

BIYOİSTATİSTİK

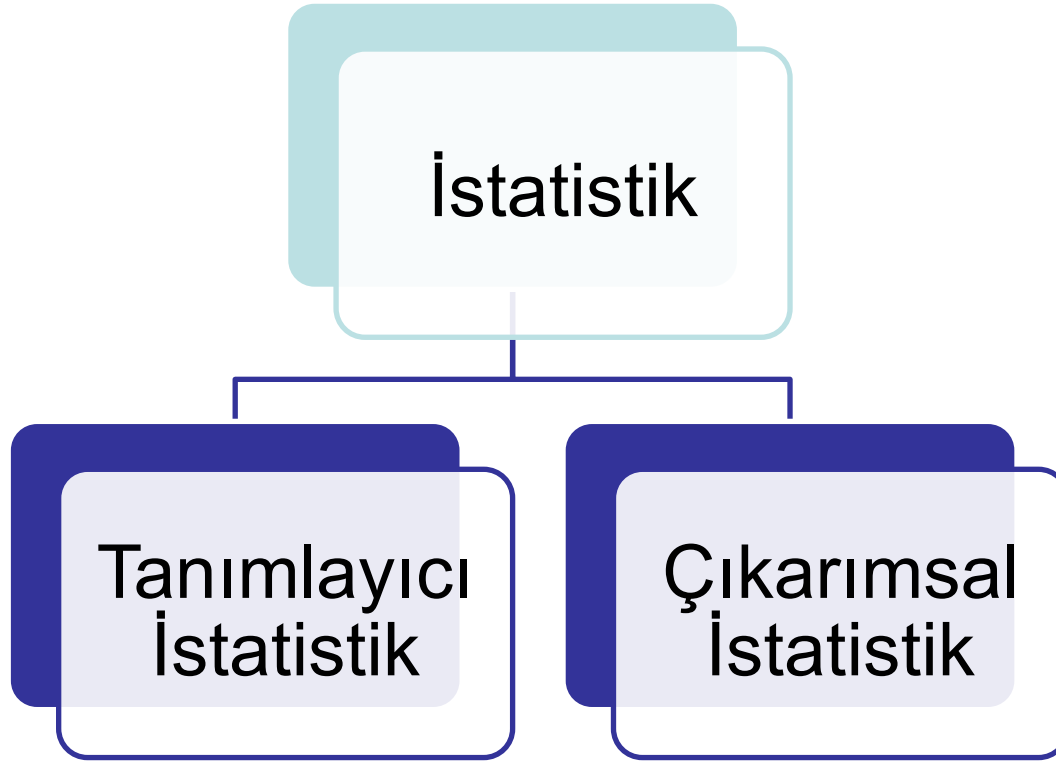
İstatistiksel Tahminleme ve Hipotez Testlerine Giriş

Dr. Öğr. Üyesi Aslı SUNER KARAKÜLAH

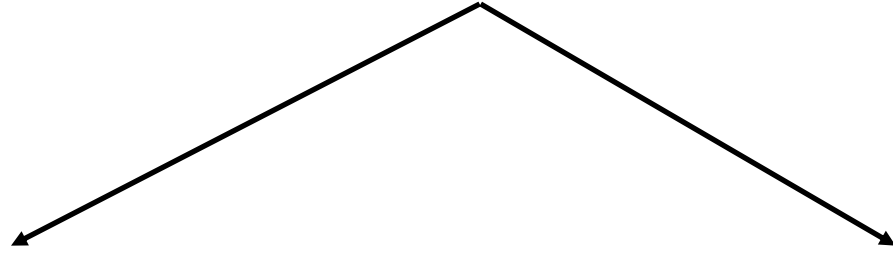
Ege Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim AD.

