İnönü Üniversitesi Tıp Fakültesi Dergisi
14(4) 251-255 (2007)

Yaşam Verisi İçin Homojenlik Analizi

Nihal Ata*

*Hacettepe Üniversitesi Fen Fakültesi, İstatistik Bölümü, Ankara

Amaç: Bu çalışma çok değişkenli istatistiksel yöntemlerden homojenlik analizinin yaşam verilerinde kullanımını göstermek amacıyla yapılmıştır.
Gereç ve Yöntem: Değişken sayısının ikiden çok olması durumunda hastalara ait özellikler arasındaki ilişkiler incelenirken ki-kare analizi homojenlik analizi kullanılmıştır. Hastalığı nüks eden akciğer kanseri hastalarına ait özellikler belirlenmeye çalışılmıştır.
Bulgular: Yaşı 59'dan küçük, sigara tüketimi 30 paket yıldan az olan, patolojik evresi I olan, tümörünün boyutu 40 mm'den az olan bireylerin hastalıkları nüks etmez iken; 60 yaşından büyük, sigara tüketimi 31 paket yıl ve üzeri olan, patalojik evresi II, III ve IV olan, tümörünün boyutu 41 mm ve daha fazla olan bireylerin hastalıklarının nüks ettiği sonucuna ulaşabiliriz.
Sonuç :Homojenlik analizi kullanılarak bireylerin özellikleri hakkında yorumlar yapılabileceği gösterilmiştir.

Anahtar Kelimeler : Akciğer kanseri, Homojenlik analizi, Yaşam verisi.

Homogeneity Analysis For Survival Data

Objective: This study was designed to demonstrate that one of the multivariate statistical methods, homogeneity analysis, can be used for survival data.
Material and Method: If there are more than two variables, the relation between the the characteristics of patients can be examined by homogeneity analysis instead of chi-square analysis. The characteristics of lung cancer patients whose disease had relapsed has been tried to be determined.
Results: Disease of individuals who are younger than 59, have cigarette consumption less then 30 package year, have pathological stage I and have a tumour whose dimension is smaller than 40mm does not relapse whereas disease of individuals who are older than 60, have cigarette consumption higher then 31 package year, have pathological stage II, III or IV and have a tumour whose dimension is higher than 41mm relapse.
Conclusion: It is concluded that interpretations about the survival of individuals can be done by homogeneity analysis.

Key Words: Lung cancer, Homogeneity analysis, Survival data.

İstatistiksel analizlerde değişkenler arasında ilişkilerin olup olmadığı ve ilişkinin olması durumunda ise bunların yorumlanması oldukça önemlidir. İki değişken arasındaki ilişki incelendiğinde bu değişkenlerden birinin ya da her ikisinin birden diğer değişkenlerle ilişkili olabileceği tamamen göz ardı edilmektedir. Elimizde iki nitel değişken olduğu durumlarda ki-kare çözümlemesine alternatif olan ve daha güvenilir olduğu düşünülen uygunluk analizi uygulanabilir. Böylece değişkenler ve düzeyleri arasındaki ilişkiler iki boyutlu bir tabloda görsel olarak ifade edilebilmekte ve daha kolay bir biçimde yorumlanabilmektedir. Ancak ikiden fazla nitel değişken olması durumunda ise optimal ölçekleme tekniklerinden homojenlik analizi kullanılabilmektedir. [1]

Çalışmada günümüzde sık görülen kanser türlerinden biri olan akciğer kanseri hastalığına yakalanan hastalara ait yaşam verileri incelenerek, homojenlik analizinin yaşam verileri için kullanımının gösterilmesi amaçlanmıştır.

GEREÇ VE YÖNTEM

Homojenlik analizi doğrusal olmayan çok değişkenli istatistiksel yöntemlerden biridir. Analizde, sayısal olmayan çok değişkenli veri yapısını göstermek amaçlanmaktadır. Nesnelere ve değişkenlerin kategorilerine skorlar atanmaktadır.

