



T.C.
ONDOKUZ MAYIS ÜNİVERSİTESİ

FEN EDEBİYAT FAKÜLTESİ
İSTATİSTİK BÖLÜMÜ

İST.482 PARAMETRİK OLMAYAN
İSTATİSTİKSEL YÖNTEMLER

PROF. DR. YÜKSEL ÖNER

1. Hafta

BÖLÜM 1

1.1 TEMEL KAVRAMLAR

Bu bölümde hatırlatma amaçlı olarak istatistiğin bazı temel kavramları tanıtılacak ve bu kavramlarla ilgili kısa bilgilere yer verilecektir.

Tanım 1.1 (Kitle) Araştırmacının ilgi alanı içerisine giren birimler topluluğuna veya amaca uygun olarak belirli özelliğe/özelliklere sahip olan birimler topluluğuna kitle denir.

Araştırmacının amacı ilgilendiği özellik ya da özellikler açısından kitle hakkında bilgi edinmektir. Kitle sahip olduğu birim sayısı bakımından iki sınıfa ayrılır.

1. Somut Kitle: Sonlu sayıda birime sahip kitleye denir.

2. Soyut Kitle: Kapsadığı birim sayısı sonsuz olan kitleye denir.

- ✓ Ondokuz Mayıs Üniversitesine hali hazırda kayıtlı olan öğrenciler (somut kitle)
- ✓ Tıp Fakültesi hastanesine muayene olmak amacıyla Ağustos ayında başvuran hastalar (somut kitle)
- ✓ Karadeniz Bölgesinde yetişen A-türü bitkiler (soyut kitle)
- ✓ Ondokuz Mayıs Üniversitesine geçmişte kayıt yaptırmış, hali hazırda kayıtlı olan ve gelecekte kayıt yaptıracak olan öğrenciler (soyut kitle)

Deneyisel araştırmalarda geçerli olan bir kitle türü de **hipotetik kitledir**. Bu tür kitlede birimler gerçekte yoktur, fakat zihinde canlandırılabilir. Örneğin yeni geliştirilen bir ilaç tesadüfi olarak seçilen n birimlik hasta grubuna uygulandığında, sanki hastaların tamamına bu ilacın uygulandığı ve n birimlik örneğin bu yapıda bir kitleden çekildiği kabul edilir.

Tanım 1.2 (Örnek) Kitledeki birimler içerisinden tesadüfi olarak seçilen ve kitlenin özelliklerini taşıdığı kabul edilen az sayıda birimden oluşan gruba örnek denir.

Kitleden örnek çekmenin amacı, örnekteki bilgileri kullanarak kitle ile ilgili karakteristikler hakkında tahmin veya hipotez testi yoluyla bilgi edinmektir.

Tanım 1.3 (Parametre) Kitleyi tanımak amacıyla bir ya da daha fazla değişken için kitledeki birimler üzerinde yapılan ölçümlerden elde edilen sonuçları kullanarak bazı aritmetik işlemler sonucunda bulunan sayısal karakteristik değerlere **parametre** denir.

Tanım 1.4 (İstatistik) Bir kitleden çekilen rastgele bir örnekteki birimlerin ölçüm değerlerinden hesaplanarak elde edilen örneklem karakteristiklerine **istatistik (tahmin edici)** denir.

Gerçekte parametreler bilinmeyen, değişmez değerler alan sabitler iken, istatistikler örneklemden örnekleme değişen değerler alan değişkenlerdir. Parametreler bilinmedikleri zaman örnekteki bilgi ile tahmin edilirler. μ, σ^2, π ve ρ birer kitle parametresi iken, \bar{X}, S^2, P ve r söz konusu parametrelerin tahmin edicileri olan istatistiklerdir.

Tanım 1.5 (Değişken) İstatistiksel araştırmalarda birimden birime değişen değerler alan özelliklere değişken denir.

Değişkenler farklı kriterler bakımından sınıflandırılabilir.

