ANADOLU ÜNİVERSİTESİ BİLİM VE TEKNOLOJİ DERGİSİ –B
Teorik Bilimler

ANADOLU UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY –B
Theoretical Sciences
Cilt/Vol.:1-Sayı/No: 1 : 43-48 (2011)

KORELASYON MATRİSLERİNİN EŞİTLİĞİ TESTİNDE PERMÜTASYON TESTİ

Ufuk EKİZ [1], Meltem EKİZ[1]

ÖZ

Bu çalışmada, farklı yığınlardaki p tane tesadüfi değişken arasındaki korelasyon matrislerinin eşitliği hipotezinin test edilmesinde permütasyonların oluşturulmasına dayalı Shipley'in önerdiği test istatistiğinin I. tip hata olasılıkları Monte-Carlo yöntemiyle hesaplanmıştır. Normal ve karma-normal dağılımlı üç yığın için ayrı ayrı incelenen deneysel I.tip hata olasılıkları, örnek çapları 3 ile 10 arasında değişkenlik gösterirken ve anlamlılık düzeyi 0.05, 0.10 ve 0.20 iken elde edilmiştir. Simülasyon sonuçlarına göre, normal dağılan veri için tahmin edilen deneysel olasılık değerleri beklenen α'nın çok yakınında değerler vermiştir. Karma-normal veri için sonuçların tümünün α'dan küçük değerler olması, test istatistiğinin normal dağılmayan (karma-normal) veri için de sağlam olduğunu göstermiştir.

Anahtar Kelimeler : Korelasyon matrislerinin eşitliği, Permütasyon testi, Değişebilirlik.

TESTING THE EQUALITY OF THE CORRELATION MATRICES WITH THE PERMUTATION TEST

ABSTRACT

In this study, the emprical rejection levels of the Shipley's test based on permutations are calculated by using the Monte Carlo procedure in order to test the hypothesis of the equality of correlation matrices of different populations, involving p variables. These values are obtained for normal and mixtures of normal distributed three population and are investigated when the expected rejection level is 0.05, 0.10 and 0.20 and the sample size differs from 3 to 10. In respect of the simulation results, the emprical rejection levels of the normal data seems to be very close to the expected rejection levels. Separately, since all the results are smaller than α in case of mixtures of normal data, the test statistic appears to be robust for non-normal (mixtures of normal) data too.

Keywords: Equality of correlation matrices, Permutation test, Exchangeability.

[1,] Gazi Üniversitesi Fen Edebiyat Fak. İstatistik Bölümü 06500 Beşevler/Ankara.
Tel: 0312 202 14 77; Fax: 0312 212 22 79; E-posta:ufukekiz@gazi.edu.tr

Geliş: 9 Kasım 2009; **Düzeltme:** 15 Ocak 2010; **Kabul:** 30 Mart 2010

1. GİRİŞ

Farklı yığınlardaki p tane tesadüfi değişken için elde edilecek korelasyon, kovaryans ya da kısmi korelasyon matrisleri arasında anlamlı bir fark olup olmadığı ile ilgili birçok çalışma vardır. Manly ve Rayner (1987) çok değişkenli normal dağılım varsayımının sağlandığı ve yeterince büyük çaplı örnekler için kovaryans matrislerinin eşitliği hipotezi için kullanılabilecek olabilirlik oran testini önermiştir. Korelasyon matrislerinin eşitliği hipotezinin testi problemlerinde, korelasyon matris elemanlarının standartlaştırılarak kullanıldığı testlerde, olabilirlik oran testindeki varsayımlar ihmal edilmiştir (Krzanowski, 1993). Riska (1985) Jacknife yöntemini, Krzanowski (1993) permütasyon testini kullanarak korelasyon matrislerinin eşitliği hipotezinin test edilmesi üzerine çalışmalar yapmışlardır. Krzanowski'nin önerdiği test, korelasyon matrislerinin ortak aygen vektör ve aygen değerlerine dayalıdır ve normal dağılmayan veri için de Monte Carlo simülasyonları iyi sonuçlar vermiştir. Ayrıca Shipley (2000), p değişkenli yığınların korelasyon matrislerinin eşitliği hipotezinin test edilmesinde, korelasyon matrisinin yalnızca köşegen üst elemanlarının kullanılmasını ve permütasyonların oluşturulmasını önermiştir.

