T.C.

ONDOKUZ MAYIS ÜNİVERSİTESİ

FEN EDEBİYAT FAKÜLTESİ
İSTATİSTİK BÖLÜMÜ

ÇOK DEĞİŞKENLİ İSTATİSTİK I DERSİ

PROF. DR. YÜKSEL ÖNER

11. Hafta

özenilen üniversite

ii) Σ matrisinin bilinmediğini kabul edelim.

Varyans kovaryans matrisinin bilindiği durumda görüldüğü üzere hem olabilirlik oran testi, hem de bileşim-kesişim testi aynı test istatistiğini vermektedir. Ancak; bileşim-kesişim testi H_0 hipotezinin ret edilmesi halinde çoklu karşılaştırma veya bireysel ortalamaların değerlendirilmesi imkânı sağladığından burada sadece bileşim-kesişim testi yaklaşımı verilecektir.

$\underline{X} \sim N_p\left(\underline{\mu}, \Sigma\right)$ dağılımı verilsin ve Σ matrisi bilinmiyor olsun. $\underline{\mu}_0 \in IR^p$ bilinen bir kolon vektörü iken test edilecek hipotezler:

$$H_0: \underline{\mu} = \underline{\mu}_0 \qquad\qquad H_0: \underline{a}\,\underline{\mu} = \underline{a}\,\underline{\mu}_0$$

$$H_1: \underline{\mu} \neq \underline{\mu}_0 \qquad \text{veya} \qquad H_1: \underline{a}\,\underline{\mu} \neq \underline{a}\,\underline{\mu}_0 \tag{4.31}$$

şeklinde oluşturulur. Test istatistiğinin elde edilmesi amacı ile $\underline{X} \sim N_p\left(\underline{\mu}, \Sigma\right)$ dağılımından rastgele olarak çekilen n birimlik örnek; $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$ olsun. $\underline{a} : 1 \times p$ boyutlu bilinen bir vektör olmak üzere $Y = \underline{a}\,\underline{X}$ doğrusal dönüşümü ile çok değişkenli kitleyi tek değişkenli kitleye dönüştürelim. Bu durumda dönüşüm kitlesinin dağılımı için $Y = \underline{a}\,\underline{X} \sim N\left(\underline{a}\,\underline{\mu}, \sigma_Y^2 = \underline{a}\,\Sigma\,\underline{a}'\right)$ öyleki $\sigma_Y^2 = \underline{a}\,\Sigma\,\underline{a}'$ bilinmiyordur. Bu dönüşüme göre dönüşüm kitlesine ait n birimlik örnek; $Y_1 = \underline{a}\,\underline{X}_1, Y_2 = \underline{a}\,\underline{X}_2, \ \dots \ , Y_n = \underline{a}\,\underline{X}_n$ olup, i=1, 2, … ,n için $Y_i = \underline{a}\,\underline{X}_i \sim N\left(\underline{a}\,\underline{\mu},\ \underline{a}\,\Sigma\,\underline{a}'\right)$ ve bağımsızdır. Örnek ortalama istatistiği ile bu istatistiğin örnekleme dağılımı;

$$\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n} \underline{a}\,\underline{X}_i = \underline{a}\left(\frac{1}{n}\sum_{i=1}^{n} \underline{X}_i\right) = \underline{a}\,\overline{\underline{X}} \sim N\left(\underline{a}\,\underline{\mu}, \frac{\underline{a}\,\Sigma\,\underline{a}'}{n}\right)$$

şeklinde elde edilir. Burada $\sigma_Y^2 = \underline{a}\,\Sigma\,\underline{a}'$ bilinmediğinden örnekten tahmin edilir. Kitle varyansının yansızlık özelliğine sahip en çok olabilirlik tahmin edicisi;

$$\hat{\sigma}_Y^2 = S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\underline{a}\,\underline{X}_i - \underline{a}\,\overline{\underline{X}}\right)^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left[\underline{a}\left(\underline{X}_i - \overline{\underline{X}}\right)\right]^2$$

$$= \frac{1}{n-1}\sum_{i=1}^{n} \underline{a}\left(\underline{X}_i - \overline{\underline{X}}\right)\left(\underline{X}_i - \overline{\underline{X}}\right)'\underline{a}' = \underline{a}\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(\underline{X}_i - \overline{\underline{X}}\right)\left(\underline{X}_i - \overline{\underline{X}}\right)'\right]\underline{a}' = \underline{a}\,S\underline{a}' \quad \text{olur.}$$

Buna göre örnek ortalama istatistiği $\overline{Y} = \underline{a}\,\overline{\underline{X}}$ istatistiğinin varyansının tahmini; $S_{\overline{Y}}^2 = \frac{\underline{a}\,S\underline{a}'}{n}$ dir. Test istatistiği ise;

$$t = \frac{\overline{Y} - E\left(\overline{Y}\right)}{S_{\overline{Y}}} = \frac{\underline{a}\,\overline{\underline{X}} - \underline{a}\,\underline{\mu}}{\left(\frac{\underline{a}\,S\underline{a}'}{n}\right)^{1/2}} = \frac{\sqrt{n}\,\underline{a}\left(\overline{\underline{X}} - \underline{\mu}\right)}{\left(\underline{a}\,S\underline{a}'\right)^{1/2}} \sim t_{n-1} \tag{4.32}$$

veya

$$T^2 = n\left(\overline{\underline{X}} - \underline{\mu}\right)' S^{-1}\left(\overline{\underline{X}} - \underline{\mu}\right) \tag{4.33}$$

şeklinde tanımlanan tek örnek Hotelling T^2 istatistiğidir. Bu istatistiğin örnekleme dağılımı yaklaşık olarak;

$$F = \frac{n-p}{p(n-1)}\ T^2 \sim F_{p\,;n-p} \tag{4.34}$$

şeklinde elde edilir.

