

MONOTON TREND İÇİN ÇOKDEĞİŞKENLİ PARAMETRE DIŞI BİR TEST VE UYGULAMA

A NONPARAMETRIC MULTIVARIATE TEST FOR MONOTONE TREND

Dr.Müslim EKNİ, Dr.Hamza GAMGAM

Gazi Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü
Gazi Tıp Dergisi 2 : 19-24, 1991

ÖZET : Bir ilacın belirli bir hastaya uygulanmasında değişik ölçümlerin alınması, çoğu zaman belirli zaman aralıklarında olur. Bu tür gözlemlerde "normal değerler (normal range)" yönteminin kullanılmasının bazı durumlarda sakıncalı olduğu iddia edilmektedir. Bir ilaç deneyi süresince, kanı oluşturan belirli bileşenlerdeki değişimler için Kendall'ın γ sına ve Spearman'ın r_s sıra korelasyonuna dayalı olarak monoton trend için yeni bir çokdeğişkenli test önerilmektedir. Bu testin özelliği gerçekçi olmayan varsayımlardan ve normal değerler yönteminin sakıncalarından arınmış olmasıdır.

Anahtar Kelimeler : Normal Değerler, Parametre Dışı İstatistiksel Testler, Trend.

SUMMARY : Upon application of a drug to a specific patient, measurements are usually made at regular intervals. For such observations it is said that using normal range method has disadvantages in some cases. For any drug trial a new multivariate test has been recommended for the monotone trend based on Kendall's γ and Spearman's r_s rank correlation to determine the changes in specific components of the blood. Not having any unrealistic assumptions or disadvantages of the normal range method are the important features of this test.

Key Words : Normal Range, Nonparametric Statistical Test, Trend.

GİRİŞ

Bir ilacın uygulanmasında her hastanın durumu belirli zaman aralıklarında kan testleriyle kontrol edilebilir. Bu ölçüm sonuçları, ilgili hastadan alınan kan ölçümlerini sağlıklı bir kişide bulunması gereken değerlerle karşılaştıran normal değerler yönteminde kullanılabilir. Sağlıklı kişiler yığımından alınmış "n" birimlik bir örneğin ortalaması \bar{x} ve standart sapması s ise normal değerler $\bar{x} \pm ks$ arasında olacaktır. k ise belirli bir anlamlılık düzeyinde student t dağılımından bulunan değerdir. Sağlıklı kişilerin oluşturduğu yığında kanı

oluşturan değişkenlerden her birinin normal dağılıma sahip olması durumunda k değerleri kolayca bulunabilir. Genel olarak $n > 30$ olduğu zaman 0.95 güven katsayısında k değeri yaklaşık olarak 2 alınabilir. Her ilaç deneyinde, kanı oluşturan bileşenlerin her ölçüm zamanında hasta üzerinden elde edilen ölçüm değerleri normal değerlerle mukayese edilir.

Normal değerler kavramı bu tür vakalar için uygulamada eleştirilere maruz kalmaktadır. Sağlıklı kişiler tanımı ve sağlıklı kişiler yığımından yeterli kadar kan ölçümünün alınmasındaki zorluk eleştiri-

$$x = \begin{bmatrix} \overline{X_{11}} & X_{12} & \dots & \overline{X_{1p}} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \overline{X_{n1}} & X_{n2} & \dots & \overline{X_{np}} \end{bmatrix} \text{ olsun.}$$

Burada 1, 2, ..., n zaman aralıklarında gözlenmiş p tane sürekli değişken dizisi bulunmaktadır. Çokdeğişkenli durum için H_0 ve H_1 hipotezleri aşağıdaki gibi olsun.

H_0 : P tane değişkenin herbirinin gözlem değerlerinin sıralanması tesadüfidir.

H_1 : P tane değişkenin herbirinin gözlem değerlerinin sıralanması tesadüfi değildir.

