T.C.

ONDOKUZ MAYIS ÜNİVERSİTESİ

FEN EDEBİYAT FAKÜLTESİ
İSTATİSTİK BÖLÜMÜ

ÇOK DEĞİŞKENLİ İSTATİSTİK I DERSİ

PROF. DR. YÜKSEL ÖNER

12. Hafta

4.4 Çok Değişkenli Normal Dağılıma Sahip İki Bağımsız Grubun Karşılaştırılması

Birinci grup $\underline{X}_1 \sim N_p\left(\underline{\mu}_1\,,\ \Sigma_1\right)$ ve ikinci grup $\underline{X}_2 \sim N_p\left(\underline{\mu}_2\,,\ \Sigma_2\right)$ dağılımlı olmak üzere, grupların birbirinden bağımsız ve ortak kitle varyans kovaryans matrisine sahip olduğunu kabul edelim. Gruplar için ortak kitle varyans kovaryans matrisini $\Sigma_1 = \Sigma_2 = \Sigma$ ile gösterelim. Bu durumda grupların karşılaştırılması kitle ortalama vektörleri yönünden yapılacaktır. Bu karşılaştırma kitle ortalama vektörleri üzerinde bir hipotez testi ile gerçekleştirilebilir. Bu hipotez testinde gerekli olan test istatistiğinin elde edilmesi Σ matrisinin bilinip bilinmemesine bağlıdır.

i) Σ matrisi biliniyor olsun

Test edilecek hipotezler;

$$H_0: \underline{\mu}_1 = \underline{\mu}_2, \quad \left(\underline{\mu}_1 - \underline{\mu}_2 = \underline{0}\right)$$

$$H_1: \underline{\mu}_1 \neq \underline{\mu}_2\,, \left(\underline{\mu}_1 - \underline{\mu}_2 \neq \underline{0}\right) \tag{4.37}$$

şeklinde oluşturulur. Bu hipotezleri test etmede kullanılacak olan test istatistiğinin elde edilebilmesi için her iki gruptan rastgele olarak örnekler çekilir. Birinci gruptan rastgele çekilen n_1 birimlik bir örnek $\underline{X}_{11}, \underline{X}_{12}, \dots , \underline{X}_{1n_1}$ ve ikinci gruptan rastgele çekilen n_2 birimlik bir örnek $\underline{X}_{21}, \underline{X}_{22}, \dots, \underline{X}_{2n_2}$ olsun. Bu örneklere göre; $\underline{\mu}_1$ parametresinin en çok olabilirlik tahmin edicisi $\overline{\underline{X}}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}\underline{X}_{1i}$ örnek ortalama vektörü istatistiği olup, bu istatistiğin örnekleme dağılımı $\overline{\underline{X}}_1 \sim N_p\left(\underline{\mu}_1\,,\ \frac{1}{n_1}\Sigma\right)$ dır. Benzer şekilde $\underline{\mu}_2$ parametresinin en çok olabilirlik tahmin edicisi $\overline{\underline{X}}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2}\underline{X}_{2i}$ örnek ortalama vektörü istatistiği olup, bu istatistiğin örnekleme dağılımı da $\overline{\underline{X}}_2 \sim N_p\left(\underline{\mu}_2\,,\ \frac{1}{n_2}\Sigma\right)$ olur. Böylece $\underline{\mu}_1 - \underline{\mu}_2$ parametresinin en çok olabilirlik tahmin edicisi olan istatistik $\overline{\underline{X}}_1 - \overline{\underline{X}}_2$ örnek ortalama vektörleri farkı istatistiğidir. Bu istatistiğin örnekleme dağılımı ise $\overline{\underline{X}}_1 - \overline{\underline{X}}_2 \sim N_p\left(E\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right),\ Cov\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right)\right)$ olacaktır, öyle ki burada; $E\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right) = E\left(\overline{\underline{X}}_1\right) - E\left(\overline{\underline{X}}_2\right) = \underline{\mu}_1 - \underline{\mu}_2$ ve

$$Cov\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right) = Cov\left(\overline{\underline{X}}_1\right) + Cov\left(\overline{\underline{X}}_2\right), \left(\overline{\underline{X}}_1\ ve\ \overline{\underline{X}}_2 \text{ bağımsız olduğundan}\right)$$

$= \frac{1}{n_1}\Sigma + \frac{1}{n_2}\Sigma = \frac{n_1+n_2}{n_1 n_2}\Sigma$ olur. Böylece $\overline{\underline{X}}_1 - \overline{\underline{X}}_2 \sim N_p\left(\underline{\mu}_1 - \underline{\mu}_2\,,\ \frac{n_1+n_2}{n_1 n_2}\Sigma\right)$ olup, Σ matrisi bilindiğinden H_0 hipotezini test etmek için kullanılacak olan test istatistiği bu dağılımdan türetilecek olan karesel form olacaktır. Bu karesel form ve karesel formun örnekleme dağılımı;

$$\chi^2 = \left[\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right) - \left(\underline{\mu}_1 - \underline{\mu}_2\right)\right]' \left(\frac{n_1+n_2}{n_1 n_2}\Sigma\right)^{-1} \left[\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right) - \left(\underline{\mu}_1 - \underline{\mu}_2\right)\right] \sim \chi^2_p$$

$$= \frac{n_1 n_2}{n_1+n_2}\left[\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right) - \left(\underline{\mu}_1 - \underline{\mu}_2\right)\right]' \Sigma^{-1} \left[\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right) - \left(\underline{\mu}_1 - \underline{\mu}_2\right)\right] \sim \chi^2_p \tag{4.38}$$

şeklindedir. H_0 hipotezi doğru iken test istatistiğinin alabileceği değer;

$$\chi_h^2 = \frac{n_1 n_2}{n_1+n_2}\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right)' \Sigma^{-1}\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right) \tag{4.39}$$

olacaktır.

Karar: H_1 hipotezine göre α önem seviyesinde kritik değer $\chi_t^2 = \chi_{p;\alpha}^2$ olmak üzere, eğer;

$\chi_h^2 \leq \chi_t^2$ ise H_0 hipotezi ret edilemez

$\chi_h^2 > \chi_t^2$ ise H_0 hipotezi ret edilir.

ii) Σ matrisi bilinmiyor olsun

Test edilecek hipotezler; Eşitlik (4.37)'de verilenle aynıdır. Bu hipotezleri test etmede kullanılacak olan test istatistiğinin elde edilebilmesi için her iki gruptan yine rastgele olarak örnekler çekilir. Birinci gruptan rastgele çekilen n_1 birimlik bir örnek $\underline{X}_{11}, \underline{X}_{12}, \ldots, \underline{X}_{1n_1}$ve ikinci gruptan rastgele çekilen n_2 birimlik bir örnek $\underline{X}_{21}, \underline{X}_{22}, \ldots, \underline{X}_{2n_2}$ olsun. Bu örneklere göre; $\underline{\mu}_1$ parametresinin en çok olabilirlik tahmin edicisi $\overline{\underline{X}}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1}\underline{X}_{1i}$ örnek ortalama vektörü istatistiği olup, bu istatistiğin örnekleme dağılımı $\overline{\underline{X}}_1 \sim N_p\left(\underline{\mu}_1, \frac{1}{n_1}\Sigma\right)$ dır. Benzer şekilde $\underline{\mu}_2$ parametresinin en çok olabilirlik tahmin edicisi $\overline{\underline{X}}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2}\underline{X}_{2i}$ örnek ortalama vektörü istatistiği olup, bu istatistiğin örnekleme dağılımı da $\overline{\underline{X}}_2 \sim N_p\left(\underline{\mu}_2, \frac{1}{n_2}\Sigma\right)$ olur. Böylece $\underline{\mu}_1 - \underline{\mu}_2$ parametresinin en çok olabilirlik tahmin edicisi olan istatistik $\overline{\underline{X}}_1 - \overline{\underline{X}}_2$ örnek ortalama vektörleri farkı istatistiğidir. Bu istatistiğin örnekleme dağılımı;