H_0 doğru iken T^2 istatistiğinin alabileceği değer $T_h^2 = n\left(\underline{\bar{X}} - \underline{\mu}_0\right)' S^{-1}\left(\underline{\bar{X}} - \underline{\mu}_0\right)$ ve böylece F istatistiğinin alabileceği değer $F_h = \frac{n-p}{p(n-1)}\ T_h^2$ olacaktır.

Karar:

α önem seviyesinde H_1hipotezine göre kritik değer; $F_t = F_{p\,;n-p\,;\,\alpha}$ veya $T_t^2 = \frac{p(n-1)}{n-p} F_t$ olmak üzere; eğer $F_h > F_t$ (veya $T_h^2 > T_t^2$) ise H_0 hipotezi ret edilir, eğer $F_h \leq F_t$ (veya $T_h^2 \leq T_t^2$) ise H_0 hipotezi ret edilemez.

Yorum: H_0 hipotezi ret edilememiş ise bu durumda örneklemin çekildiği kitlenin ortalama vektörünün $\underline{\mu}_0$ vektörü olduğu söylenir.

H_0 hipotezi ret edilmiş ise o zaman örneklemin çekildiği kitlenin ortalama vektörü $\underline{\mu}_0$'dan farklı bir vektördür. Bu durumda H_0 hipotezinin ret edilme kararı üzerinde etkili olan değişken ya da değişkenleri belirlemek mümkündür. Bunun için $\underline{a}\,\underline{\mu}$ parametresine ait $(1-\alpha)$ güven katsayılı güven aralığı kullanılarak;

$$\begin{aligned} &H_0: \underline{a}\,\underline{\mu} = \underline{a}\,\underline{\mu}_0 \\ &H_1: \underline{a}\,\underline{\mu} \neq \underline{a}\,\underline{\mu}_0 \end{aligned} \tag{4.35}$$

şeklindeki hipotezler test edilir. Kullanılacak olan güven aralığı;

$$P\left[\underline{a}\,\underline{\bar{X}} - \sqrt{\frac{\underline{a}\,S\underline{a}'}{n} \times T_t^2} \leq \underline{a}\,\underline{\mu} \leq \underline{a}\,\underline{\bar{X}} + \sqrt{\frac{\underline{a}\,S\underline{a}'}{n} \times T_t^2}\right] = 1 - \alpha \tag{4.36}$$

eşitliği ile verilir.

$j = 1, 2, \ldots, p$ için $\underline{a} = [0 \;\; 0 \;\; \ldots 1 \ldots 0]$ (j-nci bileşeni 1 diğer bileşenlerin hepsini 0) seçilirse bu durumda;

$\underline{a}\,\underline{\mu} = \mu_j$, $\underline{a}\,\underline{\mu}_0 = \mu_{j0}$, $\underline{a}\,\underline{\bar{X}} = \bar{X}_j$, $\underline{a}\,S\underline{a}' = s_{jj}$ olacaktır. Buna göre (4.35)'de verilen hipotezler;

$$\begin{aligned} &H_0: \mu_j = \mu_{j0} \\ &H_1: \mu_j \neq \mu_{j0} \quad, j = 1, 2, \ldots, p \end{aligned}$$

şekline dönüşür. Bu hipotezler Eşitlik (4.36)'de verilen güven aralığı ile test edilir. Eğer güven aralığı μ_{j0}, $(j = 1, 2, \ldots, p)$ değerini kapsıyor ise $H_0: \mu_j = \mu_{j0}$ hipotezi kabul edilir ve X_j, $(j = 1, 2, \ldots, p)$ değişkeni $H_0: \underline{\mu} = \underline{\mu}_0$ hipotezinin ret edilmesinde etkili olmamıştır.

Eğer güven aralığı $\mu_{j0}, (j = 1, 2, \ldots, p)$ değerini kapsamıyor ise $H_0: \mu_j = \mu_{j0}$ hipotezi ret edilir ve bu durumda $X_j, (j = 1, 2, \ldots, p)$ değişkeni $H_0: \underline{\mu} = \underline{\mu}_0$ hipotezinin ret edilmesinde etkili olmuştur denir.

Diğer taraftan **uygun olması halinde** p tane değişken için ikişer ikişer değişken ortalamaları arasında fark olup olmadığı yine Eşitlik (4.36)'da verilen güven aralığı ile incelenebilir. Bu amaçla: $j \neq k = 1, 2, \ldots, p$ için $\underline{a} = [0 \quad 0 \quad \ldots 1 \ldots -1 \quad \ldots . 0]$ (j-nci bileşeni 1, k-ncı bileşeni -1 ve diğer bileşenlerin hepsini 0) seçilirse bu durumda;

$\underline{a}\,\underline{\mu} = \mu_j - \mu_k$, $\underline{a}\,\overline{\underline{X}} = \overline{X}_j - \overline{X}_k$, $\underline{a}\,S\underline{a}' = s_{jj} + s_{kk} - 2s_{jk}$ olacaktır. Buna göre (4.35)'de verilen hipotezler;

$H_0: \mu_j = \mu_k$
$H_1: \mu_j \neq \mu_k \quad , j \neq k = 1, 2, \ldots, p$

şekline dönüşür. Bu hipotezlerin sayısı $\binom{p}{2}$ kadar olup, her bir hipotez Eşitlik (4.36)'de verilen güven aralığı ile test edilir. Eğer güven aralığı "0" değerini kapsıyor ise $H_0: \mu_j = \mu_k$ hipotezi kabul edilir, aksi takdirde ret edilir.

Örnek 4.4 $\underline{X} \sim N_2\left(\underline{\mu}, \Sigma\right)$ dağılımından çekilen 25 birimlik bir örneklemden örnek ortalama vektörü $\overline{\underline{X}} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$ ve kareler ve çarpımlar toplamı matrisi $W = \begin{bmatrix} 144 & 72 \\ 72 & 96 \end{bmatrix}$ istatistikleri hesaplanmıştır. Buna göre;

a) Kitle ortalama vektörünün $\underline{\mu}_0 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ olup olmadığına %5 önem seviyesinde karar veriniz?
b) Farklılığa sebep olan değişken varsa belirleyiniz?
c) Kitlede iki değişken ortalaması farkı için $(\mu_1 - \mu_2)$ %95 güven katsayılı güven aralığını oluşturunuz?