Bu skorlar ise verideki bağımlılıkların geometrik bir gösterimini oluşturmak için kullanılmaktadır.[4]

k_j kategorili J tane kategorik değişkene sahip N nesne (birey, ürün, ülke, vb.) alınsın. G_j, j Є J olmak üzere j değişkene karşılık gelen N x k_j gösterge matrisi olsun. i=1,…,N, t=1,…,k_j olmak üzere gösterge matrisi elemanları, i.nesne t kategorine ait ise g_{it}=1, diğer kategorilere ait ise $g_{it} = 0$ olan birim matristir. Homojenlik ilkelerine göre, maksimum homojenliği elde etmek için değişkenlere dönüşüm uygulanmaktadır.[4]

Y_j çoklu kategori niceliklerini (multiple category quantifications) içeren k_j x p matrisi olsun. X ise p tane optimal ölçekleri (resulting p optimal scales) içeren N x p matris olsun. X matrisinin elemanları aynı zamanda nesne skorları olarak adlandırılmaktadır. Genelde, mükemmel bir sonuç bulmak mümkün değildir, tam homojenliği sergileyen Y_j'ler ve X belirlenmelidir. Bu nedenle, Eşitlik (1)'de verilen Gifi kayıp fonksiyonu kullanılarak ölçülen tam homojenlikten ayrılışları minimize etmek gerekir;

$$\sigma(X; Y_1,..., Y_j) = J^{-1}\sum_{j=1}^{J} SSQ(X - G_j Y_j) \quad (1)$$

Burada $SSQ(H) = tr(H'H)$ H matrisinin Frobenius normudur (H matrisinin elemanlarının kareler toplamıdır). Her j Є J için X=0, and Y_j=0'a karşılık gelen açık çözümden kaçınmak için, aşağıdakilere verilen normalizasyon kısıtlamaları yapılır;

$$X'X = NI_p \quad (2)$$

$$u'X = 0 . (3)$$

Burada I, pxp boyutlu birim matrisi ve u tüm elemanları bir olan bir sütun vektörüdür. Minimizasyon problemlerinin çözümü dalgalı en küçük kareler (Alternating Least Squares, ALS) algortiması ile bulunur. Algoritmanın birinci adımında, Eşitlik (1) sabit X için Y_j e göre minimize edilir ve Eşitlik (4)'teki sonuç elde edilir;

$$\hat{Y}_j = D_j^{-1} G_j' X , \quad j \in J . (4)$$

Burada $D_j = G_j' G_j$ j değişkenin tek değişken marjinallerini içeren $k_j \times k_j$ köşegen matrisidir. Algoritmanın ikinci adımında, Eşitlik (1) sabit Y_j'ler için X e göre minimize edilir ve sonuç Eşitlik (5) ile verilir;

$$\hat{X} = J^{-1}\sum_{j=1}^{J} G_j Y_j . (5)$$

Algoritmanın üçüncü adımında ise X matrisi sütunlarda bir araya getirilmiş ve Gram-Schmidt süreci ile birimdikeyleştirilmiştir. Böylece normalleştirme kısıtları Eşitlik (2) ve Eşitlik (3) ile sağlanmış olur. Bu adımlar, algoritma global minimuma yakınsayana kadar tekrar edilir. ALS algoritması, Eşitlik (1) ile verilen probleme istenilen çözümü bulmaktadır. Bu çözüm literatürde HOMALS (dalgalı en küçük kareler yöntemi ile homojenlik analizi, homogeneity analysis by means of alternating least squares) çözümü olarak bilinmektedir.[2]

ALS algoritması $\hat{Y}_j' D_j \hat{Y}_j = \hat{Y}_j' D_j (D_j^{-1} G_j' \hat{X}) = \hat{Y}_j G_j' \hat{X}$ durumu kullanılarak yakınsarsa, Gifi kayıp fonksiyonu Eşitlik (6)'daki gibi yazılabilir:

$$J^{-1}\sum_{j=1}^{J} tr(X - G_j Y_j)'(X - G_j Y_j) =$$

$$Np - J^{-1}\sum_{j=1}^{J} tr(\hat{Y}_j' D_j \hat{Y}_j) . (6)$$

$Y_j' D_j Y_j$ matrisinin köşegen elemanlarının toplamı çözüme uygunluğu göstermektedir. s boyuttaki j değişkenlerinin ayrışım ölçüleri (discrimination measures) ise aşağıdaki gibi verilir:

$$\eta_{js}^2 \equiv Y_j'(s) D_j Y_j(s) / N, \; j \in J, s = 1,..., p . (7)$$

Burada, $Y_j(s)$ Y_j matrisinin s. sütununu göstermektedir ve çözümün s. boyutunda j değişkeni için nicelleştirmeyi belirtir.[3]