H_1 hipotezi P tane değişkenden en az biri için monoton trend olduğunu ifade eder. Tek değişkenli teste olduğu gibi, çokdeğişkenli test gözlemlerin sıra sayılarına dayanır. x'in sıralamaya dayalı R matrisi şöyledir :

$$R = \begin{bmatrix} \overline{R_{11}} & R_{12} & \dots & \overline{R_{1p}} \\ R_{21} & R_{22} & \dots & R_{2p} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \overline{R_{n1}} & R_{n2} & \dots & \overline{R_{np}} \end{bmatrix}$$

X matrisinin her sütunundaki gözlem değerleri, büyüklüklerine göre 1, 2, ..., n sıra sayıları verilerek R matrisi elde edilir. Bu nedenle, bu matrisin her sütunu 1, 2, ..., n'in bir permütasyonudur.

Çokdeğişkenli test istatistiği, daha önce tanımladığımız K_x istatistiğinin bir fonksiyonudur. Çokdeğişkenli test istatistiği;

$$K_g = \sum_{i < j} \text{sign}(x_{jg} - x_{ig}), \quad g = 1, \dots, p$$

olsun. H_0 hipotezinin doğruluğu altında herbir

$$\frac{K_g}{\sigma} \sim (0,1) \text{ dir.}$$

Burada $\sigma^2 = n(n-1)(2n+5)/18$ 'dir. K vektörü aşağıdaki gibi tanımlansın.

$$K = (K_1, K_2, \dots, K_p)^T$$

n verilmişken P vektör ($R_{i1}, R_{i2}, \dots, R_{ip}$), $i = 1, \dots, n$ olsun. Şayet H_0 doğru ise R matrisinin satır itibarıyla sıralanması tamamen tesadüfidir.

Bu sıralanma n! kadar farklı durumda ve eşit olasılıkla ortaya çıkar.

$\sigma_{gh} = E(K_g K_h / R$ nin herhangi bir satır vektörü) ifadesi K_g ve K_h arasındaki şartlı kovaryansı gösterir. j # g, h için K_g ve K_h , R_{ij} 'den bağımsız olduğundan

$$\sigma_{gh} = E(K_g K_h / R_{ig}, R_{ih}), \quad i = 1, \dots, n$$

olur. Bu beklenen değer zaman periyodundan bağımsız olarak R_{ig} ile R_{ih} eşleşmesine bağlıdır.

Teorem 1 : K_g ve K_h arasındaki şartlı kovaryans σ_{gh} aşağıdaki gibi tanımlanmıştır (Jasquelin ve Timoty, 1981).

$$\sigma_{gh} = K_{gh} / 3 + (n^3 - n) r_{gh} / 9$$

K_{gh} : g. ve h. koordinat için yani $X_{1g}, X_{2g}, \dots, X_{ng}$ ve $X_{1h}, X_{2h}, \dots, X_{nh}$ için Kendall istatistiğidir.

r_{gh} : g. ve h. koordinat için yani $X_{1g}, X_{2g}, \dots, X_{ng}$ ve $X_{1h}, X_{2h}, \dots, X_{nh}$ için Spearman istatistiğidir.

Bu durumda $\gamma_{gh} = 2K_{gh} / n(n-1)$ ve $\infty = 3/(2n+5)$ olmak üzere, K_g / σ ve K_h / σ arasındaki şartlı kovaryans;

$$\frac{\sigma_{gh}}{\sigma^2} = \infty \gamma_{gh} + (1 - \infty) r_{gh}$$

olur (Jasquelin ve Timorty, 1981). Ayrıca σ_{gh} / σ^2 değeri r_{gh} 'a yaklaşır. r_{gh} değeri de P_{gh} 'a yaklaştığından, şartlı varyans kovaryans matrisi olan

$\Sigma = \text{Kov}(K / \sigma / R$ nin herhangi bir satır vektörü)

için aşağıdaki ifade yazılabilir (Jasquelin ve Timoty, 1981).

$$\Sigma = \begin{bmatrix} 1 & P_{12} & \dots & P_{1p} \\ P_{12} & 1 & \dots & P_{2p} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ P_{1p} & P_{2p} & \dots & 1 \end{bmatrix} = \lim_{n \rightarrow \infty} \text{Kov} (K/\sigma)$$

Bu ise şartsız kovaryans matrisidir.

