3.2. İKİ KİTLENİN KARŞILAŞTIRILMASI

3.2.1. İki Kitle Ortalaması Arasındaki Farkın Önem Kontrolü

3.2.1.1. Kitle Varyansı σ_1^2 ve σ_2^2 Biliniyor

Birbirinden bağımsız iki normal kitleden alınan bağımsız örneklemler için örneklem ortalamaları arasındaki farkın dağılımı da normaldir. Buna göre,

$$\overline{X}_1 - \overline{X}_2 \sim N\left((\mu_1 - \mu_2), (\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}) \right),$$

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})}} \sim N\left(0,1\right), P\left(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}\right) = 1-\alpha$$

$$P((\overline{X}_1 - \overline{X}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \leq (\mu_1 - \mu_2) \leq (\overline{X}_1 - \overline{X}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}) = 1-\alpha$$

$1-\alpha$ güven düzeyinde $\mu_1 - \mu_2$ için güven

dir.

1) Hipotez kurulur.

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 < \mu_2, \mu_1 > \mu_2, \mu_1 \neq \mu_2$$

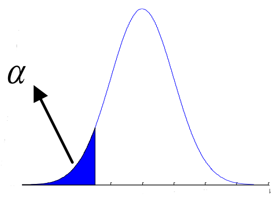
2) Test istatistiği hesaplanır.

$$z_h = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

3) Kritik bölgeye göre hipotez reddedilir yada reddedilemez.

$H_0 : \mu_1 = \mu_2$

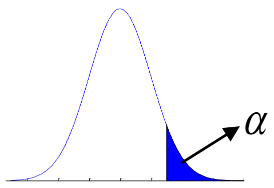
$H_1 : \mu_1 < \mu_2$



$z_h < -z_\alpha$ ise H_0 reddedilir

$H_0 : \mu_1 = \mu_2$

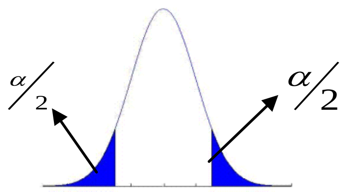
$H_1 : \mu_1 > \mu_2$



$z_h > z_\alpha$ ise H_0 reddedilir

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$



$z_h < -z_{\alpha/2}$ veya $z_h > z_{\alpha/2}$ ise H_0 reddedilir

3.2.1.2. Kitle Varyansı σ_1^2 ve σ_2^2 Bilinmiyor

Kitle varyanslarının bilinmediği durumda kitle varyanslarının birbirine eşit olup olmamasına göre hipotez testi yapılır.

a) $\sigma_1^2 = \sigma_2^2 = \sigma^2$ (Birbirlerine eşit)

Kitle varyansları bilinmiyor fakat kitle varyanslarının birbirine eşit olduğu biliniyor olabilir. Bu durumda örneklemlerin varyanslarından yararlanarak ortak bir varyans hesaplanır ve bu varyans kullanılarak hipotez testi yapılır.

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \sim t_{(n_1+n_2-2)} \ , S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

$$P((\overline{X}_1 - \overline{X}_2) - t_{(\alpha/2, n_1+n_2-2)} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq (\mu_1 - \mu_2) \leq (\overline{X}_1 - \overline{X}_2) + t_{(\alpha/2, n_1+n_2-2)} S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}) = 1 - \alpha$$

$1-\alpha$ güven düzeyinde $\mu_1 - \mu_2$ için güven

1) Hipotez kurulur.

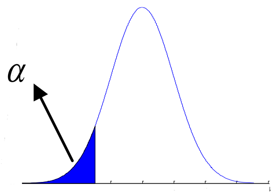
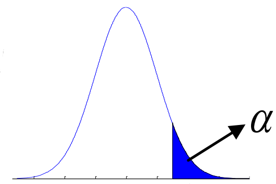
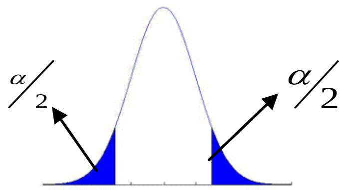
$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 < \mu_2, \mu_1 > \mu_2, \mu_1 \neq \mu_2$

2) Test istatistiği hesaplanır.

$$t_h = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

3) Kritik bölgeye göre hipotez reddedilir yada reddedilemez.

