



TÜRKİYE CUMHURİYETİ
MARMARA ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

**TERS NORMAL DAĞILIM (INVERSE GAUSSIAN
DISTRIBUTION) VE SAĞKALIM ANALİZİ İLE BİR
UYGULAMA**

PINAR GÜNEL
YÜKSEK LİSANS TEZİ

BİYOİSTATİSTİK ANABİLİM DALI

DANIŞMAN
Doç. Dr. Nural BEKİROĞLU

İSTANBUL-2009

ÖNSÖZ VE TEŞEKKÜR

Bu tez çalışmasının, (yaşam süresi verileri, dağılımları üzerinde çalışan) Ters Normal Dağılım ve Sağkalım Analizi (Kaplan-Meier Tahmin Yöntemi) konusunda bilgiye ihtiyacı olan araştırmacılara faydalı olacağını umuyorum.

Tez konusunu bana öneren, hazırlanması esnasında her türlü yardımı esirgemeyen ve yüksek lisans öğrenimim boyunca desteğini ve ilgisini eksik etmeyen değerli hocam Doç. Dr. Nural Bekiroğlu'na, tez çalışmasında kullanılan veri tabanına ulaşmamı sağlayan Biyoistatistik Uzmanı Sayın Rana Konyalıoğlu'na, hayatımın her aşamasında sürekli olarak destek veren aileme sonsuz teşekkürlerimi sunarım.

Ayrıca yüksek lisans öğrenimim esnasında maddi desteğinden dolayı TÜBİTAK'a teşekkürlerimi bir borç bilirim.

İÇİNDEKİLER

ÖNSÖZ/TEŞEKKÜR	iii
İÇİNDEKİLER	iv
SİMGE LİSTESİ	vi
ŞEKİL LİSTESİ	vii
TABLO LİSTESİ	ix

ÖZET	1
SUMMARY	2
1. GİRİŞ, AMAÇ VE KAPSAM	3
2. GENEL BİLGİLER	5
2.1. Yaşam Süresi Verileri için Üstel Dağılım Ailesi	5
2.1.1. Yaşam süresi dağılımı (Ömür dağılımı)	5
2.1.2. Üstel dağılım	7
2.1.3. Gama dağılım	10
2.1.4. Weibull dağılım	13
2.1.5. Log-normal dağılım	16
3. TERS NORMAL DAĞILIM (INVERSE GAUSSIAN DISTRIBUTION)	20
3.1. Ters Normal Dağılım Tarihçe	20
3.2. Ters Normal Dağılımın Karakteristik Özellikleri	21
3.2.1. Ters Normal dağılımın olasılık yoğunluk fonksiyonu	22
3.2.2. Ters Normal dağılımın birikimli dağılım fonksiyonu	25
3.2.3. Ters Normal dağılımın sağkalım fonksiyonu	25
3.2.4. Ters Normal dağılımın risk (hazard) fonksiyonu	26
3.3. Ters Normal Dağılım Parametre Tahmini	27
3.3.1. Ters Normal dağılım parametreleri için EÇO tahminleri	29
3.3.2. Ters Normal dağılım parametreleri için	

güven aralıkları	29
3.4. Ters Normal Dağılım Uyum İyiliği (Uyumluluk) Testleri	30
3.4.1. Kolmogorov-Smirnov testi	31
3.4.2. Anderson-Darling testi	32
3.4.3. Çarpıklık Basıklık katsayısı yaklaşımı ile uyum iyiliği (uyumluluk) testi	33
3.4.3.1. Ters Normal dağılım çarpıklık katsayısı testi	34
3.4.3.2. Ters Normal dağılım basıklık katsayısı testi	36
4. SAĞKALIM ANALİZİ	39
4.1. Sağkalım Fonksiyonu (Sağkalım Eğrisi)	39
4.2. Sağkalım Fonksiyonu'nun Kaplan-Meier Tahmini	43
5. YÖNTEM VE UYGULAMA	46
5.1. Tanımlayıcı İstatistikler	46
5.2. Ters Normal Dağılım ve Sağkalım Analizi Sonuçları (Bulguları)	48
6. TARTIŞMA VE ÖNERİLER	55
KAYNAKLAR	59
ÖZGEÇMİŞ	63

SİMGE LİSTESİ

X	Rasgele sürekli bir değişken
x	Rasgele sürekli değişkenin herhangi bir değeri
$f(x)$	Olasılık yoğunluk fonksiyonu
$F(x)$	Birikimli dağılım fonksiyonu
$S(x)$	Sağkalım fonksiyonu
$h(x)$	Risk (hazard) fonksiyonu
$\Gamma(\gamma)$	Gama fonksiyonu
$\ln(x)$	e tabanında logaritma
e^x	Üstel fonksiyon
$\Phi(.)$	Standart normal dağılımın birikimli dağılım fonksiyonu
α	Tip 1 hata
$B(t)$	Brownian hareket (Wiener süreci) fonksiyonu
δ_1	Ters Normal dağılım çarpıklık katsayısı
δ_2	Ters Normal dağılım basıklık katsayısı
$E(x)$	Beklenen değer
p	Olasılık
n	Örnek büyüklüğü

ŞEKİL LİSTESİ

Şekil 1.	Farklı λ değerleri için, Üstel dağılımın olasılık yoğunluk fonksiyonunun grafiksel gösterimi	8
Şekil 2.	Farklı λ değerleri için, Üstel dağılımın birikimli dağılım fonksiyonunun grafiksel gösterimi	8
Şekil 3.	$\lambda = 2$ değeri için, Üstel dağılımın sağkalım fonksiyonunun grafiksel gösterimi	9
Şekil 4.	$\lambda = 1$ değeri için, Üstel dağılımın risk (hazard) fonksiyonunun grafiksel gösterimi	9
Şekil 5.	Farklı γ değerleri için, Gama dağılımının olasılık yoğunluk fonksiyonunun grafiksel gösterimi	10
Şekil 6.	Farklı γ değerleri için, Gama dağılımının birikimli dağılım fonksiyonunun grafiksel gösterimi	11
Şekil 7.	Farklı γ değerleri için, Gama dağılımının sağkalım fonksiyonunun grafiksel gösterimi	12
Şekil 8.	Farklı γ değerleri için, Gama dağılımının risk (hazard) fonksiyonunun grafiksel gösterimi	12
Şekil 9.	Farklı α değerleri için, Weibull dağılımının olasılık yoğunluk fonksiyonunun grafiksel gösterimi	14
Şekil 10.	Farklı α değerleri için, Weibull dağılımının birikimli dağılım fonksiyonunun grafiksel gösterimi	14
Şekil 11.	Farklı α değerleri için, Weibull dağılımının sağkalım fonksiyonunun grafiksel gösterimi	15
Şekil 12.	Farklı α değerleri için, Weibull dağılımının risk (hazard) fonksiyonunun grafiksel gösterimi	15
Şekil 13.	Farklı σ değerleri için, Lognormal dağılımının olasılık yoğunluk fonksiyonunun grafiksel gösterimi	17
Şekil 14.	Farklı σ değerleri için, Lognormal dağılımının birikimli dağılım fonksiyonunun grafiksel gösterimi	17

Şekil 15.	Farklı σ değerleri için, Lognormal dağılımının sağkalım fonksiyonunun grafiksel gösterimi	18
Şekil 16.	Farklı σ değerleri için, Lognormal dağılımının risk (hazard) fonksiyonunun grafiksel gösterimi	19
Şekil 17.	Farklı μ değerleri için, Ters Normal dağılımının olasılık yoğunluk fonksiyonunun grafiksel gösterimi	23
Şekil 18.	Farklı λ değerleri için, Ters Normal dağılımının olasılık yoğunluk fonksiyonunun grafiksel gösterimi	24
Şekil 19.	Farklı μ değerleri için, Ters Normal dağılımının sağkalım fonksiyonunun grafiksel gösterimi	26
Şekil 20.	Farklı μ değerleri için, Ters Normal dağılımının risk (hazard) fonksiyonunun grafiksel gösterimi	27
Şekil 21.	1974'te Amerika Birleşik Devletleri'ndeki bireylere ait sağkalımın birikimli dağılımının bir tahminini gösteren grafik	39
Şekil 22.	1974'te Amerika Birleşik Devletleri'ndeki bireylere ait sağkalım eğrisi	41
Şekil 23.	Sağkalım analizi çalışmasına farklı zamanlarda giren hastaların çalışmaya girdikleri süreleri gösteren grafik	43
Şekil 24.	Sağkalım analizi çalışmasına farklı zamanlarda giren hastaların çalışmaya girdikleri sürelerin çalışmanın başlama noktasına çekildiğini gösteren grafik	43
Şekil 25.	Sağkalım eğrisi	44
Şekil 26.	Ölen meme kanserli hastaların izlem süreleri dağılımının olasılık yoğunluk fonksiyonu grafiği	50
Şekil 27.	Ölen meme kanserli hastaların izlem süreleri Ters Normal dağılımının risk (hazard) fonksiyonu grafiği	51
Şekil 28.	Ölen meme kanserli hastaların izlem sürelerinin Lognormal dağılımının risk (hazard) fonksiyonu grafiği	53

TABLO LİSTESİ

Tablo 1.	Çalışmada kullanılan veri setine ait bazı önemli değişkenlerin dağılımı	46
Tablo 2.	Tanı konduktan sonra ölen meme kanserli hastaların ölümüne kadar geçen süreler	48
Tablo 3.	EasyFit paket programı Ters Normal dağılım uyum iyiliği test sonuçları	49
Tablo 4.	Logaritması alınan verilerin Kolmogorov-Smirnov normallik testi	52
Tablo 5.	EasyFit paket programı Lognormal dağılım uyum iyiliği test sonuçları	52
Tablo 6.	İzlem sürelerine göre hastaların sağkalım olasılıklarının bir kısmı	54

ÖZET

Yaşam ve insana ait veriler genellikle sağa çarpık ve pozitif ölçümlerden oluşur. Bu tür verilerin hangi dağılıma ait olduğuna karar verilmesi için verilerin uygun olabileceği dağılımların özelliklerinin bilinmesi gerekir. Yaşam süresi dağılımları için incelenmesi gereken en önemli özellik risk (hazard) oranıdır. Bu tez çalışmasında yaşam süresi dağılımları temel özellikleri ile ele alınmıştır. Bir meme kanseri (n=1499) veri tabanındaki, meme kanseri nedeniyle ölen hastalara ait yaşam süresi verilerinin hangi dağılıma uygunluk gösterdiği tartışılmıştır. Ölüm riskinin ilk dönemlerde yüksek olduğu (ilk dönemde hızlı bir yükseliş ve zirveden sonraki ilk dönemde hızlı düşüş) düşünülerek verilerin Ters Normal dağıldığı varsayılmıştır. Çalışmada kullanılan paket programı sonuçları doğrultusunda verilerin Ters Normal dağılıma uygun olduğu ortaya çıkmıştır. Çarpıklık ve basıklık katsayısı yaklaşımı ile yapılan uyum iyiliği testleri sonucunda da verilerin Ters Normal dağıldığı kanıtlanmıştır. Ancak verilerin risk (hazard) fonksiyonu grafiği Ters Normal dağılımın özelliğini yansıtmamaktadır. Risk (hazard) fonksiyonunun şekli sebebiyle verilerin Lognormal dağılıma uygun olacağı öngörülmüştür. Gerekli hesaplamalar yapılarak verilerin Lognormal dağıldığı belirlenmiştir. Daha sonra, ele alınan verilerin ortalaması ile Sağkalım analizinden elde edilen sonuçlar birleştirilerek yorumlanmıştır. Buna göre uzun dönemde sağkalım olasılığı yüksektir. Verilere ait risk fonksiyonunun Ters Normal Dağılıma uygunluk göstermemesinin nedeni hastalığın tedavisinde epey yol kat edilmiş olması şeklinde açıklanabilir. Ters Normal dağılımın, hastalığın ilk dönemlerinde hızlı risk artışı gösteren başka kanser türlerine (serviks, akciğer, karaciğer gibi) uygun olacağını düşünmekteyiz.

