T.C.

ONDOKUZ MAYIS ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

İST.616 ÇOK DEĞİŞKENLİ İSTATİSTİK II
4. Hafta

Prof. Dr. YÜKSEL ÖNER

SAMSUN 2020

özenilen üniversite

3.4 Varyans Kovaryans Matrisleri ile İlgili Testler

3.4.1 Tek Kitle İçin Bartlett Testi

$\underline{X} \sim N_p\left(\underline{\mu}, \Sigma\right)$ dağılımı verilsin ve varyans kovaryans matrisi (Σ) bilinmiyor olsun. $\Sigma_0: p \times p$ simetrik, pozitif tanımlı ve bilinen bir matris olmak üzere test edilecek hipotezler:

$$H_0: \Sigma = \Sigma_0$$

$$H_1: \Sigma \neq \Sigma_0 \tag{3.42}$$

şeklinde oluşturulur. H_0 hipotezini test etmek için gerekli olan test istatistiğini elde etmek için verilen kitleden rastgele çekilen n birimlik bir örnek $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n$ olsun. Bu durumda kitle parametrelerinin en çok olabilirlik tahmin edicileri $\overline{\underline{X}}$ ve S istatistikleridir. Olabilirlik oran testi yaklaşımı uygulandığında test istatistiği olan olabilirlik oran istatistiği (Λ);

$$\Lambda = \left[\left(\frac{n-1}{n}\right)^p \frac{|S|}{|\Sigma_0|}\right]^{n/2} exp\left\{-\frac{n-1}{2}\left[\dot{I}z(S\Sigma_0^{-1}) - np\right]\right\} \tag{3.43}$$

olarak elde edilir. Test istatistiği örnek hacmine bağlı olarak Λ istatistiğinin örneklem dağılımı olarak ifade edilen Ki-Kare istatistiği olacaktır. Buna göre test istatistiği $n \geq 50$ iken;

$$L = (n-1)\left[ln|\Sigma_0| - ln|S| + \dot{I}z(S\Sigma_0^{-1}) - p\right] \sim \chi^2_{\frac{p(p+1)}{2}} \tag{3.44}$$

şeklinde, $n < 50$ iken;

$$L' = \left[1 - \frac{1}{6(n-1)}\left(2p + 1 - \frac{2}{p+1}\right)\right] L \sim \chi^2_{\frac{p(p+1)}{2}} \tag{3.45}$$

şeklindedir.

Karar: $\boldsymbol{\alpha}$ önem seviyesi ve $\chi_t^2 = \chi^2_{\frac{p(p+1)}{2};\alpha}$ kritik değer olmak üzere, eğer $L > \chi_t^2$ (veya $L' > \chi_t^2$) ise H_0 hipotezi ret edilir. Aksi takdirde ret edilemez.

Yorum: H_0 hipotezi ret edilirse n birimlik örneklemin çekildiği çok değişkenli kitlenin varyans kovaryans matrisi Σ_0 matrisi olamaz.

H_0 hipotezi ret edilemezse n birimlik örneklemin çekildiği çok değişkenli normal kitlenin varyans kovaryans matrisi Σ_0 matrisidir.

3.4.2 İki Ve Daha Fazla Kitle İçin Box M Testi

Kabul edelim ki g: kitle sayısı ($g \geq 2$) olsun. $k = 1, 2, \ldots, g$ olmak üzere, k-ncı kitle ve dağılımı $\underline{X}_k \sim N_p\left(\underline{\mu}_k, \Sigma_k\right)$ ile verilsin. Bu kitlelerin birbirinden bağımsız gruplar olduğunu kabul edelim. Bu durumda bu grupların varyans kovaryans matrislerinin aynı olup olmadığı, diğer bir ifadeyle g- tane birbirinden bağımsız çok değişkenli normal dağılıma sahip grupların dağılımlarının homojenlik gösterip göstermediği incelenebilir. Bu amaçla kullanılan test

Bartlett testinin ikiden fazla gruba genelleştirilmesi olan Box M testidir. Boz M testinde test edilecek hipotezler varyans kovaryans matrislerinin eşitliği üzerine kurulur.

H_0: $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_g = \Sigma$

H_1: $\exists\Sigma_k$ diğerlerinden farklı (3.46)

Test istatistiğini elde etmek için her bir gruptan n_k birimlik örnekler çekilir. k-ncı grup için örnek birimleri $\underline{X}_{k1}, \underline{X}_{k2}, \dots, \underline{X}_{kn_k}$, $\left(k = \overline{1, g}\right)$olmak üzere, $\underline{\mu}_k$ ve Σ_k parametrelerinin yansız tahmin edicileri sırasıyla; $\underline{\overline{X}}_k = \frac{1}{n_k}\sum_{i=1}^{n_k}\underline{X}_{ki}$ ve $S = \frac{1}{n_k-1}\sum_{i=1}^{n_k}\left(\underline{X}_{ki} - \underline{\overline{X}}_k\right)\left(\underline{X}_{ki} - \underline{\overline{X}}_k\right)'$ istatistikleridir.