Bu çalışmada, Shipley'in önerdiği test istatistiği tanıtılarak, Monte Carlo simülasyon yöntemiyle, deneysel I.tip hata olasılıklarının beklenen I.tip hata olasılıklarına olan yakınlığı normal ve karmanormal veri için incelenmiştir. Ayrıca kullanılan yöntemin gerçek veri üzerinde uygulaması da yapılmıştır.

2. TEST İSTATİSTİĞİ

Permütasyon testi, tesadüfi değişkenlerin değişebilirliğine (exchangeability) dayalı parametrik olmayan bir test yöntemidir (Sakaori, 2002). Bu yöntemin, yokluk hipotezi altında kesin ve sapmasız sonuçlar ürettiği ve en az parametrik testler kadar güçlü olduğu bilinmektedir.

Permütasyon testinin geçerliliği için şart olan tesadüfi değişkenlerin değişebilirliği aşağıdaki gibi tanımlanmıştır. p tesadüfi değişken sayısını, n gözlem sayısını ifade etmek üzere, $i = 1, 2, \ldots, n$ ve $X_i = \left(X_1, X_2, ..., X_p\right)$ şeklinde tanımlansın. Eğer

$$Pr\left(\bigcap_{i=1}^{n}\left(X_i \le x_i\right)\right) = Pr\left(\bigcap_{i=1}^{n}\left(X_i \le x_{c_i}\right)\right)$$

olasılığı sağlanıyorsa, $X_1, X_2, ..., X_n$ tesadüfi değişkenleri değişebilirdir denir (Sakaori, 2002). Burada $\left(c_1, c_2, ..., c_n\right)$ n tane gözlemin mümkün tüm permütasyonlarından herhangi birini ifade etmektedir. X_i tesadüfi değişkenleri bağımsız ve aynı dağılımlı ise değişebilirlik özelliği sağlanıyor demektir.

Yığın sayısı g olmak üzere bu g tane bağımsız yığına ilişkin korelasyon matrislerinin eşitliği hipotezi göz önünde bulundurulsun. X_{ijk}, k yığındaki i birime ilişkin j tesadüfi değişken olmak üzere, k yığındaki j tesadüfi değişkenin ortalaması μ_{jk} ve varyansı da σ_{jk}^2 olsun. $\left(i = 1, 2, \ldots, n_k, \text{ j} = 1, 2, \ldots, p, \; k = 1, 2, \ldots, g\right)$. k yığındaki p tane tesadüfi değişkenden oluşan $X_{ik} = \left(X_{i1}, X_{i2}, ..., X_{ip}\right)$ vektörüne ilişkin varyans kovaryans matrisi,

$$\sum_k = \begin{bmatrix} \sigma_{1k}^2 & \sigma_{12k} & \ldots & \sigma_{1pk} \\ \sigma_{21k} & \sigma_{2k}^2 & \ldots & \sigma_{2pk} \\ . & . & \ldots & . \\ . & . & \ldots & . \\ . & . & \ldots & . \\ \sigma_{p1k} & \sigma_{p2k} & \ldots & \sigma_{pk}^2 \end{bmatrix}$$

şeklinde tanımlansın. Yığınlar arasında tesadüfi değişken vektörleri X_i'lerin değişebilirliği, X_{jk}'nın μ_{jk} ve σ_{jk} ile standartlaştırılması sonucu sağlanır. Eğer bu yığın parametreleri bilinmiyorsa bunların yerine örnekten elde edilen tahmin edicilerinin kullanılması ile standartlaştırma yapılarak ta değişebilirliğin sağlandığı ileri sürülmektedir (Krzanowski, 1993). k. yığındaki i. tesadüfi değişkenlerin standartlaştırılması, $X_{ik} = \left(X_{i1}, X_{i2}, ..., X_{ip} \right)$ olmak üzere,

$$Y_{ik} = \left(\frac{X_{i1} - \mu_{1k}}{\sigma_{1k}}, \frac{X_{i2} - \mu_{2k}}{\sigma_{2k}}, ..., \frac{X_{ip} - \mu_{pk}}{\sigma_{pk}} \right)$$

