ANKARA ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

NORMALLİK TESTLERİNİN KARŞILAŞTIRILMASI

Arif ÖZER

ZOOTEKNİ ANABİLİM DALI

ANKARA
2007

ÖZET

Yüksek Lisans Tezi

NORMALLİK TESTLERİNİN KARŞILAŞTIRILŞMASI

Arif ÖZER

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
Zootekni Anabilim Dalı

Danışman: Prof. Dr. Zahide KOCABAŞ

Bu tez çalışmasında, Normallik testlerinden Eğrilik, Diklik, D' Agostino-Pearson, Jargue-Bera, Kolmogorov-Smirnov, Lilliefors, Ki-kare, Anderson-Darling ve Shapiro-Wilk testlerini I. tip ve güç performansları bakımından karşılaştırılmıştır. Yapılan karşılaştırmalarda I. tip hata olasılığı %5 olarak kararlaştırılmıştır ve söz konusu karşılaştırmalar, simulasyon tekniği ile, normal, simetrik ve çeşitli eğriliklerdeki dağılımlarda ve çeşitli örnek genişliklerinde yapılmıştır. 100.000 simulasyon denemesi sonunda ele alınan testlerden I. tip hata olasılığı bakımından Jargue-Bera testi iyi sonuçlar verirken testin gücü bakımından genellikle Shapiro-Wilk testi diğer testlerden daha güçlü olarak gözlemlenmiştir. Normal ve normal olmayan dağılımlar birlikte göz önünde tutulduğunda Shapiro-Wilk testi diğer testlere göre en iyi sonuçları vermiştir. Bunun yanı sıra Anderson-Darling, Eğrilik ve D'Agostino-Pearson testlerinin de güçlü testler olduğu sonucuna varılmıştır.

Haziran 2007, 45 sayfa

Anahtar Kelimeler: Normallik testleri, varyans analizi, I. tip hata, testin gücü, eğrilik, diklik, D' Agostino-Pearson, Jargue-Bera, Kolmogorov-Smirnov, Lilliefors, ki-kare, Anderson-Darling ve Shapiro-Wilk

ABSTRACT

Master Thesis

Comparison of Normality Tests

Arif ÖZER

Ankara University
Graduate School of Natural and Applied Sciences
Department of Animal Science

Supervisor: Prof. Dr. Zahide KOCABAŞ

In this study, Skewness, Kurtosis, D' Agostino-Pearson, Jargue-Bera, Kolmogorov-Smirnov, Lilliefors, Chi-Square, Anderson-Darling and Shapiro-Wilk tests of normality were compared in terms of type I error and power of tests. At this comparisons, type I error probability was decided as %5 and simulation were done for normal, symmetric and various skew distributions and for various sample sizes. At the end of 100.000 simulation experiment, it was determined that in terms of type I error probability Jargue-Bera test gave the best results in all tests but for power of tests, Shapiro-Wilk test gave generally the most powerful results. Shapiro-Wilk test gave more powerful the others to take into consideration normal and nonnormal distributions. Besides, Anderson-Darling, Skewness and D' Agostino-Pearson Tests were powerful tests.

June 2007, 45 pages

Key Words: Normality tests, analysis of variance, type I error, power of test, skewness, kurtosis, D' Agostino-Pearson, Jargue-Bera, Kolmogorov-Smirnov, Lilliefors, chi-square, Anderson-Darling and Shapiro-Wilk

TEŞEKKÜR

Çalışmalarımı yönlendiren, araştırmalarımın her aşamasında ve her defasında büyük bir sabırla, bilgi, öneri ve yardımlarını esirgemeyerek akademik ortamda olduğu kadar beşeri ilişkilerde de engin fikirleriyle yetişme ve gelişmeme katkıda bulunan danışman hocam sayın Prof. Dr. Zahide KOCABAŞ'a, tezin simülasyon çalışmalarında yakın ilgi ve önerileri ile bana yardımcı olan ve hiç bir sorumu yanıtsız bırakmayan değerli hocalarım, Prof. Dr. Fikret GÜRBÜZ, Doç. Dr. Muhip ÖZKAN ve Yard. Doç. Dr. Sıddık KESKİN'e, çalışmalarım süresince maddi manevi desteklerini esirgemeden ve hiç bir zaman benim sorularımdan bıkmayarak bana daima yol gösteren Araş. Gör. Dr. Özgür KOŞKAN'a, ayrıca, çalışmalarımda göstermiş olduğu yakın ilgiden dolayı Araş. Gör. Yeliz KAŞKO ve Araş. Gör. Hasan MEYDAN'a, her konuda olduğu gibi bu konuda da her zaman manevi destekleriyle benim yanımda olan ve hiçbir zaman benden desteklerini esirgemeyen sevgili Annem ve Babam'a sonsuz sevgi ve saygılarımı sunar, en derin duygularımla teşekkür ederim.

Arif ÖZER
Ankara, Haziran 2007

İÇİNDEKİLER

SİMGELER DİZİNİ

AD	Anderson-Darling
DP	D'Agostino-Pearson
J-B	Jargue-Bera
K-S	Kolmogorov-Smirnov
W	Shapiro-Wilk
α	Deneme sonunda geçekleşen I. tip hata olasılığı
β	II. tip hata olasılığı

ŞEKİLLER DİZİNİ

ÇİZELGELER DİZİNİ

1. GİRİŞ

Parametrik yöntemler verilerin normal dağılmış olduğu varsayar. Bu yüzden normallik testleri istatistikte önemli bir yere sahiptir ve çok fazla sayıda metot bulunmuştur. Parametrik testlerin uygulanabilmesi için, varyans analizinin ön şartlarından biri olan gözlemlerin normal dağılması ön şartının yerine gelmesi gerekmektedir. Normal dağılım ön şartı; gözlemlerin denemede üzerinde durduğumuz özellik bakımından normal dağılım gösteriyor olması demektir. Bu gözlemler ölçmek, tartmak veya herhangi bir şekilde analiz etmek suretiyle elde edilirler. Bu yaklaşımla elde edilen gözlemlerin hemen hemen hepsi normal dağılımı andıran bir görünüm vermesine rağmen bazı özelliklerin gösterdiği dağılımlar normal dağılıma benzememektedir.

Üzerinde durulan özelliğin normal dağılım gösterip göstermediği ya daha önceden özellikle konuyla ilgili yapılmış araştırmalardan ya da araştırıcının bizzat kendisi tarafından yapılan ön denemelerle belirlenebilir. Sağlıklı ve yetişken bir insanın vücut sıcaklığı buna örnek verilebilir. Ancak yapılacak olan ön denemelerde de oldukça fazla sayıda deney ünitesine ihtiyaç vardır. Elde edilen gözlemlerin teorik dağılıma uygun dağılıp dağılmadıkları istatistik yöntemlerle test edilebilir.

Verilerin normal dağılıma uymama nedenleri aşağıdaki biçimde sıralamak mümkündür;

- Üzerinde durulan özellik bakımından elde edilen gözlem değerleri geniş bir varyasyon gösterip dağılımın bazı bölgelerinde değer alamaması, yani; sürekliliğin bozulduğu durumlar,
- Örnek genişliği 30' dan küçük olduğu durumda dağılım normalden uzaklaşması,
- Ortalama, ortanca değer ve tepe değerinin birbirine eşit olmadığı durumlarda dağılım sola yatık ya da sağa yatık olarak gözlenebilmesiGözlemlerin büyük bir bölümünün normal dağılmasına karşın, az sayıda da olsa bazı gözlemlerin aykırı değer (outliers) durumunda bulunması.

Yukarıda sözü geçen nedenlerle dağılımın normalden sapma gösterebileceği göz önünde tutularak, istatistiksel testlerin uygulanmasına geçilmeden önce verilere normallik analizi yapmak gereği ortaya çıkmaktadır.

Dağılımın normal olup olmadığı grafiksel ve istatistik analiz yöntemleri ile kontrol edilebilir. Elde edilen verilerin histogramının çizilmesi ya da normal olasılık grafiğinin çizilmesi ile dağılımın normalden sapıp sapmadığı kabaca da olsa anlaşılabilir. Yani, grafiksel yöntemlerle normalden sapmanın şekli ve büyüklüğü hakkında bir ön fikir edinilebilir. Ancak, söz konusu sapmanın dağılımın normal olarak kabul edilmemesi için önemli bir sapma olup olmadığının belirlenmesi mümkün olamamaktadır. Bu durum ise elde edilen verilerin normal dağılım gösterip göstermediklerinin belirlenmesinde hipotez testi yapılması gereğini ortaya çıkartmaktadır.

Diğer bir ifade ile söz konusu verilerin normal dağılıma uyup uymadıkları hakkında grafiksel yöntemlerle edinilen izlenimin istatistik yöntemlerle de test edilmesi gerekir. Bunun için çok sayıda normallik testinden yararlanılabilir. Verilerde normallik ön şartının yerine gelip gelmediğini test etmek için en etkin yöntemin kullanılması gerekir. Mevcut testlerin değişik deneme koşullarında I. tip hata ve testin gücü bakımından karşılaştırılması yapılmalı ve söz konusu koşullarda bu testlerin performansları ortaya konulmalıdır. Aynı hipotezin test edilmesinde kullanılan testlerden hangisinin gücü daha yüksek ise ve hangisi başlangıçta kararlaştırılan I. tip hatayı deneme sonunda da aynı düzeyde koruyorsa normallik ön şartının sağlanıp sağlanmadığını test etmede söz konusu test tercih edilmelidir.

Bu tez çalışmasında verilerin normal dağılıma uygun olup olmadığını ortaya koymak amacıyla kullanılan Eğrilik, Diklik, D'Agostino-Pearson (DP), Jargue-Bera (J-B), Kolmogorov-Smirnov (K-S), Lilliefors, Ki-kare, Anderson-Darling (AD) ve Shapiro-Wilk (W) testlerinin karşılaştırılması amaçlanmaktadır. Bu karşılaştırmalarda çeşitli örnek hacimleri, normal ve normal olmayan çeşitli dağılım şekilleri dikkate alınarak normallik testlerinin gerçekleşen I.tip hataları ve güçleri irdelenecektir.

2. KAYNAK ÖZETLERİ

Shapiro *et al.* (1968), W test istatistiğini diğer normallik testleri ile bu testlerin hassasiyetlerini belirlemek için çok sayıda çeşitli dağılımlarda ve örnek genişliklerinde değerlendirmişlerdir. Çalışmalarında, W (Shapiro-Wilk), $\sqrt{b_1}$, b_2, KS (Kolmogorov-Smirnov), CM (Cramer-Von Mises), WCM (tartılı CM), D (modifiye edilmiş KS), CS (Ki-Kare) ve u (Studentized range)' testlerini kullanmışlar ve W testinin bu testlerden daha güçlü olduğunu, Uzaklık testlerinin (KS, CM, WCM, D) genel olarak çok hassas testler olmadığını, U istatistiğinin simetrik dağılımlara özellikle de kısa kuyruklu dağılımlara karşın çok iyi bir test olmasına rağmen asimetrik dağılımlarda neredeyse duyarsız olduğunu, $\sqrt{b_1}$ ve b_2 testlerinin duyarlı testler olmasına karşın W testinin etkisi bu testlerden daha iyi olduğunu bildirmişlerdir.

Shapiro and Francia (1972), Shapiro-Wilk (W) testinin bir modifikasyonu olan (W') testi önermişlerdir. Bu testte, Shapiro and Wilk'in W testinde verdikleri katsayıların elde edilmesinde kullanılan normal sıra istatistiklerin kovaryans matrisleri yerine normal sıra istatistiklerinin beklenen değerlerine bağlı katsayıları kullanmışlardır. Bu iki testin normal olmayan dağılımlarda duyarlılıklarını belirlemek için yaptıkları karşılaştırmalarda da her iki testin birbirine eş değer göründüğünü bildirmişlerdir.