$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 < \mu_2$	$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 > \mu_2$	$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$
		
$t_h < -t_{n_1+n_2-2,\alpha}$ ise H_0 reddedilir	$t_h > t_{n_1+n_2-2,\alpha}$ ise H_0 reddedilir	$t_h < -t_{n_1+n_2-2,\alpha/2}$ veya $t_h > t_{n_1+n_2-2,\alpha/2}$ ise H_0 reddedilir

b) $\sigma_1^2 \neq \sigma_2^2$ (Birbirinden farklı)

Kitle varyansları bilinmiyor fakat kitle varyanslarının birbirinden farklı olduğu biliniyor ise ayrı ayrı örneklemlerin varyansları hesaplanır ve bu varyanslar kullanılarak hipotez testi yapılır.

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_{(\upsilon)}, \quad \upsilon = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\left(\frac{\left(\frac{S_1^2}{n_1}\right)^2}{n_1+1} + \frac{\left(\frac{S_2^2}{n_2}\right)^2}{n_2+1}\right)} - 2$$

$$P\left((\overline{X}_1 - \overline{X}_2) - t_{(\alpha/2,\upsilon)}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \leq (\mu_1 - \mu_2) \leq (\overline{X}_1 - \overline{X}_2) + t_{(\alpha/2,\upsilon)}\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}\right) = 1 - \alpha$$

$1-\alpha$ güven düzeyinde $\mu_1 - \mu_2$ için güven

1) Hipotez kurulur.

$H_0 : \mu_1 = \mu_2$

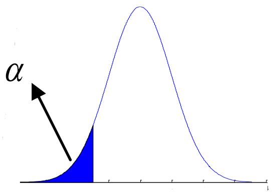
$H_1 : \mu_1 < \mu_2, \mu_1 > \mu_2, \mu_1 \neq \mu_2$

2) Test istatistiği hesaplanır.

$$t_h = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

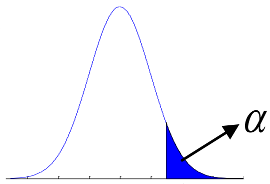
3) Kritik bölgeye göre hipotez reddedilir yada reddedilemez.

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 < \mu_2$



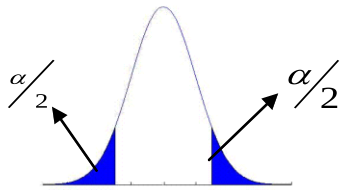
$t_h < -t_{(\upsilon,\alpha)}$ ise H_0 reddedilir

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 > \mu_2$



$t_h > t_{(\upsilon,\alpha)}$ ise H_0 reddedilir

$H_0 : \mu_1 = \mu_2$
$H_1 : \mu_1 \neq \mu_2$



$t_h < -t_{(\upsilon,\alpha/2)}$ veya $t_h > t_{(\upsilon,\alpha/2)}$ ise H_0 reddedilir

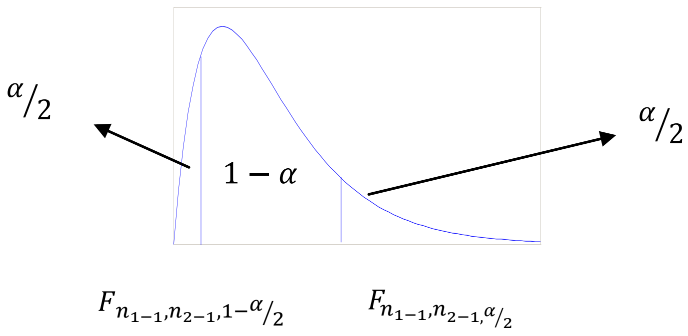
3.2.2. İki kitle varyans oranı için güven aralığı ve Hipotez testi

