

BİYOİSTATİSTİK



Moleküler Biyoloji ve Genetik Bölümü
2020-2021 Eğitim Öğretim Yılı

Doç. Dr. Pelin KASAP
Ondokuz Mayıs Üniversitesi, İstatistik Bölümü

Dersin Amacı

- Bu dersin amacı öğrencilere veri toplama, gruplandırma ve analiz etme kabiliyetini kazandırmaktır. Ayrıca, öğrencilerin karşılaştıkları bilimsel problemlerin çözümünde doğru istatistiksel analiz yöntemlerini seçmelerini ve sonuçları yorumlamalarını sağlamaktır.

Dersin İeriđi

- Biyoistatistiđe Giriř, temel kavramlar ve tanımlar
- Verilerin toplanması, özetlenmesi ve sunulması
- Tanımlayıcı İstatistikler-Merkezi Eğilim Ölüleri
- Tanımlayıcı İstatistikler-Deđişim Ölüleri
- Bazı önemli olasılık dađılımları ve Biyoistatistik'teki uygulamaları
 - Binom
 - Poisson
 - Normal
- Merkezi Limit Teoremi, Hipotez testi ve güven aralığı
- Regresyon ve Korelasyon Analizi

Dersin İeriđi

- Tek ve iki rneklem testleri ve Biyoloji alanında uygulamaları
- Bir-ynl Varyans Analizi (ANOVA) ve Biyoloji alanında uygulamaları
- iki-ynl Varyans Analizi (ANOVA) ve Biyoloji alanında uygulamaları
- Parametrik Olmayan Testler;
 - Ki-Kare testi
 - Mann-Whitney U testi
 - Kruskall Wallis testi

Kaynaklar

- ▶ Daniel, W.W.(1999). BIOSTATISTICS, A Foundation for Analysis in the Health Sciences, Seventh Edition, John Wiley&Sons, New York.
- ▶ Spiegel, M.R., Stephens, L.J.; Esin A ve Çelebioğlu, S. (1999). Teori ve Problemlerle İstatistik, Üçüncü baskıdan çeviri, Nobel Yayın.
- ▶ Er, F. ve Peker, K.Ö. (2009). Biyoistatistik, TC Anadolu Üniversitesi Yayını, Eskişehir.
- ▶ Apaydın, A., Kutsal, A. ve Atakan, C. (2002). Uygulamalı İstatistik, Klavuz Pazarlama, Ankara.
- ▶ Karagöz, Y. (2015). SPSS 22 Uygulamalı Biyoistatistik, Tıp, Eczacılık, Diş Hekimliği ve Sağlık Bilimleri için, Nobel Yayınevi.
- ▶ Özdamar, K. (2015). SPSS ile Biyoistatistik, 10.Baskı, Nisan kitabevi Yayınları.
- ▶ Sümbüloğlu, K. ve Sümbüloğlu, V. (2002). Biyoistatistik, Hatipoğlu Yayınları, Ankara.
- ▶ Öztürk, F. (2011). Olasılık ve İstatistiğe Giriş I-II.
- ▶ Akdeniz, F. (2006). Olasılık ve İstatistik