Teorem 2 : H_0 hipotezinin doğruluğu altında K/σ , p değişkenli normal dağılıma sahiptir. Σ tam ranklı ise $K^T \Sigma^{-1} K/\sigma^2$ karesel formu p serbestlik derecesinde bir π^2 dağılımına sahiptir. Eğer Σ 'nın rankı q ($q < p$) ise, yani Σ tam ranklı değilse, $K^T \Sigma^{-1} K/\sigma^2$ karesel formu, Σ^- matrisi Σ 'nın bir genelleştirilmiş tersi olmak üzere, q serbestlik derecesinde π^2 dağılımına sahiptir (Jasquelin ve Timoty, 1981).

Sonuç olarak H_0 için tanımlanmış karesel form, dağılımdan bağımsız bir test istatistiğine yaklaşır.

AYNI DEĞERLİ GÖZLEMLER

X matrisinin g . sütununda aynı değerli gözlemler varsa, bunlara ortalama sıra sayısı verilir. Bu durumda H_0 doğru iken

$$\sigma_g^2 = [n(n-1) (2n+5) - \sum_{j=1}^m t_j (t_j - 1) (2t_j + 5)] / 18$$

K_g istatistiğinin varyansı

olacaktır (Hollander ve Wolfe, 1973). Burada m ve t_j aşağıdaki gibi tanımlanır:

m : X matrisinin g . sütunundaki aynı değerli gözlem gruplarının sayısı.

$\frac{K}{\sigma} \sim N(0, \Sigma)$ olarak yazılabilir. Bu durumda

t_j : aynı değerli gözlemlerin j . grubundaki gözlem sayısı.

$K \sim N(0, \sigma^2 \Sigma)$ olur. Dolayısıyla

$$K^T (\sigma^2 \Sigma)^{-1} K \sim \pi_p^2 \text{ dir.}$$

Bunun sonucunda $\sigma^2 \Sigma = S$ olarak alınabilir.

$$S = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1p} \\ S_{12} & S_{22} & \dots & S_{2p} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ S_{1p} & S_{2p} & \dots & S_{pp} \end{bmatrix}$$

S matrisi tam ranklı ise bu durumda $K^T S^{-1} K$ asimptotik olarak p serbestlik dereceli π^2 dağılımına uyar (Jasquelin ve Timoty, 1981).

UYGULAMA

Bir hastanın kanını oluşturan bilirubin, BUN (üre), Kreatinin ve Alkalen Fosfataz bileşenlerinin iki yıllık periyot içinde ve yaklaşık birer aylık zaman aralıklarında elde edilen ölçüm değerleri aşağıda verilmiştir (Jasquelin ve Timoty, 1981).

H_0 : Kanı oluşturan dört bileşene ait gözlem değerlerinin sıralanması tamamen tesadüfidir.

H_1 : Kanı oluşturan dört bileşenden en az birinde gözlem değerlerinin sıralanması tesadüfi değildir.

Burada $n = 24$ ve $p = 4$ 'dür. Yapılan hesaplamalardan

$$K = [17 \ 80 \ -43 \ -112]^T$$

olarak bulunmuştur. Her değişkende aynı değerli gözlemler olduğundan, ilgili formüller kullanılarak varyans ve kovaryanslar hesaplanmış ve K istatistiğinin şartlı varyans - kovaryans matrisi

$$S = \begin{bmatrix} 1319.667 & -371.333 & 226.000 & 65.667 \\ & 1563.333 & -181.667 & -623.333 \\ & & 1403.667 & 274.000 \\ & & & 1616.667 \end{bmatrix}$$