İki bağımsız normal kitle ile ilgilenildiği varsayılsın. İki kitle için ortalama ve varyanslar sırası ile μ_1, μ_2 ve σ_1^2, σ_2^2 ve bilinmiyor olsunlar. Bu kitlelerde, $\frac{\sigma_1^2}{\sigma_2^2}$ oranı için güven aralığı oluşturulmak istensin. Bu nedenle birinci kitleden n_1 ve ikinci kitleden n_2 birimlik örneklem seçilsin ve bu örneklemlere ilişkin varyanslar S_1^2, S_2^2 olarak elde edilmiş olsun. Bu durumda;

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F_{n_1-1, n_2-1}$$

olup F dağılımının yoğunluk fonksiyonunun eğrisi simetrik olmadığından tablo değerlerini bulabilmek için aşağıdaki özellik kullanılır,

$$F_{n_1-1, n_2-1, 1-\alpha/2} = \frac{1}{F_{n_2-1, n_1-1, \alpha/2}}$$



Buna göre, aşağıdaki gibi gerekli düzenlemeler yapılarak iki kitle varyans oranı için aralık tahmini elde edilir.

$$P\left(F_{n_1-1,n_2-1,1-\alpha/2} < \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} < F_{n_1-1,n_2-1,\alpha/2}\right) = 1-\alpha$$

$$P\left(\frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1,n_2-1,\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}F_{n_2-1,n_1-1,\alpha/2}\right) = 1-\alpha$$

biçiminde yazılır.

1) Hipotez kurulur.

H_0: $\sigma_1^2 = \sigma_2^2$
H_1: $\sigma_1^2 < \sigma_2^2$ ya da $\sigma_1^2 > \sigma_2^2$ ya da $\sigma_1^2 \neq \sigma_2^2$

2) Test istatistiği

$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

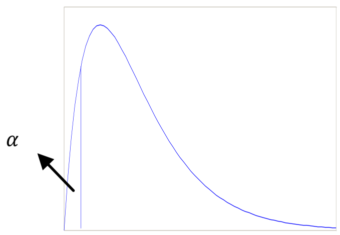
olmak üzere

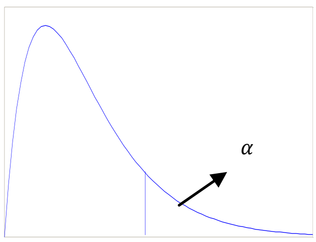
$$F_h = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2}$$

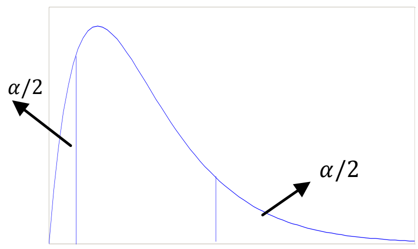
değeri hesaplanır.

3) Kritik bölgeye göre hipotez reddedilir ya da reddedilemez.

$H_0: \sigma_1^2 = \sigma_2^2$	$H_0: \sigma_1^2 = \sigma_2^2$	$H_0: \sigma_1^2 = \sigma_2^2$
$H_1: \sigma_1^2 < \sigma_2^2$	$H_1: \sigma_1^2 > \sigma_2^2$	$H_1: \sigma_1^2 \neq \sigma_2^2$







$F_h < F_{n_1-1,n_2-1,1-\alpha}$ ise H_0 reddedilir	$F_h > F_{n_1-1,n_2-1,\alpha}$ ise H_0 reddedilir	$F_h < F_{n_1-1,n_2-1,1-\alpha/2}$ ya da $F_h > F_{n_1-1,n_2-1,\alpha/2}$ ise H_0 reddedilir

3.2.3. İki kitle oranı arasındaki fark için güven aralığı ve hipotez testi

$p_1 - p_2$ farkı için güven aralığı oluşturmak amacıyla ilk olarak birinci kitleden n_1 ve ikinci kitleden n_2 birimlik örneklem seçilir. Birinci örneklemden oran tahmini, $P_1 = \frac{X_1}{n_1}$ ve ikinci örneklemden oran tahmini $P_2 = \frac{X_2}{n_2}$ olmak üzere, büyük örneklem çapları için

$$P_1 - P_2 \sim N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

olması varsayımı altında iki kitle oranı arasındaki fark için güven aralığı,

$1-\alpha$ güven düzeyinde
$p_1 - p_2$ için güven aralığı

$$P\left((P_1 - P_2) - z\alpha/_2\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}} < p_1 - p_2 < (P_1 - P_2) + z\alpha/_2\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}\right) = 1-\alpha$$

biçiminde elde edilir.