$\overline{\underline{X}}_1 - \overline{\underline{X}}_2 \sim N_p\left(\underline{\mu}_1 - \underline{\mu}_2, \frac{n_1+n_2}{n_1 n_2}\Sigma\right)$ olacaktır. Burada Σ matrisi bilinmediğinden $Cov\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right)$ değerini elde edebilmek için Σ matrisinin bir yansız tahmin edicisi olan ve örneklem ortak varyans kovaryans matrisi olarak bilinen S matrisi elde edilmelidir. Bu S matrisi;

$$S = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1+n_2-2} = \frac{W_1+W_2}{n_1+n_2-2}\left(veya\ S = \frac{n_1 S_1 + n_2 S_2}{n_1+n_2}, n_1 ve\ n_2\ çok\ büyükken\right) \tag{4.40}$$

eşitliği ile hesaplanır. Burada $k = 1, 2$ için S_k , k-ncı örnekleme ait örnek varyans kovaryans matrisi olup;

$$S_k = \frac{1}{n_k-1}\sum_{i=1}^{n_k}\left(\underline{X}_{ki} - \overline{\underline{X}}_k\right)\left(\underline{X}_{ki} - \overline{\underline{X}}_k\right)', k = 1, 2 \tag{4.41}$$

eşitliği ile hesaplanır ve $k = 1, 2$ için Σ_k parametresinin bir yansız tahmin edicisidir. Böylece test istatistiği iki örneklem Hotelling T^2 istatistiği olarak bilinen;

$$T^2 = \left[\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right) - \left(\underline{\mu}_1 - \underline{\mu}_2\right)\right]' \left(\frac{n_1+n_2}{n_1 n_2}S\right)^{-1} \left[\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right) - \left(\underline{\mu}_1 - \underline{\mu}_2\right)\right]$$

$$= \frac{n_1 n_2}{n_1+n_2}\left[\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right) - \left(\underline{\mu}_1 - \underline{\mu}_2\right)\right]' S^{-1} \left[\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right) - \left(\underline{\mu}_1 - \underline{\mu}_2\right)\right] \tag{4.42}$$

istatistiği kullanılmaktadır. H_0 hipotezi doğru iken test istatistiğinin alabileceği değer;

$$T_h^2 = \frac{n_1 n_2}{n_1+n_2}\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right)' S^{-1}\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right) \tag{4.43}$$

olur. İki örneklem Hotelling- T^2 istatistiğinin örnekleme dağılımı;

$$F = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)} T^2 \sim F_{p;n_1+n_2-p-1} \quad (4.44)$$

olup, H_0 hipotezi doğru iken F istatistiğinin alabileceği değer ise; $F_h = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)} T_h^2$ olur.

Karar: H_0 hipotezi hakkında karar verme ya F ya da T^2 istatistiğine göre verilebilir. H_1 hipotezine göre α önem seviyesindeki kritik değer $F_t = F_{p;\, n_1+n_2-p-1;\alpha}$ veya $T_t^2 = \frac{p(n_1+n_2-2)}{n_1+n_2-p-1} F_t$ olmak üzere, eğer;

$F_h \leq F_t$ $(veya\ T_h^2 \leq T_t^2)$ ise H_0 hipotezi ret edilemez

$F_h > F_t$ $(veya\ T_h^2 > T_t^2)$ ise H_0 hipotezi ret edilir.

Yorum: H_0 hipotezi kabul edilsin. Bu durumda gruplara ait kitle ortalama vektörlerinin $\left(\underline{\mu}_1\ ve\ \underline{\mu}_2\right)$ aynı olduğu ya da n_1 ve n_2 birimlik örneklerin %(1-α) güvenle aynı kitleden çekildikleri söylenir.

H_0 hipotezi ret edilsin. Bu durumda örneklerin çekildiği kitleler farklı kitle ortalama vektörlerine sahiptir. Ancak; bu farklılığın kaynağının ne olduğunun veya hangi değişken/değişkenler olduğunun araştırılması gerekir. Bu amaçla: $\underline{a} : 1 \times p$ bilinen bir vektör ve $\underline{\delta} = \underline{\mu}_1 - \underline{\mu}_2$ kitle ortalamalar farkı olmak üzere $\underline{a}\,\underline{\delta}$ parametresi için güven aralığından yararlanılır. Bu güven aralığı;

$$H_0: \underline{a}\,\underline{\delta} = 0$$

$$H_1: \underline{a}\,\underline{\delta} \neq 0 \quad (4.45)$$

hipotezlerinin test edilmesi amacıyla kullanılır. $\underline{a} = [0\ \ldots 1 \ldots 0] : 1 \times p$, (j-nci bileşen 1 ve diğer bileşenlerin hepsi "0" seçiliyor) ve $\underline{\delta} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_p \end{bmatrix} = \begin{bmatrix} \mu_{11} - \mu_{12} \\ \mu_{21} - \mu_{22} \\ \vdots \\ \mu_{P1} - \mu_{P2} \end{bmatrix}$ olmak üzere $\underline{a}\,\underline{\delta} = \mu_{j1} - \mu_{j2}$, $(j = 1, 2, \ldots, p)$ olup, p- tane değişken için karşılaştırma yapılır. Bu karşılaştırmaları yapmak için kullanılacak güven aralığı;

i) Σ matrisi biliniyorsa;

$$P\left[\underline{a}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) - \sqrt{\chi^2_{p;\alpha}} \sqrt{\frac{n_1+n_2}{n_1 n_2} \underline{a}\, \Sigma\, \underline{a}'} \leq \underline{a}\,\underline{\delta} \leq \underline{a}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) + \sqrt{\chi^2_{p;\alpha}} \sqrt{\frac{n_1+n_2}{n_1 n_2} \underline{a}\, \Sigma\, \underline{a}'}\ \right] = 1 - \alpha \quad (4.46)$$

ii) Σ matrisi bilinmiyorsa;

$$P\left[\underline{a}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) - \sqrt{T_t^2} \sqrt{\frac{n_1+n_2}{n_1 n_2} \underline{a}\, \mathrm{S}\, \underline{a}'} \leq \underline{a}\,\underline{\delta} \leq \underline{a}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) + \sqrt{T_t^2} \sqrt{\frac{n_1+n_2}{n_1 n_2} \underline{a}\, \mathrm{S}\, \underline{a}'}\ \right] = 1 - \alpha \quad (4.47)$$

şeklindedir. Eğer güven aralığı sıfırı kapsıyorsa H_0 kabul edilir, kapsamıyorsa H_0 ret edilir.

Örnek 4.5 İstatistik bölümü 2016 girişli öğrenciler kitlesi ile 2017 girişli öğrenciler kitlesinden rastgele olarak 8 ve 10 birimlik örnekler çekilmiştir. Bu öğrencilerin Olasılık Teorisi I (X_1), Matematiksel İstatistik (X_2) ve Örnekleme I (X_3) derslerine ait başarı notları aşağıdaki gibi gözlenmiştir. Her bir kitleye ait dağılımların çok değişkenli normal ve kitleler için ortak varyans kovaryans matrisinin $\Sigma = \begin{bmatrix} 25 & 6 & 2 \\ 6 & 9 & -2 \\ 2 & -2 & 4 \end{bmatrix}$ olduğunu kabul edelim. Buna göre %1 önem seviyesinde 2016 ve 2017 kodlu öğrenciler için bu derslere ait başarı ortalama vektörleri arasında fark olup olmadığına karar veriniz?

