ASİMETRİK VE SİMETRİK MARJİNAL DAĞILIMLARDA ÇOK DEĞİŞKENLİ NORMALLİK

T.C. Marmara Üniversitesi
I.I.B.F. Dergisi
YIL 2004, CILT XIX, SAYI 1

A. Mete Çilingirtürk
acilingi@marmara.edu.tr
Marmara Üniversitesi

Dilek Altaş
dilekaltas@marmara.edu.tr
Marmara Üniversitesi

ÖZET

Pek çok sosyal problemin çözümünde kullanılan istatistiksel analizlerin uygulanabilmesi, veri setinin çok değişkenli normal dağılıma uygunluğu varsayımının geçerli olmasına bağlıdır. Çok değişkenli normal dağılım, diğer dağılımlarla ilgili ayrıntılı bilgiye sahip olunmaması ve matematiksel yaklaşımların mümkün olmaması nedeniyle uygulaması en kolay olan bir dağılımdır. Ayrıca normal dağılım, anakütledeki dağılımların asimtotik yapısına uygun olması ve çok değişkenli istatistiklerin örneklem dağılımlarının merkezi limit teoreminden dolayı normallik göstermesi nedeniyle en çok tercih edilen dağılımdır. Gerçek dünyadaki pek çok problemin incelenmesinde normal dağılım varsayımlarının kullanılması oldukça tutarlı bir yaklaşım olup, bu dağılım, çok değişkenli istatistiksel analizlerin uygulanmasında özel bir öneme sahiptir.

Çok değişkenli normalliğin sınanmasında farklı yöntemler geliştirilmiş olmakla birlikte, Mardia'nın (1970) çoklu asimetri ve basıklık ölçülerine göre ileri sürdüğü test istatistiği, halen en sık kullanılan ve güçlü yöntemdir. Sosyal bilimlerde yapılan araştırmalarda anketler birincil kaynak olarak kullanılmakta ve kişilerin duygu, düşünce, davranış ve algılamaları sıklıkla likert ölçeğinde ölçülmektedir. Likert ölçeğinin kullanılmasının nedeni, elde edilen verilerin aralıklı ölçek varsayılabilmesi ve dolayısıyla nicel veri analizi yöntemlerinin kullanılmasına imkan sağlamasıdır.

Çalışmanın amacı, çok değişkenli normal dağılımın sınanabilmesi için uygun bir örnek hacmini belirlemektir. Bu nedenle binom ve normal dağılıma uygun bağımsız değişkenler türetilerek oluşturulan veri setlerine farklı örnek hacimleri için (n=30,65,100,.....,220) Mardia testi uygulanmıştır. Elde edilen test istatistikleri ve kuyruk olasılıkları karşılaştırılarak yorumlanmıştır.

Anahtar kelimeler ve deyimler: *Çoklu normallik, Mardia testi, Likert ölçeği*

1. GİRİŞ

Sosyal bilimlerde ölçmeye ve araştırmalara konu olan ve bu nedenle gereğince ölçülmesi gereken değişkenlerden biri olan tutum, belirli nesne, durum, kurum, kavram veya diğer insanlara karşı olumlu veya olumsuz tepkide bulunma eğilimidir (Tezbaşaran, 1997).

Tutumların ölçülmesinde kullanılan en önemli yaklaşım, söz konusu tutuma ilişkin bir ölçeğin hazırlanarak uygulanmasıdır. Bir tutum ölçeği ölçülmek istenen tutum konusu ile ilgili bir dizi ifadeyi içermekte ve kullanılan ölçekle ilgili süreklilik, tek boyutluluk ve doğrusallık gibi bazı temel özelliklerin sağlanması gerekmektedir (Tezbaşaran,1997). Tek boyutlu ölçeklemeden başlayarak çok boyutlu ölçeklemeye kadar çeşitli yöntemler geliştirilmiş olup, bu tekniklerden daha ekonomik olması nedeniyle en yaygın olarak kullanılanı Likert'in (1932) modelidir. Likert bu çalışmasında, ölçülmek istenen tutumla ilişkili çok sayıda olumlu ve olumsuz ifadenin çok sayıda cevaplayıcıya uygulandığını, ifadelerin 3,5 veya 7 seçenekli olduğunu ve her bir ifadenin oransal olarak anlamlılığının istatistiksel analizlerle yapılabileceğini belirtmiştir (Biographical Dictionary of Management). Likert ölçeğinden elde edilen puanlar sıralama ölçeği tipindedir. Bu nedenle bu puanları kullanarak bireyler arasındaki tutum farklılıklarını ortaya çıkarmak zor olduğundan, sıralama yoluyla elde edilen sıralama ölçeği tipindeki puanlar, aralıklı ölçek tipinde bilgi veren puanlara dönüştürülebilir (Clason, Dormody, 1994).

