

BİYOİSTATİSTİK

Tek Örneklem ve İki Örneklem Hipotez Testleri

Dr. Öğr. Üyesi Aslı SUNER KARAKÜLAH

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

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

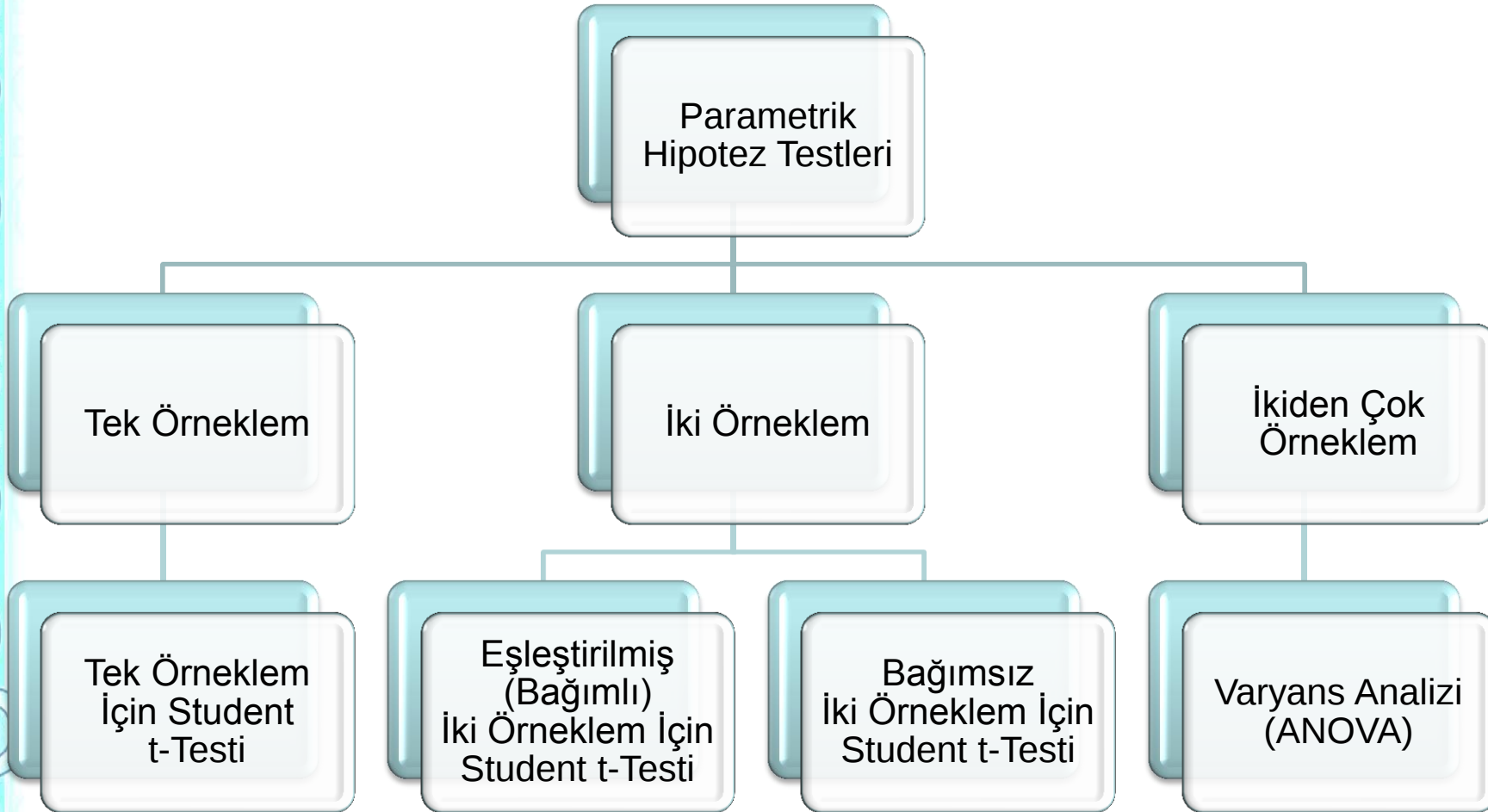
Hipotez Testi Aşamaları

1. Verinin ölçeği, gruptaki birey sayısı, grupların bağımlı ya da bağımsız olması ve varsayımlar dikkate alınarak uygun istatistiksel test yöntemine karar verilir.
2. H_0 ve H_1 hipotezleri belirlenir.
3. Test istatistiği hesaplanır.
4. Önem düzeyi-hata (yanılma) payı saptanır.
5. Tablolardan yanılma düzeyi ve serbestlik derecesindeki tablo değeri bulunur.
6. Hesaplanan değer ile tablo değeri karşılaştırılır.
7. Karşılaştırma sonucuna göre karar verilip sonuç yorumlanır.

**ORTALAMA KARŞILAŞTIRMAYA
YÖNELİK PARAMETRİK HİPOTEZ
TESTLERİ**

- Bu bölümde, sürekli verilerin yani nümerik (interval ya da ratio) skalada elde edilmiş değişkenlerin *ortalamalarını* karşılaştırmaya yönelik, *parametrik* hipotez testi yöntemleri incelenecektir.

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı



TEK ÖRNEKLEM TESTLERİ

- Bu bölümdeki testler, bir değişkenin (X şd.) ortalama değerinin, bilinen bir kitlenin ortalaması ile karşılaştırılmasında kullanılır.

A. TEK ÖRNEKLEM İÇİN STUDENT T-TESTİ

- Varsayımlar:
 - Değişkenimiz sürekli veri yani nümerik skalada elde edilmiş olmalıdır.
 - $n < 30 \Rightarrow X \sim N(\mu, \sigma^2)$, şans değişkeninin **normal dağılıma** ait bir kitleden alındığı varsayımı önemlidir.
 - $n \geq 30 \Rightarrow$ Normal dağılım varsayımı önemsizdir.