Geometrik olarak ayrışım ölçümleri, *p*-boyutlu uzayın orijinine kategori nicelleştirmelerinin (marjinal frekanslarla ağırlıklandırılan) ortalama kareleri alınmış uzaklığını verir. Bununla birlikte, ayrışım ölçümlerinin, (kayıp veri olmadığını varsayarak) optimal olarak nicelleştirilen değişken $G_j Y_j(s)$ ve nesne skorları sütununa karşı gelen X(s) arasındaki kareleri alınmış korelasyona eşit olduğu gösterilebilir. Böylece, kayıp fonksiyonu aşağıdaki gibi yazılabilir;

$$n\left(p - \frac{1}{m}\sum_j \sum_s \eta_{js}^2\right) = n\left(p - \sum_s \gamma_s\right) (8)$$

Burada $\gamma_s = m^{-1}\sum_j n_{js}^2$ özdeğerler olarak adlandırılır ve ayrışım ölçümlerinin ortalamasına karşı gelir ve analizde elde edilen özdeğerler, *p*-boyutsallığın her birinde türetilen çözümün uyumunun tam bir ölçümünü verir.[5]

HOMALS çözümünün bazı temel özellikleri aşağıdaki gibi verilebilir:

- Kategori niceleştirmeleri ve nesne skorları ortak bir uzayda noktalar olarak sunulur.
- Bir kategori noktası bu kategoriye ait olan nesnelerin merkezidir.
- Aynı cevap örüntüsü ile nesneler, özdeş nesne skorlarını kabul eder. Genelde, iki nesne arasındaki uzaklık onların profilleri arasındaki "benzerlik" ile ilişkilidir.
- Düşük marjinal frekanslara sahip kategori noktaları ortak uzayın orijininden daha uzakta, yüksek marjinal frekanslara sahip kategori noktaları ise orijine daha yakın yer alır.
- Kategori tek bir nesneye tek olarak uygulanırsa, nesne noktası ve kategori noktası çakışacaktır.[2-5]

UYGULAMA

Uygulamada, 236 akciğer kanseri hastasına ait bilgiler kullanılmıştır. Hastalar, ameliyat olduktan sonra hastalıklarının ilk nüks etmesine kadar geçen süre (min=1 ay, max=93 ay) boyunca izlenmiştir. "Hastalığın nüks etmesi" sonuç değişkeni ve yaş (YŞ), sigara tüketimi (paket yıl olarak, ST), tümörün boyutu (mm olarak, BY) ve patalojik evre (E) açıklayıcı değişkenler olarak ele alınmıştır. Bu değişkenler ve düzeylerine ait bilgiler Tablo 1'de verilmiştir. Hastaların izlenme süresi sona erdiğinde 236 hastadan 94'ünde (%39.8) hastalığın nüks ettiği (başarısızlık) ve 142'sinde (%60.2) ise hastalığın nüks etmediği (durdurma) gözlenmiştir.

Değişkenler arasındaki ilişki homojenlik analizi kullanılarak ortaya konulmaya çalışılmıştır. Bunun için, parantez içinde kategori sayıları olmak üzere, hastalığın nüks etmesi (2), yaş(4), sigara tüketimi (4), tümörün boyutu (4) ve patalojik evre (4) değişkenleri iki boyutlu bir grafikte, kategorileri kombinasyonlarının nasıl olacağını görmek amacıyla homojenlik analizi uygulanmış ve hastalara ait 2x5x4x4x4 çok yönlü tablonun analizi sonucunda her bir değişkenin ve her bir boyutun ayrışım ölçüleri Tablo 2'de verilmiştir. Bununla birlikte analiz sonucunda elde edilen özdeğerler ise λ_1= 0.2900 ve λ_2 =0.2828'dir. Homojenlik analizinde özdeğerler, gerçek grafik ile elde edilen iki boyutlu grafik arasındaki uyumunun tam bir ölçümünü vermektedir. Bu doğrultuda, gerçek grafik ile elde edilen iki boyutlu grafik arasındaki uyumun 0.5728 iyi olduğu söylenebilir.