Çok değişkenli normallik varsayımı pek çok istatistiksel analizlerin yapılabilmesi için gerekli en önemli varsayımlardan biridir. Çoklu normalliğin sağlanmasını gerektiren yöntemler kısaca sıralanmışlardır. Likert'in ölçek tanımında, madde puanları sürekli değişken olduğundan madde puanları ile ölçek puanları arasındaki korelasyon, Pearson Korelasyon katsayısı ile hesaplanmalıdır (Tezbaşaran, 1997). Pearson korelasyon katsayısı iki boyutlu normal dağılım varsayımı gerektirir. Likert ölçek tipinde üst gruptaki cevaplayıcıların madde puanları ile alt gruptaki cevaplayıcıların madde puanları ortalaması arasındaki farkın anlamlılığı t testi ile sınanır. Çok değişkenli hipotez testlerinde, örneklemler parametreleri normal dağılımlı bir anakütleden çekilmiştir (Tatlıdil, 2002). Normal dağılıma sahip olmayan bir örnekte örnek hacminin Hotelling T^2 istatistiği üzerindeki etkisi incelenmiş, bu istatistiğin asimetriliğe duyarlı olduğu belirlenmiştir (Mardia, 1970; Srivastava, Mudholkar,2001). Homoskedasite için kullanılan standart LR(benzerlik oranı) test istatistiği normallikten sapmalardan çok etkilenmektedir (Hawkins, 1981). Faktör analizinde, normallik varsayımı faktörlerin anlamlılığının sınanmasında kullanılan istatistiksel testler için gereklidir (Hair, Anderson, Tatham, Black, 1998). Kümeleme analizinde verilerin normal dağılımlı olması varsayımı olmakla birlikte normallik varsayımı prensipte kalmakta, sadece uzaklık değerlerinin normalliği yeterli görülmektedir (Tatlıdil, 2002). Diskriminant analizinde veri matrisinin normal dağılımlı olması varsayımlardan biridir (Hair, Anderson, Tatham, Black, 1998). Kanonik korelasyon analizinde, her bir kanonik fonksiyonun anlamlılığının testi için çok değişkenli normallik varsayımının sağlanması gereklidir (Hair, Anderson, Tatham, Black, 1998).

2. ÇOK DEĞİŞKENLİ NORMAL DAĞILIM SINAMALARI

Çok değişkenli eğiklik ve basıklık ölçüleri, t-istatistiğinin sapmasızlığının incelendiği bazı çalışmalar ile gelişmiştir. Pearson'un temellerine sahip tek değişkenli asimetri ve basıklık ölçülerinden yola çıkarak, Mardia (1970) çok değişkenli normalliği tanımlamış ve çok değişkenli asimetri ve basıklık ölçülerinin asimtotik dağılımlarına dayanan bir çoklu normal dağılım uygunluk testi geliştirmiştir. Hawkins (1981), Malkovich ve Afifi tarafından 1973'de tanımlanan $V_{ij}^{*}=(X_{ij}-X_i)'S_i^{-1}(X_{ij}-X_i)$ örneklem dağılım fonksiyonunun Hotelling T^2 fonksiyonuna ikamesinin binomial dönüştümünün (0, 1) aralığında düzgün dağıldığını göstermiştir. Bu şekilde elde edilen örnek dağılım fonksiyonu ile Anderson-Darling test istatistiğinin değişen varyans ve çoklu normallik üzerine yeterli bilgiler sağlayacağı ileri sürülmüştür. Mardia'nın (1974) önerdiği test kullanılarak 3 farklı dağılıma[1] sahip 5 ve 10 değişkenli 100 gözleme kadar sahip örnekler üzerinde sonuçlarını incelemiştir. Bu yöntemin avantajı, standart paket programlar yardımıyla çoklu normalliğin test edilebileceğidir. Machado (1983), Malkovich ve Afifi tarafından önerilen çok değişkenli normalliğe dayalı iki test istatistiğinin, 4 değişkene ve 50 gözleme kadar veriler için simulasyon ile dağılım özelliklerini incelemiş ve 25'in üzerindeki örnek hacimleri için yaklaşımların yerinde olduğunu belirtmiştir. Csörgö (1986) çalışmasında, Murota ve Takeuchi tarafından geliştirilen tek değişkenli normallik testini, deneysel değişkenlerin t dönüşümlerinin çok değişkenli karakteristik fonksiyonun asimtotik davranışları ile çok boyut için geliştirerek, çok değişkenli normal dağılıma uygunluk testi önermiştir. Testin simulasyon ile üst limitlerinin bulunması dışında iki ve dört değişkenli verilere[2] uygulayarak Mardia ve Rincon-Gallardo testleri ile aynı sonuçlara ulaşmışlardır. Cox ve Wermuth (1994) bağımlılık yapısı olan değişkenlerde doğrusal bağımlılığın analizinde kullanılan standart regresyon analizlerinde elde edilen katsayı testlerinin sıralı istatistiklerinin beklenen değerlerini, normal dağılımın sıralı istatistiklerinin beklenen değerleri ile karşılaştırarak, doğrusallıktan sapmaları belirlemişlerdir. Sayısal veriler üzerinde yaptıkları uygulama yanında sıralı ölçek değişkenlerde kutupsal ön kodlamanın doğrusallaştırılabileceğini göstermişlerdir. Çalışmalarında değişkenlerin medyana göre iki kutuplu değişkene dönüştürülmesi ile de doğrusal bağımlılığın kontenjans tablolarında yer alan frekanslar kullanılarak (MacFadden, 1955) belirlenebileceğini göstermişlerdir. Huffer ve Park (2002), benzer bir çalışma[3] ile çoklu normal dağılıma uygunluğun test edilebileceğine işaret etmişlerdir. Değişkenlerin birbirinden bağımsız aynı boyuta sahip transformasyonundan sonra, her bir değişkenin eşit gözlemli alt gruplara ayrılarak ortak dağılıma ait frekansların çok boyutlu kontenjans tablosunda toplanması sağlanmıştır. Ki-kare test istatistiğinde