Çözüm:

a) $\underline{X} \sim N_2\left(\underline{\mu}, \Sigma\right)$ ve Σ matrisi bilinmiyor. Hipotezler:
$H_0: \underline{\mu} = \underline{\mu}_0$
$H_1: \underline{\mu} \neq \underline{\mu}_0 ; \left(\underline{\mu}_0 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}\right)$)

Test istatistiği Σ matrisi bilinmediğinden dolayı $T^2 = n\left(\overline{\underline{X}} - \underline{\mu}\right)' S^{-1}\left(\overline{\underline{X}} - \underline{\mu}\right)$ olmak üzere $F = \frac{n-p}{p(n-1)} T^2 \sim F_{p;n-p}$ dir. H_0 doğru iken test istatistiğinin alabileceği değer:

$W = (n-1)S$ ve $n = 25$ olduğundan $S = \frac{1}{n-1} W = \frac{1}{24}\begin{bmatrix} 144 & 72 \\ 72 & 96 \end{bmatrix} = \begin{bmatrix} 6 & 3 \\ 3 & 4 \end{bmatrix}$ olup;

$T^2 = 25\left(\begin{bmatrix} 1 \\ 4 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix}\right)' \begin{bmatrix} 6 & 3 \\ 3 & 4 \end{bmatrix}^{-1} \left(\begin{bmatrix} 1 \\ 4 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \end{bmatrix}\right) = 25[-1 \quad 2]\frac{1}{15}\begin{bmatrix} 4 & -3 \\ -3 & 6 \end{bmatrix}\begin{bmatrix} -1 \\ 2 \end{bmatrix} =$
$\frac{5}{3}[-10 \quad 15]\begin{bmatrix} -1 \\ 2 \end{bmatrix} = \frac{200}{3} = 66{,}67$ bulunur. Böylece; $p = 2$ iken
$F = \frac{23}{2 \times 24}(66{,}67) = 31{,}946$ elde edilir.

Karar:$\alpha = 0{,}05$ için kritik değer $F_t = F_{p;n-p;\alpha} = F_{2;23;0{,}05} = 3{,}42$ veya $T_t^2 = \frac{p(n-1)}{n-p} F_t = \frac{2\times 24}{23}(3{,}42) = 7{,}137$ olup, $\begin{cases} 31{,}946 > 3{,}42 \\ F > F_t \end{cases}$ (veya $\begin{cases} 66{,}67 > 7{,}137 \\ T^2 > T_t^2 \end{cases}$) olduğundan H_0 hipotezi ret edilir. Buna göre örneklemin çekildiği kitlenin ortalama vektörü $\underline{\mu}_0 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ vektöründen farklıdır.

b) $\boldsymbol{X_1}$ <u>değişkeni için inceleyelim</u>: $\underline{a} = [1 \quad 0]$ seçersek, $\underline{a}\,\underline{\mu} = \mu_1$ ve $\underline{a}\underline{\mu}_0 = 2$ olacağından hipotezler;

$H_0 : \mu_1 = 2$

$H_1 : \mu_1 \neq 2$ olur. Güven aralığı Eşitlik (4.36) gereğince

$$P\left[\underline{a}\,\overline{\underline{X}} - \sqrt{\frac{\underline{a}\,S\underline{a}'}{n} \times T_t^2} \leq \underline{a}\,\underline{\mu} \leq \underline{a}\,\overline{\underline{X}} + \sqrt{\frac{\underline{a}\,S\underline{a}'}{n} \times T_t^2}\right] = 1 - \alpha \text{ dir.}$$

$\underline{a}\,\overline{\underline{X}} = 1$, $\underline{a}\,S\underline{a}' = s_{11} = 6$ ve $T_t^2 = 7{,}137$ olduğundan;

$$P\left(1 - \sqrt{\frac{6}{25}(7{,}137)} \leq \mu_1 \leq 1 + \sqrt{\frac{6}{25}(7{,}137)}\right) = 0{,}95$$

$P(-0{,}31 \leq \mu_1 \leq 2{,}31) = 0{,}95$ bulunur. Güven aralığı $\mu_{10} = 2$ değerini kapsadığından H_0 hipotezi ret edilemez. Yani X_1 değişkeni $H_0 : \underline{\mu} = \underline{\mu}_0$ hipotezinin ret edilmesinde etkili değildir.

$\boldsymbol{X_2}$ <u>değişkeni için inceleyelim</u>: $\underline{a} = [0 \quad 1]$ seçersek, $\underline{a}\,\underline{\mu} = \mu_2$ ve $\underline{a}\underline{\mu}_0 = 2$ olacağından hipotezler;

$H_0 : \mu_2 = 2$

$H_1 : \mu_2 \neq 2$ olur. Güven aralığı Eşitlik (4.36) gereğince

$\underline{a}\,\overline{\underline{X}} = 4$, $\underline{a}\,S\underline{a}' = s_{22} = 4$ ve $T_t^2 = 7{,}137$ olduğundan;

$$P\left(4 - \sqrt{\frac{4}{25}(7{,}137)} \leq \mu_2 \leq 4 + \sqrt{\frac{4}{25}(7{,}137)}\right) = 0{,}95$$

$P(2{,}93 \leq \mu_2 \leq 5{,}07) = 0{,}95$ bulunur. Güven aralığı $\mu_{20} = 2$ değerini kapsamadığından H_0 hipotezi ret edilir. Yani X_2 değişkeni $H_0 : \underline{\mu} = \underline{\mu}_0$ hipotezinin ret edilmesinde etkilidir

c) Kitlede iki değişken ortalaması farkı için $(\mu_1 - \mu_2)$ %95 güven katsayılı güven aralığı;