şeklinde olur. $i = 1, ..., n$ ve $n_1 + n_2 + ... + n_k = n$ iken $Y_{11}, Y_{21}, ..., Y_{n_1 1}$; $Y_{12}, Y_{22}, ..., Y_{n_2 2}$; ...; $Y_{1g}, Y_{2g}, ..., Y_{n_g g}$ standart tesadüfi değişkenleri g tane bağımsız yığın için değişebilirlik özelliğini sağlar. $k = 1, 2, \ldots, g$ ve $X_{ik} = \left(X_{i1}, X_{i2}, ..., X_{ip} \right)$ olmak üzere, R_k, k yığına ilişkin $p \times p$'lik korelasyon matrisi, $\hat{R}_k$ da R_k'nın tahmin edicisi olsun. r_k, R_k'nın köşegenin üstünde yer alan $p(p-1)/2$ tane korelasyon katsayısından meydana getirilen bir vektör olarak tanımlansın. $r_k = \left[R_{12}, R_{13}, ..., R_{1p}; R_{23}, R_{24} ..., R_{2p}; \ ... \ ; R_{(p-1)p} \right]$ iken g tane yığına ilişkin korelasyon matrislerinin eşitliği için tanımlanan yokluk hipotezi,

$$H_0; \ r_1 = r_2 = \ ... \ = r_g$$

ile ifade edilir. $\hat{r}_k$, r_k'nın tahmin edicisi olmak üzere bu hipotezi test etmek için kullanılacak test istatistiği,

$$S = \sum_{i=1}^{g-1} \sum_{j=i+1}^{g} \left(\hat{r}_i - \hat{r}_j \right)' \left(\hat{r}_i - \hat{r}_j \right)$$

dır (Shipley, 2000). Permütasyon testi yöntemiyle H_0 hipotezinin testinde, örnekten S istatistiği hesaplanır. Sonra n tane gözlemden her bir gruba $n_1, n_2, \ ... \ , n_g$ birim düşecek şekilde oluşturulması gereken tüm mümkün permütasyonların herbiri için S^* (çok sayıda permütasyonun oluşması durumunda tesadüfi olarak belirli sayıda permütasyonu seçmek yoluna gidilebilir) hesaplanır. Permütasyonlardan hesaplanan S^*'ların içerisinden S'den büyük olanların oranı başta belirlenen anlamlılık düzeyinden küçük ise H_0 hipotezi red kararı verilir.

3. SİMÜLASYON ÇALIŞMASI

H_0 hipotezinin doğruluğu altında I.tip hata olasılıklarının belirlenmesine ilişkin simülasyon çalışması için Matlab programında yazmış olduğumuz program kullanılmıştır. Bu program tesadüfi değişkenlerin dağılımlarının normal ve karma-normal olması durumları için iki farklı şekilde tasarlanmıştır. Tesadüfi değişkenlerin çok değişkenli normal dağılıma sahip olması durumu için,

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ & r_{22} & r_{23} \\ & & r_{33} \end{bmatrix}$$

olmak üzere, $H_0 : R_1 = R_2 = R_3$ olsun. H_0 hipotezinin doğruluğu koşulu altında, varyans-kovaryans matrisi

$$\sum = \begin{bmatrix} 5 & 1 & 2 \\ 1 & 4 & 3 \\ 2 & 3 & 8 \end{bmatrix}$$

aynı olmak üzere merkezi parametreleri

$$\mu_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \ \mu_2 = \begin{bmatrix} 2 \\ 5 \\ 7 \end{bmatrix}, \ \mu_3 = \begin{bmatrix} 6 \\ 1 \\ 8 \end{bmatrix}$$

olan çok değişkenli normal dağılımlardan n_1, n_2 ve n_3 çaplı örnekler üretilmiştir. α 'nın herhangi bir değeri için 1000 tekrarın sonucunda, gerçekte doğru olan H_0 hipotezinin permütasyon testi sonuçlarına göre hangi oranda red edildiği tahmin edilmiştir. Elde edilen sonuçlar α 'nın ve n_1, n_2, n_3 'ün çeşitli değerleri için Tablo 1'de verilmiştir.