H_0 doğru iken gruplar ortak kitle varyans kovaryans matrisine(Σ) sahip olacağından, bir parametre olan bu matrisin yansız bir tahmin edicisi örnek için ortak varyans kovaryans matrisi olarak bilinen;

$$S = \frac{\sum_{k=1}^{g}(n_k-1)S_k}{\sum_{k=1}^{g}(n_k-1)} = \frac{\sum_{k=1}^{g}W_k}{N-g} \tag{3.47}$$

istatistiğidir. Böylece test istatistiği;

$$\chi^2 = MC^{-1} \sim \chi^2_{\frac{p(p+1)(g-1)}{2}} \tag{3.48}$$

olup, burada;

$$M = \sum_{k=1}^{g}(n_k - 1)ln|S| - \sum_{k=1}^{g}(n_k - 1)ln|S_k| \tag{3.49}$$

$$C^{-1} = \begin{cases} 1 - \frac{2p^2+3p-1}{6(p+1)(g-1)}\left[\sum_{k=1}^{g}\frac{1}{n_k-1} - \frac{1}{N-g}\right] & , n_k'\text{lar farklı ise} \\ 1 - \frac{(2p^2+3p-1)(g+1)}{6(p+1)g(n-1)} & , n_1 = n_2 = \cdots = n_g = n \text{ ise} \end{cases} \tag{3.50}$$

eşitlikleri ile bulunur.

Karar: α önem seviyesi ve $\chi_t^2 = \chi^2_{\frac{p(p+1)(g-1)}{2};\alpha}$ kritik değer olmak üzere, eğer $\chi^2 > \chi_t^2$ ise H_0 hipotezi ret edilir. Aksi takdirde ret edilemez.

Yorum: H_0 hipotezi ret edilirse grupların varyans kovaryans matrislerinin eşit olmadığı, bir diğer ifade ile kovaryans matrisleri yönünden gruplarının dağılımının homojen olmadığı söylenir.

H_0 hipotezi ret edilemezse grupların varyans kovaryans matrislerinin eşit olduğu, bir diğer ifade ile kovaryans matrisleri yönünden gruplarının dağılımının homojen olduğu söylenir.

Örnek 3.6 $\underline{X} \sim N_3\left(\underline{\mu}, \Sigma\right)$ dağılımı verilsin. Bu dağılımdan rastgele çekilen 20 birimlik bir örneklem için örnek varyans kovaryans matrisi;

$S = \begin{bmatrix} 3{,}42 & 2{,}6 & 1{,}89 \\ 2{,}6 & 8{,}0 & 6{,}5 \\ 1{,}89 & 6{,}5 & 9{,}62 \end{bmatrix}$ bulunmuştur. Bu örneklemin çekildiği kitle varyans kovaryans matrisinin;

$\Sigma_0 = \begin{bmatrix} 4 & 3 & 2 \\ 3 & 6 & 5 \\ 2 & 5 & 10 \end{bmatrix}$ olup olamayacağına %5 önem seviyesinde karar veriniz?

Çözüm: Hipotezler:

$H_0: \Sigma = \Sigma_0$

$H_1: \Sigma \neq \Sigma_0$

Test istatistiği $n = 20 (< 50)$ olduğundan; $L' = \left[1 - \frac{1}{6(n-1)}\left(2p + 1 - \frac{2}{p+1}\right)\right] L \sim \chi^2_{\frac{p(p+1)}{2}}$ olup, burada $L = (n-1)[ln|\Sigma_0| - ln|S| + \text{İ}z(S\Sigma_0^{-1}) - p]$ dir.