Stephens (1974), Ampirik dağılım fonksiyonunu (EDF) temel alan Kolmogorov-Smirnov (D), Cramer-Von Mises (W^2), Kuiper (V), Watson (U^2) ve Anderson-Darling (A^2) gibi testlerle Shairo-Wilk testini çeşitli örnek genişliklerinde, çeşitli dağılımlar kullanarak normallik bakımından güç performanslarını test etmiştir. Bu karşılaştırma sonucunda EDF testlerinin W istatistiğine karşı düşük güç değerlerine sahip olduğunu ancak Anderson- Darling testi ve W^2 testinin W testine yakın güç değerleri verdiğini bildirmiştir.

Pearson *et al.* (1977), normallik testlerini güçleri bakımından karşılaştırmışlar ve W test istatistiğinin yatık (skewed) dağılımlara karşı en güçlü test olduğunu ayrıca normal olmayan dik (kurtosis, $B_2 \neq 3$) ve simetrik dağılımlarda da daha güçlü olduğu sonucuna varmışlardır.

Pettitt (1977), son yıllarda yapılan çalışmalarda ortalama ve varyansın bilinmediği durumlarda normallik testlerinden Anderson-Darling testinin güçlü olduğunu bildirmiştir. Çalışmasında küçük örnekler için kritik bölgelerin başlangıç değerleri için tablolar oluşturmuştur. Bu kritik bölgelerin başlangıcını temsil eden değerleri, asimtotik değerler kullanarak pratikte kullanılacak değerlere yuvarlanarak elde etmiştir. Bu değerleri, ortak bir normallik değerlendirmesi yapmak için Fisher'in metodunu kullanan çeşitli bağımsız örnekleri temel alarak elde etmiştir. Bu çalışmada güç bakımından karşılaştırmaların Shapiro-Wilk istatistiğini temel alan metotla yapıldığı bildirilmiştir. Bu iki farklı istatistiğe dayanan bu metotlar arasında çok küçük farklılıkların görüldüğü bildirilmiştir.

Saniga and Miles (1979), yaptıkları çalışmada, 6 adet normallik testini güçleri bakımından asimetrik sürekli dağılımlar için karşılaştırmışlardır. Karşılaştırma sonucunda $\sqrt{b_1}$, b_2, W, D'Agostino D, U ve müşterek $\sqrt{b_1}$, b_2 testlerinden ilk dört testin diğer iki testten daha duyarlı olduğunu ve testlerin ve kritik değerlerin n≥5 ve I. tip hatalarda hesaplanmasının kolaylığı bakımından $\sqrt{b_1}$, b_2 testlerinin diğer testlere göre daha uygun olduğunu bildirmişlerdir.

Linnet (1988), Anderson-Darling, Cramer-Von Mises ve Kolmogorov-Smirnov testlerinin dönüştürülmüş verilerde kullanımını dikkate almıştır. Simulasyon çalışmalarının, test istatistiklerinin normal dağılım durumunda parametre değerlerinin bağımsız olduğunu gösterdiğini ve böylece ortalama, standart sapma ve dönüştürme parametresini bilinmediği durumlarda, uygun olan kritik değerlerin kullanılması için bu çalışmada tablo haline getirildiğini ve uyum iyiliği testlerinin dönüştürülmüş edilmiş değerlere uygulanabileceğini bildirmiştir.

Çoğu çalışmalarda, araştırmalar, çok sayıdaki alternatif dağılımlarda testlerin istatistiksel güçlerinin üzerinedir ve bu testlerin kullanılmaları için önerilmeleri durumunda, Shapiro-Wilk W testi, üçüncü örnek momenti ($\sqrt{b_1}$), dördüncü örnek momenti testleri (b_2) ve D'Agostino-Pearson K^2 testleri çok güçlü testler olarak ortaya çıkmaktır. Özellikle Shapiro-Wilk testi ve D'Agostino-Pearson K^2 testleri birçok normal olmayan dağılımları belirlemede çok güçlü testlerdir. $\sqrt{b_1}$ ve b_2 testleri sırasıyla normal olmayan eğriliği ve dikliği belirlemede çok iyi özelliklere sahiptir. χ^2 (chi-square) ve Kolmogrov-Smirnov testi normalliği belirlemede zayıf güç özelliklerine sahiptir (D'Agostino *et al*. 1990).

D'Agostino (1973,1986) ve Zar (1999), Normallik için D'Agostino Pearson testinin, çok yaygın olarak kullanılan Kolmogorov-Smirnov testi ve Ki-kare testinden uyum iyiliğini değerlendirmede daha etkili olduğunu bildirmişlerdir (Sheskin 2000).

Gan and Koehler (1990), çeşitli normallik testlerini, güç tahminlerinde 1000'lik Monte Carlo örneklerini temel alarak tamamen değişik dağılımlarda karşılaştırmışlardır. Bu karşılaştırmalar sonucunda, normallik için Shapiro-Wilk testini genelde en iyi test olarak bulmuşlardır. Fakat bütün dağılımlar göz önüne alındığında hepsinde çok güçlü olmadığını sonucuna varmışlardır (Seier 2002).

Rahman and Govindarajulu (1997), Shapiro-Wilk testini modifiye ederek bütün örnek genişliklerine uygulanabilen yeni bir test önermişlerdir. Bu testi, Shapiro-Wilk ve Shapiro-Francia testleri ile karşılaştırmışlardır. Bu Karşılaştırmalarda uniform, lojistik, lognormal, 1,2,4,10 serbestlik dereceli ki-kare ve beta (2,1) gibi alternatif dağılımlar kullanmışlardır. Yaptıkları bu güç çalışması denemelerinde 10,20,35,50,75 ve 99 örnek genişliklerinde % 1, % 5, % 10 α önem seviyesinde yürütmüşler ve 5000 deneme sonunda simetrik olmayan (skewed) dağılımlar için bu testin Shapiro-Wilk ve Shapiro-Francia testine benzer sonuçlar verdiğini ya da daha iyi olduğunu bildirmişlerdir. Eğer dağılımın şekli simetrik ve leptokurtic ($B_2>3$) durumunda ise Shapiro-Francia testinin daha iyi olduğunu, Lojistik dağılımda ise her üç testinde düşük güç performansına sahip olduğunu bildirmişlerdir. Bu nedenle örnek genişliği küçük (n=10) olduğunda Shapiro-Wilk testinin modifiye edilmiş versiyonunu önermişlerdir.

Cho *et al.* (2002), yaptıkları çalışmada, Geary'nin Skewnes ve Kurtosis istatistiklerini kullanan bir normallik testi önermişlerdir. Önerdikleri bu testin, J-B testi gibi kolay hesaplanmakta ve küçük örnek hacimlerinde asimtotik ölçüye daha yakın olduğunu bildirmişlerdir. Normallik testi için yaygın bir şekilde kullanılan Jargue-Bera testini bu önerilen test ile karşılaştırdıklarında, Jargue-Bera testinin kısa kuyruklu dağılımlara karşı zayıf güçte olduğunu ve genelde özellikle orta hacimli örnekler için yapılan güç karşılaştırmalarında J-B testine karşı daha güçlü olduğu sonucuna varmışlardır. Ayrıca, çalışmalarında 600.000 örneklemeyle % 1, % 5 ve % 10'luk hata olasılıklarının kritik değerlerini belirlemişlerdir.

Seier (2002), normallik testlerini, I. tip hata ve güçlerini karşılaştırmıştır. Her bir testin I. tip hata ve gücünü bilgisayar simulasyonları ile küçük (20), orta (50) ve büyük (100) örnek genişliklerinde çok sayıda, çeşitli dağılımlarda tahmin etmiştir. Yapılan çalışmada Anderson-Darling (A^2) testinin Kolmogorov-Smirnov (D) testinden daha güçlü olduğu sonucuna varmıştır. QH* (Chen-Shapiro, 1995), W ve K^2 testlerinin küçük, orta ve büyük örnek genişliklerinde en yüksek güce sahip olduklarını bildirmiştir.

Mendeş ve Pala (2003), Shapiro-Wilk, Lilliefors ve Kolmogorov-Smirnov testlerini I. tip hata ve güçlerini karşılaştırmışlardır. Yaptıkları simülasyon çalışmasında farklı örnek genişliklerinde ve farklı dağılımlarda 100000 kez bu testleri denemişlerdir. Bu simülasyon çalışması sonunda Shapiro-Wilk testi bu üç test arasında en güçlü test olarak ortaya çıktığını ve bunu Lilliefors testinin izlediğini, Kolmogorov-Smirnov testinin ise bu üç test arasında en zayıf olduğunu, ayrıca bu üç testi üssel dağılımda benzer ve güçlü olarak bildirmişlerdir.

Brys *et al.* (2004), Jargue-Bera (J-B) testinin normallikten aşırı sapan değerler olduğunda, normalliği ortaya çıkaramayacağını bildirmişler ve bu nedenle, bu durum için daha güçlü üç adet normallik testini test etmişlerdir. Bu testler güç bakımından incelendiğinde; sola ve sağa aşırı sapan değerlerde MC1 ve MC2 testlerinin MC3 testinden (% 5 α önem seviyesi dışında) daha güçlü olduğunu bildirmişlerdir.

Dong and Giles (2004), normallik için 5 adet testi çeşitli örnek genişliklerinde % 10, % 5, % 2 ve % 1 önem seviyelerinde karşılaştırmışlardır. Bu testler ELR (beklenen olasılık (olabilirlik) oranı), Jargue-Bera testi, D'Agostino's testi (D testi), Pearson's X^2 uyum iyiliği testi ve X^{2^*} düzeltilmiş uyum iyiliği testleridir. Bu karşılaştırma sonuçlarına göre Lognormal dağılımda ELR testinin, özellikle küçük örnek genişliklerinde, % 5, % 2 ve % 1 önem seviyelerinde bütün testler arasında en güç test olduğu, J-B testinin de ELR testine yakın değerler verdiği, 2 serbestlik dereceli ki-kare dağılımında, bu 5 test göz önünde tutulduğunda çeşitli örnek genişliklerinde ELR testinin en güçlü test olduğu, çeşitli simetrik dağılımlarda ise J-B testinin en güçlü test olduğu sonucuna varmışlardır.

Poitras (2005), yaptığı simülasyon çalışmasında, Jargue-Bera testi ile moment testlerini karşılaştırmış ve Jargue-Bera testinin küçük ve orta örnek genişliklerinde kullanımına karşı D'Agostino Pearson K^2 gibi moment testlerinin kullanılması yönünde tavsiyede bulunmuştur.

Arcones and Wang (2005), U-işlemlerini temel alan iki yeni normallik testi sunmuşlardır. Bu iki yeni testi $(D_{n,m}$ ve $\tilde{D}_{n,m})$, Lilliefors (1967), Shapiro-Wilk (1968), Csörg*ö* (1986), Epps-Pulley (1983) (BHEP testi) gibi çeşitli normallik testleri ile α=0.05 önem seviyesinde güç performansları bakımından çeşitli dağılımlarda karşılaştırmışlardır. Yapılan simulasyon çalışmasında 10000 deneme sonunda bu iki yeni testin arasında çok fazla bir fark olmadığını, Lilliefors testinin bütün testlerden güç değerleri bakımından daha küçük değerler aldığını bildirmişlerdir. Csörgö testinin tutarsız olduğu, kimi dağılımlarda en güçlü kiminde de güçlü olmadığı ve en güçlü üç testin W, BHEP ve $D_{n,m}$ testleri olduğu sonucuna varmışlardır.

Keskin (2006), üçüncü moment (g_1), dördüncü moment (g_2), Shapiro-Wilk ve D'Agostino Peaorson testlerini I. tip ve II. tip hata bakımından karşılaştırmıştır.

Karşılaştırmada çeşitli örnek genişliklerinde populasyondan her defasında 100000 adet örnek çekerek söz konusu testleri % 5 seviyesinde performanslarını değerlendirmiştir. Bu çalışmanın sonucunda, Shapiro-Wilk testinin, çeşitli şartlar göz önüne alındığında en uygun performansı sergilediğini bildirmiştir.

