarak verilen basit doğrusal regresyon denkleminde regresyon doğrusunun eğimi hangi değere eşittir?

a. (23)/(7,23)
b. 7,23
c. 23
d. (7,23)/(23)
e. Hesaplanamaz.

5. Bir araştırmada basit doğrusal regresyon denklemi y=(1,41)-6x olarak tespit edilmiştir. Denkleme göre x'in değeri 2,75 olursa y'nin değeri ne olur?

a. 7,41
b. 12
c. -4,59
d. 15
e. -15,09

6. Bir araştırmada basit doğrusal regresyon denklemi y=-15+(4,50)x olarak tespit edilmiştir. Denkleme göre y'in değeri 8,55 olursa x'nin değeri ne olur?

a. 5,23
b. 23,55
c. -6,45
d. -3,33
e. 1,90

7. n=6 ve hata kareler toplamı $(e_i)^2$=2,3 için hesaplanacak olan regresyon doğrusu değişkenliği s_e hangi değere eşittir?

a. 26,125
b. 9,53
c. 1,110
d. 0,758
e. 47

8. Evren β değerini tahmin etmek için kullanılan b regresyon katsayısının değeri 5,23 ve bu katsayının standart hata değeri s_b=1,635 olarak verildiğinde sıfır hipotezini test etmek için kullanılan test istatistiği t hangi değere eşittir?

a. 3,198
b. 43
c. 115
d. 1,635
e. 238,56

9. Bir regresyon probleminde açıklayıcı değişken ya da değişkenlerin bağımlı değişkendeki değişkenliğin ne kadarını açıkladığını hesaplamak için hangi katsayıya ihtiyaç duyulur?

a. Korelasyon katsayısı
b. Doğrunun eğimi
c. Belirlilik katsayısı
d. Kontenjans katsayısı
e. Aritmetik ortalama

10. Basit doğrusal regresyon denklemi parametre tahminlerini bulmak için kullanılan eşitlikler hangi teknik yardımı ile geliştirilmiştir?

a. Korelasyon analizi
b. En küçük kareler tekniği
c. Varyans analizi
d. Hipotez testi.
e. Belirlilik katsayısı

Yaşamın İçinden

“Copier Sales of America şirketi Kanada ve Amerika Birleşik Devletlerinde çeşitli büyüklükteki şirketlere fotokopi makinesi satan bir şirkettir. Marcy Bancer, Copier Sales of America şirketinde ulusal satış yöneticiliğine henüz yükselmiş satış yöneticisidir. Yaklaşan satış toplantısına ülke genelindeki tüm satış temsilcileri katılacaktır. Marcy Bancer, satış temsilcilerine satış için her gün müşterileri fazladan aramanın önemini anlatmak istemektedir. Bu amaçla Marcy Bancer, satış için müşterileri arama sayısı ile fotokopi makinesi satış sayısı arasındaki ilişkiyi gösteren bilgiyi satış temsilcileriyle paylaşmaya karar verir. Satış temsilcileri içinden rassal (tesadüfi) olarak seçtiği 10 satış temsilcisinin, son bir ayda yapmış oldukları müşterileri fazladan arama sayılarını ve fotokopi makinesi satış sayılarını ele alır. Acaba rassal olarak belirlenen 10 satış temsilcisinin son bir ayda yapmış oldukları müşterileri fazladan arama sayısı ile fotokopi makinesi satış sayısı arasındaki ilişkinin yönü ve derecesi nedir? Müşterileri fazladan arama sayısı ve fotokopi satış sayısı değişkenlerinden hangisi bağımlı hangisi bağımsız değişkendir? Bağımsız değişken bağımlı değişkeninin ne kadarını açıklamaktadır? Rassal (tesadüfi) olarak seçilen 10 satış temsilcisinin performasına göre gelecek dönem veya dönemler için bir tahminde bulunulabilir mi?”

Kaynak: Lind, D.A., Marchal, W.G. ve Wathen, S.A. (2005). Statistical Techniques in Business and Economics. McGrawn-Hill /Irwin Series. New York.