1. Nitelik-Nicelik Bakımından: Bir değişken ile ilgili yapılan gözlem ya da ölçüm sonuçları için; birimlerin alabileceği değerler sahip oldukları niteliklerle ifade ediliyorsa bu değişkene nitel değişken, birimlerin alabileceği değerler sayısal değerlerle ifade ediliyorsa bu değişkene de nicel değişken adı verilir. Örneğin; boy uzunluğu, ağırlık, aylık gelir, bölümlerdeki öğrenci sayısı, istatistik dersi başarı notu, kan serumunda bulunan albumin miktarı, v.s. birer nicel türden değişken iken, göz rengi, cinsiyet, tercih edilen siyasi parti, öğrencilerin kayıtlı olduğu bölüm, ikamet edilen yer v.s. birer nitel değişkendir.

2. Kesiklilik- Süreklilik Bakımından: Nicel türden değişkenler için kullanılan bir sınıflama çeşididir. Nicel türden bir değişkenin alabileceği değerler sonlu sayıda ya da sayılabilir sonsuzlukta ise bu değişkene kesikli değişken, bir değişkenin alabileceği değerler sonsuz sayıda ise veya belirlenen bir aralıktaki her bir değeri alabiliyorsa bu değişkene sürekli değişken adı verilir. Örneğin; bölümlerdeki öğrenci sayısı, ailelerdeki çocuk sayısı, cep telefonuna kayıtlı kişi sayısı v.s. birer kesikli değişkendir. Ancak; boy uzunluğu, ağırlık, başarı notu, sıcaklık, kan serumunda bulunan albumin miktarı v.s. birer sürekli değişkendir.

3. Bağımlılık- Bağımsızlık Bakımından: Çoğu araştırmalarda bazı değişken/değişkenler başka değişkenler tarafından açıklanmaya çalışılır. Ya da değişkenler neden-sonuç ilişkisi içerisinde olabilir. Araştırmacının kontrol edebildiği, ilgisini yoğunlaştırdığı nicel ya da nitel türden olabilen değişken bağımsız değişken; kontrol edemediği ve bağımsız değişken tarafından açıklanmaya çalışılan değişken ise bağımlı değişken adını alır. Örneğin; aylık gelir-ödenen aylık kira ikilisi için bağımlı değişken ödenen aylık kira iken bağımsız değişken aylık gelirdir. Başarı notu-ders çalışma süresi ikilisi için başarı notu bağımlı değişken, ders çalışma süresi bağımsız değişkendir. Uyuşturucu bağımlılığı- eğitim düzeyi ikilisi için uyuşturucu bağımlılığı bağımlı değişken iken eğitim düzeyi bağımsız değişkendir.

4. Ölçme Düzeyi Bakımından: Değişkenleri ölçme de dört tür ölçek kullanılmaktadır.

a) Sınıflama (Nominal) Ölçme Düzeyi: Üzerinde ölçüm yapılan değişken için birimler sahip oldukları niteliklere göre kategorilere ayrılabilir ya da sınıflanabiliyorsa bu tür değişkenler sınıflama ölçme düzeyine sahiptir. Örneğin; cinsiyet, doğum yeri, medeni hal, öğrencilerin kayıtlı olduğu bölüm, gözlük kullanma durumu v.s. değişkenleri için uygun olan ölçek sınıflama ölçeğidir. Sınıflama düzeyinde ölçülen değişkenlerde birimlerin aldığı değerlerin sayılarla ifade edilmesi durumunda, bu sayılar sadece ilgili niteliği ya da kategoriye tanımlamak için kullanılan bir gösterge değeridir. Bu sayılar matematiksel işlemler (toplama, çıkarma, çarpma, bölme veya büyüklük ya da küçüklük yönünden kıyaslama gibi) yapmaya elverişli değildir.

b) Sıralama (Ordinal) Ölçme Düzeyi: Üzerinde ölçüm yapılan değişken için birimlerin alabileceği değerleri gösteren kategoriler arasında bir kıyaslama veya derecelendirme yapılabilir ya da bu tür değişkenler sıralama ölçme düzeyine sahiptir. Örneğin; aylık gelir, başarı notu, eğitim düzeyi, kan serumunda bulunan albumin miktarı v.s. Sıralama düzeyinde ölçülen değişkenlerde birimlerin aldığı değerlerin sayılarla ifade edilmesi durumunda, bu sayılar ilgili niteliği ya da kategoriye tanımladığı gibi aynı zamanda kategoriler arasındaki derecelendirmeyi ya da kıyaslamayı da göstermektedir. Bu sayılar matematiksel işlemler (toplama, çıkarma, çarpma, bölme gibi) yapmaya elverişli değil, ancak; büyüklük-küçüklük karşılaştırması için elverişlidir.