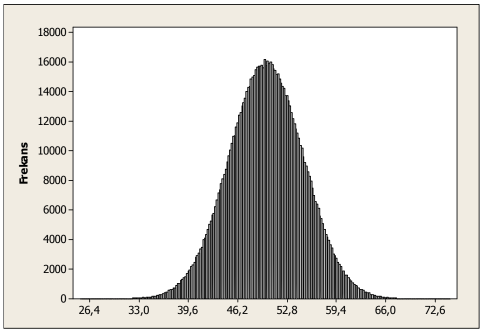
Öztuna vd. (2006), Lilliefors, Shapiro-Wilk, D'Agostino Pearson ve Jargue-Bera testlerini I. tip ve II. tip hata bakımından karşılaştırmışlardır. Bu karşılaştırmada yapılan simulasyon çalışmasının 23 farklı örnek hacimleri ve 8 farklı dağılım için 1000 defa uygulandığını ve bu karşılaştırmaların sonucunda en küçük I. tip hata oranı nedeniyle normal dağılım ve standart normal dağılımlar için Jargue-Bera testinin en iyi sonuç verdiğini bildirmişlerdir. Normal olmayan dağılımlar için küçük örnek hacimlerinde yeterli gücü sağlaması bakımından Shapiro-Wilk testini en güçlü test olarak belirlemişlerdir.

3. MATERYAL VE YÖNTEM

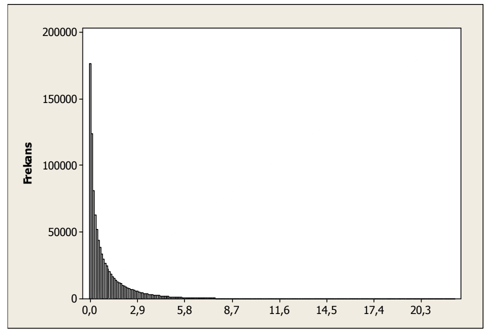
Bu tez çalışmasında bilgisayar simülasyon programı için Monte Carlo teknikleri kullanılmıştır. Bu tezin materyalini, Microsoft Power Station Developer Studio ve IMSL Library yardımıyla üretilen tesadüf sayıları oluşturmaktadır. Bu çalışma için gerekli olan bütün hesaplamalar FORTRAN programlama dilinde yazılmış programlar yardımıyla, Intel Pentium III işlemcili bir bilgisayarda yürütülmüştür.

Eğrilik, Diklik, D'Agostino-Pearson, Jargue-Bera, Kolmogorov-Smirnov, Lilliefors, Ki-kare, Anderson-Darling ve Shapiro-Wilk testlerinin karşılaştırılmasında I. tip hata olasılığı başlangıçta α=0.05 olarak kararlaştırılmıştır. Ortalaması 50, standart sapması 5 olan normal dağılım gösteren bir populasyondan örneklerdeki gözlem sayıları n=7, 8, 9, 10, 20, 30, 40, 50, 100 olan örneklerden her defasında 100000 adet örnek çekilmiştir. Bu aşamadan sonra, ele alınan testlerin gerçekleşen I. tip hata olasılıkları; 100000 simulasyon denemesi sonunda ret edilen H_0 hipotezi adetleri sayıldıktan sonra % ye dönüştürülmesi sonucunda elde edilmiştir.

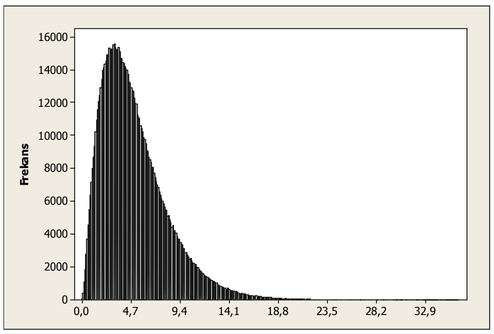
Söz konusu testlerin, testin gücü (1-β) bakımından karşılaştırılmasında ise normal olmayan $X_{(1)}^2$, $X_{(5)}^2$, Beta (22,3), Beta (2,5), Uniform, Lognormal dağılımları kullanılmıştır. Bu dağılımları gösteren populasyonlardan örnek genişliği n=7, 8, 9, 10, 20, 30, 40, 50, 100 olan örneklerden her defasında 100000 adet örnek çekilmiştir. Bu aşamadan sonra, ele alınan testlerin gerçekleşen güç olasılıkları; 100000 simulasyon denemesi sonunda ret edilen H_0 hipotezi adetleri sayıldıktan sonra % ye dönüştürülmesi sonucunda elde edilmiştir. Bu çalışmada kullanılan dağılımların şekilleri Şekil 3.1-3.8'de verilmiştir.



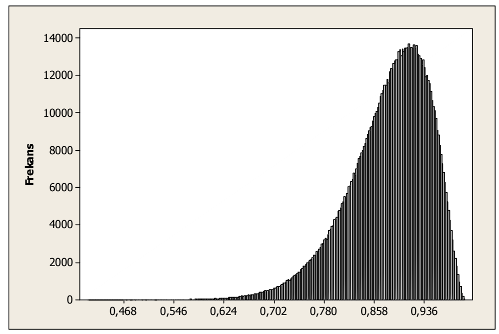
Şekil 3.1 Normal Dağılım



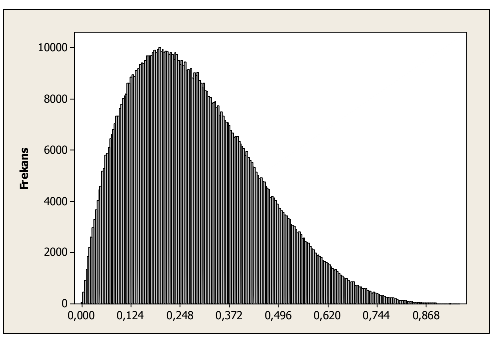
Şekil 3.2 $X_{(1)}^2$ Dağılımı



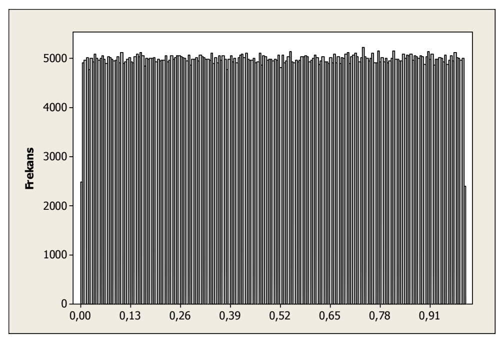
Şekil 3.3 Şekil 1.2 $X_{(5)}^2$ Dağılımı



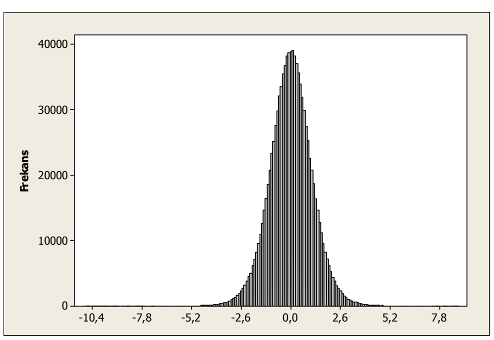
Şekil 3.4 Beta (22,3) Dağılımı



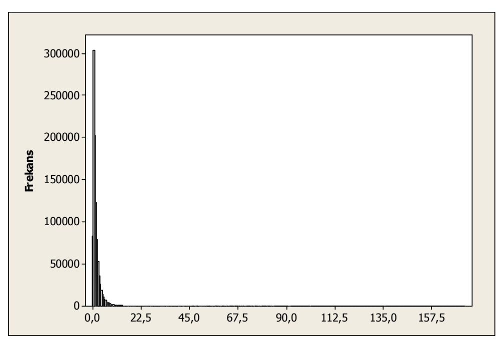
Şekil 3.5 Beta (2,5) Dağılımı



Şekil 3.6 Uniform Dağılım



Şekil 3.7 $t_{(10)}$ Dağılımı



Şekil 3.8 Lognormal Dağılım

4. NORMALLİK TESTLERİ

4.1 Ki-Kare (χ^2) Uyum Testi

Parametrik olmayan testlerden olan Ki-Kare testinin uygulanmasının kolay ve pratik olması nedeniyle araştırmacılar tarafından yaygın olarak kullanıldığı için bu çalışmada yer alması uygun görülmüştür.

Ki-Kare testi ilk kez Pearson (1900) tarafından ortaya atılmış en eski ve en bilinen uygunluk testidir. Bu test, gözlenen ve beklenen veri frekansları arasındaki farkları temel alır. θ'nin parametre vektörü olduğu, bilinmeyen $F(y,\theta)$ ortak dağılımına sahip n,örnek genişliğinde $y_1, y_2, ..., y_n$ bağımsız gözlemlerden oluşan bir örnek olduğu kabul edilirsin, $F(y)$, y tesadüf değişkenin gerçek ama bilinmeyen kümülatif dağılımın fonksiyonu ve $F_o(y)$ teorik kümülatif normal dağılım fonksiyonu olmak üzere;

Hipotez takımı,

$H_o : F(y,\theta) = F_o(y,\theta)$,

$H_1 : F(y,\theta) \neq F_o(y,\theta)$

şeklinde oluşturulmaktadır (Dong and Giles 2004). Sözel olarak tanımlamak gerekirse,

H_0: Dağılım, normal olasılık yoğunluk fonksiyona uygun dağılmaktır.

H_1: Dağılım, normal olasılık yoğunluk fonksiyonuna uygun olarak dağılmamaktadır.

biçimindeki hipotezlerin test edilmesi söz konusudur.

Eğer y_i gözlemleri μ ve σ örnek ortalaması ve standart sapması olmak üzere $x_i = \dfrac{y_i - \mu}{\sigma}$ şeklinde dönüştürülürse F_o, $N(0,1)$ olarak belirtilir.

p_{oi}, i .nci sınıfa düşen gözlemlerin beklenen olasılıklarını,

np_{oi} ,beklenen frekansları ve

k , sınıf sayısını,

n_i , gözlenen frekansları ifade eder. Burada $i = 1,2,...,k$ olmak üzere;

Test istatistiği,

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_{oi})^2}{np_{oi}}$$

$$= \frac{1}{n} \sum_{i=1}^{k} \frac{n_i^2}{p_{oi}} - n$$

elde edilir. H_o hipotezi doğru ise $\chi^2_{(k-3)}$ sınırlayıcı dağılıma sahiptir (Dong and Giles 2004).

Yukarıdaki eşitlik kullanılarak hesaplanan istatistiğin ki-kare dağılımı göstermesi için sınıf sayısının yeterli olması ve sınıflar için hesaplanacak beklenen frekansın 5'ten küçük olmaması gerekir (Kesici ve Kocabaş 1998).

Ki-kare uygunluk testi tek taraflı bir testtir. Çünkü $n_i - np_{oi}$ farklarının kareleri alınarak χ^2 test istatistiği hesaplanır. Fark büyüdükçe, farkların kareleri pozitif yönde sonsuza doğru büyür.

Kritik değer (K.D.), α önem seviyesi ve s.d=k–1-m serbestlik derecesine göre hazırlanmış χ^2 kritik değerler tablosundan belirlenir. Burada m tahmin edilen parametre sayısıdır. Örneğe ait verilerin dağılımının belli bir teorik dağılıma uyup uymadığının belirlenmesi için populasyon parametreleri bilinmediğinde örnekten tahmin edilen değerleri kullanır.

Bu durumda serbestlik derecesi aşağıdaki şekilde hesaplanır. Örnek olarak; normal dağılım için tahmin edilen parametreler μ ve σ olduğundan m=2 alınır. Bu sebeple, kritik değer,

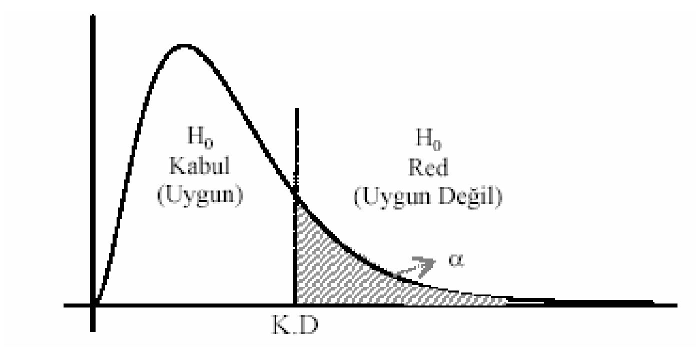
K.D.= $\chi^2_{\alpha;k-1-m}$

olarak sembolize edilir ve bu değer K.D.= $\chi^2_{\alpha;k-1-2} = \chi^2_{\alpha;k-3}$ şeklinde hesaplanır.