Web: www.biyoistatistik.med.ege.edu.tr

Biyostatistik ve Tıbbi Bilişim Anabilim Dalı



İSTATİSTİK



Tanımlayıcı yöntemler (Descriptive)

- Verilerin sınıflandırılması
- Frekans dağılımları
- Tanımlayıcı istatistikler (ortalamalar, çeyrek ve yüzdelikler, standart sapma gibi ölçüler)
- Tablo
- Grafik

Çıkarsamaya yönelik yöntemler (Inference)

- Tahminleme
- Hipotez testi

Olasılık teorisi yardımı ile

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

Burada anlatılan yöntemler, bilinmeyen KİTLE PARAMETRE değeri hakkında;

- TAHMİN yapmaya,
- KARAR vermeye,

yönelik olmak üzere iki grupta incelenmektedir.

Parametre (Tanım):

Kitlenin sayısal olarak ölçülebilen herhangi bir özelliğine '*Kitle Parametresi*' denir.

Kitle parametre deęerleri genellikle bilinemedięinden, örneklemden gidilerek TAHMİN edilmeye alıřılırlar.

Örneklem	Tahminleyici	Kitle
\bar{x}	$\frac{\sum_{i=1}^n x_i}{n}$	μ
S	$\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$	σ
\hat{p}	$\frac{X}{n}$	p

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

- Örneklemden hesaplanan,

$$\bar{x}, S \text{ ve } \hat{p}$$

sayısal değerlerine ‘*Örneklem İstatistiği*’ denir ve bunlar ilgili kitle parametreleri için birer TAHMİN’dir.

Biyostatistik ve Tıbbi Bilişim Anabilim Dalı

- Parametreler SABİT değerlerdir.
- Örneklem istatistikleri;
 - incelediğimiz değişkenler (X ş.d.) gibi birer ŞANS DEĞİŞKENİ'dir,
 - X ş.d. için olan kitle dağılımları gibi birer örnekleme dağılımına sahiptirler.

ÖRNEKLEME DAĞILIMI VE MERKEZİ LİMİT TEOREMİ

- Ortalaması μ ve varyansı σ^2 olan bir kitleden her seferinde n birey içerecek şekilde çekilen örneklemelerden hesaplanan, ÖRNEKLEM ORTALAMA' larının,
 - Ortalaması, kitle ortalaması μ 'ye
 - Varyansı, (σ^2 / n) 'ye eşit olur. (Standart hata σ / \sqrt{n} , örnek ortalamasının standart sapması)
 - Eğer n yeterince büyükse, \bar{x} yaklaşık olarak NORMAL DAĞILIM gösterir. Kitle normal dağılıyorsa n 'den bağımsız olarak normal dağılır. (Merkezi Limit Teoremi)

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

- $n \geq 30$, σ^2 **biliniyor**, X 'in dağılımı ne olursa olsun

$$\bar{x} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ ve } \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

- $n \geq 30$, σ^2 **bilinmiyor**, X 'in dağılımı ne olursa olsun

$$\bar{x} \sim N\left(\mu, \frac{S^2}{n}\right) \text{ ve } \frac{\bar{x} - \mu}{S / \sqrt{n}} \sim N(0,1)$$

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

- $n < 30$, σ^2 **bilinmiyor**, X 'in dağılımı **NORMAL**

$$\bar{x} \sim N\left(\mu, \frac{S^2}{n}\right) \text{ ve } \frac{\bar{x} - \mu}{S/\sqrt{n}} \sim t_{(sd=n-1)}$$

Student-t Dağılımı

TAHMİNLEME

- **Nokta Tahmini:**

$$\bar{x} \rightarrow \mu$$

$$S^2 \rightarrow \sigma^2$$

$$\hat{p} \rightarrow p$$

- **Aralık Tahmini:** Bilinmeyen kitle parametresinin belirli bir güven ile (ya da hata payı ile), içinde bulunması muhtemel olduğu aralıktır.

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

Aralık Tahmini için $(1-\alpha)*100$ 'lük güven aralığı

$$\bar{x} \pm z_{\alpha/2} * \sigma / \sqrt{n}$$

$$(1-\alpha) = 0.90 \quad z_{\alpha/2} = 1.645$$

$$= 0.95 \quad z_{\alpha/2} = 1.96$$

$$= 0.99 \quad z_{\alpha/2} = 2.575$$

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

- X ş.d. dağılımı ve σ^2 **bilinmiyor**, $n \geq 30$:

$$\bar{X} \pm z_{\alpha/2} * S / \sqrt{n}$$

- X ş.d. **normal dağılmış**, $n < 30$ ve σ^2 **bilinmiyor**:

$$\bar{X} \pm t_{(\alpha/2;sd=n-1)} * S / \sqrt{n}$$



Örnek 2.6:

11 kişilik bir hasta grubunda plazmadaki yağ asidi (X ş.d.) (100 ml./mgr.) aşağıdaki şekilde ölçülmüştür.