c) Eşit Aralıklı (Interval) Ölçme Düzeyi: Bir değişken için birimler üzerinde ölçüm yapıldığında değişkenin alabileceği değerler kabul edilen bir ölçüm birimi ile birlikte ifade ediliyorsa veya ölçmeyi yapana ya da kullanılan ölçme aracına göre göreceli bir başlangıç (sıfır) noktası tespit edilerek ve eşit aralıklara bölünerek sayısallaştırılabiliyorsa, bu tür değişken eşit aralıklı ölçme düzeyine sahiptir. Burada göreceli olarak alınan ve sıfır ile tanımlanan başlangıç noktası, üzerinde ölçüm yapılan birimin olmadığı anlamına gelmez, yani eşit aralıklı ölçme düzeyinde ölçülen herhangi bir birim sıfır değerini de alabilir. Örneğin; başarı notu (puan), tutum puanı (puan), sıcaklık (°C), kardeş sayısı(adet) belli bir ürüne yapılan harcama (TL) v.s. değişkenleri eşit aralıklı ölçme düzeyine sahiptir. Bu ölçme düzeyine sahip değişkenlere ait verilerle bölme işlemi dışında tüm matematiksel işlemler yapılabilir ve bu işlemler anlamlıdır.

d) Oranlama Ölçme Düzeyi: Bir değişken eşit aralıklı ölçme düzeyinin bütün özelliklerine sahip iken, eğer sadece başlangıç noktası olarak alınan sıfır noktası gerçekten ölçülen özelliğe sahip hiçbir birimin olmadığını gösteriyorsa, bu tür değişken oranlama ölçme düzeyine sahiptir. Örneğin; boy uzunluğu(cm), ağırlık(kg), yaş(yıl), partilere çıkan oy sayısı(adet), SPA sonuçları ($\mu g/L$) v.s. değişkenleri oranlama ölçme düzeyine sahiptir. Bu ölçme düzeyine sahip değişkenlere ait veriler tüm matematiksel işlemler için uygundur.

1.2 ÖRNEKLEME DAĞILIMI

İstatistikler örneklemeden örnekleme değişen değerler alan değişkenler olarak ifade edilmektedir. Bu sebeple örnekleme dağılımı genel olarak bir istatistiğin olasılık dağılımı olarak bilinmektedir. Gerek parametrik tekniklerde gerekse parametrik olmayan tekniklerde hipotez testlerinde kullanılan örnek istatistiğinin veya test istatistiğinin örnekleme dağılımları dikkate alınarak karar kuralları oluşturulmaktadır. Bu sebeple bir istatistiğin örnekleme dağılımı oluşturulurken izlenecek olan yol şu şekilde olmalıdır:

- ✓ İlgili istatistiğin hesaplanabilme şartları altında kaç farklı durumda hesaplanabileceği belirlenir.
- ✓ İlgili istatistiğin hesaplanabileceği her bir durumda alabileceği değerler hesaplanır.
- ✓ İlgili istatistiğin alabileceği değerleri alma olasılıkları bulunur.

Örnek 1.1 $N = 5$ birimlik bir kitlede kitle birimlerinin X değişkeni bakımından aldıkları değerler X_i : 10, 14, 16, 20, 24 olsun. Bu kitleden yerine koymama yöntemi ile $n = 2$ birimlik bir örnek tesadüfi olarak çekilsin.

a) Örnek ortalaması (\bar{X}) ve örnek varyansı (S^2) istatistiklerinin örnekleme dağılımlarını oluşturunuz?

b) Bu istatistiklerin sırası ile beklenen değerlerini ve varyanslarını hesaplayınız?

c) Kitlenin ortalama ve varyansını hesaplayınız?

d) $P(\bar{X} \geq 18)$, $P(\bar{X} < 18)$, $P(S^2 > 18)$ ve $P(S^2 \leq 18)$ ihtimallerini hesaplayınız?