Yapılan açıklamalardan anlaşılacağı üzere, eğer populasyona ait ortalama ve standart sapma biliniyorsa serbestlik derecesi hesaplanırken sınıf sayısından bir çıkarılır.

Normal dağılım için serbestlik derecesi hesaplanırken sınıf sayısı (k), 4 ten küçük olmamalıdır. Aksi takdirde serbestlik derecesi 1 den az olacaktır ki bu mümkün değildir (Sheskin 2000).

Test istatistiğinde hesaplanan χ^2 değeri ile kritik $\chi^2_{\alpha;k-1-m}$ değeri, Şekil 4.1'deki karar modeline göre mukayese edilerek karar verilir. Buna göre, $\chi^2 < \chi^2_{\alpha;k-1-m}$ ise, H_0 hipotezi kabul edilerek gözlenen değerlerle beklenen değerlerin birbirine (n_i'lerin np_{oi}'lere) uygun olduğuna, görülen farklılığın önemsiz olduğuna α önem seviyesinde karar verilir.



Şekil 4.1 Ki-kare karar modeli (Bircan vd. 2003)

$\chi^2 > \chi^2_{\alpha;k-1-m}$ ise, H_0 hipotezi reddedilerek gözlenen değerlerle beklenen değerlerin birbirine (χ^2'lerin $\chi^2_{\alpha;k-1-m}$'lere) uygun olmadığına α önem seviyesinde karar verilir.

Beklenen frekansların küçük olması durumunda küçük bir beklenen frekansın χ^2 ye katkısı büyük olacaktır. np_{oi} küçüldükçe χ^2 büyüyecektir. Bu durum H_0 hipotezinin reddedilmesi ihtimalini artırır.

4.2 Kolmogorov-Smirnov Testi

χ^2 uygunluk testlerinin alternatifi olan Kolmogorov-Smirnov testi, ilk olarak 1933 yılında Kolmogorov tarafından önerilmiştir. Daha sonra 1939 yılında Smirnov tarafından geliştirilmiştir. Kolmogorov-Smirnov testi, teorik (beklenen) ve ampirik kümülatif (birikimli) dağılım fonksiyonu arasındaki maksimum mutlak farka dayanır.

Kolmogorov-Smirnov testine ilişkin hipotez takımı aşağıdaki gibi kurulabilir.

H_0: Bu örnek normal dağılım gösteren bir populasyondan alınmıştır.

H_1: Bu örnek normal dağılım gösteren bir populasyondan alınmamıştır.

Kolmogorov-Smirnov (KS) test istatistiği;

$D = \max | S_n(X) - F_0(X) |$ ifadesine göre hesaplanır. Burada,

$S_n(X)$, örnekten gözlenen X değişkenin birikimli dağılım fonksiyonudur.

$F_0(X)$, X tesadüf değişkenin kümülatif (birikimli) teorik (beklenen) dağılım fonksiyonudur.

Buradan hesaplanan test istatistiği tablo değerinden büyükse ilgili H_0 hipotezi α önem seviyesinde reddedilir ve gözlenen verilerin normal dağılıma uygun olmadığı sonucuna varılır.

Kolmogorov-Smirnov test istatistiği aşağıdaki gibi de yazılabilir.

$KS = \max [D^+, D^-]$

$D^+ = \max [i / n – u_i]$

$D^{-} = \max [(i-1) / n - u_i]$

Burada;

$i = 1, 2, \ldots n$,

n:örnekteki toplam gözlem sayısı,

$u_i = F_0(X)$ eklemeli (kümülatif) N (0, 1) dağılım fonksiyonunu gösterir.

Eklemeli (kümülatif) dağılım fonksiyonuna ait değerleri hesaplamak için örneğe ait gözlem değerleri küçükten büyüğe doğru sıralanır. Daha sonra gözlemlere ait Z değerleri ve bu Z değerlerine ait eklemeli (kümülatif) olasılıklar hesaplanır.

4.3 Lilliefors Testi

Kolmogorov-Smirnov testi, eğer dağılım tamamıyla belirli ise (sadece normal değil ayrıca ortalaması ve standart sapması ile biliniyorsa) doğru bir şekilde uygulanabilir. Populasyonun bazı parametreleri örnekten tahmin edildiğinde maksimum sapmanın dağılımı bilinmez (Massey, 1951). Bir başka deyişle dağılımın bazı parametreleri örnekten tahmin edildiğinde Kolmogorov-Smirnov testi kullanılamaz, en azından çoğunlukla kritik tablo değerleri kullanılmaz. Kolmogorov-Smirnov test istatistiğinin hesaplanmasında, populasyonun ortalaması ve standart sapması bilinmediğinde örnek ortalaması ve standart sapması kullanılır o zaman tablo değeri olarak karşılaştırmada Lilliefors'un tablo değerleri kullanılır (Lilliefors, 1967).

Lilliefors testine ilişkin hipotez takımı aşağıdaki gibi kurulabilir.

H_0: Bu örnek normal dağılım gösteren bir populasyondan alınmıştır.

H_1: Bu örnek normal dağılım gösteren bir populasyondan alınmamıştır.

4.4 Anderson-Darling Testi

Anderson and Darling (1954) tarafından önerilen bu test Komogorov-Smirnov testi gibi ampirik kümülatif dağılım fonksiyonunu temel alır.

Anderson-Darling istatistiğini (A^2) hesaplamak için, örnek değerleri küçükten büyüğe sıralanır ve standardize edilir. Kümülatif Gaussian fonksiyon değeri i. inci değişken için z_i olarak ifade edilir ve

$A^2 = -n - \frac{1}{n}\left\{\sum_{i=1}^{n}(2i-1)\left[\ln z_i + \ln(1 - z_{n+1-i})\right]\right\}$ şeklinde hesaplanır.

Stephens (1974, 1976), Anderson-Darling testi ampirik dağılım fonksiyonunu modifiye etmiştir. Modifiye edilmiş test istatistiği, $T^*=(1+0.75/n+2.25/n^2)A^2$ şeklindedir. Üst kuyruk olasılığı $\alpha=\exp(1.2937–5.709T^2+0.0186T^{*2})$ şeklinde hesaplanır. α_0 belirlenen önem seviyesi olmak üzere, eğer $\alpha< \alpha_0$ ise ilgili H_0 hipotezi reddedilir (Zhang 1999).

Anderson-Darling testine ilişkin hipotez takımı aşağıdaki gibi kurulabilir.

H_0: Bu örnek normal dağılım gösteren bir populasyondan alınmıştır.

H_1: Bu örnek normal dağılım gösteren bir populasyondan alınmamıştır.

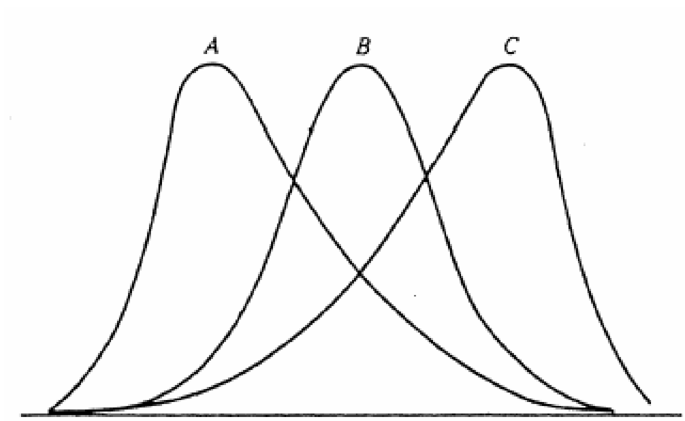
4.5 Eğrilik (Skewness) Testi

Eğrilik (Skewness), bir dağılımın onun kendi ortalaması etrafındaki asimetriklik derecesini karakterize eder.

Eğrilik (Skewness), dağılımın simetriden sapması hakkında fikir verir (Kavuncu 1995).

Bir dağılımın eğriliği (simetrikliği), ortalamaya göre üçüncü momentidir. Pozitif eğrilik, daha pozitif değerlere doğru uzayan bir asimetrik kuyruklu bir dağılım gösterir. Negatif eğrilik ise daha negatif değerlere doğru uzayan bir asimetrik kuyruklu bir dağılım gösterir.

Normal dağılım için eğrilik değeri 0'a eşittir. Normal olmayan bir populasyon, merkezi moment değerlerinin normal değerlerden farklılık göstermesi ile tanımlanabilir. Normal dağılım simetriktir. Yani $\sqrt{\beta_1} = 0$ dır. Asimetrik bir normal olmayan dağılım $\sqrt{\beta_1} \neq 0$ değerine sahiptir $\sqrt{\beta_1} \rangle 0$ ilişkisinde eğrilik sağa, $\sqrt{\beta_1} \langle 0$ ilişkisinde ise sola yatıktır (Şekil 4.2), (D'Agostino 1990).



Şekil 4.2.a. $\sqrt{\beta_1} \rangle 0$ b. $\sqrt{\beta_1} = 0$ c. $\sqrt{\beta_1} \langle 0$ durumunda dağılımların şekli (D'Agostino 1990)

D'Agostino (1970, 1986) ve D'Agostino *et al.* (1990) gibi bazı kaynaklarda eğriliğieğriiği göstermek için kullanılan $\sqrt{\beta_1}$ bir populasyon parametresi olarak gösterilmektedir (Zar 1999). $\sqrt{\beta_1}$ değerinin tahmininde örnekten hesaplanan istatistik $\sqrt{b_1}$ 'dir.

Eğrilik (Skewness) değeri,

$m_3 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^3}{n}$ eşitliği ile hesaplanır. Sheskin (2000), eğrilik (Skewness) değeri için m_3'le parametre tahmininde aşağıdaki eşitliklerin sapmasız tahmin sağladığını bildirmiştir.

$$m_3 = \frac{n\sum_{i=1}^{n}(X_i - \overline{X})^3}{(n-1)(n-2)}$$

$$m_3 = \frac{n\sum_{i=1}^{n}X_i^{\,3} - 3\sum_{i=1}^{n}X_i \sum_{i=1}^{n}X_i^{\,2} + \frac{2(\sum_{i=1}^{n}X_i)^3}{n}}{(n-1)(n-2)}$$

Eğrilik (Skewness) değerini hesaplamak için, minimum örnek genişliği n=3 olmalıdır. m_3 için hesaplanan değer üçüncü dereceden birimli olması nedeniyle, eğriliği belirtmek için genellikle populasyon parametresi γ_1'in bir tahmini olan birimsiz istatistik g_1 kullanılır.

g_1 değeri; $g_1 = \frac{m_3}{S^3}$

Burada;

$$S = \sqrt{\frac{\sum_{i=1}^{n}(Xi - \overline{X})^2}{n-1}}$$ 'dir.

Ortalama etrafında simetrik bir dağılımda g_1 değeri 0'a eşit olacaktır. g_1 değeri 0'dan önemli bir şekilde büyükse sağa yatık ve önemli bir şekilde küçük olursa sola yatık bir dağılım olacaktır. Normal dağılım simetrik olmasına rağmen ($g_1 = 0$ ile), bütün simetrik dağılımlar normal değildir. Simetrik olan t dağılımı ve $\pi_1 = .5$ olması durumunda binomiyal dağılım normal olmayan dağılım örnekleridir (Sheskin 2000).

$g_1 = \gamma_1 = 0$ ya da $\sqrt{b_1} = \sqrt{\beta_1} = 0$ durumunda olan normal dağılım her zaman simetrik olacaktır.

(D'Agostino 1970 ve D'Agostino *et al.* 1990), g_1 değerinin $\sqrt{b_1}$ istatistiğine dönüştürülmesi aşağıdaki şekilde verilmiştir (Sheskin 2000).

$$\sqrt{b_1} = \frac{(n-2)g_1}{\sqrt{n(n-1)}}$$

Eğrilik değeri için hipotez takımı;

H_0: Dağılım, normal dağılıma uygundur.