Aynı işlemler tesadüfi değişkenlerin dağılımının $\varepsilon = 0.30$ ve

$$\sum_1 = \begin{bmatrix} 5 & 1 & 2 \\ 1 & 4 & 3 \\ 2 & 3 & 8 \end{bmatrix}, \quad \sum_2 = \begin{bmatrix} 4 & 2 & 4 \\ 2 & 9 & 4 \\ 4 & 4 & 16 \end{bmatrix}$$

olmak üzere,

$$(1-\varepsilon) N(\mu_1, \sum_1) + \varepsilon N(\mu_2, \sum_2)$$
$$(1-\varepsilon) N(\mu_2, \sum_1) + \varepsilon N(\mu_3, \sum_2)$$
$$(1-\varepsilon) N(\mu_3, \sum_1) + \varepsilon N(\mu_1, \sum_2)$$

ile tanımlanan çok değişkenli karma-normal dağılımlardan n_1, n_2 ve n_3 çaplı örnekler üretilerek yapılmış ve sonuçlar Tablo 2'de verilmiştir.

Tablo 1. Normal dağılımdan üretilen veri için H_0 hipotezinin doğruluğu altında elde edilen deneysel I. tip hata olasılıkları

	α		
n_1, n_2, n_3	0.05	0.10	0.20
3,3,3	0.0440	0.1030	0.1770
4,4,4	0.0560	0.0940	0.2070
5,5,5	0.0490	0.1220	0.1980
5,6,4	0.0380	0.1000	0.1990
5,6,5	0.0470	0.0960	0.1910
5,6,6	0.0430	0.1130	0.2200
6,6,6	0.0500	0.1010	0.2020
10,10,10	0.0520	0.1070	0.1920

Tablo 2. Karma-normal dağılımdan üretilen veri için H_0 hipotezinin doğruluğu altında elde edilen deneysel I. tip hata olasılıkları

	α		
n_1, n_2, n_3	0.05	0.10	0.20
3,3,3	0.0180	0.0700	0.1770
4,4,4	0.0220	0.0600	0.1620
5,5,5	0.0260	0.0750	0.1600
5,6,4	0.0240	0.0730	0.1680
5,6,5	0.0430	0.0770	0.1540
5,6,6	0.0370	0.0690	0.1540
6,6,6	0.0120	0.0770	0.1740
10,10,10	0.0330	0.0890	0.1780

4. GERÇEK VERİYE DAYALI BİR UYGULAMA

Korelasyon matrislerinin eşitliğinde permütasyon testi yönteminin kullanımını gerçek bir veri üzerinde uygulamak amacı ile Smith vd. (1962)'de yer alan veriler kullanılmıştır. Bu veri setinde ağırlıklarına göre dört gruba sınıflandırılmış birimler ve her birimden yaş, ağırlık ve boy uzunluğu tesadüfi değişkenlerine ilişkin alınmış ölçümler yer almaktadır. Gözlenmiş bu veriler üzerinden elde edilen S istatistiğine ilişkin değer 22.1623 olarak hesaplanmıştır. Ayrıca 4 gruptan her biri için elde edilen korelasyon matris tahminleri,

$$\hat{R}_1 = \begin{bmatrix} 1 & -0.9608 & 0 \\ & 1 & -0.2774 \\ & & 1 \end{bmatrix},$$

$$\hat{R}_2 = \begin{bmatrix} 1 & 0.9955 & 0.7467 \\ & 1 & 0.7015 \\ & & 1 \end{bmatrix},$$

$$\hat{R}_3 = \begin{bmatrix} 1 & -0.9693 & -0.8195 \\ & 1 & 0.8824 \\ & & 1 \end{bmatrix}$$