- $H_0: \mu = \mu_0$
 $H_1: \mu \neq \mu_0$ ($\mu > \mu_0$, $\mu < \mu_0$)
 ($\mu_0 \rightarrow$ bilinen kitleye ait ortalama)

- Test istatistiği:

$$t_h = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}} \sim t_{[sd=n-1]}$$

- Önem Seviyesi: α
 - Araştırmacı tarafından araştırmanın başında belirlenir.
 - RED bölgesinin büyüklüğünü gösterir.
 - Genellikle 0.01, 0.05, 0.10 gibi değerler alabilir.

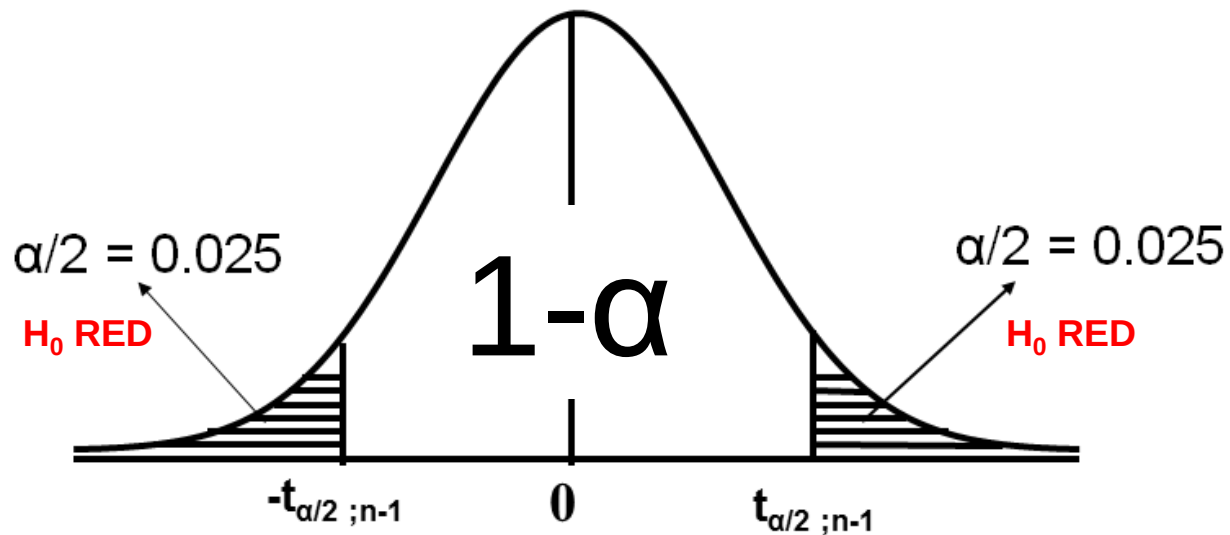
- **Karar:** α önem seviyesinde
(Çift Yönlü Hipotez için)

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

$$t_h > t_{[\alpha/2;sd=n-1]} \Rightarrow H_0 \text{ reddedilir.}$$

$$t_h < -t_{[\alpha/2;sd=n-1]} \Rightarrow H_0 \text{ reddedilir.}$$

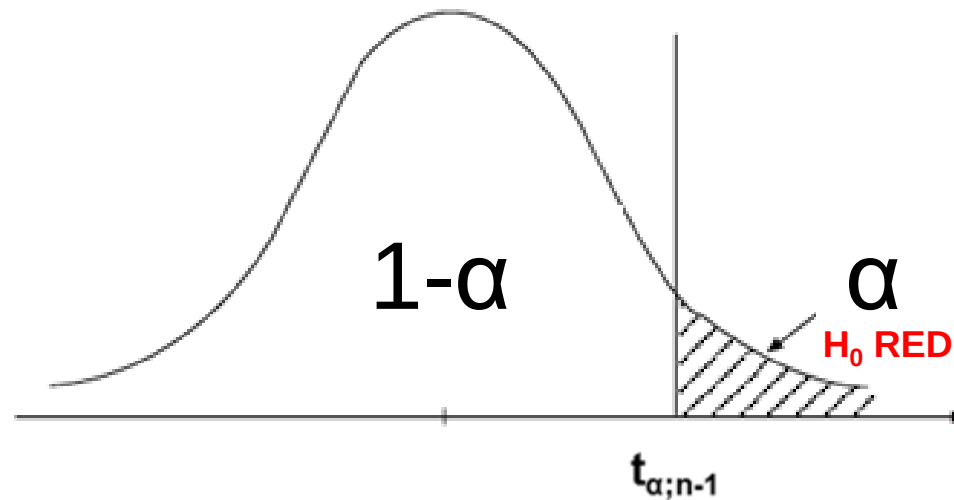


- **Karar:** α önem seviyesinde
(Tek Yönlü Hipotez için)

$$H_0 : \mu \leq \mu_0$$

$$H_1 : \mu > \mu_0$$

$$t_h > t_{[\alpha;sd=n-1]} \Rightarrow H_0 \text{ reddedilir.}$$

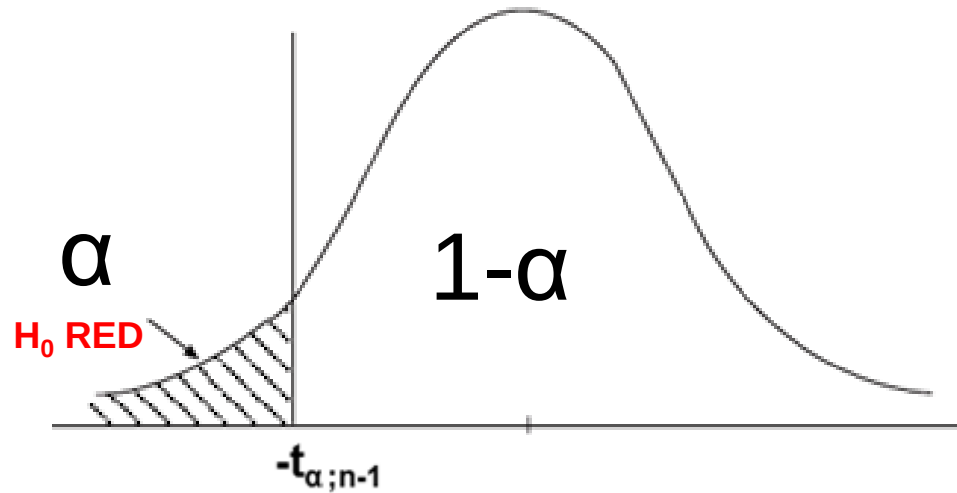


- **Karar:** α önem seviyesinde
(Tek Yönlü Hipotez için)

$$H_0 : \mu \geq \mu_0$$

$$H_1 : \mu < \mu_0$$

$$t_h < -t_{[\alpha;sd=n-1]} \Rightarrow H_0 \text{ reddedilir.}$$



Örnek 1:

Sağlıklı çocuklarda kan kolesterol düzeyinin 175 mg/dl olduğu bilinmektedir. Kalp damar hastalığından ölen erkeklerin çocuklarının da risk taşıyıp taşımadığını belirlemek amacıyla, babası kalp damar hastalığından ölen 10 erkek çocuğun kolesterol değeri aşağıdaki gibi bulunmuştur. Hipotezi $\alpha=0.05$ önem seviyesinde test ediniz.

180, 190, 175, 220, 215, 160, 230, 200, 220, 240

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

	X_i	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
1	180	-26	676
2	190	-16	256
3	175	-31	961
4	220	14	196
5	245	39	1521
6	160	-46	2116
7	230	24	576
8	200	-6	36
9	220	14	196
10	240	34	1156
Toplam	2060		7690

1. $H_0: \mu = 175 \text{ mg/dl}$

$H_1: \mu > 175 \text{ mg/dl}$

2. $\alpha = 0.05$

H_0 hipotezinin doğruluğu ve $X \sim N(\mu = 175; \sigma^2)$ varsayımı altında

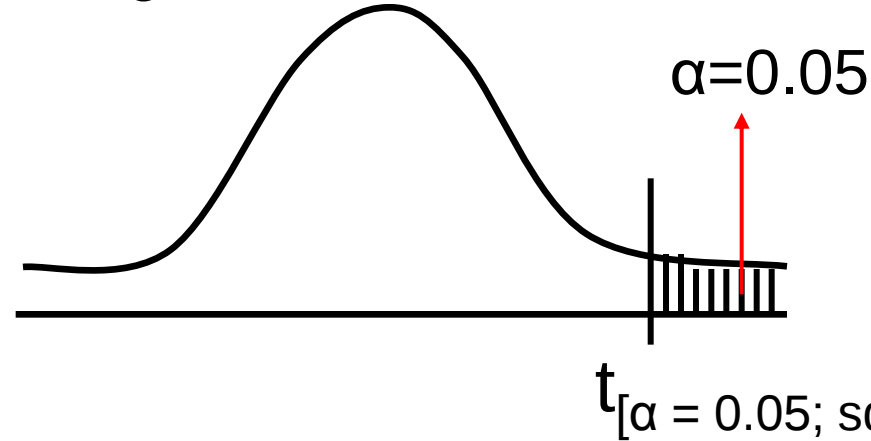
$$\bar{X} \sim N(175, S_{\bar{X}}^2 = \frac{S^2}{n})$$

test istatistiği :

$$t_h = \frac{\bar{X} - \mu_0}{S_{\bar{X}}} \sim t_{[sd=n-1]}$$

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

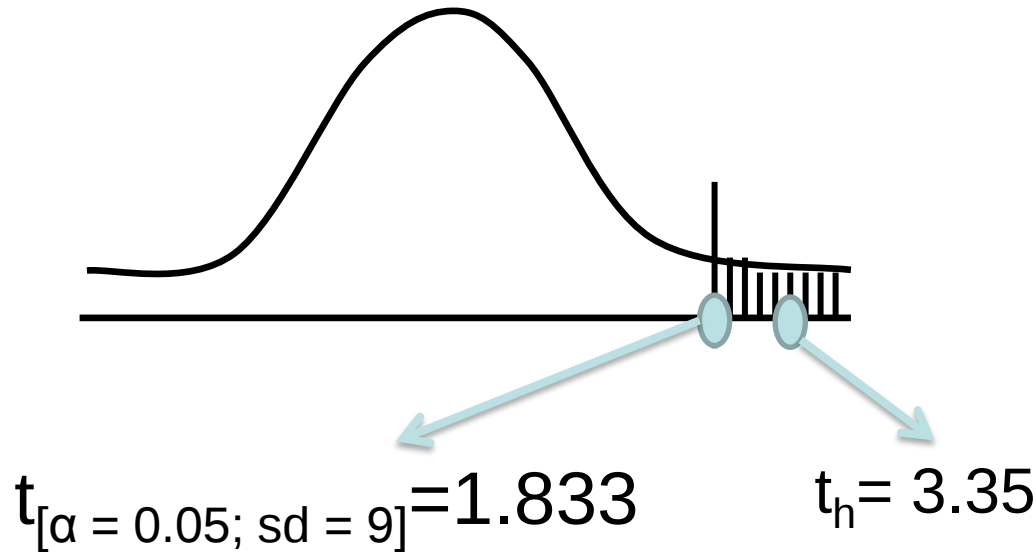
3. Karar Bölgesi:



Eğer $t_h > 1.833$ H_0 hipotezi reddedilecektir.

4. Veriler kullanılarak t_h değerinin bulunur.

$$\begin{aligned}\bar{x} &= \frac{\sum x_i}{n} = \frac{2060}{10} = 206 \text{ mg / dl} \\ S &= \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{7690}{9}} = 29.23 \\ S_{\bar{x}} &= \frac{S}{\sqrt{n}} = \frac{29.23}{\sqrt{10}} = 9.25 \\ t_h &= \frac{206 - 175}{9.25} = 3.35\end{aligned}$$



5. $t_h = 3.35 > t_{\text{tablo}} = 1.833$ H_0 hipotezi reddedilir.
6. Yorum: Babası kalp damar hastalığından ölen çocukların kolestrol düzeyi, sağlıklı çocuklarınkine göre istatistiksel olarak anlamlı derecede yüksek bulunmuştur [$p < 0.05$].