H_0: $\gamma_1 = 0$ ya da $\sqrt{\beta_1} = 0$

H_1: Dağılım, normal dağılıma uygun değildir.

H_1: $\gamma_1 \neq 0$ ya da $\sqrt{\beta_1} \neq 0$

şeklinde kurulabilir.

Simetrik bir dağılımda $\sqrt{b_1}$ değeri 0'a eşittir. $\sqrt{b_1}$ değeri 0'dan önemli bir şekilde büyükse sağa yatık ve önemli bir şekilde küçük olursa sola yatık bir dağılım olacaktır.

Tek örnekte populasyonun eğriliğinin değerlendirilmesinde, g_1 ve/veya $\sqrt{b_1}$ değerinin 0'dan önemli bir şekilde farklı olup olmadığı belirlenir. Burada $\sqrt{b_1}$'in dağılımı Johnson dönüştürmesi ile ortalaması sıfır ve varyansı bir olan standart normal dağılıma dönüştürülür (D'Agostino 1970). Bu dönüştürme aşağıdaki şekilde yapılır.

$$Y = \sqrt{b_1}\sqrt{\frac{(n+1)(n+3)}{6(n-2)}},$$

$$\beta_2(\sqrt{b_1}) = \frac{3(n^2+27n-70)(n+1)(n+3)}{(n-2)(n+5)(n+7)(n+9)},$$

$$W^2 = -1 + \left\{2(\beta_2(\sqrt{b_1})-1)\right\}^{1/2},$$

$$\delta = 1 / \sqrt{\ln W} \;,$$

$$\alpha = \left\{ 2 / \left(W^2 - 1 \right) \right\}^{1/2},$$

$$Z\left(\sqrt{b_1}\right) \;=\; \delta \ln\left(Y / \alpha + \left\{ \left(Y / \alpha \right)^2 + 1 \right\}^{1/2} \right).$$

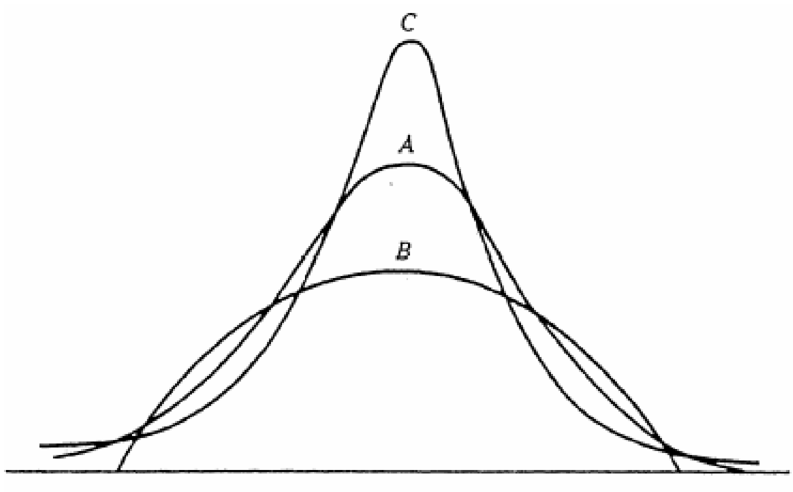
Sheskin (2000), n ≥ 9 olduğunda $Z\left(\sqrt{b_1}\right)$ değeri, $\sqrt{b_1}$ 'in örnekleme dağılımı için iyi bir tahmin sağladığını bildirmiştir. $Z\left(\sqrt{b_1}\right)$ istatistiği, Z dağılımı gösterir ve $Z_{hesap} > Z_{tablo}$ olduğunda ilgili test hipotezi reddedilir.

4.6 Diklik (Kurtosis) Testi

Diklik, normal dağılım ile mukayese edilen bir dağılımın nispi diklik ve yassılığını karakterize eder. Pozitif diklik, sivri bir şekilde pik bir dağılıma sahiptir. Negatif diklik ise yayvan bir şekilde yassı bir dağılıma sahiptir.

Diklik, yani ortalamaya göre dördüncü moment, dağılımın sivriliği, yani diklik ve basıklık hakkında fikir verir (Kavuncu 1995).

Normal dağılım için diklik değeri 3'e eşittir. Şekil 4.3, $\beta_2 \neq 3$ durumunda normal olmayan bir dağılım göstermektedir. $\beta_2 \rangle 3$ durumunda dağılımın merkezinde daha yüksek pik yapma eğilimindedir ve leptokurtic olarak tanımlanırlar. $\beta_2 \langle 3$ olması durumunda bunların yaptığı pik bakımından ise normalden daha yayvan olma eğilimindedir ve bu dağılımlar platykurtic olarak tanımlanır.



Şekil 4.3.a. $\beta_2 = 3$ b. $\beta_2 \langle 3$ c. $\beta_2 \rangle 3$ A,B,C durumunda dağılımların şekli (D'Agostino 1990)

Anscombe and Glynn (1983), D'Agostino (1986) ve D'Agostino *et al.* (1990) gibi bazı kaynaklarda dikliği göstermek için kullanılan β_2 in populasyon parametresi olarak gösterilmektedir (Zar 1999).

β_2 değerinin tahmininde örnekten hesaplanan istatistik b_2'dir. $g_2 = 0$ ve $b_2 = 3$ olan bir dağılım her zaman normal olacaktır. Yani $g_2 = \gamma_2 = 0$ ya da $b_2 = \beta_2 = 3$ durumundaki dağılım her zaman normal olacaktır.

Diklik (Kurtosis) değeri;

$$m_4 = \frac{\sum_{i=1}^{n} (Xi - \overline{X})^4}{n}$$

eşitliği ile elde edilir. Sheskin (2000), diklik diklik değeri için m_4'le parametre tahmininde aşağıdaki eşitliklerin sapmasız tahmin sağladığını bildirmiştir.

$$m_4 = \frac{[[\sum (X - \overline{X})^4 (n)(n+1)]/(n-1)] - 3[\sum (X - \overline{X})^2]^2}{(n-2)(n-3)}$$

$$m_4 = \frac{(n^3+n^2)\sum X^4 - 4(n^2+n)\sum X^3 \sum X - 3(n^2-n)(\sum X^2)^2 + 12n\sum X^2(\sum X)^2 - 6(\sum X)^4}{n(n-1)(n-2)(n-3)}$$

Diklik (Kurtosis) değerini hesaplamak için minimum örnek genişliği n = 4 olmalıdır.

m_4 için hesaplanan değer dördüncü dereceden birimli olması nedeniyle dikliğini belirtmek için, genellikle populasyon parametresi γ_2'in bir tahmini olan birimsiz istatistik g_2 kullanılır. g_2 değeri;

$g_2 = \frac{m_4}{S^4}$ eşitliği ile hesaplanır.

Mesokurtic (normal) bir dağılımda değeri g_2 nin 0'a eşit olacaktır. g_2 değeri 0'dan önemli bir şekilde büyükse leptokurtic ve önemli bir şekilde küçük olursa platykurtic bir dağılım olacaktır.

D'Agostino *et al.* 1990, g_2 değerinin b_2 istatistiğine dönüştürülmesi aşağıdaki şekilde verilmiştir (Sheskin 2000).

$$b_2 = \frac{(n-2)(n-3)g_2}{(n+1)(n-1)} + \frac{3(n-1)}{n+1}$$

Tek örnekte populasyon dikliğinin değerlendirilmesinde, g_2 ve/veya b_2 değerinin, g_2=0, b_2 = 3 değerlerinden önemli bir derecede farklı olup olmadığı belirlenir. Bunun için kullanılan Z(b_2) istatistiğidir. Burada Z(b_2) istatistiğini hesaplamak için aşağıdaki adımlar izlenir.

1. b_2'nin değeri hesaplanır.

2. b_2'nin ortalaması ve varyansı hesaplanır.

$$E(b_2) = \frac{3(n-1)}{n+1}, \qquad var(b_2) = \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)},$$

3. b_2'nin standardize edilmiş değeri hesaplanır.

$$x = (b_2 - E(b_2)) / \sqrt{\mathrm{var}(b_2)} = \frac{(n-2)(n-3)\,|\,g_2\,|}{(n+1)(n-1)\sqrt{\mathrm{var}(b_2)}},$$

4. b_2'nin standardize edilmiş üçüncü momenti hesaplanır.

$$\sqrt{\beta_1(b_2)} = \frac{6(n^2-5n+2)}{(n+7)(n+9)}\sqrt{\frac{6(n+3)(n+5)}{n(n-2)(n-3)}},$$

5. $$A = 6+\frac{8}{\sqrt{\beta_1(b_2)}}\left[\frac{2}{\sqrt{\beta_1(b_2)}}+\sqrt{1+\frac{4}{\beta_1(b_2)}}\right],$$

6. $$Z(b_2) = \left(\left(1-\frac{2}{9A}\right)-\left[\frac{1-2/A}{1+x\sqrt{2/(A-4)}}\right]^{1/3}\right)/\sqrt{2/(9A)}.$$

Sheskin (2000), $n \geq 20$ olduğunda $Z(b_2)$ değeri b_2'nin örnekleme dağılımı için iyi bir tahmin sağladığını bildirmiştir. $Z(b_2)$ istatistiği, Z dağılımı gösterir ve $Z_{hesap} > Z_{tablo}$ olduğunda ilgili test hipotezi reddedilir.

Diklik değeri için hipotez takımı;

H_0: Dağılım, normal dağılıma uygundur.

H_0: $\gamma_2 = 0$ ya da $\beta_2 = 3$

H_1: Dağılım, normal dağılıma uygun değildir.

H_1: $\gamma_2 \neq 0$ ya da $\beta_2 \neq 3$

şeklinde kurulabilir.

4.7 D' Agostino Pearson Testi

D' Agostino Pearson testi, D' Agostino Pearson *et al.* (1990) tarafından önerilmiştir.

Test istatistiği,

$K^2 = Z\left(\sqrt{b_1}\right)^2 + Z(b_2)^2$ şeklinde hesaplanır.

$Z\left(\sqrt{b_1}\right)$ ve $Z(b_2)$ istatistikleri, eğrilik ve diklik testleri normale dönüştürülmüşlerdir (Seier 2002). K^2 istatistiği, iki serbestlik dereceli ki-kare ($X^2_{(2)}$) dağılımına sahiptir (D'Agostino 1990).

D' Agostino Pearson testi için hipotez takımı;

H_0: Dağılım, normal dağılıma uygundur.

H_1: Dağılım, normal dağılıma uygun değildir.

şeklinde kurulabilir.

4.8 Jargue-Bera Testi

Jargue-Bera testi, 1980 yılında Jargue ve Bera tarafından önerilmiştir. Bu test, normal dağılım olduğu varsayılan $\{x_1, x_2,...,x_n\}$ veri setinden hesaplanan eğrilik ve diklik katsayıları arasındaki farklılığı temel alır (Dong and Giles 2004).

JB istatistiği, H_0 hipotezi altında $X^2_{(2)}$ iki serbestlik dereceli asimptotik ki kare dağılımına sahiptir (Dong and Giles 2004).

Bağımsız bir populasyondan $\{x_1, x_2,...,x_n\}$ rasgele çekilmiş bir örnek olduğu kabul edilirse, j' inci merkez örnek momenti (ortalamaya göre j' inci moment);

$m_j = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^j$, j=2,3..., olmak üzere burada;

$\bar{x}$, örnek ortalaması, bu yüzden $\hat{\sigma}^2 = m_2$ şeklindedir. Pearson'ın eğrilik (skewness) ve diklik (kurtosis) istatistikleri,

$\sqrt{b_1} = \frac{m_3}{\hat{\sigma}^3}$, $b_2 = \frac{m_4}{\hat{\sigma}^4}$ olmak üzere,

JB istatistiği, $JB = n\left[\frac{(\sqrt{b_1})^2}{6} + \frac{(b_2 - 3)^2}{24}\right]$ şeklindedir.

x' ler normal dağılımdan alındığında n örnek genişliği arttıkça JB istatistiği iki serbestlik dereceli ki kare dağılımına yaklaşır (Cho and Im 2002).