Anahtar Sözcükler: Lognormal Dağılım, Sağkalım Analizi (Kaplan-Meier Tahmin Yöntemi), Ters Normal Dağılım, Yaşam Süresi Dağılımları.

SUMMARY

An Application with Inverse Gaussian Distribution and Survival Analysis

Data that belong to life and human often consist of nonnegative measurements that exhibit positive skewness. To decide the appropriate distribution for the data, characteristics of the distribution must be known. For lifetime distributions, the most crucial characteristic which must be investigated is the hazard rate. In this dissertation, characteristics of lifetime distributions were considered. In a breast cancer database (n=1499) it was discussed that which distribution fit the lifetime data of the patients who died because of breast cancer. It was assumed that the data came from Inverse Gaussian (IG) distribution regarding that the mortality rate was high in the first period (initially increasing and then decreasing quickly). It was occurred that the data fit IG distribution according to results of package that used in the study. It was also proved that the data fit IG distribution considering results of skewness and kurtosis goodness-of-fit tests. However, the hazard function of data wasn't appropriate for IG. It was proposed that the data fit the Lognormal distribution because of the hazard function shape. After all computations it was determined that the data fit the Lognormal distribution. Then, the average of data and the results of Survival analysis were interpreted together. Accordingly the survival probability was high in long-term. The recent improvements in the treatment of breast cancer could be the reason why data doesn't fit IG distribution. We think that IG distribution will be appropriate in other cancer types (such as cervix, lung and liver cancers) that demonstrate initially increasing hazard.

Key Words: Inverse Gaussian Distribution, Lifetime Distributions, Lognormal Distribution, Survival Analysis (Kaplan-Meier Estimation Method).

1. GİRİŞ, AMAÇ VE KAPSAM

İstatistiksel dağılımlar, birçok alanda olduğu gibi tıp alanında da yaygın olarak kullanılırlar. Dağılım seçimi, dağılımın elde var olan verilere ne kadar uyum sağladığı göz önünde bulundurularak yapılır.

Yaşam ve insana ait istatistiksel veriler genellikle pozitif (sağa) çarpık, negatif olmayan ölçümlerden oluşur. (30) Tıp alanında uygulama olanağı bulan dağılımlar arasında Üstel, Gama, Weibull, Lognormal gibi üstel dağılım ailesine ait yaşam süresi dağılımları yer almaktadır. Birnbaum-Saunders dağılımı, Uç değer dağılımı (Gumbel dağılımı), Loglojistik dağılım ve Rayleigh dağılımı gibi dağılımlar da üstel dağılım ailesine ait dağılımlardır ancak bu dağılımların sahip oldukları bazı özellikler sebebiyle (negatif değer alabilme, sola çarpık olma gibi) tıp alanında kullanılması uygun olamamaktadır.

Yaşam süresi dağılımlarında risk (hazard) fonksiyonu, dağılımın belirlenmesi, adının konması için önemli bir rol oynar. (6) Güvenilirlikte (özellikle endüstride) bir aletin (ürünün) ömrünün dağılımı için genellikle Weibull dağılımı uygundur. (17, 25) Bunun sebebi Weibull dağılımının risk (hazard) fonksiyonunun monoton değişiklik göstermesidir (Monoton artan, monoton azalan veya sabit). (8) Normal şartlarda insan ömrünün de monoton risk (hazard) oranı göstermesi beklenir ancak ölümcül bir hastalığa (özellikle kanser) yakalandığında risk (hazard) oranı değişir. Üstelik söz konusu kanser olduğunda, ölüm hızı hastalık ortaya çıktıktan sonra ilk dönemde hızla yükselirken sonraki yıllarda azalmaya başlar. Bir başka deyişle ölüm, hastalığın tanısı konduktan sonraki ilk dönemlerde yükselerek bir doruk noktasına ulaşır, sonra hızla azalmaya başlar, daha sonraları ise bu düşüş hızı görece olarak yavaşlar. Risk (hazard) fonksiyonunun bu beklenen şeklini sağlayan en uygun dağılımlardan biri de üstel dağılım ailesine ait tek modlu (unimodal) bir dağılım olan Ters Normal Dağılımdır. (6) Ters Normal Dağılım bu tür verinin analizi için çok yönlü ve esnek bir model elde etmeyi sağlar. (30)

Tıpta yaşam süresi verileri için uygulanan yöntemlerden biri de Sağkalım analizidir. Sağkalım analizinde, ölümün birikimli olasılığı yerine belirli bir t

zamanına kadar veya daha fazla yaşayan bireylerin oranı ya da yüzdesi olarak ifade edilen *sağkalım fonksiyonu* ya da *sağkalım eğrisi* kullanılır. (13)

Sağkalım yerine ölümü söylemek tercih edilmeyen, can sıkıcı bir durum ise de, bir kişi ölümcül bir hastalığa yakalandığını öğrendiğinde sorduğu ilk soru genelde: “Ne kadar ömrüm kaldı?” olur. Tıpta bu soruya verilen yanıt genelde “4 yıl içinde sağ kalma olasılığınız %95’tir.” cümlesi gibidir.-

Bu tez çalışması, Ters Normal dağılım ve Sağkalım analizini kullanarak hastanın sorduğu soru ve doktorun soruya verdiği yanıtın istatistiksel analizlerini irdelemeyi amaç edinmiştir.

Çalışma, “Giriş, Amaç ve Kapsam”, “Genel Bilgiler”, “Ters Normal Dağılım”, “Sağkalım Analizi”, “Yöntem ve Uygulama” ve “Tartışma ve Öneriler” olmak üzere toplam altı bölümden oluşmaktadır.

Çalışmanın birinci bölümünde, çalışmanın amacı ve kapsamı ile model seçiminde önemli olan noktalara değinilmiştir. İkinci bölümde, yaşam süresi verileri için kullanılan bazı üstel dağılımların özellikleri anlatılmıştır. Çalışmanın üçüncü bölümünde, üstel dağılım ailesi çatısı altında bulunan Ters Normal Dağılım’ın tarihsel gelişimi ile başlanıp, dağılımın özellikleri, tanımlamaları, parametre tahminleri ve uygunluk (uyum iyiliği) testleri hakkında bilgi verilmiştir. Dördüncü bölümde, Sağkalım Analizi’nin parçası olan sağkalım fonksiyonu ve bu fonksiyonu tahmin etmede kullanılan Kaplan-Meier tahmin yönteminden söz edilmiştir. Beşinci bölümde, çalışmanın yöntemi anlatılmış ve çalışmada kullanılan veri setine ait analizler bulgular şeklinde verilmiştir. Son olarak altıncı bölümde analiz sonuçları tartışılarak önerilerde bulunulmuştur.

2. GENEL BİLGİLER

2.1. Yaşam Süresi Verileri için Üstel Dağılım Ailesi

2.1.1. Yaşam süresi dağılımı (Ömür dağılımı)

Yaşam süresi dağılımı, üretimde bir ürün kümesindeki tüm ürünlerin raf ömrü uzunluğunun dağılımı, tıpta ise bireylerin yaşam süresi uzunluğunun dağılımı olarak yorumlanabilir. Yaşam süresi uzunluğu saat, hafta, ay, yıl gibi zaman birimleri ile ifade edilir. Dağılım, ana kitleden alınan örnek bir veri setinden tahmin edilebilir. Çoğunlukla yaşam süresi dağılımının belirli bir şekle sahip olduğunu varsaymak mümkündür. Yaşam ve insana ait istatistiksel veriler genellikle pozitif olduğundan yaşam süresinin, üstel dağıldığı yaklaşımı vardır. (37)

Bu bölümde, sıklıkla kullanılan başlıca yaşam süresi dağılımları kısaca anlatılacaktır. Bu dağılımlara geçmeden önce istatistiksel dağılımları tanımlamada kullanılan; olasılık yoğunluk fonksiyonu, birikimli dağılım fonksiyonu, sağkalım fonksiyonu ve risk (hazard) fonksiyonu kavramlarına değinilecektir.

X yaşam süresi gibi rasgele bir sürekli değişken olduğunda, x bu değişkenin herhangi bir değeri ve α da olasılık olmak üzere; olasılık yoğunluk fonksiyonu, birikimli dağılım fonksiyonu, sağkalım fonksiyonu ve risk (hazard) fonksiyonu aşağıdaki gibi tanımlanırlar. (8)

Olasılık yoğunluk fonksiyonu: Bir dağılımın olasılık yoğunluk fonksiyonu, fonksiyonun x_L ve x_U aralığında tanımlı integralinin, değişkenin bu aralıkta aldığı bir değer olasılığına eşit olduğu fonksiyondur. Matematiksel gösterimi (2.1) ve (2.2)'deki gibidir.

$$\int_{x_L}^{x_U} f(x)dx = P[x_L < X \leq x_U] \quad (2.1)$$

$$f(x) = \frac{d(F(x))}{dx} \quad (2.2)$$

Birikimli dağılım fonksiyonu: Bir dağılımın birikimli dağılım fonksiyonu, X değişkeninin, değişkenin aldığı herhangi bir x değerinden küçük ya da x değerine eşit olması olasılığıdır. Matematiksel gösterimi (2.3) ve (2.4) eşitliklerindeki gibidir.

$$F(x) = P[X \leq x] = \alpha \quad (2.3)$$

$$F(x) = \int_{-\infty}^x f(u) du \quad (2.4)$$

u , burada yardımcı değişkendir.

Sağkalım fonksiyonu: Bir dağılımın sağkalım fonksiyonu, X değişkeninin, değişkenin aldığı herhangi bir x değerinden büyük olması olasılığıdır. Matematiksel gösterimi (2.5)'teki gibidir.

$$S(x) = P[X > x] = 1 - F(x) \quad (2.5)$$

Risk (Hazard) fonksiyonu (risk (hazard) oranı): Risk (Hazard) fonksiyonu, güvenilirlikte *başarısızlık (failure) fonksiyonu* ya da *oranı* olarak geçer. *Başarısızlık (hata, failure)*, endüstride (üretimde) bir ürünün bozulmasını ya da raf ömrünün sona ermesini ifade ederken, tıpta bunun yerine bir bireyin bir hastalığa yakalanması ya da ölümü için kullanılan *risk (hazard) fonksiyonu* vardır. (Dağılımların tanıtılması sırasında gerektiğinde başarısızlık (failure), gerektiğinde ise risk (hazard) fonksiyonu ifadeleri kullanılacaktır.)