Tablo 2. Herbir değişken ve herbir boyut için ayrışım ölçüleri

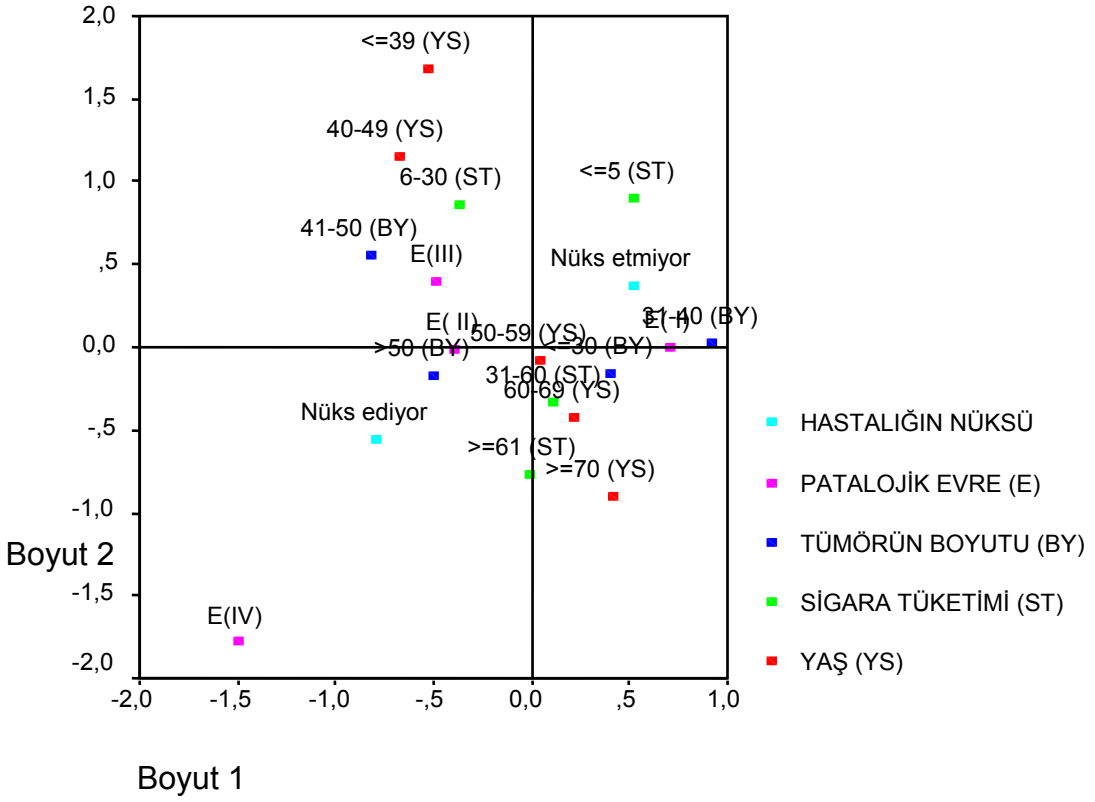
Değişken	Boyut 1	Boyut 2
Hastalığın Nüksü	0.411	0.204
Yaş	0.126	0.529
Sigara Tüketimi	0.060	0.396
Tümörün Boyutu	0.413	0.070
Patalojik Evre	0.439	0.215

Ayrışım ölçüleri kareleri alınmış korelasyonlardır. Tablo 2, incelendiğinde yaş ve sigara tüketimi değişkenlerinin Boyut 2 tarafından, tümörün boyutu, patalojik evre ve hastalığın nüksü değişkenlerinin ise Boyut 1 tarafından daha iyi açıklanabildiği sonucuna ulaşılmıştır. Ayrıca analiz sonucunda elde edilen grafik Şekil 1 de verilmiştir.