$\underline{a} = [1 \quad -1]$ seçersek, $\underline{a}\,\underline{\mu} = \mu_1 - \mu_2$ olacağından hipotezler;

$H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 \neq 0$ olur. Güven aralığı Eşitlik (4.36) gereğince

$\underline{a}\,\overline{\underline{X}} = \overline{X}_1 - \overline{X}_2 = 1 - 4 = -3$, $\underline{a}\,S\underline{a}' = s_{11} + s_{22} - 2s_{12} = 6 + 4 - 2 \times 3 = 4$ ve $T_t^2 = 7{,}137$ olduğundan;

$$P\left(-3 - \sqrt{\frac{4}{25}(7{,}137)} \leq \mu_1 - \mu_2 \leq -3 + \sqrt{\frac{4}{25}(7{,}137)}\right) = 0{,}95$$

$P(-4{,}07 \leq \mu_1 - \mu_2 \leq -1{,}93) = 0{,}95$ bulunur. Güven aralığı sıfırı kapsamadığı için H_0 hipotezi ret edilir. Yani değişkenlere ait kitle ortalamaları arasında fark vardır. Üstelik X_2 değişkenine ait kitle ortalaması (μ_2), X_1 değişkenine ait kitle ortalamasından (μ_1) daha büyüktür.

Örnek:4.5 $\underline{X} \sim N_3\left(\underline{\mu}, \Sigma\right)$ dağılımından rastgele çekilen 25 birimlik bir örneklemden $\underline{\overline{X}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ ve $S = \begin{bmatrix} 7 & 3 & 2 \\ 3 & 4 & 1 \\ 2 & 1 & 2 \end{bmatrix}$ istatistikleri hesaplanmıştır. Buna göre;

a) Kitle ortalama vektörünün $\underline{\mu}_0 = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$ olup olmadığını %5 hata ile test ediniz?

b) Farklılığa sebep olan değişken varsa belirleyiniz?

Çözüm: a) Hipotezler:

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

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

Test istatistiği : Test istatistiği Σ matrisi bilinmediğinden dolayı $T^2 = n\left(\underline{\overline{X}} - \underline{\mu}\right)' S^{-1}\left(\underline{\overline{X}} - \underline{\mu}\right)$ olmak üzere;

$F = \frac{n-p}{p(n-1)} T^2 \sim F_{p\,;n-p}$ dir. H_0 doğru iken test istatistiğinin alabileceği değer:

$n = 25, p = 3, \underline{\overline{X}} - \underline{\mu}_0 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ ve $S^{-1} = \frac{1}{|S|} Adj(S)$;

$|S| = 7 \times 7 - 3 \times 4 + 2 \times (-5) = 27$ olup, böylece $S^{-1} = \frac{1}{27}\begin{bmatrix} 7 & -4 & -5 \\ -4 & 10 & -1 \\ -5 & -1 & 19 \end{bmatrix}$ olur. Bu sonuçlar T^2'de yerlerine yazılırsa;

$T_h^2 = 25[-1 \quad 0 \quad 1]\frac{1}{27}\begin{bmatrix} 7 & -4 & -5 \\ -4 & 10 & -1 \\ -5 & -1 & 19 \end{bmatrix}\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \frac{25}{27}[-12 \quad 3 \quad 24]\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} = \frac{25 \times 36}{27} = 33{,}33$ ve

böylece $F_h = \frac{25-3}{3 \times 24}(33{,}33) = 10{,}184$ bulunur.

Karar: $\alpha = 0{,}05$ için H_1 hipotezine göre kritik değer $F_t = F_{p:n-p:\alpha} = F_{3;22;0{,}05} = 3{,}05$ veya $T_t^2 = \frac{p(n-1)}{n-p} F_t = \frac{3 \times 24}{22}(3{,}05) = 9{,}98$ bulunur. Buna göre; $10{,}184 > 3{,}05$ $(F_h > F_t)$ veya $33{,}33 > 9{,}98$ $(T_h^2 > T_t^2)$ olduğundan H_0 hipotezi ret edilir. Bu karara göre %95 güvenle örneklemin çekildiği kitlenin ortalama vektörü $\underline{\mu}_0 = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$ vektörü olamaz.

b) H_0 hipotezi ret edildiği için bu kararın verilmesinde etkili olan değişkenleri belirlemek mümkündür. Bunun için $\underline{a}\,\underline{\mu}$ parametresine ait güven aralığından yararlanılır.

i) X_1 değişkeni için inceleyelim: $\underline{a} = [1 \;\; 0 \;\; 0]$ seçelim. Bu takdirde $\underline{a}\,\underline{\mu} = [1 \;\; 0 \;\; 0]\begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} = \mu_1$

ve $\underline{a}\,\underline{\mu}_0 = [1 \;\; 0 \;\; 0]\begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} = 2$ olup test edilecek hipotezler:

$H_0: \mu_1 = 2$

$H_1: \mu_1 \neq 2$

dır. Güven aralığı ise Σ matrisi bilinmediğinden Eşitlik (4.36)'ya göre;

$$P\left[\underline{a}\,\underline{\overline{X}} - \sqrt{T_t^2}\sqrt{\frac{\underline{a}\,S\,\underline{a}'}{n}} \leq \underline{a}\,\underline{\mu} \leq \underline{a}\,\underline{\overline{X}} + \sqrt{T_t^2}\sqrt{\frac{\underline{a}\,S\,\underline{a}'}{n}}\right] = 1 - \alpha$$