---

[1] Kullanılan dağılımlar: standart normal dağılım, $U+0,1U^3$ $U\sim N(0,1)$ ve 0-1 aralığında sürekli düzgün dağılım.

[2] Yule ve Kandal'ın 1950'de kullandığı 780 gözlemli iskonto oranı ve reservlerin tasarrufa oranı verileri; Fisher tarafından 1936'da analiz edilen 50 gözlemli "iris setosa" ait dört değişkenli veri seti.

[3] Uygulamada 1986 Joint Statistical Meeting esnasında ortaya konulan 5 değişken 3848 gözlemden oluşan gizli yapı içeren yapay veri seti kullanılmıştır.

beklenen değerlerin hesabında referans olarak belli bir yapı sergilemeyecek olan çok boyutlu normal dağılım seçildiğinden sıfır hipotezinin reddi normal dağılımdan sapmayı sergileyecektir. Henze ve Wagner (1997), örnek hacminden bağımsız, tutarlı çok boyutlu asimetri ve basıklık ölçüleri ileri sürmüşlerdir. Gutjahr, Henze ve Folkers (1999), Mardia'nın örnek çok değişkenli asimetri ve basıklık ölçülerinin limit dağılımının normal dağılım altında belirlendiğini, bu nedenle Monte Carlo simulasyonları ile eliptik simetrik dağılımlarda hatalı kararlara yol açtığını göstermişlerdir. Baxter (1999), çalışmasında çoklu normal dağılım uygunluk testlerini dört başlıkta toplayarak bunların birkaçının denenmesinin uygun olacağını belirtmiştir. Beirlant, Mason ve Vynckier (1999), benzer bir çalışma yaparak testleri gruplandırmış, ve Hawkins çalışmasını geliştirerek yeni bir test önermiştir. Her iki çalışmada farklı testler veri setlerine uygulanarak sonuçları karşılaştırılmıştır. Hüsler, Liu ve Singh (2002), örnek hacmi ile ortalamaya göre maksimum Euclid uzaklığının, çok değişkenli normal dağılım ve büyüme oranı fonksiyonu ilişkisinden yola çıkarak çoklu normal dağılım kuyruk olasılığını tahmin etmişler ve normal dağılımdan sapmanın belirlenmesinde kullanılacak grafik yöntem önermişlerdir. Çalışmalarında iki ve on boyutlu normal dağılmış veri seti ile iki boyutlu üstel dağılmış veri seti kullanarak grafiksel aracı tanıtmışlardır. Olive (2003), Hawkins ile çalışmalarını geliştirerek uzaklıklara dayalı sapmasız bir test önermiştir. Klar (2002) ise çalışmasında Mardia'nın çok değişkenli sapmasız asimetri ve basıklık ölçülerine referans olarak "dağılım bağımsız" bir yaklaşım geliştirmiştir.

Bütün bu çalışmalar bazı ortak noktalarda benzerlik göstermektedirler. İlk olarak çoklu normal dağılıma uygunlukların testinde temel yaklaşımlar olduğu görülmüştür. Bunlar, simetri ve basıklık ölçülerine dayanan testler, çoklu normal dağılımın özelliklerine dayalı kutupsal veya kategorize adilmiş grup frekanslarına dayanan testler, en çok benzerlik fonksiyonuna dayanan testler, uzaklık ölçülerine dayanan yaklaşımlar ve bunların karma yöntemlerinden yola çıkan teknikler olmaktadırlar. Bir diğer ortak nokta; yazarların, istatistiklerin asimtotik veya limit dağılımlarına farklı yaklaşmaları, veya bu konuda daha az varsayım gerektiren, daha etkin ve sapmasız test istatistiğinin geliştirme çabalarıdır. Yöntemlerde ortak olan noktalardan diğeri değişkenler arasındaki ilişki yapısının elde edilecek sonuçlar üzerindeki ortak yaklaşımdır. Bu nedenle orijinal veri setleri genelde transformasyonlara tabi tutulmakta veya önceden temel bileşenler ile ilişki yapısının yok edilmesi önerilmektedir. Bütün farklılık ve benzerliklere rağmen bütün bu çalışmalarda Mardia'nın çalışmasına atıf yapılmakta, ve çok değişkenli asimetri ve basıklık ölçülerinin sapmasız tahmincisi olarak kabul edilmektedir.