Çözüm a) $N = 5$ birimlik bir kitleden $n = 2$ birimlik bir örnek yerine koymama yöntemi ile $\binom{N}{n} = \binom{5}{2} = 10$ farklı şekilde çekilebileceği için söz konusu istatistikler 10 farklı durumda

hesaplanabilirler. Örnek ortalaması ve örnek varyansı istatistiklerinin hesaplanmasında $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ ve $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ formülleri kullanılır.

Durum (i)	Örnek birim değerleri	\bar{X}_i	$(X_i - \bar{X})^2$	S_i^2
1	10,14	12	4,4	8
2	10,16	13	9,9	18
3	10,20	15	25,25	50
4	10,24	17	49,49	98
5	14,16	15	1,1	2
6	14,20	17	9,9	18
7	14,24	19	25,25	50
8	16,20	18	4,4	8
9	16,24	20	16,16	32
10	20,24	22	4,4	8

Örnek ortalaması (\bar{X}) istatistiğinin örnekleme (olasılık) dağılımı:

$\bar{X} = \bar{x}_j$	12	13	15	17	18	19	20	22
$P(\bar{X} = \bar{x}_j)$	1/10	1/10	2/10	2/10	1/10	1/10	1/10	1/10

Örnek varyansı (S^2) istatistiğinin örnekleme (olasılık) dağılımı:

$S^2 = s_j^2$	2	8	18	32	50	98
$P(S^2 = s_j^2)$	1/10	3/10	2/10	1/10	2/10	1/10

olarak elde edilir.

b) \bar{X} istatistiğinin beklenen değeri ve varyansı :

$$E(\bar{X}) = \sum_{j=1}^8 \bar{x}_j P(\bar{X} = \bar{x}_j) = \frac{1}{10} (12 + 13 + 30 + 34 + 18 + 19 + 20 + 22) = \mathbf{16,8}$$

$$\sigma_{\bar{X}}^2 = E(\bar{X}^2) - [E(\bar{X})]^2 = \sum_{j=1}^8 \bar{x}_j^2 P(\bar{X} = \bar{x}_j) - (16,8)^2 =$$

$$\frac{1}{10} (144 + 169 + 450 + 578 + 324 + 361 + 400 + 484) - 282,24 = \mathbf{8,76}$$

S^2 istatistiğinin beklenen değeri ve varyansı:

$$E(S^2) = \sum_{j=1}^6 s_j^2 P(S^2 = s_j^2) = \frac{1}{10} (2 + 24 + 36 + 32 + 100 + 98) = \mathbf{29,2}$$

$$V(S^2) = E[(S^2)^2] - [E(S^2)]^2 = \sum_{j=1}^6 (s_j^2)^2 P(S^2 = s_j^2) - (29,2)^2 =$$

$$\frac{1}{10} (4 + 192 + 648 + 1024 + 5000 + 9604) - 852,64 = \mathbf{794,56}$$

c) Kitlenin ortalama ve varyansı:

$$\mu = \frac{1}{N} \sum_{i=1}^N X_i = \frac{1}{5} (10 + 14 + 16 + 20 + 24) = \frac{84}{5} = \mathbf{16,8} ; (E(\bar{X}) = \mu)$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \mu)^2 = \frac{1}{5} [46,24 + 7,84 + 0,64 + 10,24 + 51,84] = \mathbf{23,36} ; \left(\sigma_{\bar{X}}^2 = \frac{N-n}{N-1} \frac{\sigma^2}{n} \right)$$

$$\mathbf{d) } P(\bar{X} \geq 18) = 4/10 , P(\bar{X} < 18) = 6/10 , P(S^2 > 18) = 4/10 \text{ ve } P(S^2 \leq 18) = 6/10$$

1.2 İSTATİSTİKSEL TEST TEKNİKLERİ

İstatistiksel analizde test teknikleri parametrik teknikler ve parametrik olmayan teknikler olmak üzere iki gruba ayrılır.