2016 (I. Grup)			**2017 (II.Grup)**		
Olasılık Teorisi I (X_{11})	**Matematiksel İstatistik** (X_{21})	**Örnekleme I** (X_{31})	**Olasılık Teorisi I** (X_{12})	**Matematiksel İstatistik** (X_{22})	**Örnekleme I** (X_{32})
54	62	40	35	42	48
46	48	30	65	60	52
38	34	26	80	74	64
62	50	44	40	30	26
12	26	10	36	38	12
24	30	20	74	56	38
78	62	42	52	40	35
86	78	68	28	44	35
			18	20	14
			42	56	26
Top: 400	390	280	470	460	350
$n_1 = 8$	8	8	$n_2 = 10$	10	10
$\overline{X}_{11} = 50$	$\overline{X}_{21}$ =48,75	$\overline{X}_{31}$ =35	$\overline{X}_{12}$ =47	$\overline{X}_{22}$ =46	$\overline{X}_{32}$ =35

Çözüm: I. Kitle (Grup): 2016 girişli öğrenciler.... $\underline{X}_1 = \begin{bmatrix} X_{11} \\ X_{21} \\ X_{31} \end{bmatrix} \sim N_3\left(\underline{\mu}_1, \Sigma\right)$

II. Kitle (Grup): 2017 girişli öğrenciler.... $\underline{X}_2 = \begin{bmatrix} X_{12} \\ X_{22} \\ X_{32} \end{bmatrix} \sim N_3\left(\underline{\mu}_2, \Sigma\right)$ olmak üzere, bu grupların bağımsız ve ortak varyans-kovaryans matrisine sahip olduğu varasayılmıştır. Kitleler için ortak kovaryans matrisi bilinmektedir.

Test edilecek hipotezler:

$H_0: \underline{\mu}_1 = \underline{\mu}_2, \quad \left(\underline{\mu}_1 - \underline{\mu}_2 = \underline{0}\right)$

$H_1: \underline{\mu}_1 \neq \underline{\mu}_2, \left(\underline{\mu}_1 - \underline{\mu}_2 \neq \underline{0}\right)$

Test istatistiği; Σ ortak kovaryans matrisi bilindiğinden H_0 doğru iken

$\chi^2 = \frac{n_1 n_2}{n_1+n_2}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)' \Sigma^{-1}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) \sim \chi^2_p$ olup, test istatistiğinin alabileceği değer;

$$\chi^2_h = \frac{n_1 n_2}{n_1+n_2}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)' \Sigma^{-1}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)\ ;\ n_1 = 8, n_2 = 10, p = 3$$

$\underline{\overline{X}}_1 = \begin{bmatrix} 50 \\ 48{,}75 \\ 35 \end{bmatrix}$ ve $\underline{\overline{X}}_2 = \begin{bmatrix} 47 \\ 46 \\ 35 \end{bmatrix}$ iken $\underline{\overline{X}}_1 - \underline{\overline{X}}_2 = \begin{bmatrix} 3 \\ 2{,}75 \\ 0 \end{bmatrix}$ bulunur.

$\Sigma^{-1} = \frac{1}{|\Sigma|} Adj(\Sigma) = \frac{1}{|\Sigma|} Kof(\Sigma)$, ($\Sigma$ matrisi simetrik olduğundan $Adj(\Sigma) = Kof(\Sigma)$ dır)

$$\Sigma = \begin{bmatrix} 25 & 6 & 2 \\ 6 & 9 & -2 \\ 2 & -2 & 4 \end{bmatrix},\quad |\Sigma| = 25(32) - 6(28) + 2(-30) = 572$$

$\Sigma^{-1} = \frac{1}{572}\begin{bmatrix} 32 & -28 & -30 \\ -28 & 96 & 62 \\ -30 & 62 & 189 \end{bmatrix}$ bu sonuçlar formülde yerlerine yazılırsa;

$$\chi^2_h = \frac{8\times10}{8+10}\begin{bmatrix} 3 & 2{,}75 & 0 \end{bmatrix} \frac{1}{572}\begin{bmatrix} 32 & -28 & -30 \\ -28 & 96 & 62 \\ -30 & 62 & 189 \end{bmatrix}\begin{bmatrix} 3 \\ 2{,}75 \\ 0 \end{bmatrix} = \frac{80}{10296}\begin{bmatrix} 19 & 180 & 80{,}5 \end{bmatrix}\begin{bmatrix} 3 \\ 2{,}75 \\ 0 \end{bmatrix}$$

$= 4{,}289$ elde edilir.

Karar: $\alpha = 0{,}01$ önem seviyesinde kritik değer; $\chi^2_t = \chi^2_{p;\alpha} = \chi^2_{3;0{,}01} = 11{,}34$ dür.

$4{,}289 < 11{,}34$ yani $\chi^2_h < \chi^2_t$ olduğundan H_0 hipotezi ret edilemez. Buna göre 2016 ve 2017 kodlu öğrencilerin bu üç derse ait ortalama başarıları arasında istatistiksel olarak anlamlı bir farklılık yoktur.

Örnek 4.6 49 yaşlı kişi kadın ve erkek olarak iki gruba ayrılmıştır. Birinci grup erkekler 37 kişiden ve ikinci grup kadınlar 12 kişiden oluşmaktadır. Bu iki gruba 4 değişik özelliği kapsayan bir test uygulanmıştır. Her bir grubun ilgili özellikler bakımından dağılımlarının $N_4\left(\underline{\mu}_k, \Sigma\right)$, $(k = 1, 2)$ olduğu bilinmektedir. Uygulanan testten özelliklere ait örnek ortalamaları ile örneklem için ortak varyans-kovaryans matrisi aşağıdaki gibi bulunmuştur.

Özellikler (X_j)	$\overline{X}_{j1}$	$\overline{X}_{j2}$
Tanıma Yeteneği (X_1)	12,57	8,75
Benzerlik Yeteneği (X_2)	9,57	5,33
Aritmetik Yeteneği (X_3)	11,49	8,50
Resim Yeteneği (X_4)	7,97	4,75

$$S = \begin{bmatrix} 11{,}25 & 9{,}40 & 7{,}15 & 3{,}38 \\ 9{,}40 & 13{,}53 & 3{,}38 & 2{,}55 \\ 7{,}15 & 3{,}38 & 11{,}57 & 2{,}62 \\ 3{,}38 & 2{,}55 & 2{,}62 & 5{,}81 \end{bmatrix}$$

a) Dört özellik açısından erkekler ve kadınlar kitleleri için ortalamalar arasında fark olup olmadığına %5 önem seviyesinde karar veriniz.

b) Eğer ortalamalar arasında fark varsa, bunun hangi özellikten kaynaklandığını belirleyiniz.

Çözüm a) I. Kitle: Erkekler … $\underline{X}_1 = \begin{bmatrix} X_{11} \\ X_{21} \\ X_{31} \\ X_{41} \end{bmatrix} \sim N_4\left(\underline{\mu}_1, \Sigma\right)$

II. Kitle: Kadınlar … $\underline{X}_2 = \begin{bmatrix} X_{12} \\ X_{22} \\ X_{32} \\ X_{42} \end{bmatrix} \sim N_4\left(\underline{\mu}_2, \Sigma\right)$ olup, gruplar bağımsız ve ortak varyans-kovaryans matrisine sahip, ancak Σ matrisi bilinmemektedir.

Test edilecek hipotezler:

H_0: $\underline{\mu}_1 = \underline{\mu}_2$, $\left(\underline{\mu}_1 - \underline{\mu}_2 = \underline{0}\right)$

H_1: $\underline{\mu}_1 \neq \underline{\mu}_2$, $\left(\underline{\mu}_1 - \underline{\mu}_2 \neq \underline{0}\right)$

Test istatistiği: Σ matrisi bilinmediğinden iki örneklem Hotelling- T^2 istatistiğidir.

$T^2 = \frac{n_1 n_2}{n_1+n_2}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)' S^{-1}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)$ olmak üzere $F = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)} T^2 \sim F_{p;n_1+n_2-p-1}$ dir.