Bir dağılımın risk (hazard) fonksiyonu, herhangi bir x değerinde, dağılımın olasılık yoğunluk fonksiyonunun sağkalım fonksiyonuna oranıdır. Matematiksel gösterimi (2.6) eşitliğindeki gibidir.

$$h(x) = \frac{f(x)}{S(x)} = \frac{f(x)}{[1 - F(x)]} \quad (2.6)$$

2.1.2. Üstel dağılım

Üstel Dağılım, tek parametrelidir ve yaşam süresi dağılımlarının en önemlisidir. Bunun sebebi, iki ya da üç parametrelidir diğer yaşam süresi dağılımları için temel oluşturmasıdır. (21)

Üstel dağılım, belli bir zaman aralığında meydana gelen başarısızlık (failure) olayına ait olasılığın zaman içinde değişmediği durumda, ilk başarısızlık (failure) gerçekleşinceye kadar geçen süreye ait dağılım olarak tanımlanır. Üstel dağılım ailesine ait en bilinen Poisson dağılımı da üstel dağılım gibi, herhangi bir zaman aralığındaki olayların sayısı ile olaylar arasındaki sürenin dağılımından meydana gelir. (8)

Üstel dağılım, zamana karşı dayanma modeli olarak kullanılan istatistiksel dağılımlar arasında en popüler olanıdır. Üstel dağılımın başarısızlık (failure) oranı sabittir ve bu durum matematiksel kolaylık sağlar. Ancak uygulamalarda, söz konusu sabit başarısızlık (failure) oranı yerine daha çok zamanla değişiklik gösteren başarısızlık (failure) oranıyla karşılaşılır. (8, 35)

Dağılımın ölçek parametresi $\beta > 0$ ve rasgele değişken $0 \leq x < +\infty$ olmak üzere Üstel dağılımın olasılık yoğunluk fonksiyonu (2.7)'deki gibi tanımlanır. (8) [$x < 0$ için, $f(x) = 0$ 'dır.]

$$f(x) = \frac{1}{\beta} e^{\left(\frac{-x}{\beta}\right)} \quad (2.7)$$

Üstel dağılımda;

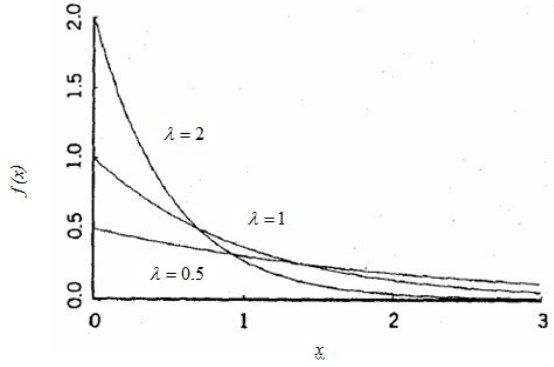
$$h(x) = \lambda = \frac{1}{\beta} \quad (2.8)$$

'dir. ($h(x)$, risk (hazard) fonksiyonu)

Bu durumda Üstel dağılımın olasılık yoğunluk fonksiyonunu (2.9)'a eşit olmaktadır. (18)

$$f(x) = \lambda e^{(-\lambda x)} \quad (2.9)$$

Üstel dağılımın olasılık yoğunluk fonksiyonuna ait grafik Şekil 1.'de verilmiştir. (8)



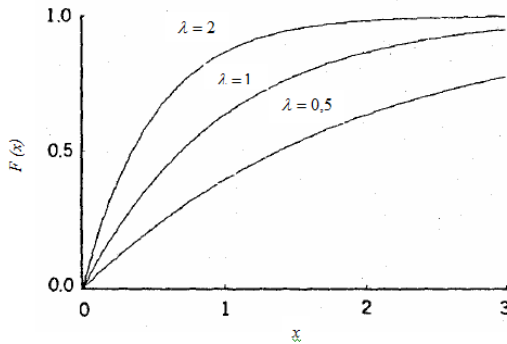
Şekil 1. Farklı λ değerleri için, Üstel dağılımın olasılık yoğunluk fonksiyonunun grafiksel gösterimi

Üstel dağılımın birikimli dağılım fonksiyonu (2.10)'da ifade edildiği gibidir.(21)

[$x < 0$ için, $F(x) = 0$ 'dır.]

$$F(x) = 1 - e^{(-\lambda x)} \quad (2.10)$$

Üstel dağılıma ait birikimli dağılım fonksiyon grafiği Şekil 2.'de verilmiştir. (8)

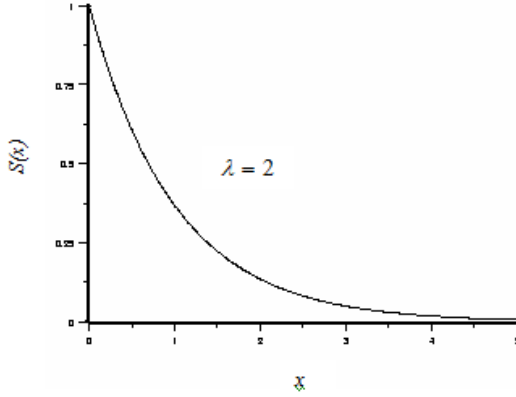


Şekil 2. Farklı λ değerleri için, Üstel dağılımın birikimli dağılım fonksiyonunun grafiksel gösterimi

Üstel dağılımın sağkalım fonksiyonu (2.11) ile elde edilir. (21)

$$S(x) = 1 - F(x) = e^{(-\lambda x)} \quad (2.11)$$

Üstel dağılımın sağkalım fonksiyonuna ait grafik Şekil 3.'te verilmiştir. (<http://www.itl.nist.gov/div898/handbook/eda/section3/eda3667.htm> , Erişim tarihi: 19.06.2009)



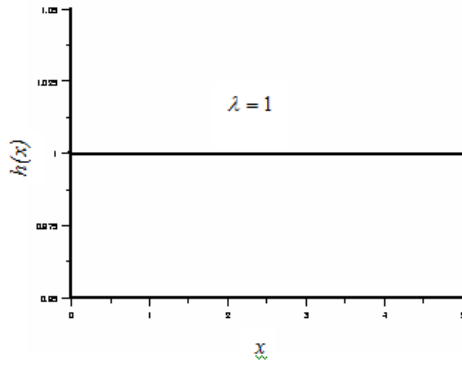
Şekil 3. $\lambda = 2$ değeri için, Üstel dağılımın sağkalım fonksiyonunun grafiksel gösterimi

Üstel dağılımın risk (hazard) fonksiyonu (2.12)'deki gibi hesaplanır ve daha önce (2.8) eşitliğindeki değer elde edilir.

$$h(x) = \frac{f(x)}{S(x)} = \frac{\lambda e^{(-\lambda x)}}{e^{(-\lambda x)}} = \lambda = \frac{1}{\beta} \quad (2.12)$$

Üstel dağılıma sahip bir yaşam süresi dağılımının başarısızlık (failure) oranı (λ) Şekil 4.'ten de görüldüğü üzere sabittir. (8, 18)

Üstel dağılımın risk (hazard) fonksiyonuna ait grafik Şekil 4.'te verilmiştir. (<http://www.itl.nist.gov/div898/handbook/eda/section3/eda3667.htm> , Erişim tarihi: 19.06.2009)



Şekil 4. $\lambda = 1$ değeri için, Üstel dağılımın risk (hazard) fonksiyonunun grafiksel gösterimi

2.1.3. Gama dağılımı

Gama dağılımı, üstel dağılımın genelleştirilmiş bir halidir ve sürekli dağılımlar içinde önemli bir yere sahip, aynı zamanda başarısızlık süresi (failure time) verilerine uygunluk gösteren bir yaşam süresi dağılımıdır. (8, 12) Yaşam süresi değerlendirmelerinde, üstel dağılımlar için gerçeğe uygun ayarlamalar yapmak üzere kullanılmışlardır. Monoton risk oranı (hazard rate, failure rate) gösteren veri seti için uygundur. (18)

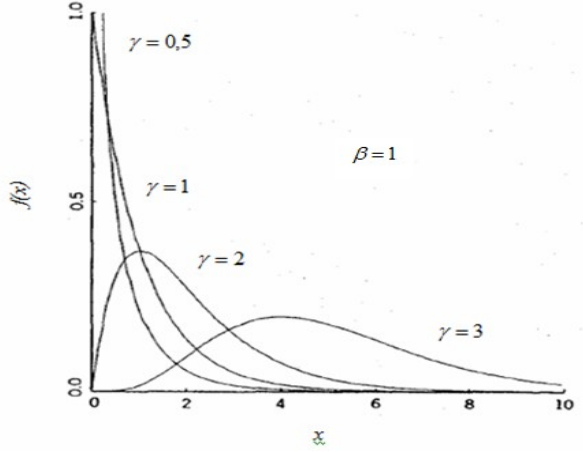
Dağılımın ölçek parametresi $\beta > 0$, şekil parametresi $\gamma > 0$ ve rasgele değişken $0 \leq x < +\infty$ olmak üzere dağılımın olasılık yoğunluk fonksiyonu (2.13)'teki gibidir. (8, 37)

$$f(x) = \left(\frac{x}{\beta} \right)^{\gamma-1} \frac{e^{-\frac{x}{\beta}}}{\beta \Gamma(\gamma)} = \frac{1}{\beta^\gamma \Gamma(\gamma)} x^{\gamma-1} e^{-\frac{x}{\beta}} \quad (2.13)$$

$\Gamma(\gamma)$, burada Gama fonksiyonu belirtir ve (2.14)'teki gibi elde edilir. (12)

$$\Gamma(\gamma) = \int_0^{\infty} x^{\gamma-1} e^{-x} dx = (\gamma - 1)! \quad (2.14)$$

Gama dağılımına ait olasılık yoğunluk fonksiyon grafiği Şekil 5.'te verilmiştir. (8)



Şekil 5. Farklı γ değerleri için, Gama dağılımının olasılık yoğunluk fonksiyonunun grafiksel gösterimi

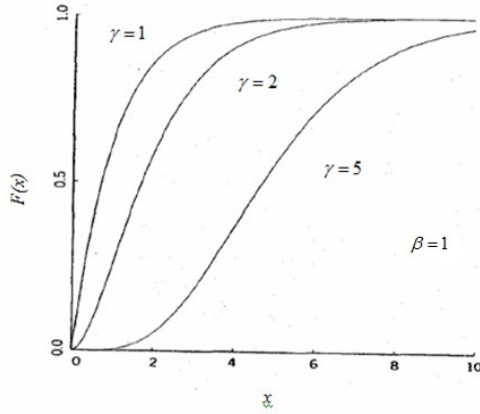
Gama dağılımının birikimli dağılım fonksiyonu (2.15)'te verilmiştir. (12)

$$F(x) = \int_0^x \frac{1}{\beta^\gamma \Gamma(\gamma)} x^{\gamma-1} e^{-\frac{x}{\beta}} dx \quad (2.15)$$

(2.15)'te γ şekil parametresinin pozitif bir tamsayı olması durumunda Gama dağılımı, Erlang dağılımı olarak bilinen dağılıma dönüşür. Bu durumda birikimli dağılım fonksiyonu (2.16)'daki gibi tanımlanır. (12, 21)

$$F(x) = 1 - \left(\sum_{n=0}^{\gamma-1} \frac{\left(\frac{x}{\beta}\right)^n e^{-\frac{x}{\beta}}}{n!} \right) \quad (2.16)$$