Kendimizi Sınayalım Yanıt Anahtarı

1. c	Yanıtınız yanlış ise "Korelasyon Analizi" bölümünü yeniden gözden geçiriniz.
2. b	Yanıtınız yanlış ise "Korelasyon Analizi" bölümünü yeniden gözden geçiriniz.
3. d	Yanıtınız yanlış ise "Korelasyon Analizi" bölümünü yeniden gözden geçiriniz.
4. b	Yanıtınız yanlış ise "Basit Doğrusal Regresyon Analizi" bölümünü yeniden gözden geçiriniz.
5. e	Yanıtınız yanlış ise "Basit Doğrusal Regresyon Analizi" bölümünü yeniden gözden geçiriniz.
6. a	Yanıtınız yanlış ise "Basit Doğrusal Regresyon Analizi" bölümünü yeniden gözden geçiriniz.
7. d	Yanıtınız yanlış ise "Basit Doğrusal Regresyon Analizi" bölümünü yeniden gözden geçiriniz.
8. a	Yanıtınız yanlış ise "Tahminin Standart Hatası" bölümünü yeniden gözden geçiriniz.
9. c	Yanıtınız yanlış ise "Korelasyon Analizi" bölümünü yeniden gözden geçiriniz.
10. b	Yanıtınız yanlış ise "Basit Doğrusal Regresyon Analizi" bölümünü yeniden gözden geçiriniz.

Sıra Sizde Yanıt Anahtarı

Sıra Sizde 1

Öncelikle değişkenlerin ortalama ve standart sapmaları hesaplanır. Bu hesaplamalar yapılmadan önce Ünite 3'ün tekrar edilmesi faydalı olacaktır. İlgilenilen değişkenlerin ortalama ve standart sapmaları,

$\bar{x} = 11{,}00 \qquad \bar{y} = 120{,}83 \qquad s_x = 5{,}13 \qquad s_y = 73{,}75$

olarak hesaplanır. Örnek 8.1'de Pearson korelasyon katsayısı 0,733 olarak hesaplanmıştır. Şimdi bu değeri ortalama ve standart sapmalar yardımıyla tekrar hesaplayalım. Bu hesaplamayla ilgili aşağıdaki tablo hazırlanmıştır.

Aylar	Yağmurlu Gün Sayısı (x)	Şemsiye Satış Miktarı (y)	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(x_i-\bar{x})(y_i-\bar{y})$
1	15	150	4	29,17	116,68
2	8	100	-3	-20,83	62,49
3	18	200	7	79,17	554,19
4	10	120	-1	-0,83	0,83
5	20	175	9	54,17	487,53
6	10	70	-1	-50,83	50,83
7	3	30	-8	-90,83	726,64
8	5	40	-6	-80,83	484,98
9	6	70	-5	-50,83	254,15
10	11	95	0	-25,83	0
11	13	110	2	-10,83	-21,66
12	13	290	2	169,17	338,34
Toplam	132	1.450			3.055

Pearson korelasyon katsayısı,

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{(n-1)s_x s_y} = \frac{3.055}{11(5{,}13)(73{,}75)} = 0{,}733$$

olarak hesaplanır.

Sıra Sizde 2

Bağımlı değişken y_i ve bağımsız değişken x_i için aritmetik ortalamaları hesaplayalım.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{14+7+3+15+11}{5} = \frac{50}{5} = 10$$

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{6+5+3+9+7}{5} = \frac{30}{5} = 6$$

Regresyon doğrusu hesaplamaları aşağıdaki gibi yapılır.

	x_i	y_i	$(x_i-\bar{x})$	$(y_i-\bar{y})$	$(x_i-\bar{x})\ (y_i-\bar{y})$	$(x_i-\bar{x})^2$	$\hat{y}_i$	$e_i = y_i - \hat{y}_i$	$(e_i)^2$
1	14	6	4	0	0	16	7,6	-1,6	2,56
2	7	5	-3	-1	3	9	4,8	0,2	0,04
3	3	3	-7	-3	21	49	3,2	-0,2	0,04
4	15	9	5	3	15	25	8	1	1
5	11	7	1	1	1	1	6,4	0,6	0,36
Toplam	50	30			40	100			4

Regresyon doğrusunun katsayıları aşağıdaki gibi hesaplanır:

$$b = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{40}{100} = 0,4$$

$a = \bar{y} - b\bar{x} = 6 - (0,4)\ (10) = 2$

Bu hesaplamalar sonucunda regresyon doğrusu denklemi aşağıdaki gibi bulunur.