160, 168, 154, 156, 172, 163, 166, 169, 150, 170, 167

Bu hastaların geldiği kitlenin ortalaması için %90 güven aralığını hesaplayın.

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

Durum Saptaması $n < 30, \sigma^2$ bilinmiyor

Bu durumda $X \sim N(\mu, \sigma^2)$ varsayımı altında Merkezi Limit Teoremi'ne göre student-t dağılımını kullanabiliriz.

$$\bar{X} \pm t_{(\alpha/2;sd=n-1)} * S / \sqrt{n}$$

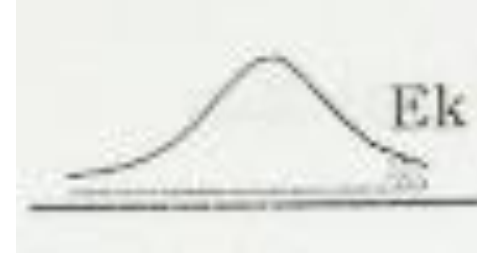
T tablosu

$$\bar{x} \pm t_{(\alpha/2;sd=n-1)} * S / \sqrt{n}$$

6 İSTATİSTİĞE GİRİŞ

Ek Tablo-3 : Student t dağılımına ilişkin kritik değerler (Tek yönlü)

Serbestlik Derecesi	Olasılık													
	0.45	0.40	0.35	0.30	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	0.0005	
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	636.619	
2	0.142	0.289	0.445	0.617	0.816	1.061	1.586	1.886	2.920	4.303	6.965	9.925	31.598	
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	12.924	
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	8.610	
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	6.869	
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	5.959	
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	5.408	
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	5.041	
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	4.781	
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	4.587	
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	4.437	
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	4.318	
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	4.221	
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	4.140	
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	4.073	
16	0.128	0.256	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.588	2.921	4.015	
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	3.955	
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	3.923	
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	3.883	
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	3.850	
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	3.819	
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	3.792	
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	3.767	
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	3.745	
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	3.725	
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	3.707	
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	3.690	
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	3.674	
29	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.758	3.659	
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	3.646	
40	0.126	0.255	0.388	0.529	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551	
60	0.126	0.254	0.387	0.527	0.679	0.848	1.046	1.296	1.671	2.000	2.400	2.680	3.400	
20	0.126	0.254	0.386	0.526	0.678	0.847	1.045	1.295	1.670	1.999	2.399	2.679	3.390	



Serbestlik Derecesi	Olasılık									
	0.45	0.40	0.35	0.30	0.25	0.20	0.15	0.10	0.05	0.01
1	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706
2	0.142	0.289	0.445	0.617	0.816	1.061	1.586	1.886	2.920	4.303
3	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182
4	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776
5	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571
6	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447
7	0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365
8	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262
10	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.228
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201

1.812

Biyostatistik ve Tıbbi Bilişim Anabilim Dalı

$$\bar{x} \pm t_{(\alpha/2; sd=n-1)} * S / \sqrt{n}$$

- $\alpha = 0.10$

$t (sd=10; \alpha/2 = 0.05) = 1.812$ 'dir (tablo değeri)

$$\bar{x} = 163.182 \quad \sqrt{S^2 / n} = \sqrt{52.364 / 11} = 2.182$$

Formüller kullanılarak 11 gözlem değerinden hesaplandı.

$$163.182 \pm 1.812 * 2.182 \quad (159.223; 167.141)$$

AGS ÜGS

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

Örnek 2.6 için sonuçların yorumlanması

- Plazmadaki yağ asidinin bilinmeyen kitle ortalaması μ için %90'lık güven sınırları (159.223; 167.141)' dir.
- Hesaplanan güven aralığının μ ' yü içerme olasılığı 0.90' dir.
- μ , %90 güvenle (%10 hata payı ile) verilen aralık içinde olabilir.