Şeklindedir. Burada $\underline{a}\,\underline{\overline{X}} = [1\ \ 0\ \ 0]\begin{bmatrix}1\\2\\3\end{bmatrix} = 1$, $\underline{a}\,S\,\underline{a}' = [1\ \ 0\ \ 0]\begin{bmatrix}7&3&2\\3&4&1\\2&1&2\end{bmatrix}\begin{bmatrix}1\\0\\0\end{bmatrix} = 7$ ve $T_t^2 = 9{,}98$ olup, bu değerler güven aralığında yerlerine yazılırsa;

$$P\left(1 - \sqrt{9{,}98}\sqrt{\frac{7}{25}} \leq \mu_1 \leq 1 + \sqrt{9{,}98}\sqrt{\frac{7}{25}}\right) = 0{,}95 \quad \Rightarrow \quad P(-0{,}672 \leq \mu_1 \leq 2{,}672) = 0{,}95$$

bulunur. Güven aralığı $\mu_{10} = 2$ değerini kapsadığından $H_0: \mu_1 = 2$ hipotezi ret edilemez. Buna göre X_1 değişkeni $H_0: \underline{\mu} = \underline{\mu}_0$ hipotezinin ret edilmesinde etkili değildir.

ii) X_2 değişkeni için inceleyelim: $\underline{a} = [0\ \ 1\ 0]$ seçelim. Bu takdirde $\underline{a}\,\underline{\mu} = [0\ \ 1\ \ 0]\begin{bmatrix}\mu_1\\\mu_2\\\mu_3\end{bmatrix} = \mu_2$

ve $\underline{a}\,\underline{\mu}_0 = [0\ \ 1\ \ 0]\begin{bmatrix}2\\2\\2\end{bmatrix} = 2$ olup test edilecek hipotezler:

$H_0: \mu_2 = 2$
$H_1: \mu_2 \neq 2$

dır. Güven aralığı ise Σ matrisi bilinmediğinden Eşitlik (4.36)'ya göre bulunur.

Burada $\underline{a}\,\underline{\overline{X}} = [0\ \ 1\ \ 0]\begin{bmatrix}1\\2\\3\end{bmatrix} = 2$, $\underline{a}\,S\,\underline{a}' = [0\ \ 1\ \ 0]\begin{bmatrix}7&3&2\\3&4&1\\2&1&2\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix} = 4$ ve $T_t^2 = 9{,}98$ olup, bu değerler güven aralığında yerlerine yazılırsa;

$$P\left(2 - \sqrt{9{,}98}\sqrt{\frac{4}{25}} \leq \mu_2 \leq 2 + \sqrt{9{,}98}\sqrt{4/25}\right) = 0{,}95 \quad \Rightarrow \quad P(0{,}736 \leq \mu_2 \leq 3{,}264) = 0{,}95$$

bulunur. Güven aralığı $\mu_{20} = 2$ değerini kapsadığından $H_0: \mu_2 = 2$ hipotezi ret edilemez. Buna göre X_2 değişkeni $H_0: \underline{\mu} = \underline{\mu}_0$ hipotezinin ret edilmesinde etkili olmamıştır.

iii) X_3 değişkeni için inceleyelim: $\underline{a} = [0\ \ 0\ \ 1]$ seçelim. Bu takdirde $\underline{a}\,\underline{\mu} = [0\ \ 0\ \ 1]\begin{bmatrix}\mu_1\\\mu_2\\\mu_3\end{bmatrix} = \mu_3$

ve $\underline{a}\,\underline{\mu}_0 = [0\ \ 0\ \ 1]\begin{bmatrix}2\\2\\2\end{bmatrix} = 2$ olup test edilecek hipotezler:

$H_0: \mu_3 = 2$
$H_1: \mu_3 \neq 2$

dır. Güven aralığı ise Σ matrisi bilinmediğinden Eşitlik (4.36)'ya göre bulunur.

Burada $\underline{a}\,\overline{\underline{X}} = [0\ \ 0\ \ 1]\begin{bmatrix}1\\2\\3\end{bmatrix} = 3$, $\underline{a}\,\Sigma\,\underline{a}' = [0\ \ 0\ \ 1]\begin{bmatrix}7 & 3 & 2\\3 & 4 & 1\\2 & 1 & 2\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix} = 2$ ve $T_t^2 = 9{,}98$ olup, bu değerler güven aralığında yerlerine yazılırsa;

$$P\left(3 - \sqrt{9{,}98}\sqrt{\frac{2}{25}} \le \mu_3 \le 3 + \sqrt{9{,}98}\sqrt{\frac{2}{25}}\right) = 0{,}95 \quad \Rightarrow \quad P(2{,}106 \le \mu_3 \le 3{,}894) = 0{,}95$$

bulunur. Güven aralığı $\mu_{30} = 2$ değerini kapsamadığından $H_0{:}\,\mu_3 = 2$ hipotezi ret edilir. Buna göre X_3 değişkeni $H_0{:}\,\underline{\mu} = \underline{\mu}_0$ hipotezinin ret edilmesinde etkili olmuştur.

Örnek:4.6 Tek örneklem Hotelling-T^2 istatistiğinin lineer dönüşümlerden etkilenmediğini ispatlayınız.

Çözüm: $\underline{X} \sim N_p\left(\underline{\mu}\,, \Sigma\right)$ dağılımı verilsin. Bu dağılımda Σ matrisi bilinmiyorken $H_0{:}\,\underline{\mu} = \underline{\mu}_0$ hipotezini test etmek için test istatistiği olarak kullanılan Tek örneklem Hotelling-T^2 istatistiği; verilen dağılımdan çekilen n birimlik örnek $\underline{X}_1, \underline{X}_2, \dots, \underline{X}_n$, bu örneğin ortalama vektörü $\overline{\underline{X}} = \frac{1}{n}\sum_{i=1}^{n}\underline{X}_i$ istatistiği ve bu istatistiğin örnekleme dağılımı $\overline{\underline{X}} \sim N_p\left(E(\overline{\underline{X}}) = \underline{\mu}\,, Cov(\overline{\underline{X}}) = \frac{1}{n}\Sigma\right)$ olmak üzere;

$$T_{\underline{X}}^2 = \left[\overline{\underline{X}} - E(\overline{\underline{X}})\right]'\left(\widehat{Cov(\overline{\underline{X}})}\right)^{-1}\left[\overline{\underline{X}} - E(\overline{\underline{X}})\right] = n\left(\overline{\underline{X}} - \underline{\mu}\right)' S^{-1}\left(\overline{\underline{X}} - \underline{\mu}\right) \dots\text{(i)}$$