$\hat{y}_i = 2 + 0,4x_i$

Tahminin standart hatası aşağıdaki gibi hesaplanır.

$$s_e = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} = \sqrt{\frac{\sum_{i=1}^{n}(e_i)^2}{n-2}} = \sqrt{\frac{4}{5-2}} = 1,15$$

Elde edilen örneklem regresyon doğrusunun sınaması için hipotezler aşağıdaki gibi yazılabilir:

$H_0 : \beta = 0$ (Regresyon doğrusu anlamlı değildir.)
$H_1 : \beta \neq 0$ (Regresyon doğrusu anlamlıdır.)

b regresyon katsayısının standart hatası aşağıdaki gibi hesaplanır:

$$s_b = s_e\sqrt{\frac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} = 1,15\sqrt{\frac{1}{100}} = 0,11$$

Sıfır hipotezini test etmek için t istatistiği aşağıdaki gibi hesaplanır:

$$t = \frac{b}{s_b} = \frac{0,4}{0,11} = 3,63$$

$\alpha=0,05$ ve n-2=5-2=2 serbestlik derecesi ile $t_{0,05;2}=3,182$ hesaplanan t istatistiği 3,63'den küçük olduğundan sıfır hipotezi reddedilecektir. Diğer bir ifadeyle evren β değerini tahmin etmek için kullandığımız b regresyon katsayısı istatistiksel olarak anlamlıdır, yani elde edilen regresyon doğrusu amaca uygun olarak kullanılabilir. Evren β değeri için güven aralığı aşağıdaki gibi aşağıdaki gibi hesaplanabilir:

$$b - t_{\frac{a}{2},n-2}\, s_b \leq \beta \leq b + t_{\frac{a}{2},n-2}\, s_b = 1 - a$$

$(0,4 - (3,182)\ (0,11) \leq \beta \leq 0,4 + (3,182)\ (0,11)) = 0,95$

$(0,049 \leq \beta \leq 0,750) = 0,95$

Buna göre evren regresyon doğrusu eğimi β'nin 0,95 olasılıkla alabileceği değerler 0,049 ile 0,750 arasında olacaktır.

Yararlanılan ve Başvurulabilecek Kaynaklar

Bowerman, B.L. ve O'Connell, R.T. (1990). **Linear Statistical Models An Applied Approach,** Second Edition, Thomson Learning.

Cook, R.D. ve Weisberg S. (1999). **Applied Regression Including Computing and Graphics,** Wiley Inter-Science.

Erar, A. (1985). **Bağlanım (Regresyon) Çözümlemesi,** Ders Notları, Ankara.

Freund, J.E. (1992). **Mathematical Statistics,** Prentice Hall International.

Govil, A.K. (1984). **Definitions and Formulae in Statistics,** The Macmillan Press Ltd.

Harnett, D.L. (1982). **Statistical Methods,** Third Edition, aAddison Wesley.

Kanji, G.P. (1993). **100 Statistical Tests,** Sage Publications.

Lind, D.A., Marchal, W.G. ve Wathen, S.A. (2005). **Statistical Techniques in Business and Economics**, McGrawn-Hill /Irwin Series, New York.

Lindeman. R.H., Merenda, P.F. ve Gold, R.Z. (1980). **Introduction to Bivariate and Multivariate Analysis,** Scott, Foresman and Company.

McDonald, J.H. (2008). **Handbook of Biological Statistics,** Sparky House Publishing.

Myers, R.H. (1986). **Classical and Modern Regression with Applications,** Duxbury Press.

Rousseeuw, P.J. ve Leroy, A.M. (1987). **Robust Regression and Outlier Detection,** John Wiley & Sons.

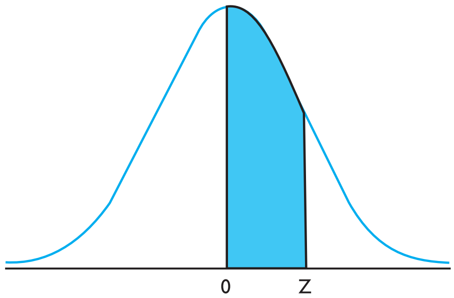
Şıklar, E. İ. (2000). **Regresyon Analizine Giriş,** Anadolu Üniversitesi Yayınları.

Cillov, H. (1993). **İstatistik Metotları,** İ.Ü. Basımevi ve Film Merkezi, İstanbul.

Ulaştırma Bakanlığı Ulaştırma ve Ulaşım Araçları Uygu-Ar Merkezi. (2005). **Ulaştırma Ana Planı Stratejisi Sonuç Raporu,** İstanbul Teknik Üniversitesi, İstanbul.