$n_1 = 37, n_2 = 12, p = 4$

$\underline{\overline{X}}_1 = \begin{bmatrix} 12{,}57 \\ 9{,}57 \\ 11{,}49 \\ 7{,}97 \end{bmatrix}$ ve $\underline{\overline{X}}_2 = \begin{bmatrix} 8{,}75 \\ 5{,}33 \\ 8{,}50 \\ 4{,}75 \end{bmatrix}$ iken $\underline{\overline{X}}_1 - \underline{\overline{X}}_2 = \begin{bmatrix} 3{,}82 \\ 4{,}24 \\ 2{,}99 \\ 3{,}22 \end{bmatrix}$ olur. S ortak kovaryans matrisinin tersi

$S^{-1} = \begin{bmatrix} 0{,}259 & -0{,}36 & -0{,}053 & -0{,}065 \\ -0{,}36 & 0{,}186 & -0{,}038 & 0{,}014 \\ -0{,}053 & -0{,}038 & 0{,}151 & -0{,}017 \\ -0{,}065 & 0{,}014 & -0{,}017 & 0{,}121 \end{bmatrix}$ olup, bu sonuçlar formülde yerlerine yazılırsa;

$$T_h^2 = \frac{37\times 12}{49}\begin{bmatrix} 3{,}82 & 4{,}24 & 2{,}99 & 3{,}22 \end{bmatrix} \begin{bmatrix} 0{,}259 & -0{,}36 & -0{,}053 & -0{,}065 \\ -0{,}36 & 0{,}186 & -0{,}038 & 0{,}014 \\ -0{,}053 & -0{,}038 & 0{,}151 & -0{,}017 \\ -0{,}065 & 0{,}014 & -0{,}017 & 0{,}121 \end{bmatrix} \begin{bmatrix} 3{,}82 \\ 4{,}24 \\ 2{,}99 \\ 3{,}22 \end{bmatrix}$$

$= 21{,}74$ bulunur. Buna göre;

$F_h = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)} T_h^2 = \frac{44}{4\times 47}(21{,}74) = 5{,}088$ elde edilir.

Karar: $\alpha = 0{,}05$ için kritik değer; $F_t = F_{p;\, n_1+n_2-p-1;\alpha} = F_{4;44;0{,}05} = 2{,}59$ ve böylece

$T_t^2 = \frac{p(n_1+n_2-2)}{n_1+n_2-p-1} \times F_t = \frac{4\times 47}{44}(2{,}59) = 11{,}066$

5,088 > 2,59 (yani $F_h > F_t$) ya da 21,74 > 11,066 (yani $T_h^2 > T_t^2$) olduğundan H_0 hipotezi ret edilir. Buna göre Erkek ve kadın grupları için söz konusu dört özellik bakımından ortalamalar arasında anlamlı bir fark vardır.

b) $H_0: \underline{\mu}_1 = \underline{\mu}_2$ hipotezi ret edildiğinden erkek ve kadın grupları arasındaki farklılığın hangi özellikten kaynaklandığını belirleyebilmek için $\underline{\delta} = \underline{\mu}_1 - \underline{\mu}_2$ kitle ortalamalar farkı olmak üzere $\underline{a}\,\underline{\delta}$ parametresi için güven aralığından yararlanılır.

i) $\boldsymbol{X_1}$: *Tanıma Yeteneği*: $\underline{a} = [1 \quad 0 \quad 0 \quad 0]$ seçelim. Bu takdirde;

$$\underline{a}\,\underline{\delta} = \underline{a}\left(\underline{\mu}_1 - \underline{\mu}_2\right) = [1 \quad 0 \quad 0 \quad 0]\begin{bmatrix}\mu_{11}-\mu_{12}\\ \mu_{21}-\mu_{22}\\ \mu_{31}-\mu_{32}\\ \mu_{41}-\mu_{42}\end{bmatrix} = \mu_{11}-\mu_{12}$$ olacağından test edilecek hipotezler:

$H_0: \mu_{11} - \mu_{12} = 0$

$H_1: \mu_{11} - \mu_{12} \neq 0$ olup, güven aralığı Σ matrisi bilinmediğinden Eşitlik (4.47) gereğince;

$$P\left[\underline{a}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) - \sqrt{T_t^2}\sqrt{\frac{n_1+n_2}{n_1 n_2}\underline{a}\,S\,\underline{a}'} \leq \underline{a}\,\underline{\delta} \leq \underline{a}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) + \sqrt{T_t^2}\sqrt{\frac{n_1+n_2}{n_1 n_2}\underline{a}\,S\,\underline{a}'}\right] = 1-\alpha$$

$$P\left[3,82 - \sqrt{11,066}\sqrt{\frac{49}{37\times 12}(11,25)} \leq \mu_{11} - \mu_{12} \leq 3,82 + \sqrt{11,066}\sqrt{\frac{49}{37\times 12}(11,25)}\right] = 0,95$$

$P\left(0,11 \leq \mu_{11} - \mu_{12} \leq 7,53\right) = 0,95$ elde edilir. Güven aralığı sıfırı kapsamadığından tanıma yeteneği bakımından erkeklerin ve kadınların ortalama değerleri birbirinden farklıdır. Üstelik %95 güvenle erkeklere ait ortalamanın kadınlara ait ortalamadan daha yüksek olduğu söylenebilir.

ii) $\boldsymbol{X_2}$: *Benzerlik Yeteneği*: $\underline{a} = [0 \quad 1 \quad 0 \quad 0]$ seçelim. Bu takdirde;

$$\underline{a}\,\underline{\delta} = \underline{a}\left(\underline{\mu}_1 - \underline{\mu}_2\right) = [0 \quad 1 \quad 0 \quad 0]\begin{bmatrix}\mu_{11}-\mu_{12}\\ \mu_{21}-\mu_{22}\\ \mu_{31}-\mu_{32}\\ \mu_{41}-\mu_{42}\end{bmatrix} = \mu_{21}-\mu_{22}$$ olacağından test edilecek hipotezler:

$H_0: \mu_{21} - \mu_{22} = 0$

$H_1: \mu_{21} - \mu_{22} \neq 0$ olup, güven aralığı Σ matrisi bilinmediğinden Eşitlik (4.47) gereğince;

$$P\left[\underline{a}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) - \sqrt{T_t^2}\sqrt{\frac{n_1+n_2}{n_1 n_2}\underline{a}\,S\,\underline{a}'} \leq \underline{a}\,\underline{\delta} \leq \underline{a}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) + \sqrt{T_t^2}\sqrt{\frac{n_1+n_2}{n_1 n_2}\underline{a}\,S\,\underline{a}'}\right] = 1-\alpha$$

$$P\left[4,24 - \sqrt{11,066}\sqrt{\frac{49}{37\times 12}(13,53)} \leq \mu_{21} - \mu_{22} \leq 4,24 + \sqrt{11,066}\sqrt{\frac{49}{37\times 12}(13,53)}\right] = 0,95$$

$P\left(0,175 \leq \mu_{21} - \mu_{22} \leq 8,305\right) = 0,95$ elde edilir. Güven aralığı sıfırı kapsamadığından benzerlik yeteneği bakımından erkeklerin ve kadınların ortalama değerleri birbirinden farklıdır. Üstelik %95 güvenle erkeklere ait ortalamanın kadınlara ait ortalamadan daha yüksek olduğu söylenebilir.

iii) X_3: $\boldsymbol{Aritmetik\ Yeteneği}$: $\underline{a} = [0 \quad 0 \quad 1 \quad 0]$ seçelim. Bu takdirde;

$\underline{a}\,\underline{\delta} = \underline{a}\,\left(\underline{\mu}_1 - \underline{\mu}_2\right) = [0 \quad 0 \quad 1 \quad 0]\begin{bmatrix}\mu_{11}-\mu_{12}\\ \mu_{21}-\mu_{22}\\ \mu_{31}-\mu_{32}\\ \mu_{41}-\mu_{42}\end{bmatrix} = \mu_{31}-\mu_{32}$ olacağından test edilecek hipotezler:

$H_0: \mu_{31}-\mu_{32} = 0$

$H_1: \mu_{31}-\mu_{32} \neq 0$ olup, güven aralığı Σ matrisi bilinmediğinden Eşitlik (4.47) gereğince;

$$P\left[\underline{a}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) - \sqrt{T_t^2}\sqrt{\frac{n_1+n_2}{n_1 n_2}\underline{a}\,\mathrm{S}\,\underline{a}'} \le \underline{a}\,\underline{\delta} \le \underline{a}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) + \sqrt{T_t^2}\sqrt{\frac{n_1+n_2}{n_1 n_2}\underline{a}\,\mathrm{S}\,\underline{a}'}\right] = 1-\alpha$$

$$P\left[2{,}99 - \sqrt{11{,}066}\sqrt{\frac{49}{37\times 12}(11{,}57)} \le \mu_{31}-\mu_{32} \le 2{,}99 + \sqrt{11{,}066}\sqrt{\frac{49}{37\times 12}(11{,}57)}\right] = 0{,}95$$

$P\left(-0{,}769 \le \mu_{31}-\mu_{32} \le 6{,}749\right) = 0{,}95$ elde edilir. Güven aralığı sıfırı kapsadığından aritmetik yeteneği bakımından erkeklerin ve kadınların ortalama değerleri birbirine benzerdir, farklılık göstermemektedir.

iv) X_4: $\boldsymbol{Resim\ Yeteneği}$: $\underline{a} = [0 \quad 0 \quad 0 \quad 1]$ seçelim. Bu takdirde;

$\underline{a}\,\underline{\delta} = \underline{a}\,\left(\underline{\mu}_1 - \underline{\mu}_2\right) = [0 \quad 0 \quad 0 \quad 1]\begin{bmatrix}\mu_{11}-\mu_{12}\\ \mu_{21}-\mu_{22}\\ \mu_{31}-\mu_{32}\\ \mu_{41}-\mu_{42}\end{bmatrix} = \mu_{41}-\mu_{42}$ olacağından test edilecek hipotezler:

$H_0: \mu_{41}-\mu_{42} = 0$

$H_1: \mu_{41}-\mu_{42} \neq 0$ olup, güven aralığı Σ matrisi bilinmediğinden Eşitlik (4.47) gereğince;

$$P\left[\underline{a}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) - \sqrt{T_t^2}\sqrt{\frac{n_1+n_2}{n_1 n_2}\underline{a}\,\mathrm{S}\,\underline{a}'} \le \underline{a}\,\underline{\delta} \le \underline{a}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) + \sqrt{T_t^2}\sqrt{\frac{n_1+n_2}{n_1 n_2}\underline{a}\,\mathrm{S}\,\underline{a}'}\right] = 1-\alpha$$

$$P\left[3{,}22 - \sqrt{11{,}066}\sqrt{\frac{49}{37\times 12}(5{,}81)} \le \mu_{21}-\mu_{22} \le 3{,}22 + \sqrt{11{,}066}\sqrt{\frac{49}{37\times 12}(5{,}81)}\right] = 0{,}95$$

$P\left(0{,}556 \le \mu_{41}-\mu_{42} \le 5{,}884\right) = 0{,}95$ elde edilir. Güven aralığı sıfırı kapsamadığından resim yeteneği bakımından erkeklerin ve kadınların ortalama değerleri birbirinden farklıdır. Üstelik %95 güvenle erkeklere ait ortalamanın kadınlara ait ortalamadan daha yüksek olduğu söylenebilir.

SORULAR

SORU:1 $\underline{X} \sim N_3\left(\underline{\mu}\,,\ \Sigma\right)$ dağılımının olasılık yoğunluk fonksiyonu

$f\left(\underline{x}\right) = kexp\left\{-\frac{1}{2}Q\left(\underline{x}\right)\right\}$ öyle ki, burada

$Q(\underline{x}) = \frac{4}{23}x_1^2 + \frac{1}{5}x_2^2 + \frac{6}{23}x_3^2 + \frac{2}{23}x_1x_3 + \frac{6}{23}x_1 - \frac{10}{23}x_3 + \frac{8}{23}$ olarak verilsin. Buna göre $\underline{X} = \begin{bmatrix} \underline{X_1} \\ \cdots \\ \underline{X_2} \end{bmatrix}$ öyle ki $\underline{X_1} = \begin{bmatrix} X_1 \\ X_3 \end{bmatrix}$ ve $\underline{X_2} = [X_2]$ parçalanması verilsin. Buna göre;

a) $\underline{X_1}$ alt vektörünün marjinal dağılımının parametreleri ile $\underline{X_2} = [1]$ iken $\underline{X_1}$ alt vektörünün şartlı dağılımının parametrelerini bulunuz?

b) Verilen dağılımdan çekilen 10 birimlik bir örneklem için örnek ortalama vektörünün $\underline{\overline{X}} = \begin{bmatrix} 1 \\ \frac{5}{2} \\ 1 \end{bmatrix}$ olup olamayacağına %95 güven seviyesinde karar veriniz?

CEVAP:1

a) Önce dağılımın parametrelerini bulalım.

$$Q(\underline{x}) = \frac{4}{23}x_1^2 + \frac{1}{5}x_2^2 + \frac{6}{23}x_3^2 + \frac{2}{23}x_1x_3 + \frac{6}{23}x_1 - \frac{10}{23}x_3 + \frac{8}{23}$$

Σ matrisinin bulunması:

$$\Sigma^{-1} = \begin{bmatrix} \frac{4}{23} & 0 & \frac{1}{23} \\ 0 & \frac{1}{5} & 0 \\ \frac{1}{23} & 0 & \frac{6}{23} \end{bmatrix} \Rightarrow \ |\ \Sigma^{-1}| = \frac{1}{5}\left(\frac{24}{23\times23} - \frac{1}{23\times23}\right) = \frac{1}{5} \times \frac{23}{23\times23} = \frac{1}{5\times23}$$

$$\Sigma = (\ \Sigma^{-1})^{-1} = \frac{5\times23}{1}\begin{bmatrix} \frac{6}{5\times23} & 0 & -\frac{1}{5\times23} \\ 0 & \frac{1}{23} & 0 \\ -\frac{1}{5\times23} & 0 & \frac{4}{5\times23} \end{bmatrix} = \begin{bmatrix} 6 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 4 \end{bmatrix}$$

$\underline{\boldsymbol{\mu}}$ **vektörünün bulunması:** $\underline{\mu} = \left[\frac{\partial Q(\underline{x})}{\partial \underline{x}} = \underline{0}\right] \Rightarrow \frac{\partial Q(\underline{x})}{\partial \underline{x}} = \begin{bmatrix} \frac{8}{23}x_1 + \frac{2}{23}x_3 + \frac{6}{23} \\ \frac{2}{5}x_2 \\ \frac{2}{23}x_1 + \frac{12}{23}x_3 - \frac{10}{23} \end{bmatrix} = \underline{0}$

$$\begin{bmatrix} \frac{8}{23} & 0 & \frac{2}{23} \\ 0 & \frac{2}{5} & 0 \\ \frac{2}{23} & 0 & \frac{12}{23} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -\frac{6}{23} \\ 0 \\ \frac{10}{23} \end{bmatrix} \Rightarrow 2\ \Sigma^{-1}\underline{x} = \begin{bmatrix} -\frac{6}{23} \\ 0 \\ \frac{10}{23} \end{bmatrix} \Rightarrow \underline{\mu} = \underline{x} = \frac{1}{2}\Sigma\begin{bmatrix} -\frac{6}{23} \\ 0 \\ \frac{10}{23} \end{bmatrix} \Rightarrow$$

$$\underline{\mu} = \begin{bmatrix} 6 & 0 & -1 \\ 0 & 5 & 0 \\ -1 & 0 & 4 \end{bmatrix}\begin{bmatrix} -\frac{3}{23} \\ 0 \\ \frac{5}{23} \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

Verilen parçalanmaya göre dağılımın parametreleri parçalanır.

$$\underline{\mu} = \begin{bmatrix} \underline{\mu_1} \\ \cdots\cdots \\ \underline{\mu_2} \end{bmatrix} \ni \ \underline{\mu_1} = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \ , \ \underline{\mu_2} = [0] \ \text{ve}\ \Sigma = \begin{bmatrix} \Sigma_{11} & . & \Sigma_{12} \\ \ldots & \ldots & \ldots \\ \Sigma_{21} & . & \Sigma_{22} \end{bmatrix} \ni \Sigma_{11} = \begin{bmatrix} 6 & -1 \\ -1 & 4 \end{bmatrix},$$

$\Sigma_{12} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $\Sigma_{21} = [0 \quad 0]$ ve $\Sigma_{22} = [5]$ olur. Buna göre;

$\underline{X_1}$ alt vektörünün marjinal dağılımının parametreleri:

$\underline{X_1} = \begin{bmatrix} X_1 \\ X_3 \end{bmatrix} \sim N_2 \left(\underline{\mu_1} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \Sigma_{11} = \begin{bmatrix} 6 & -1 \\ -1 & 4 \end{bmatrix} \right)$ dir.