ile gösterilir. Verilen dağılım üzerinde uygulanan bir lineer dönüşüm; $A{:}\,p \times p$ boyutlu ve tam satır ranklı bir matris ve $\underline{d}{:}\,p \times 1$ boyutlu bir vektör olmak üzere $\underline{Y} = A\underline{X} + \underline{d}$ olsun. Bu lineer dönüşüm çok değişkenli normal dağılıma sahip bir rastgele vektörün lineer dönüşümü olduğundan, bu lineer dönüşümün dağılımı da çok değişkenli normal dağılıma sahiptir ve $\underline{Y} = A\underline{X} + \underline{d} \sim N_p\left(E(\underline{Y})\,, Cov(\underline{Y})\right)$ şeklindedir. Burada;

$$E(\underline{Y}) = E(A\underline{X} + \underline{d}) = AE(\underline{X}) + \underline{d} = A\underline{\mu} + \underline{d} \text{ ve } Cov(\underline{Y}) = Cov(A\underline{X} + \underline{d}) = ACov(\underline{X})A' = A\,\Sigma A'$$

olup, böylece $\underline{Y} = A\underline{X} + \underline{d} \sim N_p\left(A\underline{\mu} + \underline{d}\,, A\,\Sigma A'\right)$ olur. Üstelik Σ matrisi bilinmediğinden dönüşüm kitlesine ait kovaryans matrisi $A\,\Sigma A'$ da bilinmiyordur. Bu durumda dönüşüm kitlesine ait ortalama vektörü ile ilgili $H_0{:}\,A\underline{\mu} + \underline{d} = A\underline{\mu}_0 + \underline{d}$ hipotezini test etmek için test istatistiği olarak yine Tek örneklem Hotelling-T^2istatistiği kullanılacaktır. Bu Hotelling-T^2istatistiği; dönüşüm kitlesinden çekilen n birimlik örnek $\underline{Y}_1, \underline{Y}_2, \dots, \underline{Y}_n$, bu örneğin ortalama vektörü $\overline{\underline{Y}} = \frac{1}{n}\sum_{i=1}^{n}\underline{Y}_i$ istatistiği ve bu istatistiğin örnekleme dağılımı $\overline{\underline{Y}} \sim N_p\left(E(\overline{\underline{Y}})\,, Cov(\overline{\underline{Y}})\right)$ olmak üzere;

$$T_{\underline{Y}}^2 = \left[\overline{\underline{Y}} - E(\overline{\underline{Y}})\right]'\left(\widehat{Cov(\overline{\underline{Y}})}\right)^{-1}\left[\overline{\underline{Y}} - E(\overline{\underline{Y}})\right] \dots\text{(ii)}$$

şeklinde olacaktır. Göstereceğiz ki $T_{\underline{Y}}^2 = T_{\underline{X}}^2$ dir.

Dönüşüm kitlesinden çekilen n birimlik $\underline{Y}_1, \underline{Y}_2, \dots, \underline{Y}_n$ örneğinde, her bir örnek birimi verilen dağılımdan çekilen örnekteki örnek birimlerinin, yani $i = 1, 2, \dots, n$ için $\underline{X}_i$'lerin bir doğrusal dönüşümüdür. Yani $i = 1, 2, \ldots, n$ için $\underline{Y}_i = A\underline{X}_i + \underline{d}$ şeklinde yazılabilir. Bu durumda $\underline{\overline{Y}} \sim N_p\left(E(\underline{\overline{Y}}), Cov(\underline{\overline{Y}})\right)$ dağılımının parametreleri;

$$E(\underline{\overline{Y}}) = E\left(\frac{1}{n}\sum_{i=1}^{n}\underline{Y}_i\right) = \frac{1}{n}\sum_{i=1}^{n}E(\underline{Y}_i) = \frac{1}{n}\sum_{i=1}^{n}\left(A\underline{\mu} + \underline{d}\right) = \frac{1}{n}n\left(A\underline{\mu} + \underline{d}\right) = \left(A\underline{\mu} + \underline{d}\right)$$

$$Cov(\underline{\overline{Y}}) = Cov\left(\frac{1}{n}\sum_{i=1}^{n}\underline{Y}_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}Cov(\underline{Y}_i) = \frac{1}{n^2}\sum_{i=1}^{n}(A\,\Sigma A') = \frac{1}{n^2}n(A\,\Sigma A') = \frac{1}{n}(A\,\Sigma A')$$

olup, $(A\,\Sigma A')$ bilinmediğinden $Cov(\underline{\overline{Y}}) = \frac{1}{n}(A\,\Sigma A')$ parametresinin en çok olabilirlik tahmin edicisi alınır. Bu parametrenin en çok olabilirlik tahmin edicisi $\widehat{Cov(\underline{\overline{Y}})} = \frac{1}{n}(A\,SA')$ istatistiğidir. Ayrıca;

$\underline{\overline{Y}} = \frac{1}{n}\sum_{i=1}^{n}\underline{Y}_i = \frac{1}{n}\sum_{i=1}^{n}(A\underline{X}_i + \underline{d}) = A\left(\frac{1}{n}\sum_{i=1}^{n}\underline{X}_i\right) + \frac{1}{n}\sum_{i=1}^{n}\underline{d} = A\underline{\overline{X}} + \underline{d}$ şeklinde yazılabilir. Bulunan bu sonuçlar (ii) eşitliğinde yerlerine yazılırsa;