JB değeri için hipotez takımı;

H_0: Dağılım, normal dağılıma uygundur.

H_1: Dağılım, normal dağılıma uygun değildir.

şeklinde kurulabilir.

4.9 Shapiro-Wilk Testi

İlk olarak Shapiro ve Wilk (1965) tarafından önerilen bu test, örneğe ait sıra istatistiklerinin uygun bir lineer bileşeninin karesinin, kareler toplamına bölümüyle elde edilir.

Shapiro-Wilk testi (W) için test istatistiği;

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - x)^2}$$

Burada;

$$a' = (a_1,\ a_2,\ \ldots\ a_n) = \frac{m'V^{-1}}{(m'V^{-1}V^{-1}m)^{1/2}}$$

x_i = (x_1,xy_2, ...xy_n): küçükten büyüğe doğru sıralı gözlem değerleri,

$\bar{x}$: gözlemlerin aritmetik ortalaması,

m′= (m_1,m_2, ... m_n): standart normal dağılımda N (0, 1) n adet sıra istatistiğinin beklenen değerlerinin vektörü,

V = (V_{ij}): nxn boyutlu varyans kovaryans matrisidir.

Shapiro-Wilk test istatistiğinin hesaplanmasında a' katsayısının hesaplanmasındaki zorluklar nedeniyle 2'den, 50'ye kadar a' katsayı değerlerini tablo halinde vermişlerdir. Bu tabloda a' değerlerini kullanarak b katsayısı ve W istatistiğinin hesaplanması aşağıda gösterilmiştir.

(i) Gözlemler küçükten büyüğe doğru sıralanır. $x_1 \leq x_2 \leq ... \leq x_n$

(ii) $\Sigma dx^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$ hesaplanır.

(iii) Eğer n çift sayı ise (n=2k);

$$b = \sum_{i=1}^{k} a_{n-i+1}(x_{n-i+1} - x_i)$$

Eğer n tek sayı ise (n=2k+1), burada a_{k+1}=0'dır.

$$b = a_n\ (x_n - x_1) + \ldots + a_{k+2}\ (x_{k+2} - x_k)$$

Burada; x_{k+1} yani örnek medyanı b'nin hesaplanmasına katılmaz.

(iv) W istatistiği;

$$W = \frac{b^2}{\sum d_x^2}$$ eşitliği ile hesaplanır.

(v) Hesaplanan W istatistiği $W_{hesap} < W_{tablo}$ olduğunda ilgili test hipotezi reddedilir (Shapiro-Wilk 1965).

Hipotez takımı,

H_0 : Gözlemler normal dağım gösteren bir populasyondan alınmıştır.

H_1 : Gözlemler normal dağım gösteren bir populasyondan alınmamıştır.

şeklinde kurulabilir.

5. TESTLERİN BİRBİRLERİNE OLAN BAZI ÜSTÜNLÜKLERİ VE SAKINCALARI

Kolmogorov-Smirnov istatistiği bir dizi gözlemin tam olarak belirtilmiş sürekli dağılımlardan, Fo(X) alınıp alınmadığını sınayan bir test yöntemi sunar. Bunun için genelde Ki-kare testi bir başka seçenek olabilir. Kolmogorov-Smirnov testi Ki-kare testiyle karşılaştırıldığında en az iki önemli üstünlüğe sahiptir.

1. Küçük örnek genişliklerinde Ki-kare testinin geçerliliğinin kuşku uyandırmasına karşın, Kolmogorov-Smirnov testi uygulanabilir,
2. Genelde bütün örnek genişliklerinde Ki-kare testi ile karşılaştırıldığında çok daha güçlü bir test olarak bilinmektedir (Lilliefors 1967).

Bunun dışında örnek genişliği çok düşük olduğu durumlarda frekans dağılım tablosu yapılamayacağından ki-kare testi uygulanamaz. Çünkü ki-kare testi gruplandırılmış veri gerektirir. Bu durumda Kolmogorov-Smirnov testi uygulanabilir.

Herhangi bir sınıfta yığılma varsa ki-kare bundan çok etkilenir. Kolmogorov-Smirnov testi bu durumda daha güvenilirdir.

χ^2 testinin uygulanabilmesi için beklenen frekansların 5 ve 5'ten büyük olması istenir. Kolmogorov-Smirnov testi böyle bir şarta dayanmadığı için kolayca uygulanabilmektedir. Ki-Kare testinde beklenen frekansların 5 ve 5'ten büyük olması için ya örneklerin büyük hacimli olması gerekir (bu masraflı bir iştir), ya da sınıflar birleştirilmek suretiyle beklenen frekansların 5'den büyük olması sağlanır. Bu durumda bilgi kaybı söz konusudur. Oysaki Kolmogorov-Smirnov testinde beklenen frekanslar için bir alt limit söz konusu değildir (Bircan vd. 2003).

Klasik Eğrilik ve Diklik katsayıları ortak olumsuz özelliklere sahiptir. Her ikisi de 0' (sıfır) da bozulma değerine sahiptir ve bu testler aykırı (outlier) değerlere karşı çok duyarlıdır. Bir tek aykırı (outlier) sayı bu testlerin değersiz olmasına neden olabilir ve bunun yorumlanması çok zordur (Brys *et al*. 2004).

JB testi, normallik testi için çok iyi güç özelliklerine sahip bir test olarak bilinir ve bu testin hesaplanması oldukça kolaydır. Bu testin bir sınırlayıcı özelliği sadece normallik testi için dizayn edilen bir test olmasıdır (Dong and Giles 2004).

Moment testlerinden olan Eğrilik, Diklik ve D'agostino Pearson testlerinin hesaplanabilmesi için örneklerdeki gözlem adetleri $n \geq 8$ olmalıdır. Bir diğer moment testi olan Jargue-Bera testinde ise daha küçük örnek genişlikleri için bu test istatistiği hesaplanabilse de I. tip hata ve testin gücü bakımından aldığı değerler itibariye sağlıklı sonuçlar vermemektedir. Ancak Ampirik kümülatif dağılım fonksiyonunu temel alan Kolmogorov-Smirnov, Lilliefors ve Anderson-Darling testlerinde buna benzer problemler bulunmamaktadır.

Örneklerin normal dağılıma uyumunu test etmek için W istatistiğinin hesaplanmasında lineer katsayılar yani *a* katsayılar tablosu mevcut olduğu için bu katsayılardan yararlanarak test istatistiğini hesaplamak çok kolaydır. Shapiro-Wilk testi $n<20$ küçük örnekler için bile normalliği test etmede kullanılan birçok alternatif testten daha hassastır. Shapiro-Wilk testi ile ilgili sakınca büyük örnekler için lineer katsayılar yani *a* katsayılarının hesaplanması hem de büyük örnekler için dağılımın yüzde noktalarının belirlenmesi çok zordur (Shapiro-Wilk 1965).

W 'nun hipotez dağılımının lineer katsayıları ve seçilen yüzdelik oranları örnek genişliği 3 'den 50' ye kadar tablo haline getirilmiştir.

Ancak, Royston (1982), W' in kullanımında örnek genişliğini 3' den 2000'e kadar genişletmiştir. Bu değişimde W için, (a_i) lineer katsayıları ve standart normal transformasyonuna yakın bir formül kullanmıştır.

Stephens (1974), ampirik güç çalışmalarında Anderson-Darling testi ile diğer testleri normallik bakımından karşılaştırdığını bildirmiştir. Yaptığı çalışmada Anderson-Darling ve Shapiro-Wilk testlerinin güçlü olduklarını ve bu iki test arasında güç bakımından çok küçük farklılıklar olduğunu bildirmesine rağmen W istatistiğinde örnek genişlikleri için farklı katsayılara ihtiyaç duyulurken Anderson-Darling istatistiğinin daha kolay hesaplanabildiğini bildirmiştir (Pettitt 1977).

6. ARAŞTIRMA BULGULARI

6.1. I. Tip Hata Bakımından Testlerin Karşılaştırılması

Normal dağılımdan 7,8,9,10,20,30,40,50 ve 100 örnek genişliklerinde çalışılan normallik testleri için gözlenen I. tip hata oranları Çizelge 6.1' de verilmiştir.

Çizelge 6.1 Normal dağılımdan elde edilen I. tip hata oranları

N	Eğrilik	Diklik	DP	J-B	K-S	Lilliefors	AD	Ki-Kare	W
7				0,00000	0,05291	0,05741	0,04966		0,05014
8	0,05283	0,03545	0,06061	0,00250	0,05156	0,05499	0,04955		0,04762
9	0,05267	0,03890	0,05917	0,00592	0,05061	0,05492	0,04908		0,04823
10	0,05214	0,03985	0,05839	0,00833	0,05100	0,05835	0,05035		0,05060
20	0,05075	0,04687	0,05725	0,02483	0,04828	0,05431	0,05016	0,04725	0,04933
30	0,04898	0,05017	0,05634	0,02979	0,04909	0,04393	0,04933	0,04839	0,04797
40	0,04898	0,05178	0,05668	0,03402	0,04887	0,04504	0,04932	0,04829	0,04896
50	0,04924	0,05355	0,05865	0,03788	0,04980	0,04765	0,05038	0,04886	0,05075
100	0,05044	0,05367	0,05634	0,04258	0,05034	0,05145	0,04909	0,04868	0,05032

Çizelge 6.1 incelendiğinde, bütün örnek genişliklerinde gerçekte örneklerin normal dağılım gösteren bir populasyondan alındıklarında ilgili H_0 hipotezini reddetme olasılığı yani; I. tip hata olasılığı J-B testinde diğer testlere göre en düşük çıkmıştır. Örneklerdeki gözlem sayıları n=7,8,9,10,20 olduğunda J-B testini, Diklik testi izlemiştir. Örnek genişlikleri n=8,9,30,40 olduğunda W testi, n=30,40 olduğunda ise Lilliefors, Ki-kare, Eğrilik, K-S ve AD testleri, örnek genişliği n=50 için de Lilliefors, Eğrilik, K-S ve Ki-kare testleri deneme başında kararlaştırılan I.tip hata olasılığından daha düşük değerler almıştır. AD ve W testleri ise kararlaştırılan I.tip hata olasılığında değerler alırken Diklik ve DP testleri kararlaştırılan I.tip hata olasılığından büyük değerler almıştır. Örneklerdeki gözlem sayıları n=100 olduğunda ise Eğrilik, K-S ve W testleri

kararlaştırılan I.tip hata olasılığında değer alırken, AD ve Ki-kare testleri kararlaştırılan I.tip hata olasılığından daha düşük değerlere sahiptir. DP testi bütün örnek genişliklerinde kararlaştırılan I.tip hata olasılığından daha büyük değerler aldığı gözlemlenmiştir.

6.2. Testin Gücü Bakımından Yöntemlerin Karşılaştırılması

Normallik testlerinin 1 serbestlik dereceli ki-kare dağılımından alınan farklı örnek genişliklerindeki örnekler için gözlenen güç değerleri Çizelge 6.2'de verilmiştir.

Çizelge 6.2 1 s.d.'li ki-kare dağılımından elde edilen güç değerleri

N	Eğrilik	Diklik	DP	J-B	K-S	Lilliefors	AD	Ki-Kare	W
7				0,01158	0,37647	0,38953	0,48511		0,51564
8	0,44170	0,30280	0,41058	0,14224	0,43253	0,44223	0,56102		0,58773
9	0,50742	0,32417	0,44593	0,21326	0,48645	0,50029	0,63440		0,66418
10	0,56238	0,34372	0,49591	0,27056	0,53411	0,55548	0,69556		0,72843
20	0,89446	0,55086	0,80888	0,71738	0,88217	0,89207	0,96932		0,98391
30	0,98091	0,69934	0,94182	0,91898	0,98198	0,97945	0,99834	0,34338	0,99962
40	0,99713	0,80048	0,98863	0,98463	0,99839	0,99825	0,99999	0,44878	0,99999
50	0,99964	0,86534	0,99897	0,99985	0,99839	0,99983	1,00000	0,56186	1,00000
100	1,00000	0,98308	1,00000	1,00000	1,00000	1,00000	1,00000	0,91918	1,00000

Çizelge 6.2 incelendiğinde 1 serbestlik dereceli ki-kare dağılımı gösteren populasyondan alınan örneklerde örnek genişliği arttıkça testlerin güçleri % 100 e kadar çıkmıştır. W testinin bütün örnek genişliklerinde diğer testlerden daha güçlü olduğu ya da eşit değerler aldığı görülmektedir. Bunu AD, Eğrilik, Lilliefors, K-S ve DP testi izlemektedir. Örnek genişliği $n \geq 20$ durumu için W, AD, Eğrilik, Lilliefors, K-S ve DP testlerinin güç performansları % 80 'nin üzerindedir. J-B testi örnek genişliği n=30'dan

itibaren güçlü bir test olarak ortaya çıkmaktadır. Ki-kare testinin örnek genişliği n=50 olduğu durumda bile güç bakımından zayıf bir test olduğu görülmektedir.