Gama dağılımının birikimli dağılım fonksiyon grafiği Şekil 6.'da verilmiştir. (8)

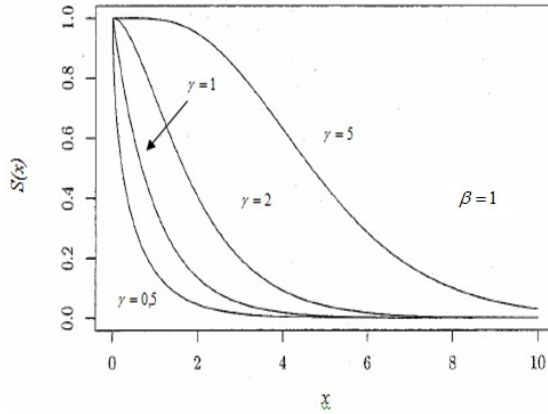


Şekil 6. Farklı γ değerleri için, Gama dağılımının birikimli dağılım fonksiyonunun grafiksel gösterimi

Gama dağılımının sağkalım fonksiyonu (2.17) ile elde edilir. (21)

$$S(x) = 1 - F(x) = \sum_{n=0}^{\gamma-1} \frac{\left(\frac{x}{\beta}\right)^n e^{-\frac{x}{\beta}}}{n!} \quad (2.17)$$

Gama dağılımına ait sağkalım fonksiyon grafiği Şekil 7.'de verilmiştir. (21)

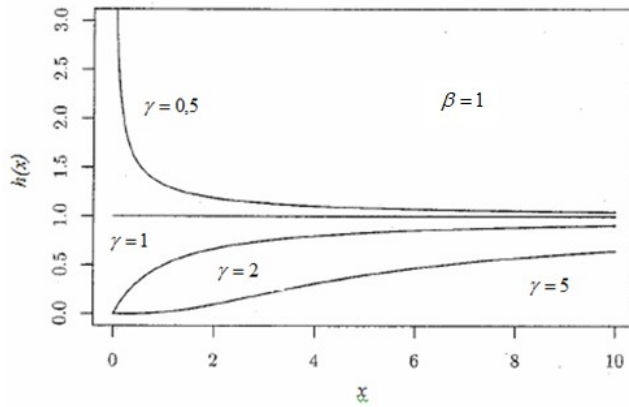


Şekil 7. Farklı γ değerleri için, Gama dağılımının sağkalım fonksiyonunun grafiksel gösterimi

Gama dağılımının risk (hazard) fonksiyonu (2.18) eşitliği ile hesaplanır. (37)

$$h(x) = \frac{f(x)}{S(x)} = \frac{\frac{1}{\beta^\gamma \Gamma(\gamma)} x^{\gamma-1} e^{-\frac{x}{\beta}}}{\sum_{n=0}^{\gamma-1} \frac{\left(\frac{x}{\beta}\right)^n}{n!} e^{-\frac{x}{\beta}}} \quad (2.18)$$

Gama dağılımına ait risk (hazard) fonksiyonunun grafiği Şekil 8.'de verilmiştir. (21)



Şekil 8. Farklı γ değerleri için, Gama dağılımının risk (hazard) fonksiyonunun grafiksel gösterimi

2.1.4. Weibull dağılımı

Weibull dağılımı, 1939'da İsveçli fizikçi Waloddi Weibull'ın alet dirençlerinin bozulma sürelerinin dağılımını göstermesinden sonra isimlendirilmiştir. (17) Weibull dağılımı çoğunlukla, sağkalım çalışmalarında yaşam süresi verilerinin analizinde, endüstride ürünlerin güvenilirlik çalışmalarında kullanılan üstel dağılım ailesine ait bir dağılımdır. (8, 24)

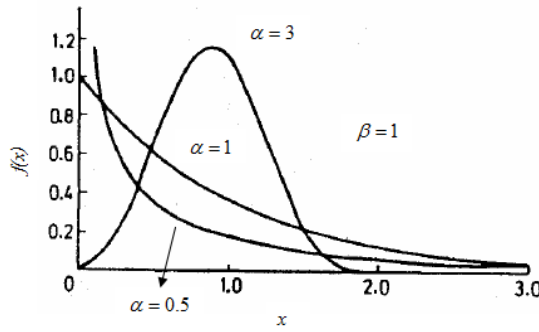
Birçok durumda, üstel dağılımın zamana karşı dayanma modeli olarak yetersiz kalması, sabit başarısızlık (failure) fonksiyonundan kaynaklanır. Bu nedenle başarısızlık (failure) olasılığının zamana bağlı olarak değiştiği Weibull dağılım gibi daha genel ve esnek bir dağılıma gereksinim duyulur. (35) Weibull dağılımı çarpık verileri modellemek için kullanılır. Monoton risk oranı (hazard rate, failure rate) gösteren veri seti için uygundur. Ancak bu dağılımı, diğer monoton risk oranı için

uygun olan dağılımlardan (gama dağılımı gibi) ayırmak için oldukça geniş veri setine ihtiyaç duyulur aksi takdirde ayırt edilemeyecek kadar benzerlik gösterirler. Buna rağmen Weibull dağılımı, esnek bir model olması bakımından güvenilirlik ve sağkalım çalışmalarında daha sık kullanılır. (23, 25)

İki parametrelili Weibull dağılımı için, dağılımın ölçek parametresi $\beta > 0$, şekil parametresi $\alpha > 0$ ve rasgele değişken $0 \leq x < +\infty$ olmak üzere dağılımın olasılık yoğunluk fonksiyonu (2.19)'daki gibidir. (8)

$$f(x) = \left(\frac{\alpha x^{\alpha-1}}{\beta^\alpha} \right) e^{-\left(\frac{x}{\beta} \right)^\alpha} \quad (2.19)$$

Weibull dağılımının olasılık yoğunluk fonksiyon grafiği Şekil 9.'da verilmiştir. (8)



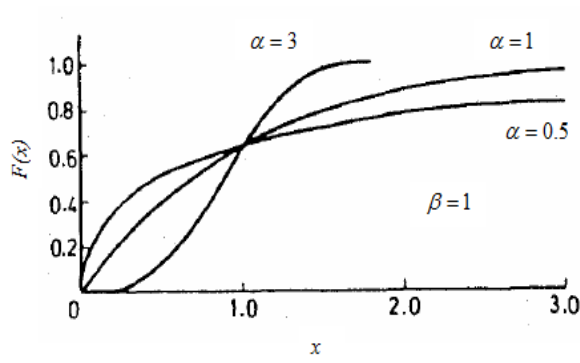
Şekil 9. Farklı α değerleri için, Weibull dağılımının olasılık yoğunluk fonksiyonunun grafiksel gösterimi

Weibull dağılımı, şekil parametresi $\alpha = 1$ için Üstel dağılıma dönüşür. $\alpha = 2.5$ için Lognormal dağılıma, $\alpha = 3.6$ için ise normal dağılıma yaklaşır. (8, 17).

Weibull dağılımının birikimli dağılım fonksiyonu (2.20)'de verilmiştir. (24)

$$F(x) = 1 - e^{-\left(\frac{x}{\beta} \right)^\alpha} \quad (2.20)$$

Weibull dağılımına ait birikimli dağılım fonksiyonu grafiği Şekil 10.'da verilmiştir. (8).

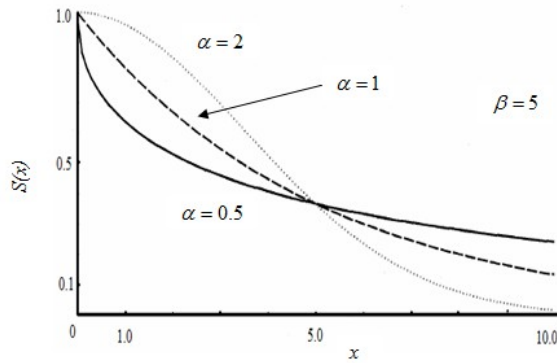


Şekil 10. Farklı α değerleri için, Weibull dağılımının birikimli dağılım fonksiyonunun grafiksel gösterimi

Weibull dağılımının sağkalım fonksiyonu (2.21)'deki gibi hesaplanır. (23)

$$S(x) = e^{-\left(\frac{x}{\beta}\right)^\alpha} \quad (2.21)$$

Weibull dağılımı sağkalım fonksiyonu Şekil 11.'de verilmiştir. (<http://statistics.unl.edu/faculty/steve/surv/anbreed1.pdf> , Erişim 19.06.2009)



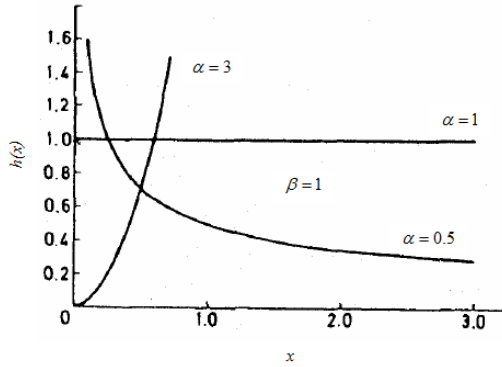
Şekil 11. Farklı α değerleri için, Weibull dağılımının sağkalım fonksiyonunun grafiksel gösterimi

Şekil 11'den görüldüğü üzere, α büyüdükçe ve süre ilerledikçe sağkalım fonksiyonu 0'a yaklaşmaktadır. (<http://statistics.unl.edu/faculty/steve/surv/anbreed1.pdf> , Erişim 19.06.2009)

Weibull dağılımın risk (hazard) fonksiyonu (2.22)'deki gibi hesaplanır. (8)

$$h(x) = \frac{\alpha x^{\alpha-1}}{\beta^{\alpha}} \quad (2.22)$$

Weibull dağılımına ait risk (hazard) fonksiyonuna ait grafik Şekil 12.'de verilmiştir. (8)



Şekil 12. Farklı α değerleri için, Weibull dağılımının risk (hazard) fonksiyonunun grafiksel gösterimi

İki parametrelili Weibull dağılımı, Şekil 12.'den de görülebileceği gibi, artan, azalan veya sabit risk (hazard) oranı gösterebilir. Ancak bu artış ve azalış sabit oranlarda (monoton) olmalıdır. (Başka bir deyişle monoton artan, monoton azalan ve sabit). (8, 23)

2.1.5. Lognormal dağılım

Lognormal dağılım, üstel dağılım ailesine ait, yaşam süresi verilerine uygulanabilir, sağa çarpık bir dağılımdır. (8)

(http://www.weibull.com/AccelTestWeb/lognormal_distribution.htm, Erişim tarihi 19.06.2009)

Monoton olmayan risk (hazard) oranı gösteren veriler için uygundur. (6) Verilere uygulanan logaritmik dönüşüm, verilerin yaklaşık olarak simetrik dağılmasını sağlar.