1) Hipotez kurulur.

H_0: $p_1 = p_2$
H_1: $p_1 < p_2$ ya da $p_1 > p_2$ ya da $p_1 \neq p_2$

2) Test istatistiği,

$$Z = \frac{(P_1 - P_2) - (p_1 - p_2)}{\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}} \sim N(0,1)$$

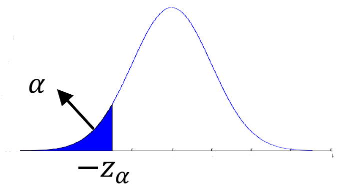
biçimindedir. P_1 ve P_2'nin örneklemden hesaplanan değerleri $\hat{p}_1$ ve $\hat{p}_2$ olmak üzere, H_0 hipotezinin doğruluğu koşulu altında test istatistiğin hesaplanan değeri

$$z_h = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

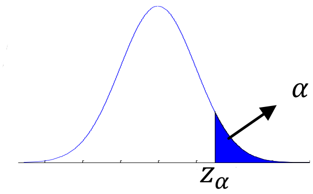
ile bulunur.

3) Kritik bölgeye göre hipotez reddedilir ya da reddedilemez.

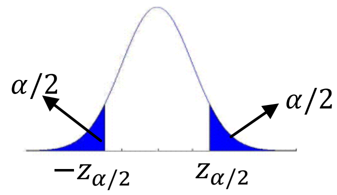
H_0: $p_1 = p_2$	H_0: $p_1 = p_2$	H_0: $p_1 = p_2$
H_1: $p_1 < p_2$	H_1: $p_1 > p_2$	H_1: $p_1 \neq p_2$



$z_h < -z_\alpha$ ise H_0 reddedilir



$z_h > z_\alpha$ ise H_0 reddedilir



$z_h < -z_{\alpha/2}$ ya da $z_h > -z_{\alpha/2}$ ise H_0 reddedilir

3.2.4 Eşler Arası Farkın Önem Kontrolü

İki kitle ortalaması arasındaki fark için hipotez testi yapılırken ve güven aralığı oluştururken çekilen örnekler bağımsız olmayabilirler. Örneğin verilen bir ilacın tansiyon üzerindeki etkisi araştırılıyorsa rasgele seçilen belli sayıdaki tansiyon hastasının ilaç almadan önce ve ilaç aldıktan sonra tansiyonları ölçülür.Yani, aynı birey üzerinde iki kez ölçüm yapılmış olur. Bu durumda örnekler bağımsız değildir.

İki bağımlı örnek için her bireyin önceki ve sonraki değerleri arasındaki fark alınarak işleme başlanır.

$$X_1 : X_{11}, X_{12}, ..., X_{1n} \sim N(\mu_1, \sigma_1^2)$$
$$X_2 : X_{21}, X_{22}, ..., X_{2n} \sim N(\mu_2, \sigma_2^2)$$

olmak üzere

$$(X_{11}, X_{21}), (X_{12}, X_{22}), ..., (X_{1n}, X_{2n})$$

ikilileri için,

$$D_j = X_{1j} - X_{2j}$$

farkları hesaplanır.