3. YÖNTEM

Çalışmanın bu kısmında Mardia'nın önerdiği çok değişkenli asimetri ve basıklık ölçüleri matris cebiri ile tanıtılacaktır. Kullanılan notasyonda veri setinin p adet değişken ve n adet gözlemden oluştuğu kabul edilmiştir. Kullanılan karakterlerin hepsi matris veya vektörlere aittir. Öncelikle "örneklem ortalama vektörü",

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{x}_1 \\ \\ \bar{x}_p \end{bmatrix}$$

içinde 1 değerleri olan nx1 sütun vektörü **1** olarak tanımlandığında,

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{\mathbf{r}=1}^{\mathbf{n}} \mathbf{x}_\mathbf{r} = \frac{1}{n}\mathbf{X}'\mathbf{1}$$

şeklinde belirlenir. Kovaryans matrisinin hesaplanması için öncelikle "merkezleştirme matrisi"

$$\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'$$

olarak belirlenir. Bu durumda örnek kovaryans matrisinin sapmasız tahmincisi

$$\mathbf{S}_\mathbf{u} = \frac{1}{n-1}\mathbf{X}'\mathbf{H}\mathbf{X}$$

olarak belirlenir. Bu durumda tek değişkenli ölçüler olarak sıklıkça kullanılan ortalamaya göre momentlerin, çok değişkenli veri setlerinde benzer olarak kullanılmasını sağlayan "invariant fonksiyonlar" matris olarak aşağıdaki şekilde belirlenir:

$$\mathbf{g}_{\mathbf{rs}} = (\mathbf{x}_\mathbf{r} - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_\mathbf{s} - \bar{\mathbf{x}})$$

Mardia, çok değişkenli asimetri ve basıklık ölçülerinin örnek tahmincilerini aşağıdaki şekilde tanımlamıştır.

$$b_{1,p} = \frac{1}{n^2}\sum_{r,s=1}^{n} g_{rs}^3$$

ve

$$b_{2,p} = \frac{1}{n}\sum_{r=1}^{n} g_{rr}^2$$

Elde edilen bu momentler affine transformasyonlar altında değişmezdirler (Mardia, Kent, Bibby, 1989). Diğer bir ifade ile ölçek ve orijinin değiştirilmesi katsayıların büyüklüğünü etkilememektedir. Çok değişkenli normal dağılımın parametreleri $\beta_{1,p}=0$ ve $\beta_{2,p}=p(p+2)$ olduğundan Mardia (1970), çok değişkenli simetri ve basıklık ölçülerinin örnek istatistiklerinin, anakütle dağılımı çoklu normal dağılım kabul edildiğinde, $n \to \infty$ iken asimtotik dağılımlarının, sırasıyla v=[p(p+1)(p+2)/6] serbestlik derecesine sahip Ki-kare dağılımına ve $b_{2,p}$'nin z transformasyonunun standart normal dağılıma uygun olacağını belirlemiştir:

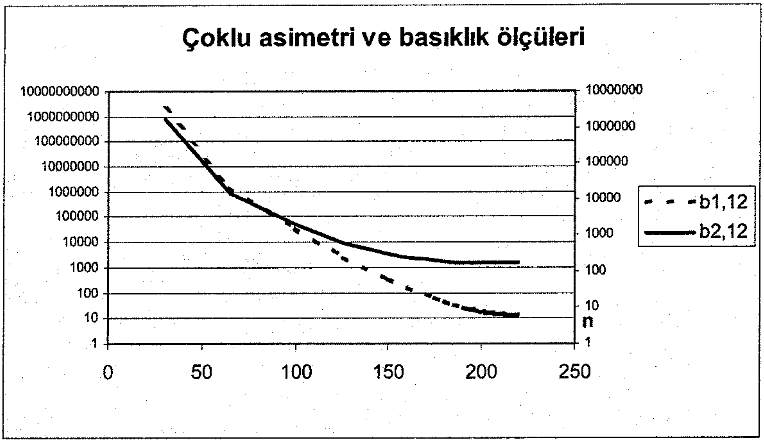
$$\frac{1}{6} n b_{1,p} \sim \chi_v^2 \qquad \text{ve} \qquad \frac{b_{2,p} - p(p+2)}{\sqrt{8p(p+2)/n}} \sim N(0,1)$$

Büyük örnekler için kabul edilen bu limit dağılımlar kullanılarak çok değişkenli normal dağılıma uygunluğun belirtildiği H_0 sıfır hipotezinin testi mümkün olmaktadır. Bu çalışmada, bu istatistiklere ilişkin dağılımlar kullanılarak yeterli örnek hacminin belirlenmesine çalışılmaktadır.