olarak elde edilir. Test istatistiğinin değeri,

Ölçüm Tarihi	Bilirubin	BUM (Üre)	Kreatinin	Alkalen Fosfataz
8.1.1974	.4	13	.7	96
5.2.1974	.4	14	.7	89
19.2.1974	.3	15	.9	67
12.3.1974	.4	13	.9	83
9.4.1974	.6	12	.9	78
14.5.1974	.3	14	.8	70
11.6.1974	.4	15	.8	65
9.7.1974	.3	16	.7	70
13.8.1974	.3	16	.8	67
10.9.1974	.4	16	.9	64
8.10.1974	.3	14	.9	74
12.11.1974	.3	15	.8	81
10.12.1974	.4	17	.9	86
14.1.1975	.5	12	.8	69
11.2.1975	.4	13	.8	65
11.3.1975	.4	16	.7	63
8.4.1975	.3	17	.5	69
13.5.1975	.4	20	.8	54
10.6.1975	.3	16	.8	65
8.7.1975	.3	15	.7	62
12.8.1975	.4	14	.7	63
9.9.1975	.4	16	.8	64
14.10.1975	.5	17	.8	68
11.11.1975	4	14	.8	72

Normal değerler bilirubin için 0-1.5, Bun (Üre) için 8-26, kreatinin için .8-1.2, ve Alkalen Fosfataz için 35-148'dir.

$$K^T S^{-1} K = [17 \ 80 \ -43 \ -112] \begin{bmatrix} 1319.667 & -371.333 & 226.000 & 65.667 \\ & 1563.333 & -181.667 & -623.333 \\ & & 1403.667 & 274.000 \\ & & & 1616.667 \end{bmatrix}^{-1} \begin{bmatrix} 17 \\ 80 \\ -43 \\ -112 \end{bmatrix}$$

$$= 21.082$$

olarak hesaplanır. π^2 dağılım tablosundan $P(\pi^2_4 > 21.082) < 0.001$ dir. Yani $P < 0.001$ olduğu kolaylıkla görülebilir. Bu durumda H_0 hipotezi red edileceğinden dört değişkenden en az birinde, gözlem değerlerinin sıralanmasının tesadüfi olmadığı, yani bir trendin etkisinin olduğu söylenebilir.

SONUÇ

Bu çalışmada, kanı oluşturan bazı bileşenler üzerinde belirli zaman aralıklarında bir ilacın etkisinin incelenmesinde normal değerler yönteminin her durumda kullanılmasının uygun olmadığı, bu tür çalışmalarda, normal dağılım ve sağlıklı kişiler yığını varsayımına gerek duymayan monoton trend için, çokdeğişkenli parametre dışı test yönteminin kullanılmasının daha uygun olacağı vurgulanmaya çalışıldı.

Yazışma Adresi : Dr.Müslim EKNİ
Gazi Üniversitesi Fen Edebiyat Fakültesi
İstatistik Bölümü
06500 Teknik Okullar
ANKARA - TÜRKİYE
Tel : 4 - 212 60 30 / 30

KAYNAKLAR

1. Elveback LR, Guillier CL and Keating FR JR : Healty, Normality, and the Ghost of Gauss" J. Amer. Med. Assoc. 211 : 69-75, 1970
2. Gibbons JD : "Nonparametric Statistical Inference, New-York, McGraw-Hill. 1971, pp. 237-238
3. Healy MJR "Normal Values From a Statistical Viewpoint". Bulletin de L'Academie Royale de Medecine de Belgique, 9 : 708-718, 1969
4. Hollander M and Wolfe DA : "Nonparametric Statistical Methods", NewYork, John Wiley. 1973, pp. 187-188
5. Jasquelin D and Timoty JK : A Nonparametric Mutlivariate Test for Monotone Trend with Pharmaceutical Applications" J. Amer. Statis. Assoc. 76 : 169-174, 1981
6. Kendall MG : "Rank Correlation Methods", London Charles W. Griffen. 1970, pp. 3-8
7. Lehmann EL : "Nonparametrics : Statistical Methods Based on Ranks" San Francisco, Holdey - Day. 1975, pp. 300-316
8. Mann HB : "Nonparametric Tests Againts Trend" Econometrica 13 : 245-259, 1945
9. Schoen I and Brooks SH : Judgment Based on 95 % Confidence Limits : A Statistical Dilemma Involving Multitest Screening and Proficiency Testing of Multiple Specimens", Am. J. Clin. Pathol. 53 : 190-193, 1970
10. Winkel P, Lyngbye J and Jörgensen K : "The Normal Region-A Multivariate Problem, "Scand. J. Clin. Lab. Invest." 30 : 339-344, 1972