6

Biyostatistiğe Giriş

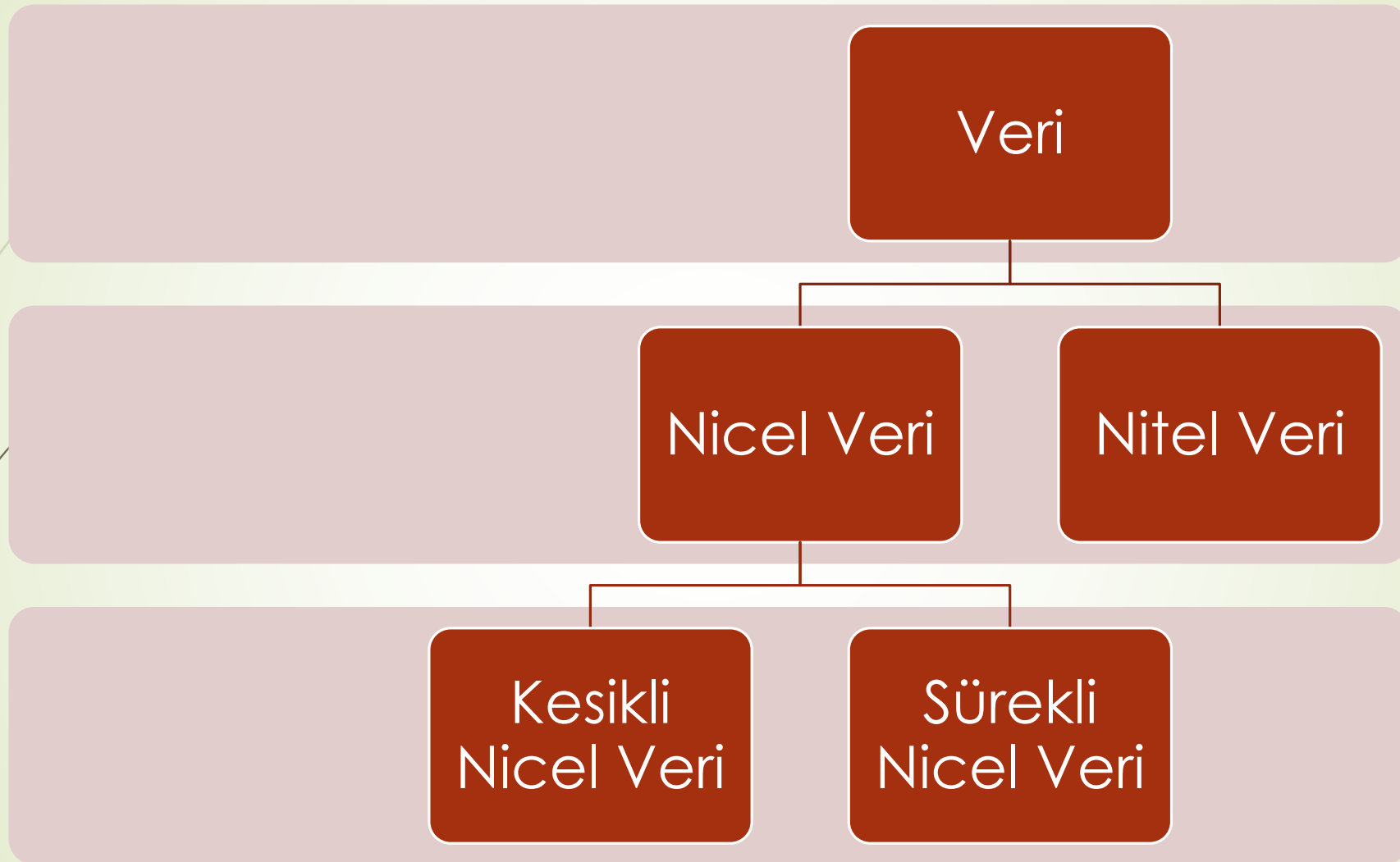
Temel Kavramlar

- Aynı özelliği taşıyan birim ya da bireylerin oluşturduğu topluluğa **kitle (population)** denir. Kitlenin büyüklüğü araştırmanın amacına bağlı olarak değişebilir. Örneğin; Türkiye'deki eczacılar ile ilgili bir araştırma yapıldığında kitle Türkiye'deki eczacıların oluşturduğu küme; Samsun'daki eczacılar ile ilgili bir araştırma yapıldığında ise kitle Samsun'daki eczacıların oluşturduğu kümedir.
- Kitleyi tanımlamak için kullanılan ölçülere *parametre* denir. Başka bir ifadeyle, kitlenin bilinmeyen özelliklerine **parametre (parameter)** denir. Kitle ortalaması (μ), kitle varyansı (σ^2), kitle korelasyonu (ρ) gibi ölçüler parametreye örnek olarak verilebilir.