$$\begin{aligned}
T^2_{\underline{Y}} &= \left[\underline{\overline{Y}} - E(\underline{\overline{Y}})\right]'\left(\widehat{Cov(\underline{\overline{Y}})}\right)^{-1}\left[\underline{\overline{Y}} - E(\underline{\overline{Y}})\right] \\
&= \left[(A\underline{\overline{X}} + \underline{d}) - (A\underline{\mu} + \underline{d})\right]'\left(\frac{1}{n}(A\,SA')\right)^{-1}\left[(A\underline{\overline{X}} + \underline{d}) - (A\underline{\mu} + \underline{d})\right] \\
&= \left[A\underline{\overline{X}} + \underline{d} - A\underline{\mu} - \underline{d}\right]'\left(\frac{1}{n}\right)^{-1}(A\,SA')^{-1}\left[A\underline{\overline{X}} + \underline{d} - A\underline{\mu} - \underline{d}\right] \\
&= n\left[A\left(\underline{\overline{X}} - \underline{\mu}\right)\right]'(A')^{-1}S^{-1}A^{-1}\left[A\left(\underline{\overline{X}} - \underline{\mu}\right)\right] \\
&= n\left(\underline{\overline{X}} - \underline{\mu}\right)'A'(A')^{-1}S^{-1}A^{-1}A\left(\underline{\overline{X}} - \underline{\mu}\right) = n\left(\underline{\overline{X}} - \underline{\mu}\right)'S^{-1}\left(\underline{\overline{X}} - \underline{\mu}\right) = T^2_{\underline{X}} \text{ elde edilir.}
\end{aligned}$$

Sonuç olarak Hotelling- T^2 istatistiği lineer dönüşümlerden etkilenmemektedir.

Örnek:4.7 Belirli bir bitki üzerinde 10 benzer deney farklı bariyerde uygulanmış ve her deneyde bitkinin kalsiyumu(X_1) ve topraktaki kalsiyum(X_2) olmak üzere iki özellik üzerinde ölçüm yapılmıştır. Bitki uzmanları ortalama değer olarak, bitkinin kalsiyumunun 2,85 birim ve topraktaki kalsiyumun 15 birim olması durumunda bahçede her şeyin yolunda olduğunu düşünmektedir. Elde edilen veriler aşağıdaki gibidir. Bu verilere dayanarak;

a) Bahçede her şeyin yolunda olup olmadığına %5 önem seviyesinde karar veriniz?

b) Bahçede işler yolunda değilse, bunun kaynağını (nedenini) araştırınız?

Bitkinin Kalsiyumu(X_1)	**Topraktaki Kalsiyum(X_2)**
2,11 2,36 2,13 2,78 1,68	10 35 2 6 2 6 6 2 2
1,76 1,68 1,89 1,98 1,76	2

Çözüm

a) $\left.\begin{matrix} X_1\text{: Bitkinin Kalsiyumu} \\ X_2\text{: Toprağın Kalsiyumu} \end{matrix}\right\} \Rightarrow \underline{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N_2\left(\underline{\mu}\,,\ \Sigma\right)$ olduğunu kabul edelim. Bitki uzmanlarının görüşlerine göre bu değişkenlere ait ortalamalar $\underline{\mu}_0 = \begin{bmatrix} 2{,}85 \\ 15 \end{bmatrix}$ iken bahçede her şeyin yolunda olduğu düşünülmektedir. Yapılan deney sonuçlarına göre bahçede her şeyin yolunda olup olmadığına karar verebilmek için verilen dağılımın ortalama vektörünün $\underline{\mu}_0$ vektörü ile aynı olup olmadığını test etmek gerekir. Buna göre test edilecek hipotezler;

$H_0\colon \underline{\mu} = \underline{\mu}_0$

$H_1\colon \underline{\mu} \neq \underline{\mu}_0$ şeklinde kurulur. Test istatistiği, kitleye ait Σ kovaryans matrisi bilinmediğinden tek örneklem Hotelling-T^2 istatistiğidir.

$T^2 = n\left(\underline{\overline{X}} - \underline{\mu}\right)' \mathrm{S}^{-1}\left(\underline{\overline{X}} - \underline{\mu}\right)$ olup, bu istatistiğin örnekleme dağılımı;

$F = \frac{n-p}{p(n-1)}\, T^2 \sim F_{p\,;n-p}$ şeklindedir. H_0 doğru iken test istatistiğinin alabileceği değeri bulalım.

$n = 10, p = 2\ \ \underline{\overline{X}} = \frac{1}{n}\sum_{i=1}^{n} \underline{X}_i = \begin{bmatrix} \overline{X}_1 \\ \overline{X}_2 \end{bmatrix} = \frac{1}{10}\begin{bmatrix} 20{,}13 \\ 73 \end{bmatrix} = \begin{bmatrix} 2{,}013 \\ 7{,}3 \end{bmatrix}$;

$S = \frac{1}{n-1}\sum_{i=1}^{n}\left(\underline{X}_i - \underline{\overline{X}}\right)\left(\underline{X}_i - \underline{\overline{X}}\right)' = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} = \begin{bmatrix} 0{,}12198 & 1{,}14102 \\ 1{,}14102 & 102{,}23333 \end{bmatrix}$

X_{1i}	X_{2i}	$X_{1i} - \overline{X}_1$	$X_{2i} - \overline{X}_2$	$(X_{1i} - \overline{X}_1)^2$	$(X_{1i} - \overline{X}_1)^2$	$(X_{1i} - \overline{X}_1)(X_{2i} - \overline{X}_2)$
2,11	10	0,097	2,7	0,009409	7,29	0,2619
2,36	35	0,347	27,7	0,120409	767,29	9,6119
2,13	2	0,117	-5,3	0,013689	28,09	-0,6201
2,78	6	0,767	-1,3	0,588289	1,69	-0,9971
1,68	2	-0,333	-5,3	0,110889	28,09	1,7649
1,76	6	-0,253	-1,3	0,064009	1,69	0,3289
1,68	6	-0,333	-1,3	0,110889	1,69	0,4329
1,89	2	-0,123	-5,3	0,015129	28,09	0,6519
1,98	2	-0,033	-5,3	0,001089	28,09	0,1749
1,76	2	-0,253	-5,3	0,064009	28,09	1,3409
20,13	73			1,09781	920,1	10,2692
2,013	**7,3**			s_{11} =**0,12198**	s_{22} =**102,23333**	s_{11} =**1,14102**