(8) x rasgele bir değişkenin bir değeri olmak üzere $Ln(x)$ 'in tanımlı olabilmesi için x pozitif olmalıdır. Bu nedenle, Lognormal dağılımın uygulanabilmesi için ilgilenilen rasgele değişkenin pozitif değerler alması gerekir. (35)

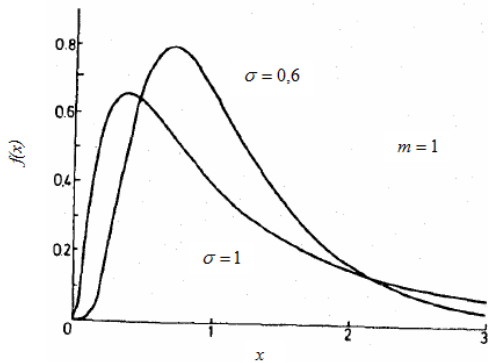
Dağılımın ölçek parametresi $m = e^\mu > 0$, şekil parametresi $\sigma > 0$ ve rasgele değişken $0 \leq x < +\infty$ olmak üzere dağılımın olasılık yoğunluk fonksiyonu (2.23) ve (2.24)'teki gibidir. (8)

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\left(\frac{-[\ln(x/m)]^2}{2\sigma^2} \right)} \quad (2.23)$$

veya

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\left(\frac{-(\ln x - \mu)^2}{2\sigma^2} \right)} \quad (2.24)$$

Lognormal dağılıma ait olasılık yoğunluk fonksiyonu grafiği Şekil 13.'te verilmiştir. (8)



Şekil 13. Farklı σ değerleri için, Lognormal dağılımının olasılık yoğunluk fonksiyonunun grafiksel gösterimi

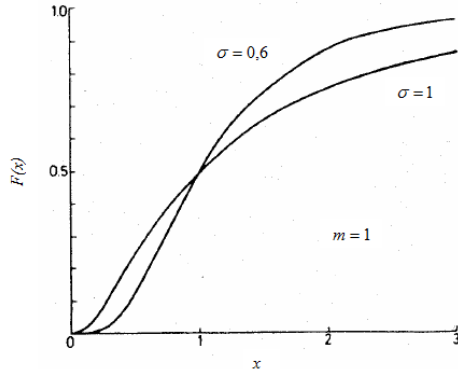
Lognormal dağılımın birikimli dağılım fonksiyonu (2.25) eşitliğindeki gibi tanımlanır. (22)

$$F(x) = \Phi \left(\frac{\ln x - \mu}{\sigma} \right) \quad (2.25)$$

Burada $\Phi(\cdot)$, Laplace integralidir. (Standart normal dağılımın birikimli dağılım fonksiyonu) (2.26)'daki gibi tanımlanır. (22)

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt \quad (2.26)$$

Lognormal dağılımın birikimli dağılım fonksiyonuna ait grafik Şekil 14.'te verilmiştir. (8)

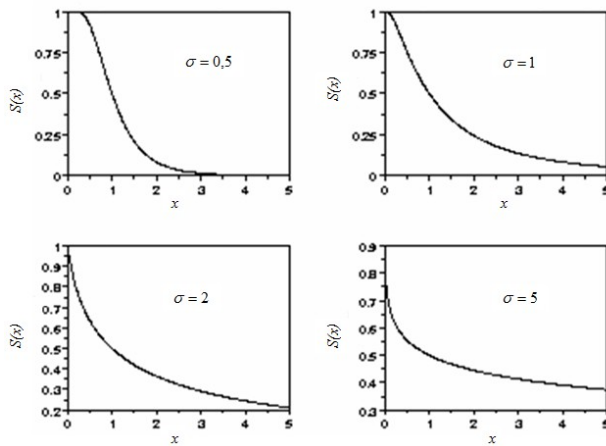


Şekil 14. Farklı σ değerleri için, Lognormal dağılımının birikimli dağılım fonksiyonunun grafiksel gösterimi

Lognormal dağılımın sağkalım fonksiyonu (2.27) ile elde edilir. (18)

$$S(x) = 1 - F(x) = 1 - \Phi\left(\frac{\ln x - \mu}{\sigma}\right) \quad (2.27)$$

Lognormal dağılıma ait sağkalım fonksiyon grafiği Şekil 15.'te verilmiştir. (<http://www.itl.nist.gov/div898/handbook/eda/section3/eda3669.htm>, Erişim tarihi 19.06.2009)

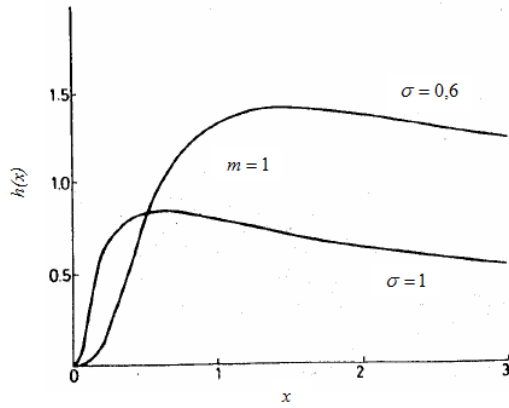


Şekil 15. Farklı σ değerleri için, Lognormal dağılımının sağkalım fonksiyonunun grafiksel gösterimi

Lognormal dağılımın risk (hazard) fonksiyonu ise (2.28)'deki eşitlik ile hesaplanır.

$$h(x) = \frac{f(x)}{S(x)} = \frac{\frac{1}{x\sigma\sqrt{2\pi}} e^{\left(\frac{-(\ln x - \mu)^2}{2\sigma^2}\right)}}{1 - \Phi\left(\frac{\ln x - \mu}{\sigma}\right)} \quad (2.28)$$

Lognormal dağılımın risk (hazard) fonksiyonuna ait grafik Şekil 16.'da verilmiştir. (8)



Şekil 16. Farklı σ değerleri için, Lognormal dağılımının risk (hazard) fonksiyonunun grafiksel gösterimi

Lognormal dağılımın risk (hazard) fonksiyonu, $x = 0$ iken 0'dır. Önce maksimum bir değere yükselir, daha sonra azalır. x arttıkça 0'a yaklaşır. (18)

3. TERS NORMAL DAĞILIM (INVERSE GAUSSIAN DISTRIBUTION)

3.1. Ters Normal Dağılım Tarihçe

Ters Normal Dağılım'ın temelleri, *Brownian tipi harekette ilk geçiş zamanı* “first passage time” dağılımının olasılık yoğunluk fonksiyonunun bağımsız türetimleri üzerinde çalışan Schrödinger (1915) ve Smoluchowsky (1915) tarafından atılmıştır. Her ikisinin yaptığı çalışmalar ters normal dağılım üzerinde bilinen ilk çalışmalar olsa da, Schrödinger ve Smoluchowsky ‘den önce Louise Bachelier, “Théorie de la spéculation” isimli doktora tezinde (1900) Ters Normal dağılımının temeli olan fonksiyonlarla Brownian önergelerin fiziksel olgusunu çalışmıştır. (31, 32)

Daha sonra 1940 yılına gelindiğinde Hadwiger, biyolojik popülasyona bağlı türev fonksiyonu incelemiş ve Ters Normal dağılımın temeli olan yoğunluk fonksiyonunu elde etmiştir. (Hadwiger’in bu çalışması Hoem ve Berge (1975) ile Jorgensen (1982) yılında ‘Hadwiger fertility (üremesi) olarak isimlendirilmiştir. (31)

Tweedie 1941 yılındaki tezi ile adı o zamanda henüz belirlenmemiş olan Ters Normal dağılımı kapsamlı bir istatistiksel bakış açısıyla incelemiştir. (31)

1944 yılında ise, ‘rasgele değişkenlerin kümülatif toplamı’ çalışmasıyla Wald, dağılımı tekrardan gündeme getirmiştir. (31)

1945’te Tweedie, *Brownian tipi harekette ilk geçiş zamanı* “first passage time” dağılımının logaritmik moment üreten fonksiyonu ile normal dağılımın ilişkisini göstermiş ve bu nedenle dağılımın ismini “Ters Normal Dağılım” olarak ön gördüğü bir makaleyi yayınlamıştır. (14, 19)

1947 yılında Wald, ters normal dağılımın standardize edilmiş versiyonuna ulaşmıştır. Bu sebeple özellikle Rus literatüründe bu versiyon ‘Wald dağılımı’ olarak da isimlendirilmiştir. (Zigangirov (1962)) (14, 31)

1977’de Chhikara ve Folks, dağılımın, olasılık yoğunluk grafiğinin diğer yaşam süresi dağılımlarına (Gama, Lognormal ve Weibull) benzerliği ve örnekleme dağılımı teorisinin elverişli olması sebebiyle yaşam süresi dağılımı olarak incelenebileceğini göstermişlerdir. Ters normal dağılımın, biyoistatistik alanında (yaşam süresi analizinde) kullanılabileceği fikri bu çerçevede düşünülmüştür. (14, 25)

Daha sonra literatürde, 1978 yılında Folks ve Chhikara, 1988 yılında Iyengar ve Patwardhan, 1993 ve 1999’da Seshadri ve 2002, 2004 yıllarına gelindiğinde de Mudholkar ve Natarajan’ın bu dağılım üzerine yer alan çeşitli çalışmalarını görmekteyiz. (14, 28, 29, 30, 31,32)

3.2. Ters Normal Dağılımın Karakteristik Özellikleri

Ters Normal (Gaussian) Dağılım, kalite ve güvenilirlik kuramı ile yaşam testlerinde kullanılan dağılımlardan biridir. (1, 2) Bununla birlikte demografi, kardiyoloji, histomorfoloji, elektrik ağları, hidroloji, yönetim bilimleri, meteoroloji, psikiyatri, psikoloji, uzaktan algılama sistemi, trafik gürültü şiddeti, pazar araştırmaları gibi bir çok alanda Ters Normal dağılımı uygulama olanağı vardır. (27, 28, 32)

Yaşam ve insana ait istatistiksel veriler genellikle pozitif çarpık ölçümlerden oluşur. Ters Normal Dağılım, bu tür verinin analizi için çok yönlü ve esnek bir model elde etmeyi sağlar. (29, 30) Ters Normal Dağılım aynı zamanda *Brownian hareketteki* ilk geçiş zamanının pozitif eğilimli dağılımı olarak da bilinir. (31) Üstel dağılım ailesine ait, tek modlu (unimodal) bir dağılımdır. Konum ve ölçek olmak üzere iki parametresi vardır. Ters Normal dağılım her ne kadar isim itibarıyla yanlış bir yönlendirmeye sebep olabilse de, en cazip özelliği, en çok bilinen Gaussian (Normal) model ile birçok benzer özelliğe sahip olmasıdır. (Örneğin, $\bar{X}$ örnek ortalaması ve μ ve λ parametrelerinin bağımsız olması gibi) (7, 26)

Ters normal dağılım, risk (hazard) fonksiyonunun monoton olmadığı durumlarda sağa çarpık pozitif verilere uygunluk sağlar. Risk (hazard) fonksiyonu ilk dönemlerde yükselerek bir doruk noktasına ulaşır ve sonra hızla düşer. Ters Normal dağılım ile hazard fonksiyonu monoton olmayan Lognormal dağılım arasındaki

ayırıcı en önemli özellik; x sonsuza giderken $(x \rightarrow \infty)$ Ters Normal dağılımın 0 olmayan bir değere yakınsamasıdır. Oysa Lognormal dağılımın risk (hazard) fonksiyonu x sonsuza giderken $(x \rightarrow \infty)$ 0'a yakınsar ki bu durumun gerçekleşmesi özellikle insan ömrü gibi fiziksel olgular için beklenemez. (Örneğin, insan ömrü için risk (hazard, ölüm) hiçbir zaman 0 olmaz.) (6)

Brownian hareket (Wiener süreci): 1828'de İngiliz Botanikçi Robert Brown, su içindeki çiçek polenlerinin düzensiz hareket ettiğini tespit etmiştir. 1905'te Albert Einstein, normal dağılımı Brownian hareketin bir modeli olarak türetmiştir. Norbert Wiener de 1923'te Brownian hareketin matematiksel temellerini oluşturmuştur. Bu çalışmalar, stokastik süreçlerin tanımlanmasında önemli bir yer tutmaktadır. Bu süreç, rassal (rasgele) yürüyüş olarak da bilinmektedir. Rassal yürüyüş bağımsız ve aynı dağılmış rassal değişkenlerin bir serisinden oluşmuş bir stokastik süreçtir. (7, 16, 38)

Brownian hareketin özelliklerini şu şekilde tanımlayabiliriz: (<http://mathworld.wolfram.com/WienerProcess.html>, Erişim tarihi: 19.06.2009)

$B(t)$, $t > 0 \in [0, T]$ rassal süreci,

- Sıfır ile başlar. $B(0)=0$.