İKİ ÖRNEKLEM TESTLERİ

- Bu bölümde yer alan testlerde,
 - deney birimleri üzerinde aynı sürekli değişken iki kez ölçülebilir (Eşleme-Pairing),
 - ölçümle belirtilen sürekli bir değişken yönünden iki bağımsız grup için karşılaştırma yapılabilir.

B. EŞLEŞTİRİLMİŞ (BAĞIMLI, PAIRED) İKİ ÖRNEKLEM İÇİN STUDENT-T TESTİ

- Tedavi öncesi ve tedavi sonrası, ilaç öncesi ve ilaç sonrası gibi çiftler halinde gözlemler alınır ve n büyüklüğü bu çiftlerin sayısı ile belirlenir.

X_1 ş.d.: İlk ölçümde elde edilen değerlere ilişkin değişken

X_2 ş.d.: Aynı değişkenin ikinci ölçümdeki değerlere ilişkin değişken

X ş.d.: $X_1 - X_2$ ya da $X_2 - X_1$ şeklinde tanımlanabilir. Özellikle H_1 hipotezini tek-yönlü olduğunda, işaretin yönü ($<$, $>$) bu farkın nasıl tanımlandığına bağlıdır.

- El ile yapılan işlemlerde

- **t testi için:**

$H_0: \mu_1 - \mu_2 = 0$ şeklinde tanımlandığında tek örnek testindeki μ_0 değeri yerine 0 değeri alır.

Örnek 2:

- Gebeliği önleyen bir oral kontraseptif ilacın kullanımının, kadınlarda sistolik kan basıncında yükselmeye neden olup olmadığını belirlemek amacıyla, 20-39 yaşlarında 15 kadının ilaç öncesi ve sonrasında sistolik kan basınçları (mm/hg) aşağıdaki gibi elde edilmiştir. $\alpha = 0.05$ önem seviyesinde hipotez testini gerçekleştiriniz.

Hasta No	İlaç Öncesi	İlaç Sonrası
1	130	130
2	120	125
3	132	137
4	142	146
5	128	130
6	129	134
7	137	140
8	128	127
9	126	126
10	127	128
11	130	135
12	120	123
13	135	133
14	140	142
15	142	147

- İncelediğimiz değişken nümerik skalada elde edilmiştir.
- $n < 30$ olduğunda $\mathbf{X}_i$ değişkeni ($\mathbf{X}_{1i} - \mathbf{X}_{2i}$ ya da $\mathbf{X}_{2i} - \mathbf{X}_{1i}$) farklarının normal dağılıma uyması gerekmektedir.
- Örneğimiz için $X = X_2 - X_1$ şeklinde tanımlandı.

1. $H_0: \mu = 0$

$H_1: \mu > 0$

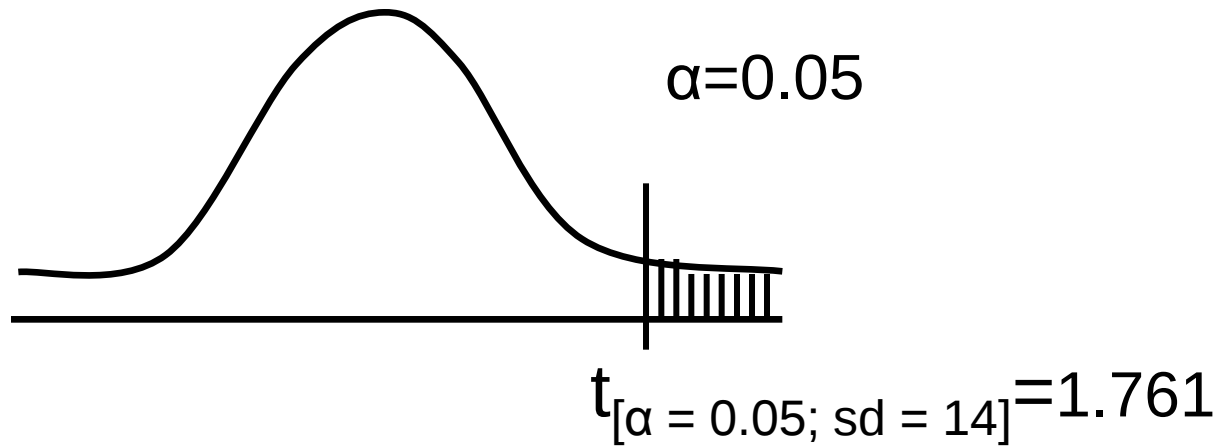
2. H_0 hipotezinin doğruluğu altında, farkların normal dağılımdan geldiği varsayıldığında

$$\bar{X} \sim N(0; S_{\bar{X}}^2) \text{ olur.}$$

Test istatistiği

$$t_h = \frac{\bar{X} - 0}{S_{\bar{X}}} \sim t_{[sd=n-1]}$$

3. Karar Bölgesi:



$t_h > 1.761$ 'den H_0 hipotezi reddedilecektir.

Biyostatistik ve Tıbbi Bilişim Anabilim Dalı

Hasta No	X1 İlaç Ö.	X2 İlaç S.	$X_i = X2_i - X1_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	130	130	0	-2.47	6.10
2	120	125	5	2.53	6.40
3	132	137	5	2.53	6.40
4	142	146	4	1.53	2.34
5	128	130	2	-0.47	0.22
6	129	134	5	2.53	6.40
7	137	140	3	0.53	0.28
8	128	127	-1	-3.47	12.04
9	126	126	0	-2.47	6.10
10	127	128	1	-1.47	2.16
11	130	135	5	2.53	6.40
12	120	123	3	0.53	0.28
13	135	133	-2	-4.47	19.98
14	140	142	2	-0.47	0.22
15	142	147	5	2.53	6.40
Toplam			37		81.73