$\underline{X_2} = [1]$ iken $\underline{X_1}$ alt vektörünün şartlı dağılımının parametreleri:

$\underline{X_1} = \begin{bmatrix} X_1 \\ X_3 \end{bmatrix} \sim N_2 \left(\underline{\mu_{1.2}} = , \Sigma_{11.2} \right)$ ve böylece;

$\underline{\mu_{1.2}} = \underline{\mu_1} + \Sigma_{12}\Sigma_{22}^{-1} \left(\underline{X_2} - \ \underline{\mu_2} \right) = \begin{bmatrix} -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \end{bmatrix} \left(\frac{1}{5} \right) (1 - 0) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ iken

$\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\, \Sigma_{21} = \begin{bmatrix} 6 & -1 \\ -1 & 4 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \left(\frac{1}{5} \right) [0 \quad 0] = \begin{bmatrix} 6 & -1 \\ -1 & 4 \end{bmatrix}$ bulunur.

b) Test edilecek hipotezler:

$H_0: \underline{\overline{X}} \in G$

$H_1: \underline{\overline{X}} \notin G$

G: Güven bölgesi olup: $P \left(Q(\underline{\overline{X}}) = n \left(\underline{\overline{X}} - \underline{\mu} \right)' \ \Sigma^{-1} \left(\underline{\overline{X}} - \underline{\mu} \right) \leq \chi^2_{p;\alpha} \right) = 1 - \alpha$

$n = 10$, $\underline{\overline{X}} - \underline{\mu} = \begin{bmatrix} 1 \\ \frac{5}{2} \\ 1 \end{bmatrix} - \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ \frac{5}{2} \\ 0 \end{bmatrix}$ olup, karesel formun değeri;

$$Q(\underline{\overline{X}}) = n \left(\underline{\overline{X}} - \underline{\mu} \right)' \ \Sigma^{-1} \left(\underline{\overline{X}} - \underline{\mu} \right) = 10 \begin{bmatrix} 2 & \frac{5}{2} & 0 \end{bmatrix} \begin{bmatrix} \frac{4}{23} & 0 & \frac{1}{23} \\ 0 & \frac{1}{5} & 0 \\ \frac{1}{23} & 0 & \frac{6}{23} \end{bmatrix} \begin{bmatrix} 2 \\ \frac{5}{2} \\ 0 \end{bmatrix} = 10 \begin{bmatrix} \frac{8}{23} & \frac{1}{2} & \frac{2}{23} \end{bmatrix} \begin{bmatrix} 2 \\ \frac{5}{2} \\ 0 \end{bmatrix} =$$

19,46 olarak bulunur. $1 - \alpha = 0{,}95$ iken $\alpha = 0{,}05$ $için$ $\chi^2_{p;\alpha} = \chi^2_{3;0{,}05} = 7{,}815$ dir. Buna göre $Q(\underline{\overline{X}}) = 19{,}46 > 7{,}815$ olduğundan H_0 hipotezi ret edilir. Böylece verilen dağılımdan çekildiği iddia edilen 10 birimlik örneklemin ortalama vektörü güven bölgesinde olmadığından, bu örneklemin ortalama vektörü $\underline{\overline{X}} = \begin{bmatrix} 1 \\ \frac{5}{2} \\ 1 \end{bmatrix}$ olamaz.

SORU:2 $\underline{X} \sim N_p \left(\underline{\mu} \ , \ \Sigma \right)$ dağılımı için Σ matrisi bilinmiyor olsun. $\underline{\mu_0}$: $p * 1$ boyutlu bilinen bir vektör olmak üzere $H_0: \underline{\mu} = \underline{\mu_0}$ hipotezini $H_1: \underline{\mu} \neq \underline{\mu_0}$ hipotezine karşı test etmek için kullanılacak olan test istatistiği, verilen kitleden rastgele çekilen n birimlik bir örneklemden elde edilen $T^2_{\underline{X}} = [\, \underline{\overline{X}} - E(\, \underline{\overline{X}})]' \left[\widehat{Cov(\, \underline{\overline{X}}\,)} \right]^{-1} [\, \underline{\overline{X}} - E(\, \underline{\overline{X}})] = n \left(\, \underline{\overline{X}} - \underline{\mu} \, \right)' S^{-1} \left(\, \underline{\overline{X}} - \underline{\mu} \, \right)$

istatistiğidir. Kabul edelim ki verilen çok değişkenli normal dağılım üzerinde $A : p * p$ boyutlu bilinen, tersi alınabilir bir matris olmak üzere $\underline{Y} = A\underline{X}$ dönüşümü uygulanarak yeni bir $\underline{Y}$ rastgele vektörü tanımlansın. Buna göre;

a) $\underline{Y}$ rastgele vektörünün dağılımını bulunuz ve yeni dağılımın kitle ortalama vektörü için test edilecek hipotezleri yukarıda tanımlanan hipotezlerden yararlanarak oluşturunuz?

b) (a) da oluşturduğunuz H_0 hipotezini test etmede kullanılabilecek olan bir test istatistiği bulunuz?

c) (b) de bulduğunuz test istatistiğini $T^2_{\underline{Y}}$ ile gösterirsek, $T^2_{\underline{Y}} = T^2_{\underline{X}}$ olduğunu ispatlayınız?

CEVAP:2

a) $\underline{X} \sim N_p\left(\underline{\mu}\,,\ \Sigma\right)$ ve $\underline{Y} = A\underline{X}$: $p * p$ boyutlu olduğundan $\underline{Y} \sim N_p\left(A\underline{\mu}\,, A\,\Sigma A'\right)$ dağılımlıdır ve Σ matrisi bilinmediğinden $\mathrm{Cov}(\underline{Y}) = A\,\Sigma A'$ matrisi de bilinmiyordur.

Yeni dağılımın kitle ortalama vektörü $A\underline{\mu}$ olup test edilecek olan hipotezler: $\underline{\mu}_0 \in IR^p$ olmak üzere

$H_0: A\,\underline{\mu} = A\underline{\mu}_0$

$H_1: A\,\underline{\mu} \neq A\underline{\mu}_0$

şeklinde olacaktır.

b) H_0 hipotezini test etmede kullanılabilecek olan bir test istatistiği $\mathrm{Cov}(\underline{Y}) = A\,\Sigma A'$ matrisi bilinmediğinden tek örnek Hotelling-T^2 istatistiği olup,

$$T^2_{\underline{Y}} = \left[\underline{\overline{Y}} - E(\underline{\overline{Y}})\right]' \left[\widehat{Cov(\underline{\overline{Y}})}\right]^{-1} \left[\underline{\overline{Y}} - E(\underline{\overline{Y}})\right]$$