HİPOTEZ TESTİ

- Modifiye edilen bir ilaç iyileşme süresini kısaltmış mıdır?
- Yüksek gürültüye maruz kalan yetişkinler normal düzeyde gürültüye maruz kalan kişilere oranla daha depresif midir?
- Yeni tanı yöntemi ile eskisi arasında hastalığı belirleme bakımından farklılık var mıdır?

Bu türden iddialara veya sorulara belirli bir hatayı göze alarak (öngörerek) yanıtlama işlevini, istatistiksel hipotez testleri ile gerçekleştiririz.

HİPOTEZ TESTİNİN UNSURLARI

Bir hipotez testinin dört temel unsuru bulunmaktadır;

- **Sıfır hipotezi (H_0)** – Test edilen hipotezdir. Genellikle eşitlik altında kurulur.
 - $H_0: \mu = 13$ (gün)
 - $H_0: \mu_g = \mu_n$ ($\mu_g - \mu_n = 0$)
 - $H_0: p = 0.75$
- **Alternatif hipotez (H_1)** – Araştırmadaki iddiadır.
- **Test istatistiği** – H_0 hipotezinin doğruluğu altındaki dağılıma bağlı olarak belirlenir.
- **Karar bölgesi** – Dağılıma ve H_1 alternatif hipotezine göre tablodan belirlenir.

Sıfır hipotezi (H_0)

- Sıfır hipotezi (H_0) yapılacak testin belirlenmesini sağladığından oldukça önemlidir.
- H_0 hipotezi standart teoriye göre beklentinin belirtilmesidir.
- Deneye başlamadan veya veriler toplanıp incelenmeden önce kurulur.
- Ortaya atılan iddia, yapılan deney ve toplanan veriler tarafından kanıtlanana kadar doğru değildir.

Sıfır hipotezi (H_0) - devam

- H_0 hipotezinin red edilip edilmemesi kararı, incelenen kitleden çekilen örneklemin taşıdığı bilgiye dayanır.
- Örneklemden hesaplanan test istatistiğinin büyüklüğü, önceden belirlenen bir değer ile karşılaştırılır.
- Eğer hesaplanan test istatistiği red bölgesinde ise H_0 hipotezi red edilir.