Ek-1

İSTATİSTİK SORULARININ CEVAPLANMASINDA GEREKLİ OLABİLECEK TABLOLAR



Tablo 1: *Normal Eğri Alanları*

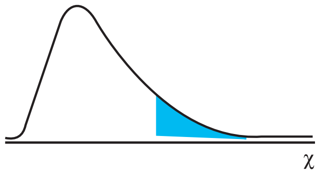
z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,3	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,4	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,5	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,6	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,7	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,8	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,9	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	0,4750	0,4756	0,4761	0,4767
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	0,4951	0,4952
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,7	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,8	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,9	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

Ek-2

Tablo 2: *Kritik t Değerleri Tablosu*

	(Anlamlılık Düzeyi)											
sd	**0,45**	**0,40**	**0,35**	**0,30**	**0,25**	**0,20**	**0,15**	**0,10**	**0,05**	**0,025**	**0,01**	**0,005**
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657
2	0,142	0,289	0,445	0,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925
3	0,137	0,277	0,424	0,584	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841
4	0,134	0,271	0,414	0,569	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604
5	0,132	0,267	0,408	0,559	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032
6	0,131	0,265	0,404	0,553	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707
7	0,130	0,263	0,402	0,549	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499
8	0,130	0,262	0,399	0,546	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355
9	0,129	0,261	0,398	0,543	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250
10	0,129	0,260	0,397	0,542	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169
11	0,129	0,260	0,396	0,540	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106
12	0,128	0,259	0,395	0,539	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055
13	0,128	0,259	0,394	0,538	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012
14	0,128	0,258	0,393	0,537	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977
15	0,128	0,258	0,393	0,536	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947
16	0,128	0,258	0,392	0,535	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921
17	0,128	0,257	0,392	0,534	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898
18	0,127	0,257	0,392	0,534	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878
19	0,127	0,257	0,391	0,533	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861
20	0,127	0,257	0,391	0,533	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845
21	0,127	0,257	0,391	0,532	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831
22	0,127	0,256	0,390	0,532	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819
23	0,127	0,256	0,390	0,532	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807
24	0,127	0,256	0,390	0,531	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797
25	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787
26	0,127	0,256	0,390	0,531	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779
27	0,127	0,256	0,389	0,531	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771
28	0,127	0,256	0,389	0,530	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763
29	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756
30	0,127	0,256	0,389	0,530	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750
∞	0,1257	0,2533	0,3853	0,5244	0,6745	0,8416	1,0364	1,2816	1,6449	1,9600	2,3264	2,5759
sd; *Serbestlik Derecesi*												

Ek-3



Tablo 3: *Ki-kare Tablosu*

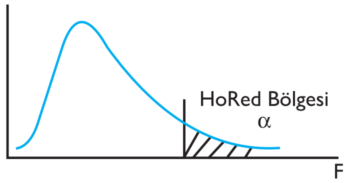
sd \ α	0,995	0,990	0,975	0,950	0,050	0,025	0,010	0,005	0,001
1	0,000039	0,0002	0,0010	0,0039	3,8415	5,0239	6,6349	7,8794	10,8276
2	0,0100	0,0201	0,0506	0,1026	5,9915	7,3778	9,2103	10,5966	13,8155
3	0,0717	0,1148	0,2158	0,3518	7,8147	9,3484	11,3449	12,8382	16,2662
4	0,2070	0,2971	0,4844	0,7107	9,4877	11,1433	13,2767	14,8603	18,4668
5	0,4117	0,5543	0,8312	1,1455	11,0705	12,8325	15,0863	16,7496	20,5150
6	0,6757	0,8721	1,2373	1,6354	12,5916	14,4494	16,8119	18,5476	22,4577
7	0,9893	1,2390	1,6899	2,1673	14,0671	16,0128	18,4753	20,2777	24,3219
8	1,3444	1,6465	2,1797	2,7326	15,5073	17,5345	20,0902	21,9550	26,1245
9	1,7349	2,0879	2,7004	3,3251	16,9190	19,0228	21,6660	23,5894	27,8772
10	2,1559	2,5582	3,2470	3,9403	18,3070	20,4832	23,2093	25,1882	29,5883
11	2,6032	3,0535	3,8157	4,5748	19,6751	21,9200	24,7250	26,7568	31,2641
12	3,0738	3,5706	4,4038	5,2260	21,0261	23,3367	26,2170	28,2995	32,9095
13	3,5650	4,1069	5,0088	5,8919	22,3620	24,7356	27,6882	29,8195	34,5282
14	4,0747	4,6604	5,6287	6,5706	23,6848	26,1189	29,1412	31,3193	36,1233
15	4,6009	5,2293	6,2621	7,2609	24,9958	27,4884	30,5779	32,8013	37,6973
16	5,1422	5,8122	6,9077	7,9616	26,2962	28,8454	31,9999	34,2672	39,2524
17	5,6972	6,4078	7,5642	8,6718	27,5871	30,1910	33,4087	35,7185	40,7902
18	6,2648	7,0149	8,2307	9,3905	28,8693	31,5264	34,8053	37,1565	42,3124
19	6,8440	7,6327	8,9065	10,1170	30,1435	32,8523	36,1909	38,5823	43,8202
20	7,4338	8,2604	9,5908	10,8508	31,4104	34,1696	37,5662	39,9968	45,3147
21	8,0337	8,8972	10,2829	11,5913	32,6706	35,4789	38,9322	41,4011	46,7970
22	8,6427	9,5425	10,9823	12,3380	33,9244	36,7807	40,2894	42,7957	48,2679
23	9,2604	10,1957	11,6886	13,0905	35,1725	38,0756	41,6384	44,1813	49,7282
24	9,8862	10,8564	12,4012	13,8484	36,4150	39,3641	42,9798	45,5585	51,1786
25	10,5197	11,5240	13,1197	14,6114	37,6525	40,6465	44,3141	46,9279	52,6197
26	11,1602	12,1981	13,8439	15,3792	38,8851	41,9232	45,6417	48,2899	54,0520
27	11,8076	12,8785	14,5734	16,1514	40,1133	43,1945	46,9629	49,6449	55,4760
28	12,4613	13,5647	15,3079	16,9279	41,3371	44,4608	48,2782	50,9934	56,8923
29	13,1211	14,2565	16,0471	17,7084	42,5570	45,7223	49,5879	52,3356	58,3012
30	13,7867	14,9535	16,7908	18,4927	43,7730	46,9792	50,8922	53,6720	59,7031
40	20,7065	22,1643	24,4330	26,5093	55,7585	59,3417	63,6907	66,7660	73,4020
50	27,9907	29,7067	32,3574	34,7643	67,5048	71,4202	76,1539	79,4900	86,6608
100	67,3276	70,0649	74,2219	77,9295	124,3421	129,5612	135,8067	140,1695	149,4493