şeklinde tanımlanabilir. Burada $i = 1,2,\dots,n$ için $\underline{Y_i}$ ler dönüşüm kitlesinden çekilen örnek birimleri olmak üzere $\underline{\overline{Y}} = \frac{1}{n}\sum_{i=1}^{n}\underline{Y_i}$ olup, $\underline{Y_i} \sim N_p\left(A\underline{\mu}\,, A\,\Sigma A'\right)$ ve bağımsız olduklarından $E(\underline{\overline{Y}}) = E\left(\frac{1}{n}\sum_{i=1}^{n}\underline{Y_i}\right) = \frac{1}{n}\sum_{i=1}^{n}E(\underline{Y_i}) = \frac{1}{n}\sum_{i=1}^{n}A\underline{\mu} = A\left(\frac{1}{n}\sum_{i=1}^{n}\underline{\mu}\right) = A\frac{1}{n}n\,\underline{\mu} = A\underline{\mu}$ ve $Cov(\underline{\overline{Y}}) = Cov\left(\frac{1}{n}\sum_{i=1}^{n}\underline{Y_i}\right) = \frac{1}{n^2}\sum_{i=1}^{n}Cov(\underline{Y_i}) = \frac{1}{n^2}\sum_{i=1}^{n}A\,\Sigma A' = A\left(\frac{1}{n^2}\sum_{i=1}^{n}\Sigma\right)A' = \frac{A\Sigma A'}{n}$ dir. Burada $Cov(\underline{\overline{Y}}) = \frac{A\Sigma A'}{n}$ bilinmediğinden en çok olabilirlik tahmin edicisi $\widehat{Cov(\underline{\overline{Y}})} = \frac{ASA'}{n}$ kullanılır. Bu değerler yerine yazılırsa;

$$T^2_{\underline{Y}} = \left[\underline{\overline{Y}} - A\underline{\mu}\right]' \left[\frac{ASA'}{n}\right]^{-1} \left[\underline{\overline{Y}} - A\underline{\mu}\right] = n\left[\underline{\overline{Y}} - A\underline{\mu}\right]' \left[ASA'\right]^{-1} \left[\underline{\overline{Y}} - A\underline{\mu}\right]$$

elde edilir.

c) Dönüşüm kitlesi $\underline{Y} = A\underline{X}$ olduğundan bu kitleden çekilen n birimlik örneklemde $i = 1,2,\dots,n$ olmak üzere i.nci örnek birimi için $\underline{Y_i} = A\underline{X_i}$ yazılabilir. Buna göre; örnek ortalama istatistiği;

$\overline{\underline{Y}} = \frac{1}{n}\sum_{i=1}^{n}\underline{Y_i} = \frac{1}{n}\sum_{i=1}^{n}A\underline{X_i} = A\left(\frac{1}{n}\sum_{i=1}^{n}\underline{X_i}\right) = A\overline{\underline{X}}$ olacaktır. Bu sonuç (c) de elde edilen test istatistiğinde yerine yazılırsa;

$$T_{\underline{Y}}^2 = n\left[\overline{\underline{Y}} - A\underline{\mu}\right]' [\mathrm{ASA}']^{-1}\left[\overline{\underline{Y}} - A\underline{\mu}\right] = n\left[A\overline{\underline{X}} - A\underline{\mu}\right]' [\mathrm{ASA}']^{-1}\left[A\,\overline{\underline{X}} - A\underline{\mu}\right]$$

$$= \mathrm{n}\left[A\left(\overline{\underline{X}} - \underline{\mu}\right)\right]' (A')^{-1}\mathrm{S}^{-1}A^{-1}\left[A\left(\overline{\underline{X}} - \underline{\mu}\right)\right] = \left(\overline{\underline{X}} - \underline{\mu}\right)' A'(A')^{-1}\mathrm{S}^{-1}A^{-1}A\left(\overline{\underline{X}} - \underline{\mu}\right)$$

$$= \mathrm{n}\left(\overline{\underline{X}} - \underline{\mu}\right)' \mathrm{S}^{-1}\left(\overline{\underline{X}} - \underline{\mu}\right) = T_{\underline{X}}^2$$ elde edilir.

SORU:3 İki farklı bölgede yaşayan bireyler üzerinde yapılan bir çalışmada bireylerin kolestrol (X_1) ve HDL (X_2) değerleri mg/dL cinsinden ölçülmüştür. İlgili değişkenler bakımından bu bölgelere ait dağılımların çok değişkenli normal ve ortak varyans-kovaryans matrisli olduğu kabul edilmiştir. Bu bölgelerden rastgele seçilen 25'er bireye ait ölçümlerden aşağıdaki istatistikler hesaplanmıştır.

BÖLGELER	**ORTALAMA VEKTÖRLERİ**	**VARYANS- KOVARYANS MATRİSLERİ**
I	$\overline{\underline{X}}_1 = \begin{bmatrix} 262 \\ 39{,}16 \end{bmatrix}$	$S_1 = \begin{bmatrix} 2347{,}92 & -83{,}25 \\ -83{,}25 & 73{,}97 \end{bmatrix}$
II	$\overline{\underline{X}}_2 = \begin{bmatrix} 184{,}2 \\ 60{,}4 \end{bmatrix}$	$S_2 = \begin{bmatrix} 632{,}333 & 14{,}875 \\ 14{,}875 & 59{,}583 \end{bmatrix}$

Buna göre;

a) İlgili özellikler bakımından bölgeler arasında fark olup-olmadığına %5 önem seviyesinde karar veriniz.
b) Eğer bölgeler arasında farklılık varsa bu farklılığın kolestrolden mi yoksa HDL'den mi kaynaklandığını belirleyiniz?

Cevap:3 a) Grup I.:Bölge I….$\underline{X_1} = \begin{bmatrix} X_{11} \\ X_{21} \end{bmatrix} \sim N_2\left(\underline{\mu_1}, \Sigma_1\right)$

Grup II.:Bölge II….$\underline{X_2} = \begin{bmatrix} X_{12} \\ X_{22} \end{bmatrix} \sim N_2\left(\underline{\mu_2}, \Sigma_2\right)$ ve $\Sigma_1 = \Sigma_2 = \Sigma$ veriliyor.

Hipotezler:

$H_0: \underline{\mu_1} = \underline{\mu_2}\ ;\ H_1: \underline{\mu_1} \neq \underline{\mu_2}$

Test istatistiği: $T^2 = \frac{n_1 n_2}{n_1+n_2}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)' S^{-1}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right)$

$n_1 = n_2 = 25$; $\underline{\overline{X}}_1 - \underline{\overline{X}}_2 = \begin{bmatrix} 262 \\ 39,16 \end{bmatrix} - \begin{bmatrix} 184,2 \\ 60,4 \end{bmatrix} = \begin{bmatrix} 77,8 \\ -21,24 \end{bmatrix}$;

$S = \frac{(n_1-1)S_1+(n_2-1)S_2}{n_1+n_2-2} = \frac{24S_1+24S_2}{48} = \frac{1}{2}\left(\begin{bmatrix} 2347,92 & -83,25 \\ -83,25 & 73,97 \end{bmatrix} + \begin{bmatrix} 632,333 & 14,875 \\ 14,875 & 59,583 \end{bmatrix}\right) =$
$\begin{bmatrix} 1490,1265 & -34,1875 \\ -34,1875 & 66,7765 \end{bmatrix}$

$T_h^2 = \frac{625}{50}[77,8 \quad -21,24]\begin{bmatrix} 0,00068 & 0,00035 \\ 0,00035 & 0,01515 \end{bmatrix}\begin{bmatrix} 77,8 \\ -21,24 \end{bmatrix} =$
$\frac{625}{50}[0,04547 \quad -0,294556]\begin{bmatrix} 77,8 \\ -21,24 \end{bmatrix} = 122,424$ bulunur. Böylece

$F_h = \frac{n_1+n_2-p-1}{p(n_1+n_2-2)}T_h^2 = \frac{47}{2\times 48}(122,424) = 59,94$ elde edilir.

Karar: $\alpha = 0,05$ için kritik değer; $F_t = F_{p;\, n_1+n_2-p-1;\alpha} = F_{2;47;0,05} = 3,18$ ve böylece

$T_t^2 = \frac{p(n_1+n_2-2)}{n_1+n_2-p-1} \times F_t = \frac{2\times 48}{47}(3,18) = 6,495$ bulunur.