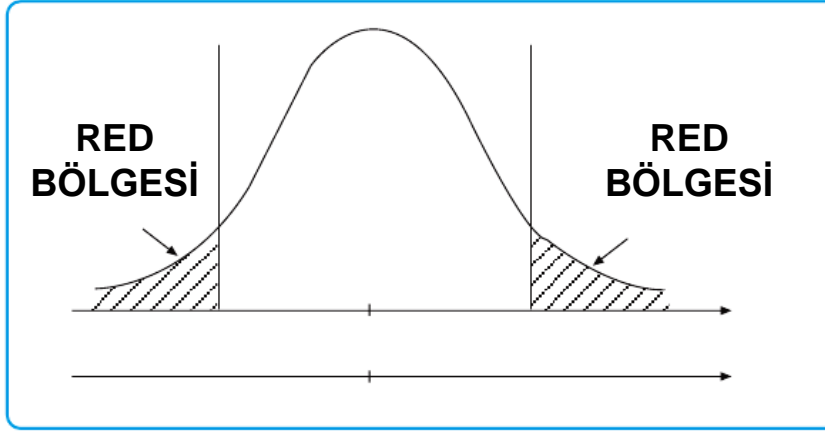
HİPOTEZ TESTİ AŞAMALARI

- **Alternatif Hipotez tipleri**

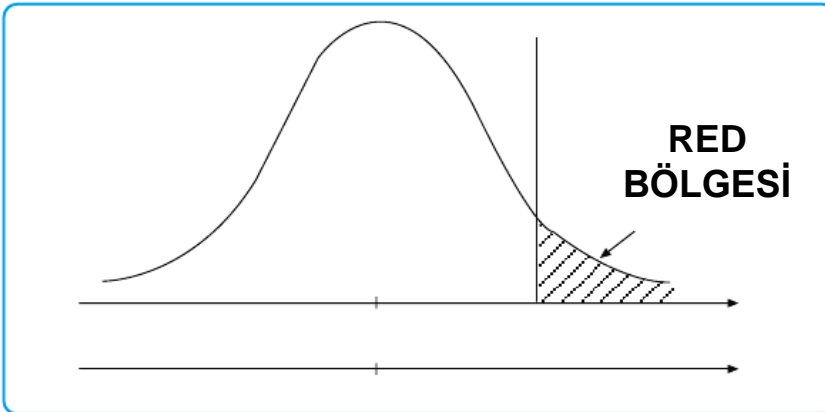
Araştırmacının iddiasını ortaya koyduğu formüldür.

- $H_1: \mu < 13$ (gün) (tek yönlü)
- $H_1: \mu_g > \mu_n$ ($\mu_g - \mu_n > 0$) (tek yönlü)
- $H_1: p \neq 0.75$ (çift yönlü)

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı



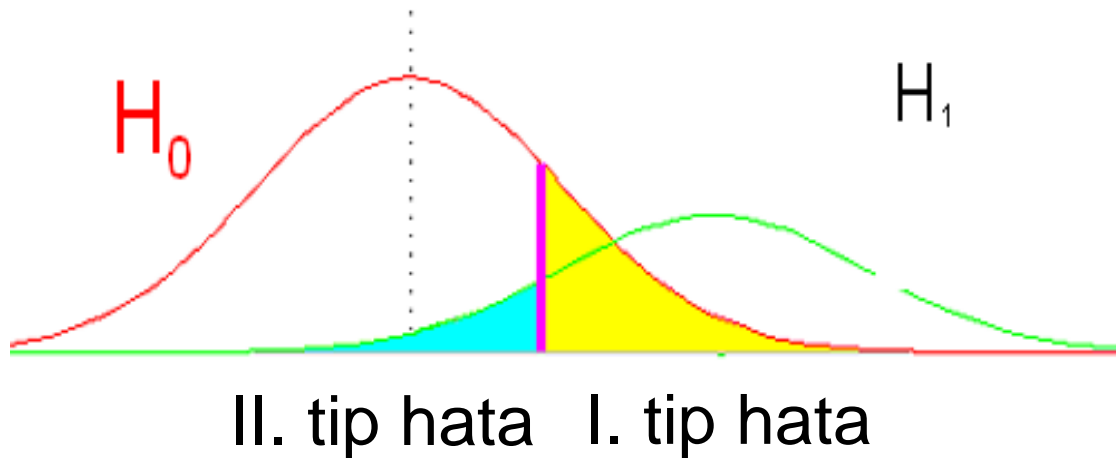
**ÇİFT
YÖNLÜ
HİPOTEZ**



**TEK
YÖNLÜ
HİPOTEZ**

KARAR SÜRECİNDE HATA TİPLERİ

Karar	H_0 Hipotezi	
	Doğru	Yanlış
H_0 Red	1.Tip hata (α)	Doğru Karar ($1-\beta$) Testin Gücü
H_0 Kabul	Doğru Karar ($1-\alpha$) Anlamlılık Düzeyi	2.Tip hata (β)



$$H_1 : \mu_0 < \mu_1$$

Örnek 2.7:

Normal değer 205 olduğu bilinen bir enzimin belirli bir tip diyet sonucunda değişip değişmediğini merak eden bir diyetisyen, diyeti uygulayan 10 kişide aşağıdaki değerleri ölçmüştür