$\underline{\overline{X}} - \underline{\mu}_0 = \begin{bmatrix} 2{,}013 - 2{,}85 \\ 7{,}3 - 15 \end{bmatrix} = \begin{bmatrix} -0{,}837 \\ -7{,}7 \end{bmatrix}$ ve $S^{-1} = \begin{bmatrix} 9{,}1537 & -0{,}1022 \\ -0{,}1022 & 0{,}0109 \end{bmatrix}$ dir. Bu değerler T^2 istatistiğinde yerlerine yazılırsa, H_0 doğru iken test istatistiğinin alabileceği değeri

$T_h^2 = 10[-0,877 \quad -7,7]\begin{bmatrix} 9,1537 & -0,1022 \\ -0,1022 & 0,0109 \end{bmatrix}\begin{bmatrix} -0,837 \\ -7,7 \end{bmatrix} =$
$10[-7,241 \quad 0,0057]\begin{bmatrix} -0,837 \\ -7,7 \end{bmatrix} = 60,168$ bulunur. Böylece $F_h = \frac{n-p}{p(n-1)} T_h^2 = \frac{8}{2\times 9}(60,168) = 26,741$ olur.

Karar: $\alpha = 0,05$ önem seviyesinde kritik değer $F_t = F_{p;n-p;\alpha} = F_{2;8;0,05} = 4,46$ iken $T_t^2 = \frac{p(n-1)}{n-p} F_t = \frac{2\times 9}{8}(4,46) = 10,035$ dir. Sonuç olarak; $26,741 > 4,46$, yani $F_h > F_t$ (veya $60,168 > 10,035$, yani $T_h^2 > T_t^2$) olduğundan H_0 hipotezi ret edilir. Buna göre %95 güven seviyesinde bahçede her şeyin yolunda olduğu söylenemez. Bu örnek grubu, bitki uzmanlarının görüşünün desteklendiğini göstermemektedir.

b) Bahçede işlerin yolunda olmamasının nedeni ya bitkinin kalsiyum değerinden ya toprağın kalsiyum değerinden ya da her ikisinden de kaynaklamaktadır. Bunu belirlemek için sırasıyla her bir değişken için inceleme yapalım.

X_1: Bitkinin Kalsiyumu için : $\underline{a} = [1 \quad 0]$ olmak üzere $\underline{a}\,\underline{\mu}$ parametresi için güven aralığını oluşturarak;

$H_0: \underline{a}\,\underline{\mu} = \underline{a}\,\underline{\mu}_0 \qquad H_0: \mu_1 = 2,85$

$H_1: \underline{a}\,\underline{\mu} \neq \underline{a}\,\underline{\mu}_0 \Rightarrow \qquad H_1: \mu_1 \neq 2,85$ hipotezlerini test edelim. Güven aralığı;

$$P\left[\underline{a}\,\overline{\underline{X}} - \sqrt{\frac{\underline{a}\,S\underline{a}'}{n} \times T_t^2} \leq \underline{a}\,\underline{\mu} \leq \underline{a}\,\overline{\underline{X}} + \sqrt{\frac{\underline{a}\,S\underline{a}'}{n} \times T_t^2}\right] = 1 - \alpha$$ dir. Burada;

$\underline{a}\,\underline{\mu} = \mu_1$, $\underline{a}\,\overline{\underline{X}} = \overline{X}_1 = 2,013$, $\underline{a}\,S\underline{a}' = s_{11} = 0,12198$ ve $T_t^2 = 10,035$ değerleri yerlerine yazılırsa;

$$P\left(2,013 - \sqrt{\frac{0,12198}{10}(10,035)} \leq \mu_1 \leq 2,013 + \sqrt{\frac{0,12198}{10}(10,035)}\right) = 0,95$$

$P(1,663 \leq \mu_1 \leq 2,363) = 0,95$ bulunur. Güven aralığı $\mu_{10} = 2,85$ değerini kapsamadığından $H_0: \mu_1 = 2,85$ hipotezi ret edilir. Buna göre bahçede işlerin yolunda olmamasının bir kaynağı bitkinin kalsiyumunun düşük olmasıdır. Çünkü bitki kalsiyumu ortalaması için düşülen değer 2,85 olup, üst güven sınırından büyüktür (2,85>2,363).

X_1: Toprağın Kalsiyumu için : $\underline{a} = [0 \quad 1]$ olmak üzere $\underline{a}\,\underline{\mu}$ parametresi için güven aralığını oluşturarak;

$H_0: \underline{a}\,\underline{\mu} = \underline{a}\,\underline{\mu}_0 \qquad H_0: \mu_2 = 15$

$H_1: \underline{a}\,\underline{\mu} \neq \underline{a}\,\underline{\mu}_0 \Rightarrow \qquad H_1: \mu_2 \neq 15$ hipotezlerini test edelim. Burada;

$\underline{a}\,\underline{\mu} = \mu_2$, $\underline{a}\,\overline{\underline{X}} = \overline{X}_2 = 7,3$, $\underline{a}\,S\underline{a}' = s_{22} = 102,23333$ ve $T_t^2 = 10,035$ değerleri yerlerine yazılırsa;

$$P\left(7,3-\sqrt{\frac{102,23333}{10}(10,035)} \le \mu_2 \le 7,3+\sqrt{\frac{102,23333}{10}(10,035)}\right)=0,95$$

$P(-2,829 \le \mu_2 \le 17.429) = 0,95$ bulunur. Güven aralığı $\mu_{20} = 15$ değerini kapsadığından $H_0: \mu_2 = 15$ hipotezi ret edilemez. Buna göre toprağın kalsiyumu bahçede işlerin yolunda olmamasının nedeni olamaz. Çünkü toprağın kalsiyum ortalaması için düşülen değer 15 olup, güven sınırları arasında kalmaktadır.