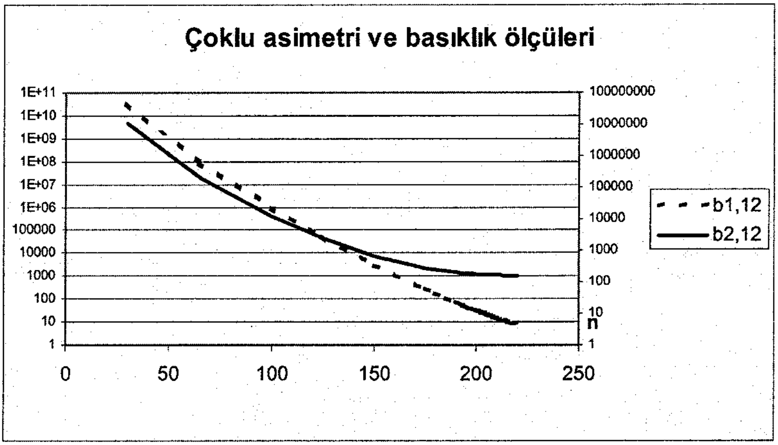
4. VERİLERİN ANALİZİ

Çalışmanın amacı, çok değişkenli normal dağılımın sınanabilmesi 5'li seviyede ölçülen tutum ölçekleri için çok değişkenli asimetri ve basıklık ölçülerinin kararlı olduğu seviyede uygun örnek hacmini belirlemektir. Bu nedenle binom ve normal dağılımlı 12 değişkenli 220 gözlemli veriler (p=12; n=220) türetilmiştir. Veri setleri içindeki her bir değişken diğerlerinden bağımsız üretildiği için marjinal dağılımlar bilinmekte, ancak ortak dağılım sınanmak istenmektedir. Birinci ve ikinci veri gurupları, Binom dağılımına uygun olarak sırasıyla 0.20 ve 0.50 parametreler ile tek değişkenli sağa asimetrik ve simetrik tesadüfi sayılardan oluşmaktadır. Üçüncü veri kümesi ise ortalaması 2 ve varyansı 2/3 olan normal dağılıma uygun N(2, 0.67) olarak türetilmiştir. Türetilen veri kümelerinde değişkenlerin aldıkları değerler 0-4 arasında değerler alacak şekilde ayarlanmıştır. Bunun sebebi 5'li ölçek sorularının genelde –2-2, 0-4 veya 1-5 aralığında tam sayılar olarak kodlanması ve analizlerin bu subjektif kodlamalara göre yapılmasıdır. Mardia'nın önerdiği çok değişkenli asimetri ve basıklık ölçülerinin hesabında orijinal veriler merkezileştirildikleri için kodlamanın başlangıç değerleri test sonuçlarına etki etmeyecektir. Diğer taraftan veri kümelerindeki değişkenlerin rassal üretilmesi, bu yöntemin eleştirilmesi ve alternatifler oluşturulmasına sebep olan, değişkenler arası ilişki yapısının oluşmamasını sağlamaktadır. Her üç veri kümesinde farklı örnek hacimleri için (n=30,65,100,.....,220) Mardia çok değişkenli asimetri ve basıklık ölçüleri hesaplanmıştır. Bu istatistiklerin önerilen limit dağılıma göre test istatistikleri ve kuyruk olasılıkları hesaplanmıştır. Elde edilen test istatistikleri ve kuyruk olasılıkları karşılaştırılmıştır. Çoklu asimetri ve basıklık istatistiklerine ve test istatistiklerine ait grafiklerde asimetri ve basıklık ölçülerini temsil eden eksenler logaritmik ölçekle düzenlenerek, elde edilen çok büyük ve küçük değerlerin aynı grafik üzerinde özetlenmesi sağlanmıştır. Bütün veri kümeleri için Ki-kare dağılımına uyan asimetri ölçüleri v=364 serbestlik derecesinde %5 anlam düzeyinde 409 kritik değerine sahiptir. Normal dağılan basıklık ölçüsünün kritik değeri ise z=1.645 olarak belirlenmiştir.

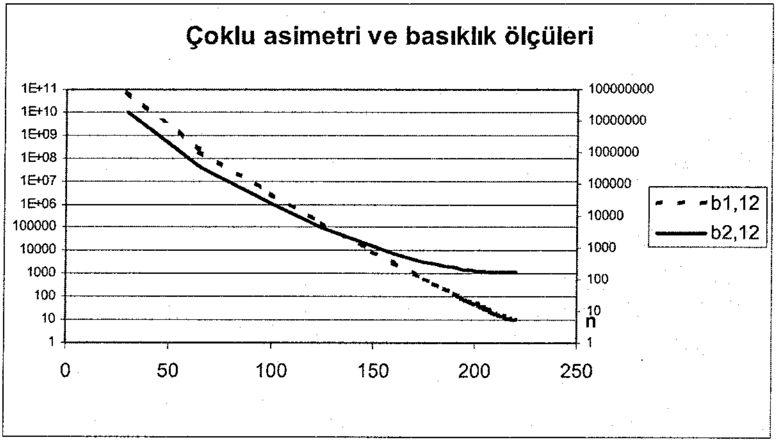
Binom ve normal dağılıma uygun türetilmiş veri kümelerinden elde edilen sonuçlar aşağıda sunulmaktadır.