1) Hipotez kurulur.

$$H_0 : \mu_D = 0$$
$$H_1 : \mu_D < 0, \mu_D > 0, \mu_D \neq 0$$

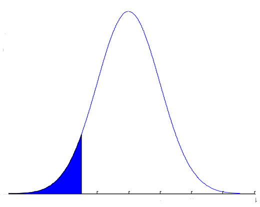
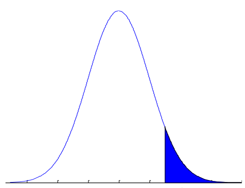
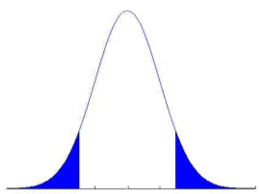
2) Test istatistiği hesaplanır.

$$t_h = \frac{\overline{D} - \mu_D}{S_D / \sqrt{n}} \sim t_{n-1}$$

$$D = X_{11} - X_{21}, X_{12} - X_{22}, ..., X_{1n} - X_{2n}$$

$$\overline{D} = \frac{\sum_{i=1}^{n} D_i}{n}, S_D = \sqrt{\frac{\sum_{i=1}^{n} D_i^2 - \frac{(\sum_{i=1}^{n} D_i)^2}{n}}{n-1}}$$

3) Kritik bölgeye göre hipotez red edilir yada red edilemez.

$H_0 : \mu_D = 0$ $H_1 : \mu_D < 0$	$H_0 : \mu_D = 0$ $H_1 : \mu_D > 0$	$H_0 : \mu_D = 0$ $H_1 : \mu_D \neq 0$
		
$t_h < -t_{n-1,\alpha}$ ise H_0 red edilir	$t_h > t_{n-1,\alpha}$ ise H_0 red edilir	$t_h < -t_{n-1,\alpha/2}$ veya $t_h > t_{n-1,\alpha/2}$ ise H_0 red edilir

Güven aralığı;

$$P(\overline{D} - t_{n-1,\alpha/2} \frac{S_D}{\sqrt{n}} \leq \mu \leq \overline{D} + t_{n-1,\alpha/2} \frac{S_D}{\sqrt{n}}) = 1 - \alpha$$

Örnek. 9 adet tansiyon hastasının ilaç başlanmadan önce büyük tansiyonları ölçülmüştür. İlaca başladıktan bir süre sonra tekrar büyük tansiyonları ölçülmüş ve sonuçlar aşağıda verildiği gibi bulunmuştur. İlacın tansiyon hastalarında tansiyon düşürmeye etkisi olduğu iddia edilmiştir. İddiayı $\alpha = 0.05$ anlam düzeyinde test ediniz.

	1	2	3	4	5	6	7	8	9
Önce	172	157	171	184	166	176	182	185	179
Sonra	167	154	168	176	157	175	179	182	175
D=Önce-Sonra	5	3	3	3	9	1	3	3	4

1) $H_0 : \mu_D = 0$ (İlacın tansiyonu düşürme etkisi yoktur.)

$H_1 : \mu_D > 0$ (İlacın tansiyonu düşürme etkisi vardır.)

2) $t_h = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}}$

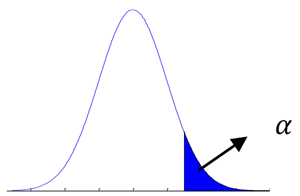
$$\bar{D} = \frac{\sum_{i=1}^{n} D_i}{n},$$

$$\sum_{i=1}^{n} D_i = 34, \ \sum_{i=1}^{n} D_i^2 = 168, \ \bar{D} = \frac{\sum_{i=1}^{n} D_i}{n} = \frac{34}{9} = 3.78$$

$$S_D = \sqrt{\frac{\sum_{i=1}^{n} D_i^2 - \frac{\left(\sum_{i=1}^{n} D_i\right)^2}{n}}{n-1}} = \sqrt{\frac{168 - -\frac{34^2}{9}}{9-1}} = \sqrt{4.94} = 2.22$$

$$t_h = \frac{\bar{D} - \mu_D}{S_D / \sqrt{n}} = \frac{3.78}{2.22 / \sqrt{9}} = 5.11$$

3)



$t_{n-1,\alpha} = 1.86$

$t_h = 5.11 > t_{n-1,\alpha} = 1.86$ olduğundan H_0 reddedilir.

Yorum: İlacın tansiyonu düşürmeye etkisi olduğu $\alpha = 0.05$ anlam düzeyinde söylenebilir.