- Kitleye ulaşmak fazla zaman alması ve maliyetli olması gibi sebepler nedeniyle her zaman mümkün değildir. Bu durumda kitleyi temsil eden ve kitleyle aynı özellikleri taşıyan bir alt küme seçilir. Çekildiği kitleyi temsil eden ve kitleye göre daha az sayıda gözlem ya da bireyden oluşan alt kümeye **örnek ya da örneklem (sample)** denir. Kitleden örneklem çekmek için kullanılan yöntemler topluluğuna **Örnekleme (Sampling)** denir.
- Örneklemin bir fonksiyonuna **istatistik** denir. İstatistik, örneklemi belirleyen ya da örneklemin özelliklerini tanımlayan sayısal değerlerdir. Örneklem ortalaması \bar{X} , örneklem varyansı s^2 ve örneklem korelasyonu r ile gösterilir. Kitle ortalaması μ , örneklemden hesaplanan \bar{X} yardımıyla tahmin edilir. Kitle varyansı σ^2 ve kitle korelasyonu ρ ise sırasıyla örneklemden hesaplanan s^2 ve örneklemden hesaplanan r ile tahmin edilir.

- Değişik değerler alan herhangi bir özelliğe **değişken (variable)** denir. Bir değişken sayısal değerlerle ölçülebiliyorsa, bu değişkene **nicel değişken** denir. Örneğin; bir kişinin boy uzunluğu, kilosuna, iyi kolesterol değeri vb. nicel değişken olarak adlandırılır. Nicel değişkenler kesikli ve sürekli değişkenler olmak üzere ikiye ayrılır. Belirli bir aralıkta her değeri alabilen değişkenler **sürekli değişkenler**, her değeri alamayan değişkenler **kesikli değişkenler** olarak adlandırılır. Bir ailedeki kişi sayısı kesikli değişkene, bu kişilerin boy uzunlukları ise sürekli değişkene örnektir. Bir değişken sayısal değerlerle ölçülemiyorsa **nitel değişken** adını alır. Kategorik değişken olarak da adlandırılır. Bir kişinin göz rengi, kan grubu, medeni hali vb. kategorik değişkene örnektir.

- İlgilenilen herhangi bir konu ile ilgili sayılarla ya da nitelikle ifade edilen bilgilere **veri (data)** denir. Veriler genellikle nicel (sayısal) veriler ve nitel veriler olmak üzere ikiye ayrılır. Nitel verilere cinsiyet (Erkek-Kadın), eğitim durumu (İlkokul-Ortaokul-Lise-Üniversite-Lisansüstü), sigara kullanma durumu (kullanıyor-kullanmıyor) vb. örnekler verilebilir. Nicel veriler kesikli nicel veriler ve sürekli nicel veriler olmak üzere ikiye ayrılır. Kesikli nicel veriler belirli bir aralıktaki tam sayıları ifade eder. Bir ildeki sağlık çalışanlarının sayısı (Doktor, Hemşire, Diş hekimi, Eczacı, vb.); bir ildeki salgın hastalıkların günlere/aylara/yıllara göre gözlenme sayıları; hastanelerde günlere/aylara/yıllara göre muayene edilen hasta sayısı kesikli nicel verilere örnektir. Sürekli nicel veriler bir aralıktaki sadece tam sayıları değil, tüm değerleri alan verilerdir. Hastaların boy uzunluğu, kilosunu, kan basıncı, iyi kolesterol, kötü kolesterol vb. kan örneğinden elde edilen sayısal değerler vb. örnekler sürekli nicel verilere örnektir.



- Günümüzde bir bilgiye ulaşmak için öncelikle veri toplanması gerekir. Veriler genellikle anketlerden, deneylerden, sayımlardan, gözlem yoluyla, düzenli olarak tutulan kayıtlardan ve arşivlerden, basılı ve elektronik ortamdaki kaynaklardan vb. elde edilebilir.
- İlgilenilen herhangi bir konu ile ilgili olarak verilerin toplanması, düzenlenmesi, özetlenmesi, analiz edilmesi, sonuçların yorumlanması, parametre tahmini yapılması gibi konuları kapsayan ve kitleye ilişkin bilimsel bir sonuca ulaşmayı sağlayan bilim dalına **İstatistik (Statistics)** denir. İstatistik, az maliyetle ve az zamanda doğru bilgiye ulaşmayı sağlamaktadır.