Normallik testlerinin 5 serbestlik dereceli ki-kare dağılımından alınan farklı örnek genişliklerindeki örnekler için gözlenen güç değerleri Çizelge 6.3'de verilmiştir.

Çizelge 6.3 5 s.d'li ki-kare dağılımından elde edilen güç değerleri

N	Eğrilik	Diklik	DP	J-B	K-S	Lilliefors	AD	Ki-Kare	W
7				0,00009	0,11030	0,11755	0,12469		0,13115
8	0,15643	0,10281	0,15954	0,01893	0,12242	0,12780	0,14509		0,15078
9	0,17211	0,11262	0,16875	0,04109	0,13300	0,14089	0,16309		0,17318
10	0,19427	0,12324	0,18453	0,06174	0,14556	0,15813	0,18334		0,19938
20	0,40520	0,20150	0,34590	0,24056	0,26649	0,28253	0,37930		0,43725
30	0,59024	0,26360	0,48835	0,40419	0,38477	0,36733	0,55839	0,09725	0,64598
40	0,73433	0,31886	0,61361	0,55307	0,49700	0,48170	0,70054	0,13464	0,80198
50	0,83268	0,36396	0,71901	0,67734	0,59576	0,58677	0,80610	0,16416	0,89800
100	0,99045	0,55028	0,97646	0,97191	0,89101	0,89323	0,98654	0,33600	0,99803

5 serbestlik Ki-kare dağılımında ise Çizelge 6.3'deki sonuçlara göre W testi örnek genişliği $n \geq 40$ dan itibaren güçlü bir test olduğu görülmektedir. Eğrilik ve AD testi $n \geq 50$ durumunda ise W testinden sonra güçlü testlerdir. Örnek genişliği n=100 durumunda testlerin güç performansları W, Eğrilik, AD, DP, J-B, Lilliefors ve K-S şeklinde sıralanmaktadır. Diklik ve Ki-kare testlerinin güç performansı ise diğer testlerden daha zayıf olduğu görülmektedir. W testi, bütün örnek genişliklerinde de söz konusu testlerden daha güçlüdür. Bu sırayı Eğrilik ve AD testleri izlemektedir.

Normallik testlerinin 22 ve 3 parametreli Beta dağılımından alınan farklı örnek genişliklerindeki örnekler için gözlenen güç değerleri Çizelge 6.4'de verilmiştir.

Çizelge 6.4 Beta (22,3) dağılımından elde edilen güç değerleri

N	Eğrilik	Diklik	DP	J-B	K-S	Lilliefors	AD	Ki-Kare	W
7				0,00008	0,08291	0,08846	0,08676		0,09120
8	0,10570	0,06986	0,11143	0,00901	0,08901	0,09368	0,10857		0,10055
9	0,11620	0,07726	0,11753	0,02128	0,09437	0,10017	0,10857		0,11437
10	0,12728	0,08420	0,12531	0,03245	0,09923	0,10944	0,11769		0,12694
20	0,25181	0,12735	0,21832	0,13281	0,16215	0,17535	0,22391		0,26321
30	0,37912	0,15856	0,30665	0,22800	0,22881	0,21485	0,33470	0,10444	0,40717
40	0,50400	0,18202	0,39568	0,32692	0,29977	0,28669	0,44337	0,10348	0,55067
50	0,61071	0,20433	0,47918	0,41884	0,36514	0,35744	0,54096	0,13822	0,67428
100	0,91715	0,28831	0,84183	0,81932	0,66133	0,66541	0,87440	0,51494	0,95583

22 ve 3 parametre değerlerine sahip Beta dağılımında Çizelge 6.4’de görüldüğü üzere testlerin güç performansları sırasıyla W, Eğrilik, AD, DP ve J-B şeklindedir. Burada da W testi güçlü bir test olarak ortaya çıkmaktadır.

Normallik testlerinin 2 ve 5 parametreli Beta dağılımından alınan farklı örnek genişliklerindeki örnekler için gözlenen güç değerleri Çizelge 6.5'de verilmiştir.

Çizelge 6.5 Beta (2,5) dağılımından elde edilen güç değerleri

N	Eğrilik	Diklik	DP	J-B	K-S	Lilliefors	AD	Ki-Kare	W
7				0,00003	0,06670	0,07215	0,06816		0,07119
8	0,07162	0,04878	0,07669	0,00408	0,07052	0,07484	0,07855		0,08079
9	0,07392	0,05374	0,07646	0,00988	0,07235	0,07782	0,07855		0,08079
10	0,07625	0,05506	0,07631	0,01451	0,07640	0,08496	0,08498		0,08819
20	0,13136	0,07896	0,11290	0,05034	0,11480	0,12549	0,14845		0,17019
30	0,19689	0,09618	0,15495	0,08090	0,15889	0,14703	0,22471	0,06854	0,27802
40	0,27011	0,10395	0,19622	0,11594	0,20724	0,19574	0,30874	0,08761	0,41634
50	0,35092	0,10997	0,24805	0,15833	0,25464	0,24728	0,39227	0,10796	0,55523
100	0,70232	0,11350	0,60845	0,50689	0,50455	0,50915	0,75571	0,21144	0,92877

2 ve 5 parametreli Beta dağılımında ise Çizelge 6.5'deki sonuçlara göre W testinin diğer testlerden daha güçlü olduğu görülmektedir. Ancak örnekteki gözlem sayısının n=7 olduğu durumda ise Lilliefors testi daha iyi sonuç vermiştir.

Normallik testlerinin uniform dağılımdan alınan farklı örnek genişliklerindeki örnekler için gözlenen güç değerleri Çizelge 6.6'de verilmiştir.

Çizelge 6.6 Uniform dağılımdan elde edilen güç değerleri

N	Eğrilik	Diklik	DP	J-B	K-S	Lilliefors	AD	Ki-Kare	W
7				0,00001	0,05883	0,06309	0,06282		0,07178
8	0,02754	0,04325	0,03181	0,00128	0,05704	0,06091	0,06712		0,07090
9	0,02429	0,05904	0,02935	0,00212	0,06199	0,06720	0,07326		0,07090
10	0,01978	0,07306	0,02795	0,00197	0,06372	0,07922	0,07922		0,07640
20	0,00637	0,30084	0,15249	0,00059	0,09744	0,10785	0,17184		0,20418
30	0,00399	0,56656	0,39390	0,00029	0,14208	0,12979	0,29534	0,12058	0,42705
40	0,00248	0,76945	0,63392	0,00008	0,19719	0,18382	0,43957	0,15370	0,70191
50	0,00197	0,89037	0,80370	0,00018	0,25764	0,24885	0,57659	0,20785	0,88156
100	0,00126	0,99867	0,99764	0,56155	0,58885	0,59458	0,94913	0,44255	0,99991

Uniform dağılımında ise Çizelge 6.6'deki sonuçlara göre örnek genişliği n= 50 ve daha büyük olduğu durumlarda Diklik, W ve DP testlerinin güç performansları yüksektir ve büyükten küçüğe sırasıyla Diklik, W ve DP testi şeklindedir. Diğer testlerin güç performansları bu testlere göre oldukça zayıftır. Örnek genişliği n= 100 olduğunda ise ilk sırayı W testi almaktadır. Bunu sırasıyla Diklik, DP ve AD testleri izlemektedir.

Normallik testlerinin 10 serbestlik dereceli t dağılımından alınan farklı örnek genişliklerindeki örnekler için gözlenen güç değerleri Çizelge 6.7'de verilmiştir.

Çizelge 6.7 10 s.d. t dağılımından elde edilen güç değerleri

N	Eğrilik	Diklik	DP	J-B	K-S	Lilliefors	AD	Ki-Kare	W
7				0,00004	0,06374	0,06283	0,06283		0,06332
8	0,07844	0,05306	0,08922	0,00525	0,06639	0,07026	0,06637		0,06510
9	0,08221	0,05954	0,09209	0,01302	0,06575	0,07057	0,06939		0,07017
10	0,08549	0,06284	0,09509	0,02019	0,06722	0,07483	0,07079		0,07359
20	0,11453	0,09468	0,12790	0,07691	0,07329	0,08022	0,08915		0,09673
30	0,13325	0,11843	0,15316	0,11692	0,07768	0,07142	0,10016	0,05681	0,10904
40	0,14570	0,13919	0,17286	0,14781	0,08152	0,07579	0,10923	0,05864	0,11247
50	0,15863	0,16178	0,19395	0,17754	0,08834	0,08483	0,11949	0,05919	0,11493
100	0,18958	0,25100	0,27079	0,28547	0,10684	0,10901	0,15862	0,06799	0,09623

10 serbestlik dereceli t dağılımında ise Çizelge 6.6'deki sonuçlara göre $t_{(10)}$ dağılımı simetrik ve aynı zamanda dik bir dağılım olduğundan dolayı bütün örnek genişlerinde bütün testlerin güç performansları düşük çıkmıştır. Ancak Dikliği test edebilen moment testlerinden Diklik, DP ve J-B testleri diğer testlere nazaran daha güçlü çıkmıştır. Örnek genişliği $n \leq 50$ durumunda DP testi diğer testlerden daha iyi bir güç performansı göstermiştir. Ancak, Örnek genişliği n=100 olduğunda ilk sırayı J-B testi almıştır. Bu testi ise sırasıyla DP ve Diklik testleri izlemiştir.

Normallik testlerinin Lognormal dağılımından alınan farklı örnek genişliklerindeki örnekler için gözlenen güç değerleri Çizelge 6.8'de verilmiştir.

Çizelge 6.8 Lognormal dağılımdan elde edilen güç değerleri

N	Eğrilik	Diklik	DP	J-B	K-S	Lilliefors	AD	Ki-Kare	W
7				0,01093	0,31789	0,32943	0,38941		0,40722
8	0,41693	0,30120	0,39880	0,14656	0,37016	0,37903	0,45737		0,47385
9	0,48093	0,33206	0,43948	0,22625	0,41891	0,43073	0,52406		0,54656
10	0,53445	0,35657	0,48599	0,28347	0,46197	0,48088	0,57941		0,60676
20	0,86999	0,59806	0,80210	0,72083	0,79026	0,80237	0,90716		0,93276
30	0,97091	0,75579	0,93431	0,91175	0,93083	0,92511	0,98332	0,37964	0,99205
40	0,99434	0,84298	0,98162	0,97663	0,97946	0,97778	0,99767	0,49454	0,99945
50	0,99909	0,89624	0,99602	0,99501	0,99434	0,99409	0,99962	0,60382	0,99992
100	1,00000	0,98903	1,00000	1,00000	1,00000	1,00000	1,00000	0,93322	1,00000

Çizelge 6.8 incelendiğinde, örneklerdeki gözlem sayıları n=20 durumu için testlerin güç performansları W, AD, Eğrilik, Lilliefors, DP testleri şeklinde sıralanırken n=30 ve n=40 için ise W, AD, Eğrilik, DP, K-S, Lilliefors, J-B, Diklik testleri şeklindeyken n=50 için W, AD, Eğrilik, DP, J-B, K-S, Lilliefors ve Diklik testleri şeklinde sıralanmaktadır. Örneklerdeki gözlem sayısı n=100 olduğunda ise W, AD, Eğrilik, DP, J-B, K-S, Lilliefors testlerinin güç performansları %100' e ulaşmıştır. Diklik testi ve Ki-kare testi ise sırasıyla 0,98903 ve 0,93322 olarak bulunmuştur.