Şekil 1 Asimetrik binom (p=0,2) marjinal dağılımlı veri seti

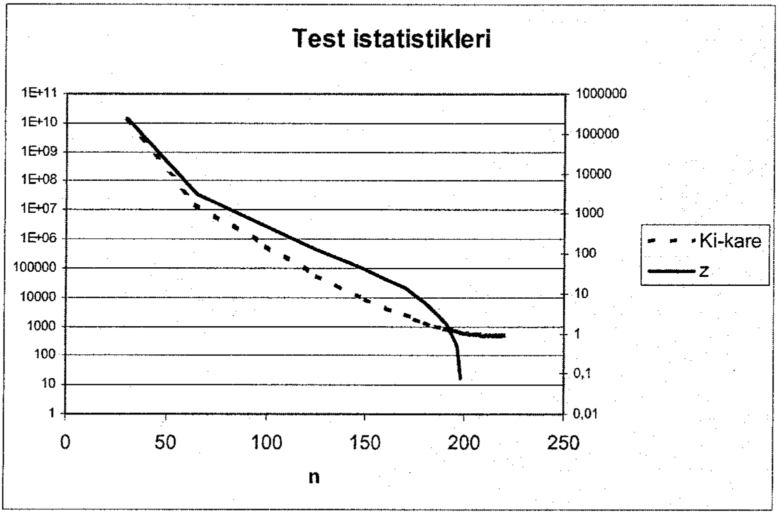


Şekil 2 Simetrik binom (p=0,5) marjinal dağılımlı veri seti



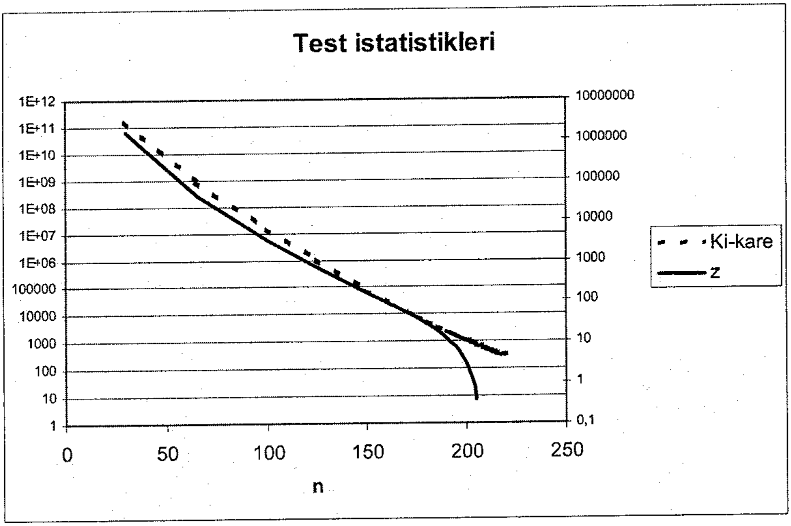
Şekil 3 Normal marjinal dağılımlı veri seti

Birinci veri kümesinden elde edilen sonuçlar, diğerlerinden farklı görülmektedir. Bunun sebebi küçük hacimli örneklerde daha küçük değerlerin çıkmasıdır. Veri kümelerinden farklı örnek hacimlerinde elde edilen test istatistiklerinin karşılaştırılmaları da mümkündür.

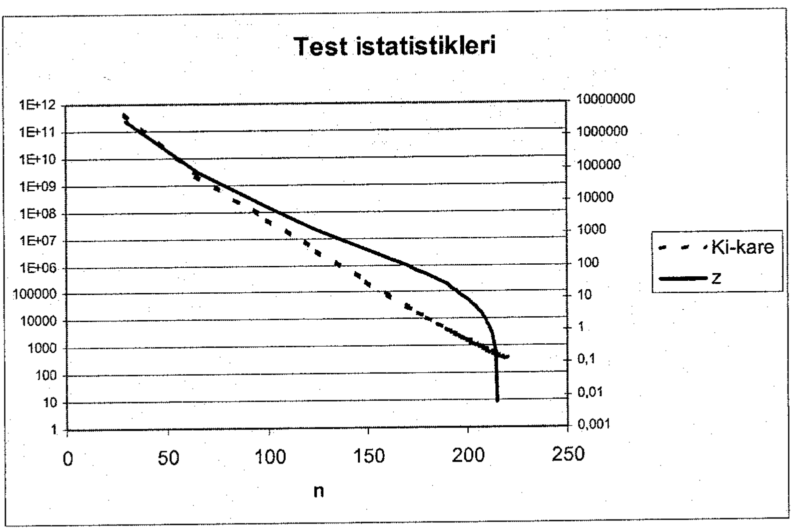


Şekil 4 Asimetrik binom (p=0,2) marjinal dağılımlı veri seti

200'den sonraki gözlem hacimleri için basıklık katsayısı negatif değerler almış, logaritmik eksene sahip grafik üzerinde gösterilememiştir. Simetrik binom dağılmış ikinci veri kümesinde ve normal dağılıma uygun türetilmiş üçüncü veri kümesinde test istatistikleri aşağıda görülmektedirler.