$$\hat{R}_4 = \begin{bmatrix} 1 & -0.8760 & -0.8926 \\ & 1 & 0.9082 \\ & & 1 \end{bmatrix}$$

olarak elde edilmiştir. Örnekten tesadüfi olarak oluşturulan 1000000 adet permütasyonun her biri için S^* değerleri hesaplanmış, $S = 22.1623$'den büyük olanların oranı 0.0295 olarak belirlenmiştir. I. tip hata olasılığı $\alpha = 0.05$ olarak alındığında, $0.0295 < 0.05$ olduğu için H_0 hipotezi red edilir. Yani dört grubun korelasyon matrislerinin eşit olduğu hipotezi red edilir.

5. SONUÇ/TARTIŞMA

Farklı yığınlardaki p tane tesadüfi değişken arasındaki korelasyon matrislerinin eşitliği hipotezinin test edilmesinde permütasyonların oluşturulmasına dayalı Shipley (2000)'in önerdiği test istatistiğine ilişkin, I. tip hata olasılıkları Monte-Carlo yöntemiyle hesaplanmıştır. Normal ve karma-normal dağılımlı üç yığın için I.tip hata olasılıkları, örnek çapları 3 ile 10 arasında değişkenlik göstermiş ve üç farklı anlamlılık düzeyi için tahmin edilmiştir. Bu uygulama ile aşağıdaki sonuçlar elde edilmiştir:

- Normal dağılan veri için tahmin edilen deneysel olasılık değerlerinin, beklenen α'nın çok yakınında değerler verdiği görülmüştür.
- Karma-normal dağılımlı veri için elde edilen sonuçların tümünün α'dan küçük olması, test istatistiğinin normal dağılmayan (karma-normal) veri için sağlam kaldığını göstermiştir.
- Veri elde etmenin pahalı ve zaman alıcı olduğu deneylerde, örnek çaplarının küçük ve farklılık göstermesi durumlarında da Shipley'in testinin korelasyon katsayılarının eşitliği hipotezinin testinde kullanılabileceği görülmüştür.

Ayrıca korelasyon matrislerinin eşitliğinde permütasyon testi yöntemi, gerçek bir veri üzerinde de uygulanarak sonuçların geçerliliği gösterilmiştir.

KAYNAKLAR

Krzanowski, W.J. (1993). Permutational tests for correlation matrices. *Statistics and Computing* 3, 37-44.

Manly, B.F., Rayner, J.C.W. (1987). The comparison of sample covariance matrices using likelihood ratio tests. *Biometrika* 74, 841-847.

Riska, B. (1985). Group size factors and geographic variation of morphometric correlation. *Ecology* 39, 792-803.

Sakaori, F. (2002). Permutation test for equality of correlation coefficients in two populations. *Commun. Statist.- Simula.* 31(4), 641-651.

Shipley, B. (2000). A Permutation procedure for testing the equality of pattern hypotheses across groups involving correlation or covariance matrices. *Statistics and Computing* 10, 253-257.

Smith, H., Gnanadesikan, R., Hughes, J.B. (1962). Multivariate analysis of variance (MANOVA). *Biometrics*. 18(1), 22-41.

Ufuk EKİZ, 1970 yılında Ankara'da doğdu. 1993 yılında Fen Edebiyat Fakültesi İstatistik Bölümü'nden mezun oldu. Yüksek Lisansını Haziran 1997, Doktorasını Mayıs 2003'te Gazi Üniversitesi Fen Bilimleri Enstitüsü'nde yaptı. 1994 yılından itibaren Gazi Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümünde görev yapmaktadır.

Meltem EKİZ, 1971 yılında Ankara'da doğdu. 1993 yılında Fen Edebiyat Fakültesi İstatistik Bölümü'nden mezun oldu. Yüksek Lisansını Kasım 1997, Doktorasını Temmuz 2005'te Gazi Üniversitesi Fen Bilimleri Enstitüsü'nde yaptı. 1994 yılından itibaren Gazi Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümünde görev yapmaktadır.