- t zamanında süreklidir.

- Bağımsız sabit artışlara sahiptir ve bu artışların dağılımı sıfır ortalama ve $t - s$ ile normal dağılır.

$$B_t - B_s \sim N(0, t - s) \quad 0 \leq s < t$$

3.2.1. Ters Normal dağılım olasılık yoğunluk fonksiyonu

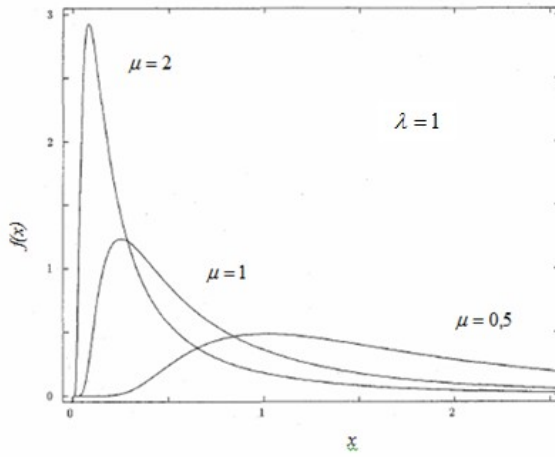
İki parametrelili Ters Normal dağılım için, dağılımın ölçek parametresi $\lambda > 0$, konum parametresi $\mu > 0$ ve rasgele değişken $x > 0$ olmak üzere dağılımın olasılık yoğunluk fonksiyonu (3.1)'deki gibi tanımlanır. (26)

$$f(x) = \sqrt{\frac{\lambda}{2\pi x^3}} e^{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}} \quad (3.1)$$

$$\text{Ters Normal dağılımın şekil parametresi } \theta = \frac{\lambda}{\mu} \quad (3.2)$$

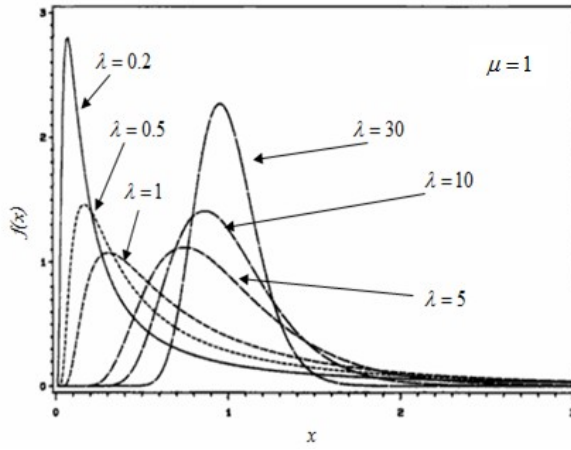
‘dir.

Ters Normal dağılımın olasılık yoğunluk fonksiyonuna ait grafik Şekil 17.’de verilmiştir. (21)



Şekil 17. Farklı μ değerleri için, Ters Normal dağılımının olasılık yoğunluk fonksiyonunun grafiksel gösterimi

Farklı λ değerleri için Ters Normal dağılımın olasılık yoğunluk fonksiyonu grafiği Şekil 18.’de verilmiştir. (7)



Şekil 18. Farklı λ değerleri için, Ters Normal dağılımının olasılık yoğunluk fonksiyonunun grafiksel gösterimi

Şekil 18'den de görüleceği üzere, λ sonsuza giderken ($\lambda \rightarrow \infty$) Ters Normal dağılım, normal dağılıma yaklaşır. (7, 26)

Ters Normal dağılımın varyansı, çarpıklık ve basıklık katsayıları (3.3), (3.4) ve (3.5) eşitlikleri ile verilmiştir. (8)

$$\text{Varyans; } \sigma^2 = \frac{\mu^3}{\lambda} \quad (3.3)$$

$$\text{Çarpıklık katsayısı; } \delta_1 = 3\sqrt{\frac{\mu}{\lambda}} \quad (3.4)$$

$$\text{Basıklık katsayısı; } \delta_2 = 3 + 15\frac{\mu}{\lambda} \quad (3.5)$$

λ 'dan bağımsız olarak çarpıklık ve basıklık katsayıları sırasıyla (3.6) ve (3.7) eşitlikleri ile hesaplanır. (30)

Çarpıklık katsayısı;

$v = E(X^{-1})$, $\mu'_2 = E(X^2)$ ve $\sigma^2 = \text{var}(X)$ olmak üzere,

$$\delta_1 = \frac{\mu'_2/\mu^2 - \mu v}{(\mu v - 1)\sqrt{\mu'_2/\mu^2 - 1}} = \frac{\mu'_2 - \mu^3 v}{(\mu v - 1)\mu \sigma} \quad (3.6)$$

Basıklık katsayısı;

$$\delta_2 = \frac{\eta^2 \mu^2}{(\nu \mu - 1)^2} + 1 \quad (3.7)$$

Burada $\eta^2 = \tau^2 + 2(1 - \mu \nu) / \mu^2 + \sigma^2 / \mu^4$ ve $\tau^2 = \text{var}(X^{-1})$ eşitlikleri ile hesaplanır.

3.2.2. Ters Normal dağılım birikimli dağılım fonksiyonu

Ters Normal dağılımın birikimli dağılım fonksiyonu (3.8)'deki gibi tanımlanır.
(7)

$$F(x) = \Phi \left(\sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} - 1 \right) \right) + e^{\left(\frac{2\lambda}{\mu} \right)} \Phi \left(- \sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} + 1 \right) \right) \quad (3.8)$$

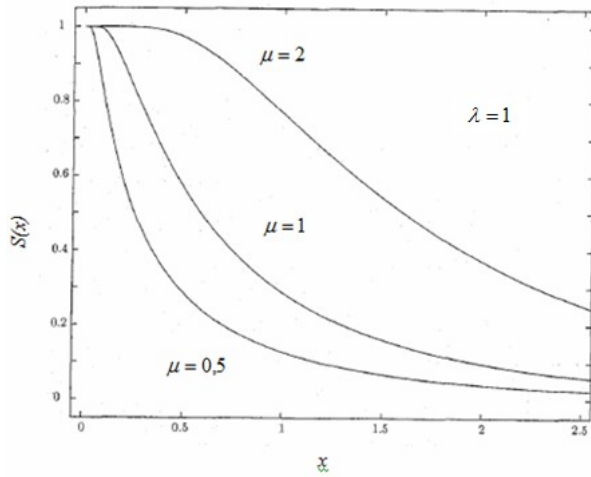
Burada $\Phi(\cdot)$, Laplace integralidir. (Standart normal dağılımın birikimli dağılım fonksiyonu) (2.26)'daki gibi tanımlanır.

3.2.3. Ters Normal dağılımın sağkalım fonksiyonu

Ters Normal dağılımın sağkalım fonksiyonu (3.9) eşitliği ile verilmiştir.

$$S(x) = 1 - F(x) = 1 - \Phi \left(\sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} - 1 \right) \right) + e^{\left(\frac{2\lambda}{\mu} \right)} \Phi \left(- \sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} + 1 \right) \right) \quad (3.9)$$

Ters Normal dağılıma ait sağkalım fonksiyonu grafiği Şekil 19'da verilmiştir. (21)



Şekil 19. Farklı μ değerleri için, Ters Normal dağılımının sağkalım fonksiyonunun grafiksel gösterimi

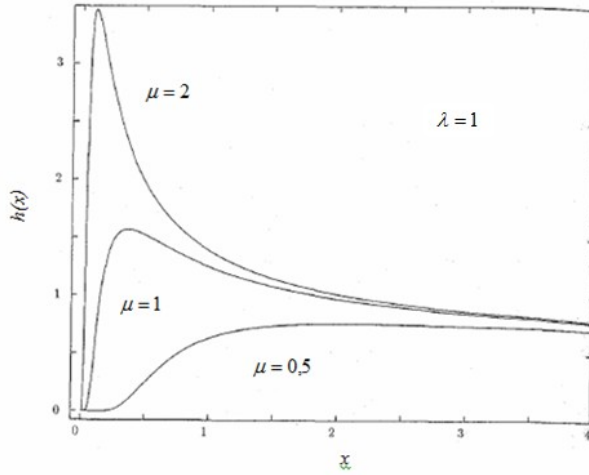
3.2.4. Ters Normal dağılım risk (hazard) fonksiyonu

Ters normal dağılım, risk (hazard) fonksiyonunun monoton olmadığı durumlarda sağa çarpık pozitif verilere uygunluk sağlar. Risk (hazard) fonksiyonu ilk dönemlerde yükselerek bir doruk noktasına ulaşır ve kavuşmazda 0 olmayan bir değere yakınsar. (6)

Ters Normal dağılımın risk (hazard) fonksiyonu (3.10) eşitliği ile elde edilir.

$$h(x) = \frac{f(x)}{S(x)} = \frac{f(x)}{1 - F(x)} = \frac{\sqrt{\frac{\lambda}{2\pi x^3}} e^{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}}}{1 - \Phi\left(\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} - 1\right)\right) + \Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} + 1\right)\right) e^{\left(\frac{2\lambda}{\mu}\right)}} \quad (3.10)$$

Ters Normal dağılıma ait risk (hazard) fonksiyonu grafiği Şekil 20.'de verilmiştir. (21)



Şekil 20. Farklı μ değerleri için, Ters Normal dağılımının risk (hazard) fonksiyonunun grafiksel gösterimi

3.3. Ters Normal Dağılım Parametre Tahmini

Bir veri setine ait dağılımın parametre tahmini, nokta veya aralık tahmini ile yapılabilir.

Başlıca nokta tahmini yöntemleri;

- En çok olabilirlik (EÇO) tahmin yöntemi (Maximum Likelihood)
- Momentler tahmin yöntemi (Method of Moments)
- En küçük hata kareler ortalaması tahmin yöntemi (Minimum Mean Square Error)
- Minimum varyans yansız tahmin yöntemi (Minimum Variance Unbiased)
- En iyi doğrusal yansız tahmin yöntemi (Best Linear Unbiased) ‘dir.