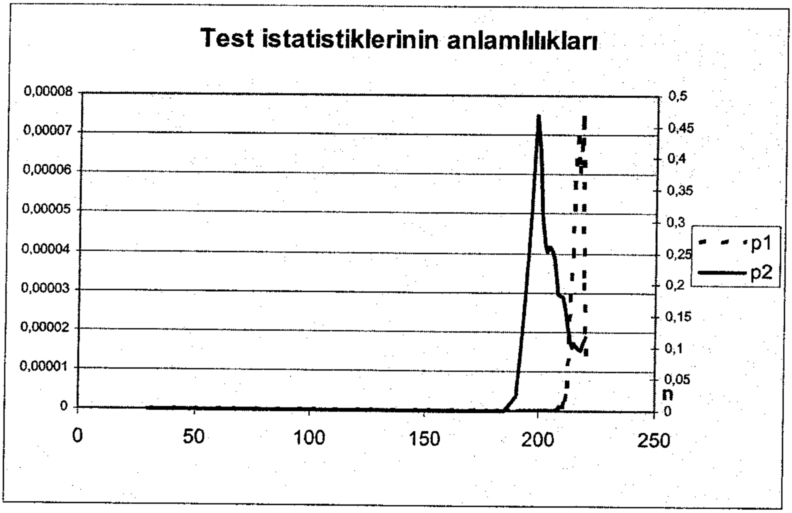


Şekil 5 Simetrik binom (p=0,5) marjinal dağılımlı veri seti



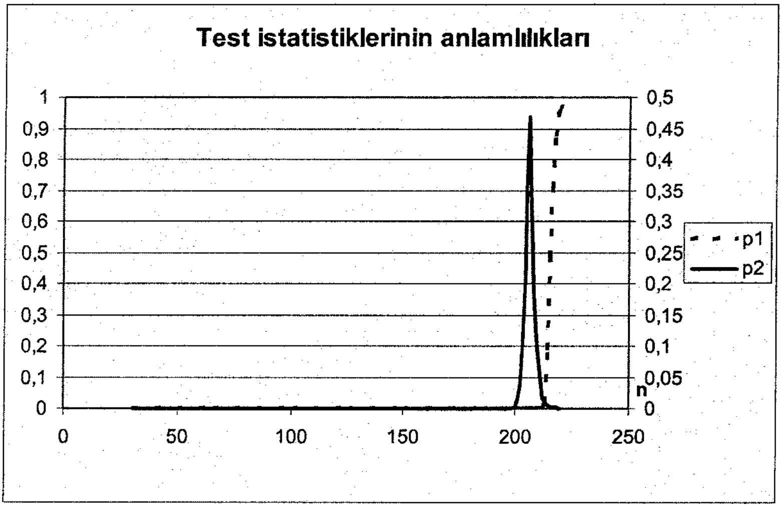
Şekil 6 Normal marjinal dağılımlı veri seti

Her üç veri kümesinde basıklık normal dağılıma yaklaşırken, birinci veri kümesinin sağa asimetrik kaldığı ve kritik değerin altına düşmediği gözlenmiştir. Ancak sonuçların daha detaylı anlaşılması için örnek istatistikleri ve test istatistiklerinden ziyade, limit dağılıma göre kuyruk olasılıklarının incelenmesi gerekir.



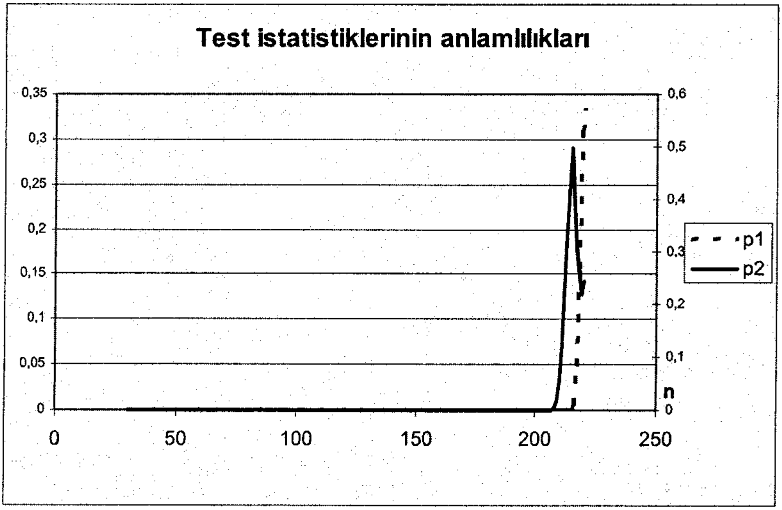
Şekil 7 Asimetrik binom (p=0,2) marjinal dağılımlı veri seti

Asimetrik binom dağılmış veri kümesinde asimetri katsayısına ilişkin kuyruk olasılığı 1/10000'in altına hiç inmemiş, dolayısıyla normal dağılıma benzerliği temsil eden sıfır hipotezi hiç kabul edilmemiştir. Basıklık katsayısı ise örnek hacminin 200'e yakın değerleri için çoklu normal dağılım özelliğine yakın değerler almaktadır. Ancak artan örnek hacmi ile bu durumdan tekrar uzaklaşmaktadır.



Şekil 8 Simetrik binom (p=0,5) marjinal dağılımlı veri seti

Her iki katsayının kuyruk olasılığı 200'ün üzerine çıkıldığında %5'i aşmakta, fakat basıklık, artan örnek hacmi ile normallikten uzaklaşmaktadır.