59,94 > 3,18 (yani $F_h > F_t$) ya da 122,424 > 6,495 (yani $T_h^2 > T_t^2$) olduğundan H_0 hipotezi ret edilir. Buna göre Bölgeler arasında ortalamalar bakımından anlamlı bir fark vardır.

b) $H_0: \underline{\mu}_1 = \underline{\mu}_2$ hipotezi ret edildiğinden bölgeler arasındaki farklılığın hangi özellikten kaynaklandığını belirleyebilmek için $\underline{\delta} = \underline{\mu}_1 - \underline{\mu}_2$ kitle ortalamalar farkı olmak üzere $\underline{a}\,\underline{\delta}$ parametresi için güven aralığından yararlanılır.

i) $\boldsymbol{X_1}$: ***Kolestrol için:*** $\underline{a} = [1 \quad 0\,]$ seçelim. Bu takdirde;

$\underline{a}\,\underline{\delta} = \underline{a}\left(\underline{\mu}_1 - \underline{\mu}_2\right) = [1 \quad 0\,]\begin{bmatrix} \mu_{11} - \mu_{12} \\ \mu_{21} - \mu_{22} \end{bmatrix} = \mu_{11} - \mu_{12}$ olacağından test edilecek hipotezler:

$H_0: \mu_{11} - \mu_{12} = 0$

$H_1: \mu_{11} - \mu_{12} \neq 0$ olup, güven aralığı Σ matrisi bilinmediğinden Eşitlik (4.47) gereğince;

$P\left[\underline{a}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) - \sqrt{T_t^2}\sqrt{\frac{n_1+n_2}{n_1 n_2}\underline{a}\,S\,\underline{a}'} \le \underline{a}\begin{bmatrix} \mu_{11} - \mu_{12} \\ \mu_{21} - \mu_{22} \end{bmatrix}\underline{\delta} \le \underline{a}\left(\underline{\overline{X}}_1 - \underline{\overline{X}}_2\right) + \sqrt{T_t^2}\sqrt{\frac{n_1+n_2}{n_1 n_2}\underline{a}\,S\,\underline{a}'}\right] =$
$1 - \alpha$

$P\left[77,8 - \sqrt{6,495}\sqrt{\frac{50}{25\times 25}(1490,1265)} \le \mu_{11} - \mu_{12} \le 77,8 + \sqrt{6,495}\sqrt{\frac{50}{25\times 25}(1490,1265)}\right] =$
0,95

$P\left(49,974 \le \mu_{11} - \mu_{12} \le 105,626\right) = 0,95$ elde edilir. Güven aralığı sıfırı kapsamadığından kolestrol bakımından bölgelerin ortalama değerleri birbirinden farklıdır. Üstelik %95 güvenle I. Bölgeye ait ortalamanın II. Bölgeye ait ortalamadan daha yüksek olduğu söylenebilir.

ii) $\boldsymbol{X_2}$: ***HDL için*** **:** $\underline{a} = [0 \quad 1]$ seçelim. Bu takdirde;

$\underline{a}\,\underline{\delta} = \underline{a}\left(\underline{\mu}_1 - \underline{\mu}_2\right) = [0 \quad 1]\begin{matrix} \mu_{11} - \mu_{12} \\ \mu_{21} - \mu_{22} \end{matrix} = \mu_{21} - \mu_{22}$ olacağından test edilecek hipotezler:

$H_0: \mu_{21} - \mu_{22} = 0$

$H_1: \mu_{21} - \mu_{22} \neq 0$ olup, güven aralığı Σ matrisi bilinmediğinden Eşitlik (4.47) gereğince;

$$P\left[\underline{a}\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right) - \sqrt{T_t^2}\sqrt{\frac{n_1+n_2}{n_1 n_2}\underline{a}\,S\,\underline{a}'} \leq \underline{a}\,\underline{\delta} \leq \underline{a}\left(\overline{\underline{X}}_1 - \overline{\underline{X}}_2\right) + \sqrt{T_t^2}\sqrt{\frac{n_1+n_2}{n_1 n_2}\underline{a}\,S\,\underline{a}'}\right] = 1 - \alpha$$

$$P\left[-21{,}24 - \sqrt{6{,}495}\sqrt{\frac{50}{25.25}(66{,}7765)} \leq \mu_{21} - \mu_{22} \leq -21{,}24 + \sqrt{6{,}495}\sqrt{\frac{49}{37\times 12}(66{,}7765)}\right] = 0{,}95$$

$P\left(-27{,}13 \leq \mu_{21} - \mu_{22} \leq -15{,}35\right) = 0{,}95$ elde edilir. Güven aralığı sıfırı kapsamadığından HDL bakımından I. Bölge ile II. Bölge ortalama değerleri birbirinden farklıdır. Üstelik %95 güvenle II.Bölgeye ait ortalamanın I. Bölgeye ait ortalamadan daha yüksek olduğu söylenebilir.

SORU:4 $\underline{X} \sim N_3(\underline{\mu}, \Sigma)$ dağılımının olasılık yoğunluk fonksiyonu

$f(\underline{x}) = k.exp\left\{-\frac{1}{2}Q(\underline{x})\right\}$ olup, $Q(\underline{x}) = \frac{5}{52}x_1^2 + \frac{7}{52}x_2^2 + \frac{5}{84}x_3^2 - \frac{3}{26}x_1x_2$

olarak veriliyor. $\underline{X} = \begin{bmatrix}\underline{X_1}\\ \cdots \\ \underline{X_2}\end{bmatrix}$ öyle ki $\underline{X_1} = [X_3]$, $\underline{X_2} = \begin{bmatrix}X_1\\ X_2\end{bmatrix}$ parçalanması verilsin. Buna göre;

a) $\underline{X_1}$ alt vektörünün marjinal olasılık yoğunluk fonksiyonunu bulunuz.
b) $\underline{X_1} = [2]$ iken $\underline{X_2}$ alt vektörünün şartlı dağılımına ait parametreleri bulunuz.
c) X_3 değişkeni ile $Y = \frac{1}{2}X_1 + \frac{1}{4}X_2$ şeklinde tanımlanan doğrusal fonksiyonun bağımsız olup olmadığını gösteriniz.

SORU:5 Marmara ve Orta Anadolu bölge Tarım İl Müdürlüklerine bağlı deneme çiftliklerinde üretilen buğday ve arpa ürün miktarlarına (milyon ton) ilişkin ortalama verimlerin bölgelere göre benzerlik gösterip göstermediği araştırılmak istenmektedir. Bu amaçla son 15 yılda üretilen, yıllara göre buğday ve arpa hasılatına ilişkin verilerden aşağıdaki istatistiksel bilgiler elde edilmiştir. Her bir bölge bazında üretilen buğday ve arpa ürün miktarlarına ait dağılımların $N_2\left(\underline{\mu}_{k.}, \Sigma\right), k = 1, 2$ dağılımına uygun olduğu varsayılmıştır. Buna göre:

a) Deneme çiftliklerinden elde edilen ortalama ürün miktarlarının bölgelere göre benzerlik gösterip göstermediğine uygun bir yöntemle %5 önem seviyesinde karar veriniz?

b) Buğday üretimi yönünden Marmara ve Güneydoğu Anadolu arasında fark olup olmadığını gösteriniz?

c) Arpa üretimi yönünden Marmara ve Güneydoğu Anadolu arasında fark olup olmadığını gösteriniz?

BÖLGE	$\underline{\bar{X}}_{k.} = \begin{bmatrix} \bar{X}_{k1} \\ \bar{X}_{k2} \end{bmatrix}$	W_k
MARMARA	$\underline{\bar{X}}_{1.} = \begin{bmatrix} 3,1 \\ 1,0 \end{bmatrix}$	$W_1 = \begin{bmatrix} 56 & -14 \\ -14 & 98 \end{bmatrix}$
ORTA ANADOLU	$\underline{\bar{X}}_{2.} = \begin{bmatrix} 7,2 \\ 2,4 \end{bmatrix}$	$W_2 = \begin{bmatrix} 168 & 42 \\ 42 & 126 \end{bmatrix}$