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

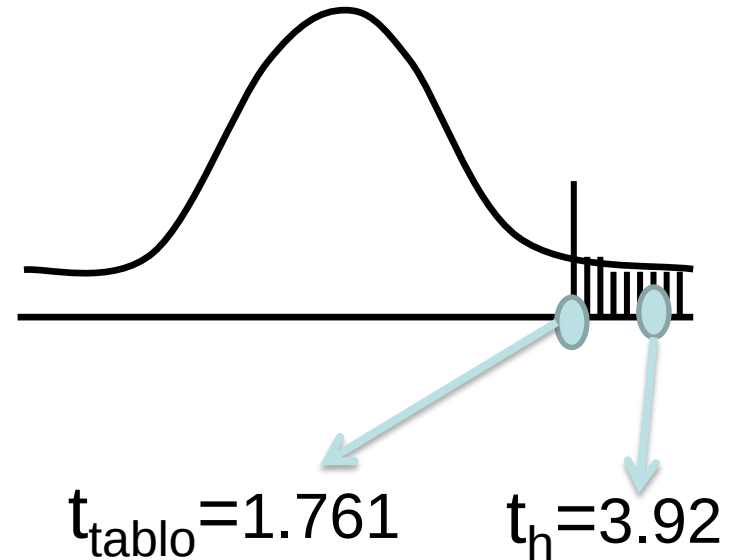
$$\bar{x} = \frac{37}{15} = 2.47 \text{ mm/hg}$$

$$S_{\bar{x}} = \frac{2.42}{\sqrt{15}} = 0.63$$

$$t_h = \frac{2.47-0}{0.63} = 3.92 > 1.761$$

$$S^2 = \frac{81.73}{14} = 5.84$$

$$S = 2.42$$



H_0 hipotezi reddedilir.

- **Yorum:** İlaç sonrası sistolik kan basıncındaki artış istatistiksel olarak anlamlı bulundu ($p < 0.05$).

C. BAĞIMSIZ (INDEPENDENT) İKİ ÖRNEKLEM İÇİN STUDENT T-TESTİ

- Değişkenler, nümerik skalada elde edilmiş olmalıdır.
- X_1 : 1. grupta elde edilen şans değişkeni
 X_2 : 2. grupta elde edilen şans değişkeni
 n_1 : 1. gruptaki birey sayısı
 n_2 : 2. gruptaki birey sayısı

- **n_1 ve $n_2 \geq 30$ ise**
 - normal dağılış varsayımı gerekmeden,
- **n_1 ve/veya $n_2 < 30$ ise**
 - sadece $X_1 \sim N(\mu_1; \sigma_1^2)$ $X_2 \sim N(\mu_2; \sigma_2^2)$ varsayımı altında bu test uygulanır.
- σ_1^2 ve σ_2^2 bilinmediğinden örneklerden hesaplanan S_1^2 ve S_2^2 kullanılır.

- $H_0: \mu_1 = \mu_2 \Rightarrow \mu_1 - \mu_2 = 0$
 $H_1: \mu_1 \neq \mu_2 \Rightarrow \mu_1 - \mu_2 \neq 0$
 $\mu_1 > \mu_2 \Rightarrow \mu_1 - \mu_2 > 0$
 $\mu_1 < \mu_2 \Rightarrow \mu_1 - \mu_2 < 0$
- $\sigma_1^2 = \sigma_2^2 = \sigma^2$ ve H_0 hipotezinin doğruluğu altında normal dağılıma uyum gösterir.

- Test İstatistiği:

$$t_h = \frac{[\bar{x}_1 - \bar{x}_2]}{S_{\bar{x}_1 - \bar{x}_2}} \sim t_{[sd=n_1+n_2-2]}$$

$$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{S^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 + n_2 - 2)}$$

- H_1 hipotezine ve α önem düzeyine göre t-tablosu kullanılarak hipotez hakkında karar verilir.

Örnek 3:

- Anne için gebelik döneminde, özel bir diyet uygulayan 12 gebe ve normal beslenme alışkanlıklarını sürdüren 10 gebenin hemoglobin değerleri (g/dl) aşağıdaki gibi elde edilmiştir. Aralarında fark olup olmadığını $\alpha=0.05$ önem düzeyinde test ediniz.

Hasta No	Beslenme Tipi	Hemoglobin
1	1	11,2
2	1	9,8
3	1	9,6
4	1	8,0
5	1	9,9
6	1	10,4
7	1	11,5
8	1	10,7
9	1	11,8
10	1	8,7
11	2	11,9
12	2	12,8
13	2	10,4
14	2	12,0
15	2	12,5
16	2	13,5
17	2	12,0
18	2	13,0
19	2	12,0
20	2	11,6
21	2	12,3
22	2	12,0

I.beslenme tipi için
Hemoglobinin

$$X_1 \sim N(\mu_1, \sigma_1^2)$$

II. beslenme tipi için
Hemoglobinin

$$X_2 \sim N(\mu_2, \sigma_2^2)$$

olduğunu varsayıyoruz.

- 
1. $H_0: \mu_1 = \mu_2 \Rightarrow \mu_1 - \mu_2 = 0$
 $H_1: \mu_1 \neq \mu_2 \Rightarrow \mu_1 - \mu_2 \neq 0$
2. H_0 hipotezinin doğruluğu ve normallik varsayımları altında test istatistiği

$$t_h = \frac{[\bar{X}_1 - \bar{X}_2] - 0}{S_{\bar{X}_1 - \bar{X}_2}} \sim t_{[sd=n_1+n_2-2]}$$

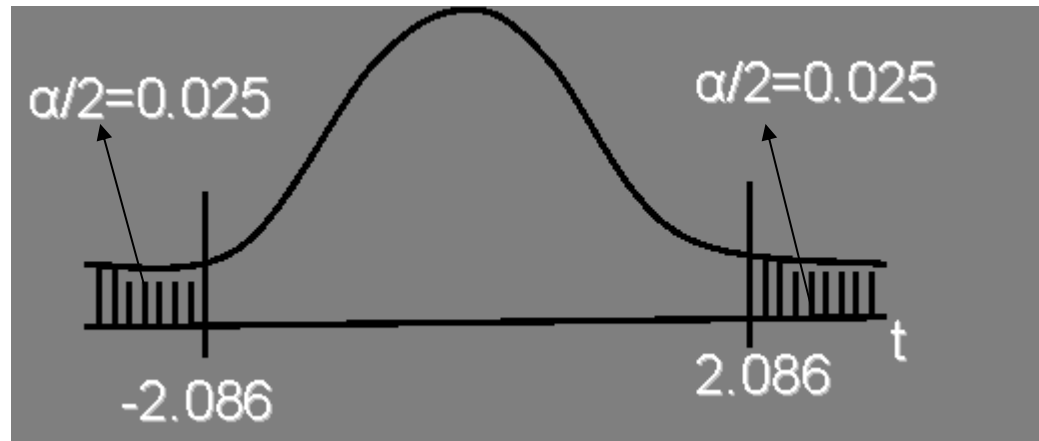
olur.