(6, 22)

Aşağıda, bu tahmin yöntemlerinden en sık kullanılan EÇO tahmin yönteminden bahsedilecektir. Ayrıca minimum varyans yansız tahmin edicisine de değinilecektir.

En Çok Olabilirlik parametre tahmini (EÇO): EÇO parametre tahmin yönteminin amacı, örnek verilerinin olasılığını (likelihood) maksimize ederek ana kitle parametreleri için tahmin ediciler bulmaktır. (4)

EÇO yöntemi birçok modele ve farklı veri tiplerine uygulanabilen çok yönlü bir tahmin yöntemidir. Yaşam süresi dağılımlarında kullanılabilirler. EÇO tahmin yönteminin avantajlarından biri, matematiksel uygulama kolaylığıdır. Birçok paket programı sık kullanılan dağılımların parametrelerinin EÇO tahmini için çok iyi algoritmalar geliştirmiştir. Böylece hesaplamadaki karmaşıklık ortadan kalkmıştır. (<http://www.itl.nist.gov/div898/handbook/apr/section4/apr412.htm> Erişim tarihi 19.06.2009)

EÇO tahmininin matematiksel gösterimi aşağıdaki gibidir:

x , rasgele sürekli bir değişkenin değeri ve $\theta_1, \theta_2, \dots, \theta_k$ k tane bilinmeyen (tahmin edilmek istenen) parametre olmak üzere dağılımın olasılık yoğunluk fonksiyonu; $f(x; \theta_1, \theta_2, \dots, \theta_k)$ şeklindedir. n adet bağımsız gözlem için $(x_1, x_2, \dots, x_n)$ olabilirlik (likelihood) fonksiyonu (3.11)'deki gibi olmaktadır. (12)

$$L(x_1, x_2, \dots, x_n | \theta_1, \theta_2, \dots, \theta_k) = L = \prod_{i=1}^n f(x_i; \theta_1, \theta_2, \dots, \theta_k). \quad (3.11)$$

Logaritmik olabilirlik fonksiyonu ise (3.12)'de verilmektedir.

$$\Lambda = \ln L = \sum_{i=1}^n \ln f(x_i; \theta_1, \theta_2, \dots, \theta_k). \quad (3.12)$$

$\theta_1, \theta_2, \dots, \theta_k$ parametrelerinin EÇO tahmini L 'nin ya da Λ 'nin maksimizasyonu ile elde edilir. Bu durumda EÇO tahmin değerleri (3.13) denkleminin çözümü ile elde edilir.

$$\frac{d(\ln L)}{d\theta_j} = \frac{d\Lambda}{d\theta_j} = 0, \quad (j = 1, 2, \dots, k) \quad (3.13)$$

EÇO tahmin edicileri, örnek büyüklüğü arttıkça “minimum varyans yansız tahmin edicisi” ne dönüşür. Yansız tahmin edici, θ anakitle parametresi olmak üzere, θ 'nın tüm örnek büyüklüklerinde beklenen değerinin kendisine eşit olduğu durumda elde edilir. ($E(\theta) = \theta$). Minimum varyanslı tahmin edici ise, en küçük

varyansa sahip olan ve güven aralığını en dar yapan tahmin edicidir. (<http://www.itl.nist.gov/div898/handbook/apr/section4/apr412.htm> Erişim tarihi: 19.06.2009), (4)

3.3.1. Ters Normal dağılım parametreleri için EÇO tahminleri

μ konum parametresi, λ ölçek parametresi olmak üzere Ters Normal Dağılım Parametreleri için EÇO tahmin edicileri (3.14) ve (3.15)'te verilmiştir. (9, 36)

$$\text{Konum parametresi; } \hat{\mu} = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (3.14)$$

$$\text{Ölçek parametresi; } \hat{\lambda}^{-1} = \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{X_i} - \frac{1}{\bar{X}} \right) \quad (3.15)$$

Ayrıca σ^2 varyans için EÇO tahmini (3.16)'da verilmiştir.

$$\text{Varyans; } \hat{\sigma}^2 = \frac{1}{n} \left[\sum_{i=1}^n \frac{\bar{X}^3}{X_i} - n\bar{X}^2 \right] \quad (3.16)$$

Ters Normal dağılım parametreleri $\hat{\mu}$ ve $\hat{\lambda}^{-1}$ Normal dağılıma benzer olarak birbirinden bağımsızdırlar. (7)

$\hat{\mu}$, aynı zamanda minimum varyans yansız tahmincisidir. Ancak λ^{-1} için aynı şey söylenemediğinden λ^{-1} için minimum varyans yansız tahmincisi (3.17)'teki gibi hesaplanır. (6)

$$\hat{\lambda}^{-1} = \frac{1}{n-1} \sum_{i=1}^n \left(\frac{1}{X_i} - \frac{1}{\bar{X}} \right) \quad (3.17)$$

3.3.2. Ters Normal dağılım parametreleri için güven aralıkları

Konum parametresi μ için güven aralığı: μ , Ters Normal dağılımın ortalamasının EÇO tahmincisi ve aynı zamanda örnek ortalaması olmak üzere, n büyüklüğündeki bir örneğe ait μ dağılımı da örnek değerleri (x 'lerin ortalaması) ile aynı ortalama sahiptir. (36)

Bu durumda Ters Normal Dağılımın konum parametresi olan μ için güven aralığı (3.18) veya (3.19)'deki gibi hesaplanır. (5)

$$V(\hat{\mu}) = \frac{\mu^3}{n\lambda} \text{ olmak üzere;}$$

$$\hat{\mu} \pm z_{1-(\alpha/2)} V(\hat{\mu}) \text{ veya } \hat{\mu} \pm t_{1-(\alpha/2)} V(\hat{\mu}) \quad (3.18)$$

$$\hat{\mu} \pm z_{1-(\alpha/2)} \frac{\mu^3}{n\lambda} \text{ veya } \hat{\mu} \pm t_{1-(\alpha/2)} \frac{\mu^3}{n\lambda} \quad (3.19)$$

Ölçek parametresi λ için güven aralığı: Ters Normal dağılım ölçek parametresi λ^{-1} , $(n\lambda)^{-1} \chi^2_{n-1}$ olarak dağılır. Bu durumda λ için güven aralığı (3.20) eşitliği ile hesaplanır. (2, 36)

$$\left(\frac{\hat{\lambda}}{n} \chi^2_{n-1, \alpha/2}, \frac{\hat{\lambda}}{n} \chi^2_{n-1, 1-\alpha/2} \right) \quad (3.20)$$

3.4. Ters Normal Dağılım Uyum İyiliği (Uyumluluk) Testleri

Uyum iyiliği testleri rasgele bir örneğin, teorik olasılık dağılım fonksiyonu ile uyumunu ölçen testlerdir. Başka bir deyişle, uyum iyiliği testleri, seçilen dağılımın elimizdeki verilere ne kadar uygun olduğunu gösterir. (22)

Örneğin Ters Normal Dağılım, Weibull ve Lognormal dağılımlarıyla çok benzeşir. Her üçü de uzun kuyruklu (long-tailed) (çarpık) verilere uygun şekillidir. Ancak uygun model seçiminde önemli olan, testin dağılımların benzer olmayan şekli arasındaki ayrımı iyi yapabilmesidir. (32)

Aşağıda, yaygın olarak uygulanan ve aynı zamanda bu tez çalışmasında uyum iyiliği için paket program olarak kullanılan EasyFit 5.0 (EasyFit 2008) paket programının da kullandığı Kolmogorov-Smirnov ve Anderson-Darling uyum iyiliği testlerinden söz edilecektir.

Ayrıca, ileri irdeleme yöntemlerinden olan çarpıklık ve basıklık katsayıları yaklaşımı ile uyum iyiliği testleri anlatılacaktır.

3.4.1. Kolmogorov-Smirnov testi

Nitel verilere uygulanan χ^2 (Ki-Kare) uyum iyiliği testinin, nicel sürekli verilere uygulanmasıdır. Kolmogorov ve Smirnov adlı Rus matematikçilerin adını taşıyan bir testtir. (4) Kolmogorov-Smirnov testi, ele alınan bir örneğin varsayılan sürekli dağılıma sahip olup olmadığına karar vermek için kullanılır. Deneysel (ampirik) birikimli dağılım fonksiyonuna dayalı bir testtir. $F(x)$ birikimli dağılım fonksiyonuna sahip $x_1, x_2, \dots, x_n$ rasgele bir örneğimiz olduğunu varsayarsak “deneysel birikimli dağılım fonksiyonu” (3.21) gibi gösterilir. (22)

$$F_n(x) = \frac{1}{n} [\text{Gözlem sayısı} \leq x] \quad (3.21)$$

Kolmogorov-Smirnov test istatistiği D , teorik birikimli dağılım fonksiyonu ile ampirik birikimli dağılım fonksiyonu arasındaki en büyük düşey farka dayanır ve (3.22) gibi gösterilir.

$$D = \max_{1 \leq i \leq n} \left(F(x_i) - \frac{i-1}{n}, \frac{i}{n} - F(x_i) \right) \quad (3.22)$$

Test hipotezleri;

H_0 : Veriler belirtilen dağılıma uymaktadır.

H_1 : Veriler belirtilen dağılıma uymamaktadır.

şeklinde ifade edilir.

Hesaplanan D değeri, teste ait kritik değerler D tablosundan elde edilen değerden daha küçükse hipotez kabul edilir. (20)

Bu testin etkileyici bir özelliği; test istatistiğinin, test edilen dağılımdan bağımsız olmasıdır. Testin başka bir avantajı da her büyüklükteki örnekleme uygulanabilir

olmasıdır. (20) Buların yanı sıra Kolmogorov-Smirnov testinin bazı sınırlamaları da vardır. (<http://www.itl.nist.gov/div898/handbook/eda/section3/eda35g.htm>, Erişim tarihi: 19.06.2009)

Bunlar:

- Test yalnızca sürekli dağılımlara uygulanır.
- Dağılımın merkezi civarında test daha hassas olma eğilimindeyken, merkezden uzaklaştıkça hassasiyet azalır.

3.4.2. Anderson-Darling testi

Anderson-Darling testi, gözlenen birikimli dağılım fonksiyonunun beklenen birikimli dağılım fonksiyonuna uyumunu karşılaştırmak için kullanılan genel bir testtir. Kolmogorov-Smirnov testinin modifiye edilmiş hali olan bu test kuyruklara Kolmogorov-Smirnov testinden daha fazla ağırlık vermektedir. (<http://www.itl.nist.gov/div898/handbook/eda/section3/eda35e.htm>, Erişim tarihi: 19.06.2009), Bu özellik, sağa çarpık olan yaşam süresi dağılımlarının ayırımı için önemli bir avantaj oluşturur.

Anderson-Darling testinde kritik değerler her dağılım için özel olarak hesaplanmalıdır. Bu bir dezavantajdır ancak bu özellik testin daha hassas olmasını sağlar. (22)

Anderson-Darling istatistiği (3.23) ile hesaplanır.

$$A^2 = n - \frac{1}{n} \sum_{i=1}^n (2i - 1) [\ln F(x_i) + \ln(1 - F(x_{n-i+1}))] \quad (3.23)$$

Anderson-Darling test hipotezi de Kolmogorov-Smirnov test hipotezinde olduğu gibi;

H_0 : Veriler belirtilen dağılıma uymaktadır.