$$p = 3,\ |\Sigma_0| = \begin{vmatrix} 4 & 3 & 2 \\ 3 & 6 & 5 \\ 2 & 5 & 10 \end{vmatrix} = 4(35) - 3(20) + 2(3) = 86,\ \Sigma_0^{-1} = \frac{1}{86}\begin{bmatrix} 35 & -20 & 3 \\ -20 & 36 & -14 \\ 3 & -14 & 15 \end{bmatrix}$$

$$|S| = \begin{vmatrix} 3{,}42 & 2{,}6 & 1{,}89 \\ 2{,}6 & 8{,}0 & 6{,}5 \\ 1{,}89 & 6{,}5 & 9{,}62 \end{vmatrix} = 88{,}9822;\ S\Sigma_0^{-1} = \frac{1}{86}\begin{bmatrix} 73{,}37 & -1{,}26 & -73{,}39 \\ -49{,}5 & 145 & -6{,}7 \\ -34{,}99 & 61{,}52 & 58{,}97 \end{bmatrix};$$

$$\text{İ}z(S\Sigma_0^{-1}) = \frac{1}{86}(73{,}37 + 145 + 58{,}97) = 3{,}2249$$

Bu sonuçlar yerlerine yazılırsa $L = 19[ln(86) - ln(88{,}9822) + 3{,}2249 - 3] = 3{,}625$ ve böylece;

$$L' = \left[1 - \frac{1}{6\times 19}\left(2\times 3 + 1 - \frac{2}{3+1}\right)\right](3{,}625) = 3{,}418$$

bulunur.

Karar: $\alpha = 0{,}05$ önem seviyesi ve $\chi^2_t = \chi^2_{\frac{p(p+1)}{2};\alpha} = \chi^2_{6;0{,}05} = 12{,}6$ kritik değer olmak üzere, ise $L' = 3{,}418 < 12{,}6$ olduğundan H_0 hipotezi ret edilemez. Buna göre %95 güvenle mevcut örneklemin çekildiği kitlenin varyans kovaryans matrisi verilen Σ_0 matrisidir.

Örnek 3.6 $\underline{X}_1 \sim N_2\left(\underline{\mu}_1, \Sigma_1\right)$ ve $\underline{X}_2 \sim N_2\left(\underline{\mu}_2, \Sigma_2\right)$ dağılımlarından 32'şer birimlik örnekler çekilmiş ve örnek varyans kovaryans matrisleri sırası ile $S_1 = \begin{bmatrix} 4{,}32 & 1{,}88 \\ 1{,}88 & 9{,}18 \end{bmatrix}$ ve $S_2 = \begin{bmatrix} 2{,}52 & 1{,}90 \\ 1{,}90 & 10{,}06 \end{bmatrix}$ olarak hesaplanmıştır. Bu örneklere göre %5 önem seviyesinde verilen dağılımların homojen kovaryans matrisli olup olmadığına karar veriniz?

Çözüm: Hipotezler

$H_0: \Sigma_1 = \Sigma_2 = \Sigma$

$H_1: \exists\Sigma_k$ diğerlerinden farklı

Test istatistiği;

$\chi^2 = MC^{-1} \sim \chi^2_{\frac{p(p+1)(g-1)}{2}}$ dir. Burada

$p = 2,\, g = 2,\, n_1 = n_2 = n = 32,\, N = n_1 + n_2 = 64$

$M = \sum_{k=1}^{g}(n_k - 1)ln|S| - \sum_{k=1}^{g}(n_k - 1)ln|S_k|$;

$|S_1| = \begin{vmatrix} 4,32 & 1,88 \\ 1,88 & 9,18 \end{vmatrix} = 36,1232$; $|S_2| = \begin{vmatrix} 2,52 & 1,90 \\ 1,90 & 10,06 \end{vmatrix} = 21,7412$ ve

$S = \frac{\sum_{k=1}^{g}(n_k-1)S_k}{\sum_{k=1}^{g}(n_k-1)} = \frac{1}{2}(S_1 + S_2) = \begin{bmatrix} 3,42 & 1,89 \\ 1,89 & 9,62 \end{bmatrix}$ olup, $|S| = \begin{vmatrix} 3,42 & 1,89 \\ 1,89 & 9,62 \end{vmatrix} = 29,3283$ bulunur. Bu sonuçlar yerlerine konursa;

$M = 62 \times ln(29,3283) - [31ln(36,1232) + 31ln(21,7412)] = 2,8198$ ve $n_1 = n_2$ olduğundan $C^{-1} = 1 - \frac{(2p^2+3p-1)(g+1)}{6(p+1)g(n-1)} = 1 - \frac{(8+6-1)(3)}{6\times3\times2\times31} = 0,965$ bulunur. Böylece test istatistiğinin alabileceği değer;

$\chi^2 = (2,8198)(0,965) = 2,721$ olarak elde edilir.

Karar: $\alpha = 0,05$ önem seviyesi ve $\chi_t^2 = \chi^2_{\frac{p(p+1)(g-1)}{2};\alpha} = \chi^2_{3\,;0,05} = 7,815$ kritik değer olmak üzere; $\chi^2 = 2,721 < 7,815$ olduğundan H_0 hipotezi ret edilemez. Buna göre verilen dağılımlar %95 güvenle homojen varyans kovaryans matrislidir.