Tablo 1. Kullanılan değişkenler ve düzeyleri

Değişken	Değişken Düzeyleri	Toplam Olay Sayısı		Başarısız Olay Sayısı	Durdurulmuş Olay Sayısı
		n	%		
Yaş	<=39	13	5.5	5	8
	40-49	38	16.1	12	26
	50-59	76	32.2	33	43
	60-69	81	34.3	32	49
	>=70	28	11.9	12	16
Sigara Tüketimi	<=5	16	6.8	4	12
	6-30	62	26.3	23	39
	31-60	123	52.1	48	75
	>=61	35	14.8	19	16
Tümör Boyutu	<=30	73	30.9	25	48
	31-40	46	19.5	12	34
	41-50	41	17.4	18	23
	>=50	76	32.2	39	37
Patolojik Evre	Evre I	102	43.2	28	74
	Evre II	61	25.8	24	37
	Evre III	60	25.4	30	30
	Evre IV	13	5.5	12	1

Şekil 1. Kategori Niceleştirmesi



Şekil 1 incelendiğinde yaşı 59'dan küçük, sigara tüketimi 30 paket yıldan az olan, patolojik evresi I olan, tümörünün boyutu 40 mm'den az olan hastaların hastalıklarının nüks etmediği sonucuna ulaşabiliriz. Hastalığı nüks etmeyenler içinde en az riskli durumda tümörünün boyutu 30 mm'den az olan, sigara tüketimi 5 paket yıldan az olan, patalojik evresi I olan, 49 yaşından genç hastalar olmaktadır. 60 yaşından büyük, sigara tüketimi 31 paket yıl ve üzeri olan, patalojik evresi II, III ve IV olan, tümörünün boyutu 41 mm ve daha fazla olan hastaların hastalıklarının nüks ettiği sonucuna ulaşabiliriz. Hastalığı nüks edenlerden içinde en riskli durumda tümörünün boyutu 50 mm'den fazla olan, sigara tüketimi 61 paket yıldan fazla olan, patalojik evresi IV olan, yaşı 70'den fazla olan hastalar olmaktadır.

SONUÇ

Bu çalışmada değişkenler arasındaki ilişkiler ki-kare analizi yerine homojenlik analizi kullanılarak incelenmiş ve akciğer kanseri hastalarının hastalıklarının tekrar nüks etmesine neden olan faktörlere ilişkin yorumlar yapılmıştır. Tıbbi araştırma konularında da homojenlik analizinden yararlanılabileceği gösterilmeye çalışılmıştır.

Akciğer kanserini etkileyen değişkenler ve değişken düzeyleri elde edilebilecek veriler eşliğinde daha iyi belirlenebildiği takdirde tıbbi açıdan daha anlamlı sonuçlar ortaya konulacaktır.

TEŞEKKÜR

Çalışmada verilerini kullandığım Sayın Dr. Ayten Kayi Cangir'e teşekkür ederim.

KAYNAKLAR

1. Tatlıdil, H., Çok Değişkenli İstatistiksel Yöntemler Ders Notları, Hacettepe Üniversitesi, İstatistik Bölümü, 2005.
2. Michailidis G. and de Leeuw, J. Constrained Homogeneity Analysis With Applications To Hierarchical Data. Technical Report. UCLA Statistics Program, Preprint 207, 1997.
3. Michailidis G. and de Leeuw, J. The Gifi System of Descriptive Multivariate Analysis Technical Report, UCLA Statistics Program, Preprint 204, 1996.
4. Michailidis, G., Leeuw, J. Multilevel homogeneity analysis with differential weighting. Computational Statistics&Data Analysis. 2000; 32 : 411-2.
5. Aytaç, M., N. Bayram, "Çoklu Karşılık Getirme Analizi Ve Öğretim Elemanları Üzerinde Bir Uygulama", V. Ulusal Ekonometri ve İstatistik Sempozyumu Bildirileri, Adana, 19-22 Eylül 2001.

Yazışma Adresi :
Dr.Nihal Ata
Hacettepe Üniversitesi
Fen Fakültesi, İstatistik Bölümü
06800 Beytepe ANKARA
Tel : 312 299 2016 -121
Fax : 312 297 7913
E-posta : nihalata@hacettepe.edu.tr