- İstatistiğin ilgilenilen konu ile ilgili verilerin toplanması, düzenlenmesi ve özetlenmesi ile ilgilenen kısmına **Tanımlayıcı İstatistikler (Descriptive Statistics)** denir.
- İstatistiğin sonuçların analiz edilmesi, test ve yorumlama, matematiksel modelleme aşaması yani verinin sadece küçük bir bölümünü inceleyerek verinin büyük bir kısmı hakkında karara ulaşmayı amaçlayan kısmına ise **Çıkarımsal İstatistik (Inferential Statistics)** denir.

- ***Biyoistatistik***, İstatistiğin Tıp, Diş Hekimliği, Biyoloji, Hemşirelik ve diğer sağlık bilimlerindeki teknolojisidir.

Verilerin Toplanması, Özetlenmesi ve Sunulması

Frekans Tabloları

- Veriler çeşitli yöntemlerle (anket, deney, gözlem, kayıtlar, arşivler vb.) elde edildikten sonra tablo haline getirilir ve çözümlenir. Verilerin düzenlendiği tablolara **frekans (sıklık) tabloları** denir. Verilerin düzenlenmeden önceki biçimine **ham ya da sınıflandırılmamış veri** denir.

Nicel verilerde frekans tabloları

Sınıf: Değişkenin değer aralığı birbirinden kesin olarak ayrılmış gruplara bölünebilir. Bu gruplara sınıf denir. Sınıf sınırları net olmalıdır ve birbirine karışmamalıdır. Sınıf sayısı k ile gösterilir ve iki şekilde belirlenir:

- Sezgisel olarak belirlenir.
- Sturges kuralına göre belirlenir. Sturges kuralına göre sınıf sayısı

$$k=1+3,3\log(n)$$

ile belirlenir. Burada n gözlem sayısıdır ve sınıf sayısının $8 \leq k \leq 15$ olması tercih edilir.

- **Dağılım Sınırları:** Örneklemdeki en küçük denek değeri ile en büyük denek değerine dağılım sınırları denir.
- **Dağılım Genişliği:** Bir örnekleme dağılımın en büyük değeri ile en küçük değeri arasındaki farka dağılım genişliği (range) denir ve R ile gösterilir.
- **Alt Sınır (As):** Bir sınıfın en küçük değeridir.
- **Üst Sınır (Üs):** Bir sınıfın en büyük değeridir.
- **Sınıf Değeri (Si):** Bir sınıfın alt sınırı ile üst sınır değerlerinin ortalamasıdır.
- **Frekans:** Bir sınıfa düşen denek sayısına frekans (sıklık) denir. Frekansların toplamı toplam denek sayısına eşit olmalıdır. Yani $\sum_{i=1}^k f_i = n$ dir.

- **Görelî Frekans (Pi):** Her sınıfa düşen denek sayısının toplam denek sayısına göre yüzdesidir. Pi: i. sınıfın görelî frekansıdır.

$$P_i = \frac{f_i}{n}, \quad i=1, \dots, k, \quad 0 < P_i < 1, \quad \sum_{i=1}^k P_i = 1$$

- **Sınıf Aralığı (c):** İki sınıf arasındaki farka sınıf aralığı denir.

$$c = \frac{\text{Dağılım Genişliği} + \text{Dağılım Genişliğinin son hanesine 1 eklenir}}{k}$$

Kümülatif Frekans Tablosu

Bazı arařtırmalarda kaç gözlem sonucunun belli bir deęerden küçük (yada büyük) olduęu arařtırılmak istenebilir. Böyle durumlarda “den daha az” yada “den daha çok” řeklindeki kümülatif frekans tablosu oluşturulur ve frekans tablosuna eklenebilir.

Nitel Verilerde Frekans Tabloları

- Nitel veriler sınıflanabilen ve sıralanabilen nitel veriler olmak üzere ikiye ayrılabilir.

Sınıflanabilen nitel verilerde frekans tabloları

Sınıflar bağımsız olarak elde edileceği için frekans tablosunda sadece sınıf, frekans ve göreceli frekans sütunları yer alır.

Sıralanabilen nitel verilerde frekans tabloları

- Veriler belli bir sıralama ölçütüne göre sıralanabilen sınıflara ayrılır.