sd; Serbestlik Derecesi

α; Anlam Düzeyi

Ek-4

Anlamlılık düzeyi (α = 0,05)



Tablo 4: *F Tablo Değerleri*

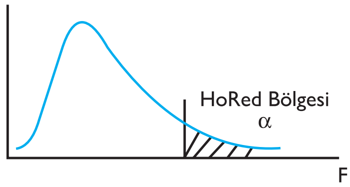
	Serbestlik Derecesi Sd_1										
Sd_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**12**	**24**	**∞**
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,68	4,53	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,00	3,84	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,57	3,41	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,28	3,12	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,07	2,90	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	2,91	2,74	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,79	2,61	2,41
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,69	2,51	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,60	2,42	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,53	2,35	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,48	2,29	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,42	2,24	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,38	2,19	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,34	2,15	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,31	2,11	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,28	2,08	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,25	2,05	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,23	2,03	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,20	2,01	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,18	1,98	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,16	1,96	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,15	1,95	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,13	1,93	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,12	1,91	1,66
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,10	1,90	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,09	1,89	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,00	1,79	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	1,92	1,70	1,39
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	1,88	1,65	1,33
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,85	1,63	1,28
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,83	1,61	1,26
∞	3,84	3,00	2,61	2,37	2,22	2,10	2,01	1,94	1,75	1,52	1,00

Sd_1: Gruplararası serbestlik derecesi

Sd_2: Gruplariçi serbestlik derecesi

Ek-5

Anlamlılık düzeyi (α = 0,01)



Tablo 5: *F Tablo De¤erleri*

	Serbestlik Derecesi Sd_1										
Sd_2	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**12**	**24**	**∞**
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	9,89	9,47	9,02
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,72	7,31	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,47	6,07	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,67	5,28	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,11	4,73	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,71	4,33	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,40	4,02	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,16	3,78	3,36
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	3,96	3,59	3,17
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	3,80	3,43	3,01
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,67	3,29	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,55	3,18	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,46	3,08	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,37	3,00	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,30	2,92	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,23	2,86	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,17	2,80	2,36
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,12	2,75	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,07	2,70	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,03	2,66	2,21
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	2,99	2,62	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	2,96	2,58	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	2,93	2,55	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	2,90	2,52	2,07
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	2,87	2,49	2,04
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	2,84	2,47	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,66	2,29	1,81
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,50	2,12	1,60
80	6,96	4,88	4,04	3,56	3,26	3,04	2,87	2,74	2,42	2,03	1,50
100	6,90	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,37	1,98	1,43
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,34	1,95	1,38
∞	6,64	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,19	1,79	1,0
Sd_1: Gruplararası serbestlik derecesi											
Sd_2: Gruplariçi serbestlik derecesi											