3. $\alpha = 0.05$ ve H_1 çift yönlü olduğundan

$$t_{[\alpha/2=0.025;sd=10+12-2]} = t_{[0.025;20]} \\ = 2.086$$

bulunur.

4. Karar Bölgesi



$t_h > 2.086$ ya da $t_h < -2.086$ olursa H_0 hipotezi reddedilecektir.

Hasta No	Beslenme Tipi	x_i	$x_i - \bar{X}$	$(x_i - \bar{X})^2$
1	1	11,2	1,04	1,08
2	1	9,8	-0,36	0,13
3	1	9,6	-0,56	0,31
4	1	8,0	-2,16	4,67
5	1	9,9	-0,26	0,07
6	1	10,4	0,24	0,06
7	1	11,5	1,34	1,80
8	1	10,7	0,54	0,29
9	1	11,8	1,64	2,69
10	1	8,7	-1,46	2,13
		101,6		13,22
11	2	11,9	-0,27	0,07
12	2	12,8	0,63	0,40
13	2	10,4	-1,77	3,13
14	2	12,0	-0,17	0,03
15	2	12,5	0,33	0,11
16	2	13,5	1,33	1,77
17	2	12,0	-0,17	0,03
18	2	13,0	0,83	0,69
19	2	12,0	-0,17	0,03
20	2	11,6	-0,57	0,32
21	2	12,3	0,13	0,02
22	2	12,0	-0,17	0,03
		146,0		6,63

$$\bar{x}_1 = \frac{101.6}{10} = 10.16 \text{ g / dl}$$

$$\bar{x}_2 = \frac{146}{12} = 12.17 \text{ g / dl}$$

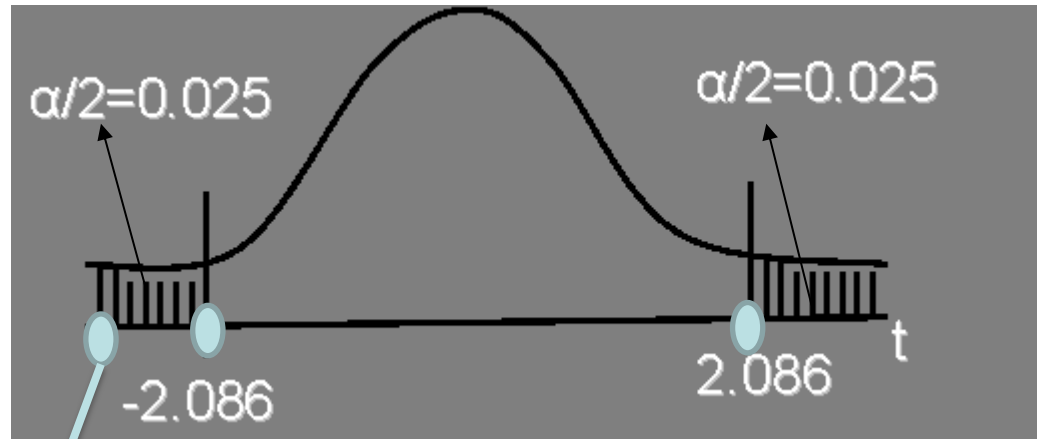
$$S_1^2 = \frac{13.22}{9} = 1.47$$

$$S_2^2 = \frac{6.63}{11} = 0.60$$

$$S^2 = \frac{(9 * 1.47) + (11 * 0.60)}{20} = \frac{19.83}{20} = 0.942$$

$$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{0.942 \left(\frac{1}{10} + \frac{1}{12} \right)} = \sqrt{0.942 * 0.18} = \sqrt{0.17} = 0.41$$

$$t_h = \frac{(10.16 - 12.17) - 0}{0.41} = \frac{-2.01}{0.41} = -4.90$$



$$t_h = -4.90 < t_{\text{tablo}} = -2.086$$

H_0 hipotezi reddedilir.

Yorum: 2 grup arasındaki hemoglobin düzeyi bakımından farklılık, istatistiksel olarak anlamlı bulunmuştur ($p < 0.05$).

Özel beslenme rejimi uygulayan grupta hemoglobin düzeyleri normallere göre daha yüksektir.

Örnek 4:

- Bruksizim (Diş Gıcırdatma Alışkanlığı) hastalarının diş hasarını önlemek ve uyku düzenini sağlamak için üst dişlerin üzerine takılan özel olarak yapılmış iki ayrı aparey üretilmiştir. Bu apareylerden birisinin seçimini yapmak için 8 bireye ilk gün A apareyi, ikinci gün ise B apareyi veriliyor ve uyuma saatleri belirleniyor. Aşağıdaki bulgular yardımıyla apareyler arası farklılığı test ederek, hangi apareyin hastaların bruksizimden kaynaklı uyku sorunlarını gidermek amacıyla kullanılmasını önerirsiniz? ($\alpha=0.05$ önem düzeyinde test ediniz.)