239, 176, 235, 217, 234, 216, 318, 190, 181, 225

$\alpha=0.05$ önem seviyesinde hipotez testini gerçekleştiriniz

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

$$H_0: \mu = 205$$

$$H_1: \mu \neq 205$$

Test istatistiği

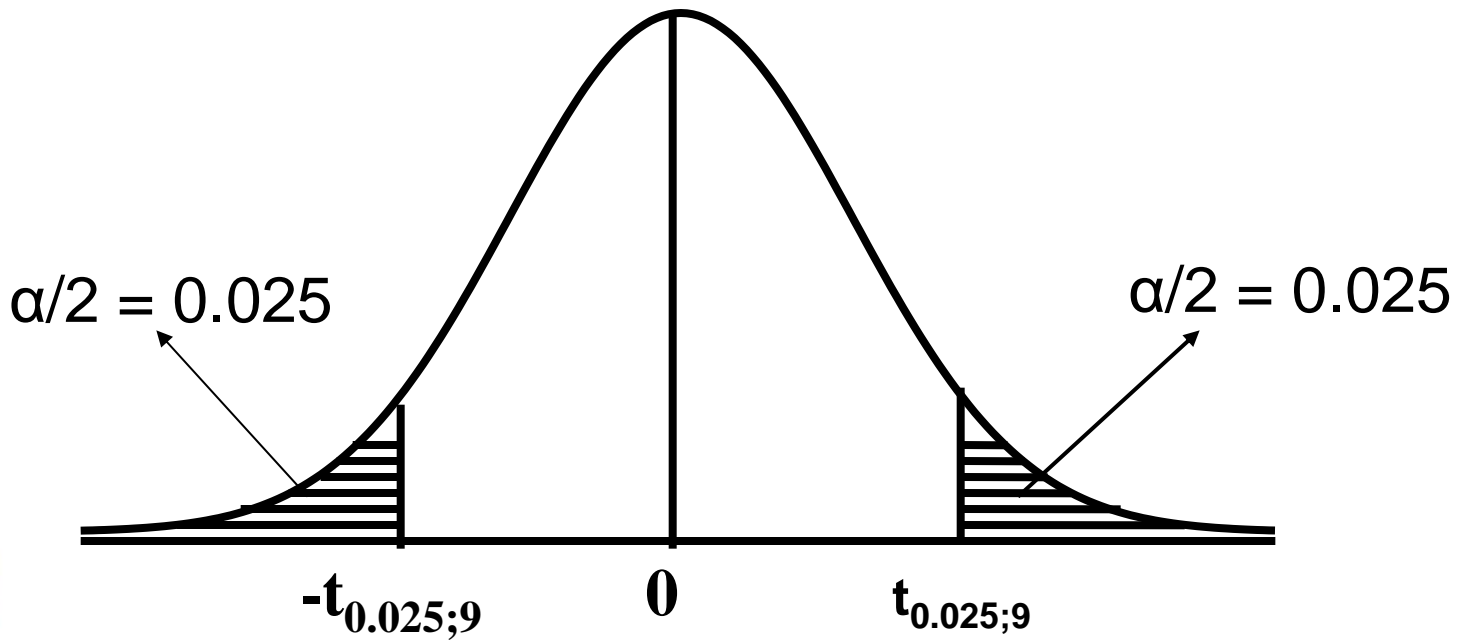
X ş.d.: Enzim düzeyi

H_0 hipotezinin doğruluğu ve $X \sim N(\mu=205, \sigma^2)$ varsayımı altında, MLT' ye göre $\bar{x} \sim N(\mu = 205, \sigma^2 / 10)$ olur.

$$\frac{\bar{x} - \mu}{S / \sqrt{n}} \sim t_{(sd=n-1)} \quad t = \frac{223.1 - 205}{40.41 / \sqrt{10}} = 1.416$$

Karar Bölgesi

$H_1: \mu = 205$ ve $\alpha = 0.05$ 'e göre aşağıdaki gibi olur



$t = 2.262$ (tablo değeri)

Biyostatistik ve Tıbbi Bilişim Anabilim Dalı

Test istatistiği: $t = 1.416$

Karar ve yorum:

$t(\text{hesap}) = 1.416 < 2.2620$ $t(\text{tablo})$ olduğundan H_0 hipotezi **red edilemez**.

Diyet sonucunda enzim düzeyindeki değişiklik, istatistiksel olarak anlamlı bulunmamıştır ($p > 0.05$).

Biyostatistik ve Tıbbi Bilişim Anabilim Dalı

- Haftaya derste anlatılacak konular...
 - Uygulama IV