Şekil 9 Normal marjinal dağılımlı veri seti

Normal dağılıma uygun türetilmiş veri kümesinde de test istatistiklerinin anlamlılıkları simetri için 217'nci ve basıklık için 210'ncu gözlemden sonra %5'in üzerine çıkmıştır.

5. SONUÇ

Elde edilen sonuçlar, simetrilik ve basıklığın örnek hacmine çok duyarlı olduğunu göstermektedir. Sonuçların ilginç bir yanı örnek hacminin artması ile çoklu normalliğe yaklaşımın yavaşlaması ve hatta çoklu normallikten uzaklaşma durumudur. Ancak bu çalışma, diğer pek çok çalışmaya göre daha yüksek bir gözlem sayısı ve daha fazla değişken ile yapılması ile farklılık göstermektedir. Özellikle anketler ile elde edilen tutum, algılama ve davranışları belirlemeye yönelik Likert ölçeği tarzındaki sorulardan elde edilen değişkenlerin, çok değişkenli analizler ile değerlendirilmesi esnasında çoklu normal dağılıma duyarlı yöntem ve istatistiklerin hesaplanması için pratik olarak en az 200 gözleme ulaşılması gerektiği görülmektedir. Çalışma tek değişkenli normalliğin kontrol edilmesinin, çoklu normalliği sağlamadığını da ortaya koymaktadır. Çalışmanın diğer bir faydası katsayıların hesap tabloları (MS Excel) yardımıyla elde edilmiş olmasıdır. Kullanılan algoritma aynı örnek verisi içerisinde farklı gözlem sayılarına göre örnek istatistiklerinin hesaplanmasına imkan sağlamaktadır.

Çalışma, simetri ve basıklık örnek istatistiklerinin, örneklemin 220'nin üzerine çıkması, değişkenlerin arasında istatistiksel bağlılık olması durumunda nasıl davranacaklarını ortaya koymamaktadır. İncelenmesi gereken diğer bir konu ise veri kümesinin boyutuna bağlı olarak örnek istatistiklerinin davranış şeklidir.

KAYNAKLAR

Baxter, M.J. (1999). On the multivariate normality of data arising from lead isotope fields. Journal of Archaeological Science, 26, 117-124.

Beirlant, J., Mason, D.M., & Vynckier, C. (1999). Goodness-of-fit analysis for multivariate normality based on generalized quantiles. Computational Statistics & Data Analysis, 30, 119-142.

Clason, Denis L.; Dormody, Thomas J.; Analyzing data measured by ındividual likert-type ıtems, Journal of Agricultural Education, Vol:35, Number:4,1994

Cox, D.R., & Wermuth, N. (1994). Test of linearity, multivariate normality and adequacy of linear scores. Applied Statistics, 43, 347-355.

Csorgo, S. (1986). Testing for normality in arbitrary dimension. Annals of Statistics, 14, 708-723.

Gutjahr, S., Henze, N., & Folkers, M. (1999). Shortcomings of generalized affine invariant skewness measures, Journal of Multivariate Analysis, 71, 1-23.

Hair,J.F., Anderson, R.E., Tatham, R., Black, C.W. (1998). Multivariate Data Analysis, Fifth Edition, Prentice-Hall, New-Jersey.

Hawkins, D.M. (1981). A new test for multivariate normality and homoscedasticity. Technometrics, 23, 105-109.

Henze, N. (1997). Extreme smoothing and testing for multivariate normality. Statistics & Probability Letters, 35, 203-213.

Henze, N., & Wagner, T. (1997). A new approach to the BHEP tests for multivariate normality. Journal of Multivariate Analysis, 62, 1-23.

Huffer, F.W., & Park, C. (2002). The limiting distribution of a test for multivariate structure. Journal of Statistical Planning and Inference, 105, 417-431.

Hüsler, J., Liu, R.Y., & Singh, K. (2002). A formula for the tail probability of a multivariate normal distribution and its applications. Journal of Multivariate Analysis, 82, 422-430.

Klar, B. (2002). A treatment of multivariate skewness, kurtosis and related statistics. Journal of Multivariate Analysis, 83, 141-165.

Liang, J., Li, R., Fang, H., & Fang K.-T. (2000). Testing multinormality based on low-dimensional projection. Journal of Statistical Planning and Inference, 86, 129-141.

Likert, R.(1932). A Technique For The Measurement of Attitudes, Archives of Psychology, New-York.

Machado, S.G. (1983). Two statistics for testing for multivariate normality. Biometrika, 70, 713-718.

Mardia, K.V., Kent, J.T., & Bibby, J.M. (1989). Multivariate Analysis. (7th. pr.). San Diego: Academic Press, Inc.

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. Biometrika, 57, 519-530.

Olive, D.J. (2003). A resistant estimator of multivariate location and dispersion. Computational Statistics and Data Analysis, Article in Press.

Tatlıdil, H. (2002). Uygulamalı Çok Değişkenli İstatistiksel Analiz, Zıraat Matb., Ankara.

Tezbaşaran, A.A. (1997). Likert Tipi Ölçek Geliştirme Kılavuzu,Türk Psikologlar Derneği Yayınları, Ankara.