7. TARTIŞMA VE SONUÇ

Simülasyon çalışmasından elde edilen sonuçlara göre normal dağılım gösteren populasyondan çekilen örneklerde, Diklik ve J-B testinde örnek genişliği arttıkça I tip hata oranı artmaktadır. W ve AD testleri ise, kararlaştırılan I. tip hata olasılığında ya da daha altında değerler almıştır. Ki-kare testinde I.tip hata olasılığı kararlaştırılan I. tip hata olasılığının biraz altında bulunmuştur. J-B testi, bütün örnek genişliklerinde normal dağılım için gerçekleşen I. tip hata değerleri en düşük olup, başlangıçta kararlaştırılan I. tip hatayı (α=0,05) koruyamamıştır.

DP testi, her ne kadar gerçekleşen güç değerleri bakımından yüksek değerler almış olsa da, başlangıçta kararlaştırılan I. tip hata oranını 0,05 seviyesinde koruyamadığı, gerçekleşen I. tip hata değerlerinin 0,05 den bir miktar yüksek olduğu (0,05634–0,06061) Çizelge 6.1' de görülmektedir. Bu durum, Seier (2002) çalışmasıyla da benzerlik göstermektedir.

1 serbestlik dereceli ki-kare dağılımından elde edilen sonuçlara göre örneklerdeki gözlem adetleri n=20 olduğunda W, AD, Eğrilik, Lilliefors, K-S ve DP testlerinin güç performansları % 80'nin üzerinde olup sırasıyla 0,98391, 0,96932, 0,89446, 0,89207, 0,88217 ve 0,80888 şeklindedir.

5 serbestlik dereceli ki-kare dağılımında ise örneklerdeki gözlem adetleri n=40 olduğunda testlerin güç gücü bakımından sadece W testi olup % 80'nin üzerinde güç performansına ulaşmakta olup 0,80198 değerini almıştır. Örnek genişliğinin n =50 olduğu durumda ise W, Eğrilik ve AD testleri %80' nin üzerinde güç performansına sahip olup sırasıyla 0,89800, 0,83268 ve 0,80610 değerlerini aldıkları gözlemlenmiştir. Örnek genişliğinin n=100 olduğu durumda ise testlerin güç performansları sırasıyla W, Eğrilik, AD, DP, J-B, Lilliefors ve K-S şeklinde bulunmuştur.

22 ve 3 Beta dağılımında örneklerdeki gözlem adetleri sadece n=100 olduğunda sırasıyla W, Eğrilik, AD, DP ve J-B testleri % 80'nin üzerinde güç performansı göstermektedir.

2 ve 5 Beta dağılımında ise örneklerdeki gözlem adetleri n=100 olduğunda yalnızca W testi 0,92877 güç olasılığı ile en iyi güç değerine sahiptir. Diğer örnek genişliklerinde bütün testlerin güç değerleri % 80'nin altında bulunduğu Çizelge 6.5' de görülmektedir.

Çizelge 6.6'daki sonuçlara göre testlerin güç performansları bakımından örnek genişliğinin ancak n=50 olduğu durumda sırasıyla Diklik, W ve DP testleri, n=100 olduğunda ise sırasıyla W, Diklik, DP ve AD testleri En iyi sonuçları vermiştir.

Lognormal dağılımdaki gözlem adetleri n=30 olduğunda Diklik ve Ki-kare testlerinde gerçekleşen testin gücü değerlerinin düşük olduğu görülmektedir. Diğer testler ise gerçekleşen güç değerleri bakımından yüksek değerler aldığı Çizelge 6.8' de görülmektedir.

Normal olmayan dağılımlardan çekilen örneklerden elde edilen sonuçlara göre özellikle $X^2_{(1)}$, $X^2_{(5)}$, Beta (2,5), Beta (22,3) ve Lognormal gibi simetrik olmayan dağılımlarda W, AD ve Eğrilik testlerinin güç performansları diğer testlerden daha yüksektir. Dağılımların $t_{(10)}$ gibi simetrik ve dik olduğu durumlarda ise testlerin güç performanslarına bakıldığında genel olarak DP, J-B ve Diklik testleri güçlü testler olarak ortaya çıkmaktadır.

Öztuna vd. (2006), yaptıkları çalışmada J-B testi için örnekteki en küçük gözlem adetini n=8 olarak almışlardır, bu çalışmada ise örnekteki en küçük gözlem adetini n=7 olarak alınmış ve elde edilen sonuçların sıfır ya da sıfıra yakın değerler olarak bulunmuştur. Ayrıca J-B testinin bütün dağılımlar için küçük ve orta hacimdeki örnek genişliklerinde güç bakımından zayıf olduğu gözlemlenmiştir.

Bu çalışmada, ki-kare testini dağılımların hepsinde güç bakımından en zayıf test olduğu sonucuna varılmıştır.

Ampirik kümülatif dağılım fonksiyonuna dayanan K-S, Lilliefors, AD ve Ki-kare testleri birbirleriyle karşılaştırıldığında AD testi, dağılımların hepsinde diğer testlerden daha güçlüdür.

Moment testleri olan Eğrilik, Diklik, DP ve J-B testleri birbirleri ile karşılaştırıldığında ise, $X^2_{(1)}, X^2_{(5)}$, Beta (2,5), Beta (22,3) ve Lognormal gibi simetrik olmayan dağılımlarda, Eğrilik testinin diğer moment testlerinden daha güçlü olduğu, bunu sırasıyla DP, J-B ve Diklik testleri izlediği gözlemlenmektedir. Uniform dağılımda ise sıralama Diklik, DP ve J-B ve Eğrilik testi şeklinde olmaktadır. t dağılımında, örnek genişliği $n \leq 50$ durumunda ilk sırayı DP testi almaktadır. Bunu J-B testi izlemektedir. Örnek genişliğinin $n > 50$ durum ise sıralama J-B, DP, Diklik ve Eğrilik testleri şeklindedir.

Sonuç olarak, normal ve normal olmayan dağılımlar birlikte göz önünde tutulduğunda W testinin bütün dağılımların genelinde diğer testlere göre en iyi sonuçları verdiği, bunun yanı sıra AD, Eğrilik ve DP testlerinin de güçlü testler olduğu sonucuna varılmıştır.

KAYNAKLAR

Arcones, M. A. and Wang, Y. 2005. Some new tests for normality based on U-processes. Statistics&Probability Letter, 76(1); 69-82

Bircan, H., Karagöz, Y. ve Kasapoğlu, Y. 2003. Ki-kare ve Kolmogorov-Smirnov Uygunluk Testlerinin Simulasyon ile Elde Edilen Veriler Üzerinde Karşılaştırılması, Cumhuriyet Üniversitesi, İktisadi ve İdari Bilimler Dergisi, Cilt 4, Sayı 1

Brys, G., Hubert, M., and Struyf, A. 2004 A Robustification of the Jargue-Bera of Normality COMSTAT'2004 Symposium, Physica-Verlag/Springer

Cho, D.W. and Im, K.S. 2002. A Test of Normality Using Geary's Skewness and Kurtosis Statistics. http://www.bus.ucf.edu/wp/content/archives/normalit.PDF. Erişim Tarihi: 15.03.2007

D'Agostino, R.B. 1970. Transformation to Normality of the Null Distribution of $g_{1,}$ Biometrika, 57 (3): 679-681.

D'Agostino, R.B., Belanger, A. and D'Agostino, R.B. Jr. 1990. A Suggestion for using Powerful and Informative Tests of Normality, The American Statistician, 44: 316-321.

Dong, L.B. and Giles, D.E. A. 2004. An Empirical likelihood Ratio Test For Normality, Department of Economics, University of Victoria Victoria, B.C., Canada V8W 2Y2,web.uvic.ca/econ/ewp0401.pdf

Kavuncu, O., 1995. İstatistik Terorisi ve Teorik Dağılımlar, T.C. Ziraat Bankası Matbaası, 179, Ankara.

Kesici, T. ve Kocabaş, Z. 1998. Biyoistatistik, Ankara Üniversitesi Eczacılık Fakültesi Yayın No:79, Ankara Üniversitesi Basımevi, Ankara

Keskin, S. 2006. Comparison of Several Univarite Normality Tests Regarding Type I Error Rate and Power of the Test in Simulation based Small Samples, Journal of Applied Science Research 2(5): 296-300

Lilliefors, H. W. 1967. Journal of the American Statistical Association, 62 (318): 399-402.

Linnet, K. 1988. Testing Normality of Transformed Data. Applied Statistics, 2: 180-186.

Massey, F. J. Jr. 1951. The Kolmogorov-Smirnov Test for Goodness of fit, Journal of the American Statistical Association, 46 (253): 68-78

Mendes, M. and Pala A. 2003. Type I error rate and power of three normality tests. Pak. J. Inform. Technol., 2: 135-139.

Öztuna, D., Elhan A.H. and Tüccar, E. 2006. Investigation of Different Normality Tests in Terms of Type 1 Error Rate and Power under Different Distributions. Turk J Med Sci, 36 (3):171-176.

Pearson, E.S., D'Agostino, R.B. and Bowman, K.O. 1977. Tests for Departure from Normalty: Comparison of Powers. Biometrika, 64 (2): 231-246.

Pettitt, A.N. 1977. Testing the Normality of Several Independent Samples Using the Anderson Darling Statistics. Applied Statistics, 26 (2): 156-161.

Poitras, G. 2005. More on the Correct Use of Ombnibus Tests for Normality, website:sfu.ca/~poitras

Rahman, M.M. and Govindarajulu, Z. 1997 A Modification of the test of Shapiro and Wilk for Normality Journal of Applied Statistics, 24 (2): 219-235

Royston, J.P. 1982. An Extension of Shapiro and Wilk's W Test for Normality to Large Samples Appl. Statist. 31 (2): 115-124

Saniga, E.M. and Miles, J.A. 1979. Power of Some standart Goodness-of-Fit Tests of Normality Against Asymetric Stable Alternatives. Journal of the American Statistical Association, 74 (368): 861-865.

Seier, E. 2002. Comparison of Tests for Univariate Normality http://interstat.statjournals.net/YEAR/2002/articles/0201001.pdf. Erişim Tarihi: 11.12.2006

Shapiro, S.S. and Wilk, M.B. 1965. An Analysis of Variance Test for Normality (Complete Samples), Biometrika, 52, Vol.3/4. 591-611.

Shapiro, S.S., Wilk, M.B. and Chen, H.J. 1968. A Comparative Study of Various Tests for Normality. Journal of the American Statistical Association, 63 (324):1343-1372.

Shapiro, S.S. and Francia, R.S. 1972. An Approximate Analysis of Variance Test for.Normality. Journal of the American Statistical Association, 67 (337): 215-216.

Sheskin, D.J. 2000. Hand of book Parametrik and Nonparametrik Statistical Procedures, Second Edition, Chapman&Hall/Crc, Boca Raton London New York Washington, D.C.

Stephens, M.A. 1974. EDF Statistics for Goodness of Fit and Some Comparisons. Journal of the America Statistical Association No.347. pp.730-737.

Zar, J.H. 1999. Biostatistical Analysis, Fourth Edition, Prentice Hall, Upper Saddle River. New Jersey 07458.

Zhang, P. 1999. Omnibus test of normality using the Q statistic. Journal of Applied Statiastics, 26 (4): 519-528

ÖZGEÇMİŞ

Adı ve Soyadı **:** Arif ÖZER

Doğum Yeri **:** ANKARA

Doğum Tarihi **:** 29/10/1978

Medeni Hali **:** Bekar

Yabancı Dili **:** İngilizce

Eğitim Durumu (Kurum ve Yıl)

Lise **:**Ankara Yenimahalle Teknik Lisesi

Lisans **:**Ankara Üniversitesi, Ziraat Fakültesi, Hayvansal Üretim Lisans Programı, Zootekni Alt Programı (1999-2003)

Yüksek Lisans **:**Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Zootekni Anabilim Dalı (Eylül 2003 - Haziran 2007)