Parametrik teknikler: Klasik teknikler olarak bilinen bu tekniklerin temel varsayımı, ilgilenilen değişken/değişkenler bakımından kitle ya da kitleler normal dağılımlı olmalıdır. İlgilenilen bağımlı değişken veya değişkenler sürekli ve en az eşit aralıklı ölçme düzeyinde ölçülmelidir. Parametrik teknikler literatürde genel olarak Z- testi, t- testi ve F- testi olarak bilinirler. Bu testlerde hipotezler kesin olarak her hangi bir kitle parametresini kapsarlar.

Parametrik olmayan teknikler: Bu teknikler kendi içerisinde iki gruba ayrılırlar.

i) Tümüyle Parametrik olmayan teknikler: Bu tekniklerde ilgilenilen değişken ya da değişkenlerin ölçme düzeyleri sınıflama veya sıralamadır. Test edilecek hipotezler genellikle bir kitle parametresi kapsamazlar. İlgili değişken bakımından kitlenin dağılımı üzerinde herhangi bir varsayım aranmaz. Ki-Kare testleri olarak bilinen bazı uyum iyiliği testleri, ilişki testleri, bağımsızlık testleri v.s bu gruba girer.

ii) Dağılıma bağlı olmayan teknikler: Bu tekniklerde ilgilenilen değişken ya da değişkenler bakımından kitle veya kitlelerin dağılımlarının normal olması varsayımı aranmaz. Değişken ya da değişkenler sürekli olup, ölçme düzeyi uygulanacak test tekniğine göre bazen en az sıralama bazen de en az eşit aralıklı düzeyde alınabilmektedir. Bu teknikler normallik varsayımının sağlanmadığı durumlarda özellikle küçük örnekler için parametrik tekniklere alternatif olarak kullanılmaktadırlar. Eğer normallik varsayımı sağlanıyorsa veya normallik varsayımı sağlanmazken büyük örneklem durumu söz konusu ise merkezi limit teorisi gereğince istatistiklerin örnekleme dağılımları normale yaklaşacağından parametrik teknikler tercih edilmelidir.

1.3 SIRALI İSTATİSTİK ve SAYI SIRALI İSTATİSTİK

Parametrik olmayan istatistikte test tekniklerine ait test istatistiklerinin örnekleme dağılımlarının oluşturulmasında sıralı istatistikten ya da sayı sıralı istatistikten yararlanılmaktadır. Bu sebeple parametrik olmayan test tekniklerine geçmeden önce bu iki kavram hakkında bilgi vermek yararlı olacaktır.

Tanım 1.6 X sürekli bir tesadüfi değişken ve F_X bu tesadüfi değişkenin dağılım fonksiyonu olsun. Bu kitleden tesadüfi olarak çekilen n birimlik bir örnek X_1, X_2, \dots, X_n ile gösterilsin. $i = 1, 2, \dots, n$ için X_i örnek birimleri içerisinde en küçük değerli olanı $X_{(1)}$ ile, ikinci en küçük değerli olanı $X_{(2)}$ ile ve bu şekilde devam ederek en büyük olanı da $X_{(n)}$ ile gösterilsin. Bu

durumda elde edilen $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ sıralı dizisine bu tesadüfi örneğin **sıralı istatistikleri** denir. $r = 1, 2, \dots, n$ olmak üzere $X_{(r)}$ 'ye **r . sıralı istatistik** denir.

Örnek birimleri ile sıra istatistikleri arasındaki en önemli fark; örnek birimleri bağımsız ve kitle ile aynı dağılıma sahip tesadüfi değişkenler iken sıralı istatistikler hem bağımsız değil hem de dağılımları kitlenin dağılımından bağımsızdır.

Tanım 1.7 X sürekli bir tesadüfi değişken ve F_X bu tesadüfi değişkenin dağılım fonksiyonu olsun. Bu kitleden tesadüfi olarak çekilen n birimlik bir örnek X_1, X_2, \dots, X_n ile gösterilsin. Bu takdirde

$$r(X_i) = \sum_{j=1}^n s(x_i - x_j) \quad , \quad i = 1, 2, \dots, n \quad (1.1)$$

ifadesine X_i **örnek biriminin sayı sıralı istatistiği** denir.