Aparey	Hasta 1	Hasta 2	Hasta 3	Hasta 4	Hasta 5	Hasta 6	Hasta 7	Hasta 8
A	10	9	7	6	9	8	7	12
B	9	5	8	7	9	10	8	5

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

- İncelediğimiz değişken nümerik skalada elde edilmiştir.
 - $n < 30$ olduğunda $\mathbf{X}_i$ değişkeni $(\mathbf{X}_{1i} - \mathbf{X}_{2i})$ ya da $\mathbf{X}_{2i} - \mathbf{X}_{1i})$ farklarının normal dağılıma uyması gerekmektedir.
 - Örneğimiz için Fark $(d_i) = X_1 - X_2$ şeklinde tanımlandı.
 - Bu örnekte iki örneklem birbirine bağımlıdır.
 - Örneklem birimleri 30'dan az olduğundan
 - Eşleştirilmiş (Bağımlı, Paired) İki Örneklem İçin Student-t Testi
- kullanılması uygundur.

1. $H_0: \mu_D = 0$

$H_1: \mu_D \neq 0$

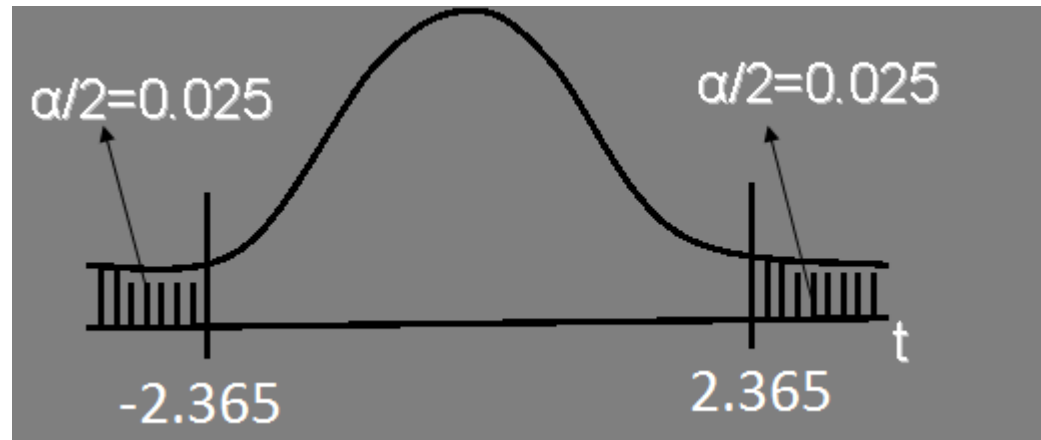
2. H_0 hipotezinin doğruluğu altında, farkların normal dağılımdan geldiği varsayıldığında

$$\bar{X} \sim N(0; S_{\bar{X}}^2) \text{ olur.}$$

Test istatistiği

$$t_h = \frac{\bar{X} - 0}{S_{\bar{X}}} \sim t_{[sd=n-1]}$$

3. Karar Bölgesi



$$t_{[\alpha/2 = 0.025; \text{sd} = 7]} = 2.365$$

$t_h > 2.365$ ya da $t_h < -2.365$ olursa H_0 hipotezi reddedilecektir.

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

Aporey	Hasta 1	Hasta 2	Hasta 3	Hasta 4	Hasta 5	Hasta 6	Hasta 7	Hasta 8
A	10	9	7	6	9	8	7	12
B	9	5	8	7	9	10	8	5
Fark (d_i)	1	4	-1	-1	0	-2	-1	7

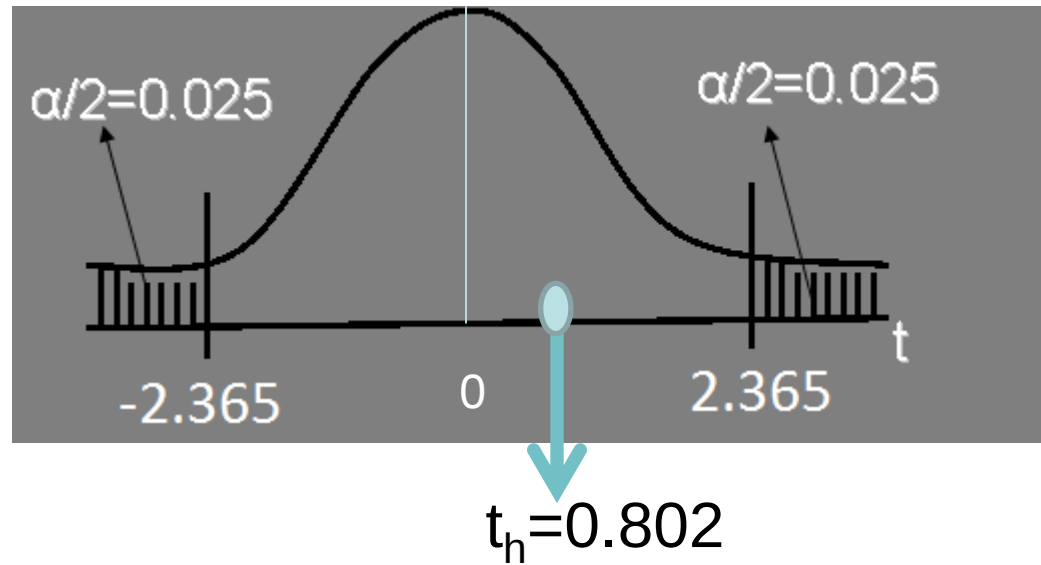
$$\bar{x} = \frac{\sum d_i}{n} = \frac{7}{8} = 0.875$$

$$S_x = \frac{3.09}{\sqrt{8}} = 1.09$$

$$S = \sqrt{\frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n-1}} = \sqrt{\frac{73 - \frac{49}{8}}{7}} = 3.09$$

$$t_h = \frac{0.875 - 0}{1.09} = 0.802$$

4. Karar



$t_h = 0.802 < 2.365$ olduğundan H_0 hipotezi red edilemez.

Yorum: A ve B apareylerinin uyku sorununu çözmede istatistiksel olarak anlamlı bir etki farklılığı bulunmamaktadır ($p > 0.05$).

Örnek 5:

- Dişeti hastalığı tanısı konmuş 12 erkek ve 8 kadın bireyin dişeti hastalığı belirtilerinin başladığı andaki yaş ortalamaları erkekler için 40, kadınlar için 36 olarak bulunmuştur. Bu bireylere ilişkin standart sapmalar ise sırasıyla erkekler için $s_E=5$, kadınlar için $s_K=4$ 'tür. Bu veriler yardımıyla erkek ve kadınlarda dişeti hastalığı başlangıç yaş ortalamaları arasında istatistiksel olarak anlamlı bir farklılık olup olmadığını araştırınız. ($\alpha=0.05$)

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

Erkeklerin Yaşları: $X_1 \sim N(\mu_1, \sigma_1^2)$

Kadınların Yaşları: $X_2 \sim N(\mu_2, \sigma_2^2)$

- Bu örnekte iki örneklem birbirinden bağımsızdır.
- Örneklem birimleri 30'dan az olduğundan

- Bağımsız (independent) İki Örneklem için Student t Testi

kullanılması uygundur.

- 
1. $H_0: \mu_1 = \mu_2 \Rightarrow \mu_1 - \mu_2 = 0$
 $H_1: \mu_1 \neq \mu_2 \Rightarrow \mu_1 - \mu_2 \neq 0$
2. H_0 hipotezinin doğruluğu ve normallik varsayımları altında test istatistiği

$$t_h = \frac{[\bar{X}_1 - \bar{X}_2] - 0}{S_{\bar{X}_1 - \bar{X}_2}} \sim t_{[sd=n_1+n_2-2]}$$

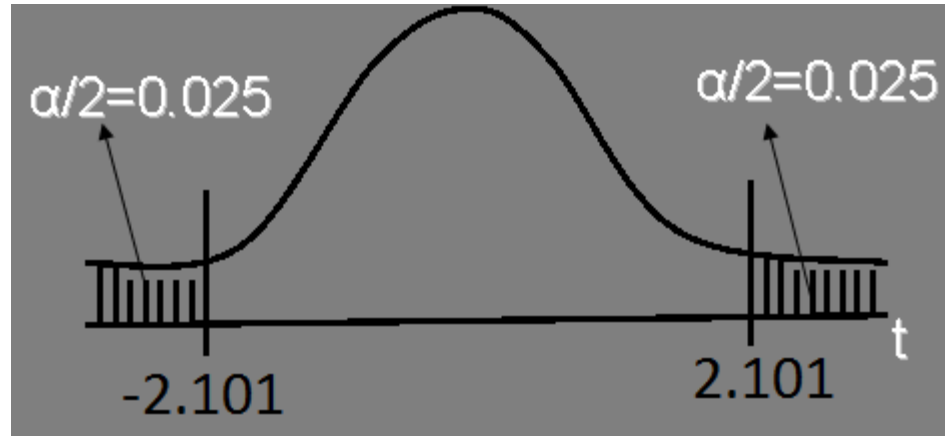
olur.

3. $\alpha = 0.05$ ve H_1 çift yönlü olduğundan, tablo değeri

$$t_{[\frac{\alpha}{2}=0,025;sd=12+8-2]} = t_{[0,025;sd=18]} \\ = 2.101$$

bulunur.

4. Karar Bölgesi



$t_h > 2.101$ ya da $t_h < -2.101$ olursa H_0 hipotezi reddedilecektir.

Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

Test istatistiği

$$t_h = \frac{[\bar{x}_1 - \bar{x}_2] - 0}{S_{\bar{x}_1 - \bar{x}_2}} \sim t_{[sd=n_1+n_2-2]}$$

$\bar{x}_1 = 40$ yaş

$s_1 = 5$

$n_1 = 12$



Erkek

$\bar{x}_2 = 36$ yaş

$s_2 = 4$

$n_2 = 8$



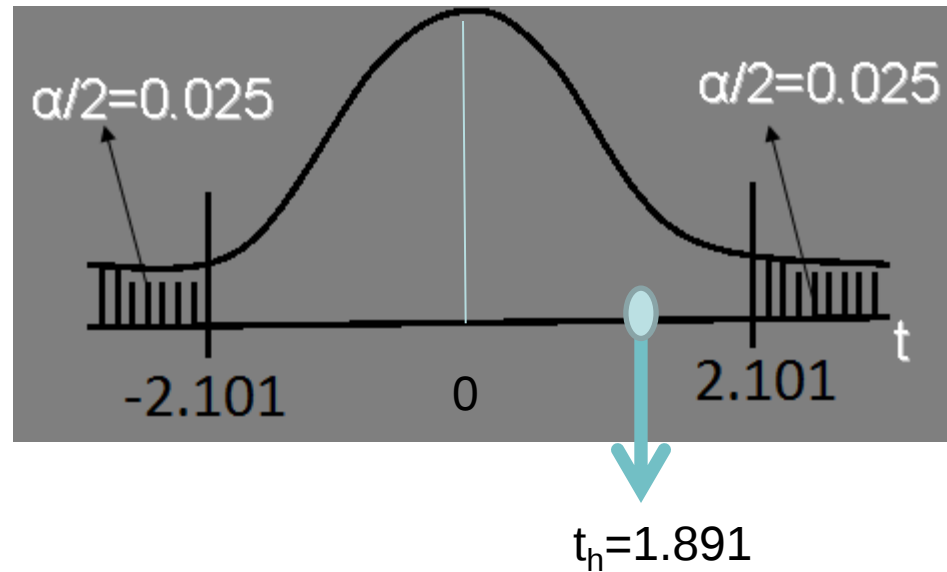
Kadın

$$s^2 = \frac{(11 * 25) + (7 * 16)}{18} = 21.5$$

$$s_{x_1 - x_2} = \sqrt{21.5 \left(\frac{1}{12} + \frac{1}{8} \right)} = 2.115$$

$$t_h = \frac{(40 - 36) - 0}{2.115} = 1.891$$

4. Karar



$1.891 < 2.101$ H_0 hipotezi red edilemez.

- **Yorum:** 2 grup arasındaki dişeti hastalığı belirtilerinin başlangıç yaşı bakımından farklılık, istatistiksel olarak anlamlı bulunmamıştır ($p>0.05$).
- Erkekler ve kadınlarda dişeti hastalığı belirtilerinin başladığı andaki yaş ortalamaları farklı değildir.

- Haftaya derste anlatılacak konular...
 - Uygulama V