Eşitlik (1.1)'de $u = x_i - x_j$ dersek, bu durumda $s(u) = \begin{cases} 1, & u \geq 0 \\ 0, & u < 0 \end{cases}$ şeklinde tanımlıdır.

Örnek 1.2 Dağılım fonksiyonu F_X olan bir kitleden tesadüfi olarak çekilen $n = 6$ birimlik bir örnek X_1, X_2, X_3, X_4, X_5 ve X_6 olsun. Bu örnek birimlerinin gözlenmiş değerleri ise $x_1 = 25, x_2 = 40, x_3 = 5, x_4 = 70, x_5 = 40$ ve $x_6 = 35$ şeklindedir. Buna göre;

a) Bu örneklemin sıra istatistiklerini oluşturunuz?

b) Bu örneklemin sayı sıralı istatistiklerini oluşturunuz?

Cözüm: a) Bu örneklemin sıra istatistikleri: $x_3 = 5 < x_1 = 25 < x_6 = 35 < x_2 = 40 \leq x_5 = 40 < x_4 = 70$ olduğundan $X_{(1)} = X_3; X_{(2)} = X_1; X_{(3)} = X_6; X_{(4)} = X_2; X_{(5)} = X_5; X_{(6)} = X_4$ olacaktır.

b) $X_i, (i = 1, 2, 3, 4, 5 \text{ ve } 6)$ için sayı sıralı istatistik değeri Eşitlik (1.1) gereğince $r(X_i) = \sum_{j=1}^n s(x_i - x_j)$ dir. Buna göre:

$$\begin{aligned} r(X_1) &= \sum_{j=1}^6 s(x_1 - x_j) = \sum_{j=1}^6 s(25 - x_j) = s(25 - 25) + s(25 - 40) + s(25 - 5) \\ &+ s(25 - 70) + s(25 - 40) + s(25 - 35) = s(0) + s(-15) + s(20) + s(-45) + s(-15) + \\ &s(-10) = 1 + 0 + 1 + 0 + 0 + 0 = 2 \Rightarrow r(X_1) = \mathbf{r(25) = 2} \end{aligned}$$

$$r(X_2) = \sum_{j=1}^6 s(x_2 - x_j) = \sum_{j=1}^6 s(40 - x_j) = 1 + 1 + 1 + 0 + 1 + 1 = 5 \Rightarrow$$

$$r(X_2) = \mathbf{r(40) = 5}$$

$$r(X_3) = \sum_{j=1}^6 s(x_3 - x_j) = \sum_{j=1}^6 s(5 - x_j) = 0 + 0 + 1 + 0 + 0 + 0 = 1 \Rightarrow$$

$$r(X_3) = \mathbf{r(5) = 1}$$

$$r(X_4) = \sum_{j=1}^6 s(x_4 - x_j) = \sum_{j=1}^6 s(70 - x_j) = 1 + 1 + 1 + 1 + 1 + 1 = 6 \Rightarrow$$

$$r(X_4) = r(70) = 6$$

$$r(X_5) = \sum_{j=1}^6 s(x_5 - x_j) = \sum_{j=1}^6 s(40 - x_j) = 1 + 1 + 1 + 0 + 1 + 1 = 5 \Rightarrow$$

$$r(X_5) = r(40) = 5$$

$$r(X_6) = \sum_{j=1}^6 s(x_6 - x_j) = \sum_{j=1}^6 s(35 - x_j) = 1 + 0 + 1 + 0 + 0 + 1 = 3 \Rightarrow$$

$$r(X_6) = r(35) = 3$$

elde edilir.

X_i :	25	40	5	70	40	35
$r(X_i)$:	2	4,5	1	6	4,5	3

(X_i)	$r(X_i)$ (Formüle göre bu şekilde olur)	$r(X_i)$ (Uygulamada bu şekilde yapılır)
25	2	2
40...4	5	4,5
5	1	1
70	6	6
40...5	5	4,5
35	3	3