H_1 : Veriler belirtilen dağılıma uymamaktadır.

şeklinde ifade edilir.

Hesaplanan test istatistiği değeri, göz önünde bulundurulan kritik değerden daha küçükse hipotez kabul edilir. (22)

3.4.3. Çarpıklık Basıklık katsayısı yaklaşımı ile uyum iyiliği (uyumluluk) testi

Çarpıklık ve basıklık katsayıları model seçimi konusunda önemli bir yer tutar. Bu katsayılar, bir olasılık dağılımının doğasını anlamada ve bu dağılımın diğer dağılımlarla karşılaştırmasında kullanılan önemli ölçütlerdir. (29, 30)

Daha önce belirtildiği üzere Ters Normal Dağılım, birçok özelliği ile Normal dağılıma benzerlik göstermektedir. Mudholkar ve Natarajan 2002 yılındaki araştırmalarında δ_1 ve δ_2 moment fonksiyonlarını Ters Normal Çarpıklığı ve Ters Normal Basıklığı olarak isimlendirmişlerdir. Bu yazılarında normal dağılıma ait klasik çarpıklık ve basıklık katsayıları $(\sqrt{\beta_1}$ ve $\beta_2)$ ile Ters Normal çarpıklık ve basıklık katsayıları $(\delta_1$ ve $\delta_2)$ arasındaki paralelliği tartışmışlardır. (29)

Natarajan ve Mudholkar 2004 yılında yayınladıkları makalede ise $(\delta_1$ ve $\delta_2)$ katsayıların tahmin değerleri olan $(d_1$ ve $d_2)$ örnek momentlerinden yola çıkarak Ters Normal dağılım için uyum iyiliği testlerini ele almışlardır. (30) Aşağıda bu çalışma içerisinde yer alan Ters Normal çarpıklık ve basıklık katsayıları açıklanacak ve bu katsayılardan hareketle uyum iyiliği testi anlatılacaktır.

Ters Normal Çarpıklığı Ters Normal Simetrisi üzerine kurulduğundan önce Ters Normal Simetrisini kısaca açıklamak gerekir.

Ters Normal simetri: $E(X) = \mu$ iken, X bir rasgele değişken olmak üzere, negatif ve pozitif sıralı $(r = \pm 1, \pm 2, \dots, k)$ tüm momentler, (3.24) eşitliğini sağladığında, μ etrafında Ters Normal simetriktir denir.

$$E\left[\left(\frac{X}{\mu}\right)^{-r}\right] = E\left[\left(\frac{X}{\mu}\right)^{r+1}\right] \quad r = 1, 2, \dots \quad (3.24)$$

Lognormal dağılım da Ters Normal simetri göstermektedir. Bu özellik (3.24) eşitliği ile kanıtlanabilir. (30)

3.4.3.1. Ters Normal dağılım çarpıklık katsayısı testi

Birinci moment olan $E(x)$ genellikle konum belirten bir değerdir. Standartlaştırılmış üçüncü moment ise $\sqrt{\beta_1}$ çarpıklık katsayısını tanımlar. Ters Normal dağılım çarpıklık katsayısı olan δ_1 , $r = 1$ için (3.24) eşitliğindeki sol ve sağ taraf arasındaki farka eşittir. (30)

$$\delta_1, \quad r = 1 \quad \text{için,} \quad E\left[\left(\frac{X}{\mu}\right)^{-1}\right] - E\left[\left(\frac{X}{\mu}\right)^2\right] \quad \text{ile hesaplanır.}$$

Bir başka deyişle δ_1 çarpıklık katsayısı, Ters Normal simetrisinin birinci formuna ait standartlaştırılmış versiyonu olarak yorumlanır.

Anakitle için Ters Normal çarpıklık katsayısı δ_1 , $\sqrt{\beta_1}$ çarpıklık katsayısına benzer olarak ve daha önce verildiği üzere (3.6)'daki gibi hesaplanır.

d_1 örnekten elde edilen Ters Normal dağılım çarpıklık katsayısının tahmin değeri olmak üzere (3.28) ile elde edilir. $\bar{X}$, örnek ortalamasıdır;

$$m_2' = \sum_{i=1}^n X_i^2 / n \quad (3.25)$$

$$s^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / n \quad (3.26)$$

$$\bar{Y} = \sum_{i=1}^n X_i^{-1} / n \quad (3.27)$$

iken,

$$d_1 = \frac{m'_2 / \bar{X}^2 - \bar{X}\bar{Y}}{(\bar{X}\bar{Y} - 1)\sqrt{m'_2 / \bar{X}^2 - 1}} = \frac{m'_2 - \bar{X}^3\bar{Y}}{s\bar{X}(\bar{X}\bar{Y} - 1)} \quad (3.28)$$

$\tilde{d}_1$, örnekten elde edilen Ters Normal dağılım çarpıklık katsayısının Jackknife tahminidir ve (3.29) ile hesaplanır.

$$\tilde{d}_1 = \sum_{i=1}^n P_{n,i}(d_1) / n \quad (3.29)$$

$P_{n,i}(d_1)$ değerleri yalancı değerlerdir (pseudo values), (3.30) ile elde edilir.

$$P_{n,i}(d_1) = nd_1 - (n-1)d_{1,i}, \quad i = 1, 2, \dots, n \quad (3.30)$$

Burada $d_{1,i}$, i 'nci gözlem silindiğinde elde edilen d_1 değeridir. $\tilde{d}_1$ 'in dağılımı, şekil parametresi (θ) ve örnek büyüklüğüne (n) bağlıdır.

$m(\tilde{d}_1)$, $\tilde{d}_1$ değeri için tahmin edilen ortalama, (3.31)'deki gibidir.

$$m(\tilde{d}_1) = \frac{-1}{n\sqrt{\hat{\theta}}} \quad (3.31)$$

$s^2(\tilde{d}_1)$, $\tilde{d}_1$ değeri için tahmin edilen varyans ise, (3.32)'deki gibidir.

$$s^2(\tilde{d}_1) = \frac{6}{n} \left(1 + \frac{3}{\sqrt{n}} e^{(-0,04\hat{\theta})} \right) \quad (3.32)$$

Şekil parametresi (θ) için tahmin değeri, (3.33)'te yer almaktadır.

$$\hat{\theta} = (n-1) / (n\bar{X}V) \quad (3.33)$$

$$V = (1/n) \sum_{i=1}^n (1/X_i - 1/\bar{X}) \quad (3.34)$$

dir.

$\tilde{d}_1$, (3.31)'de verilen ortalama ve (3.32)'de verilen varyans ile yaklaşık olarak normal dağılmaktadır. Ters Normal dağılım hipotezini test eden düşük uçlu (lower-tailed) p değeri (3.35)'de verilmektedir.

$$p = P\{Z < (\tilde{d}_1 - m(\tilde{d}_1))/s(\tilde{d}_1)\} = P\{Z < z\} = \Phi(z) \quad (3.35)$$

Ters Normal dağılım hipotezini test eden çift taraflı p değeri ise (3.36) ile elde edilir.

$$2\text{-taraflı } p \text{ degeri} = 2 \min(p, 1-p) \quad (3.36)$$

3.4.3.2. Ters Normal dağılım basıklık katsayısı testi

Ters Normal simetrisi gösteren benzer dağılımları (Ters Normal, Lognormal gibi) birbirinden ayırt etmek için kullanılan asıl ölçüt, δ_2 Ters Normal basıklık katsayısıdır. (30)

Klasik basıklık katsayısı, istatistiksel literatürde sivrilik (peakedness) ölçüsü olarak bilinir. Dağılımların, az basık, orta basık veya çok basık olduğunu tanımlamak için kullanılır. (29)

Anakitle için Ters Normal basıklık katsayısı δ_2 , β_2 basıklık katsayısına benzer olarak daha önce (3.7) verildiği gibi hesaplanır.

d_2 , örnekten elde edilen Ters Normal dağılım basıklık katsayısının tahmin değeri olmak üzere:

$$m'_2 = \sum_{i=1}^n X_i^{-2} / n \quad (3.37)$$

(3.25), (3.27) ve (3.37) eşitlikleri kullanılarak (3.38)'deki gibi hesaplanır.

$$d_2 = \frac{(m'_2 + m'_2 \bar{X}^4 - 4\bar{X}^3 \bar{Y} + 2\bar{X}^2)}{\bar{X}^2 (\bar{X} \bar{Y} - 1)^2} \quad (3.38)$$

Natarajan ve Mudholkar 2004 yılındaki çalışmalarında d_2 'nin sağa çarpık ve çok basık olmak üzere normal dağılmadığını göstermişlerdir. Bu nedenle, d_2 yerine (3.39) dönüşümünün kullanılmasını önermişlerdir. (30)

$$T_2 = \ln(d_2 - 1) \quad (3.39)$$

Ayrıca Natarajan ve Mudholkar yine 2004 yılındaki çalışmalarında Ters Normal dağılımın basıklık katsayısı δ_2 'nin Normal dağılımın basıklık katsayısı β_2 'ye benzer olarak 3 olduğunu göstermişlerdir. Bu durumda T_2 için beklenen değer $\ln(2)$ 'ye çok yakın olmaktadır. ((3.39) eşitliği kullanılarak, $\ln(3 - 1) = \ln(2)$)

$\tilde{T}_2$, örnekten elde edilen Ters Normal basıklık katsayısının Jackknife tahminidir ve (3.40) ile hesaplanır.

$$\tilde{T}_2 = \frac{\sum_{i=1}^n P_{n,i}(T_2)}{n} \quad (3.40)$$

(3.30) eşitliğine benzer olarak; $P_{n,i}(T_2)$ değerleri yalancı değerlerdir (pseudo values) ve (3.41) ile elde edilirler.

$$P_{n,i}(T_2) = nT_2 - (n-1)T_{2,i}, \quad i = 1, 2, \dots, n \quad (3.41)$$

$m(\tilde{T}_2)$, $\tilde{T}_2$ Jackknife tahmin değeri için ortalama, (3.42) olarak değerlendirilir.

$$m(\tilde{T}_2) = \ln 2 \quad (3.42)$$

$s^2(\tilde{T}_2)$, $\tilde{T}_2$ Jackknife tahmin değeri için varyans, (3.43) ile hesaplanır.

$$s^2(\tilde{T}_2) = \frac{6}{n} \left(1 + \frac{6}{n} \right) \quad (3.43)$$

$\tilde{T}_2$, (3.42)'de verilen $\ln(2)$ ortalama ve (3.43)'te verilen varyans ile yaklaşık olarak normal dağılmaktadır. Ters Normal dağılım hipotezlerini test eden düşük uçlu (lower-tailed) p değeri (3.44)'deki gibi olmaktadır.

$$p = P\{Z < (\tilde{T}_2 - \ln 2)/s(\tilde{T}_2)\} = P\{Z < z\} = \Phi(z) \quad (3.44)$$

Ters Normal dağılım hipotezini test eden çift taraflı p değeri ise (3.36) gibi elde edilir.

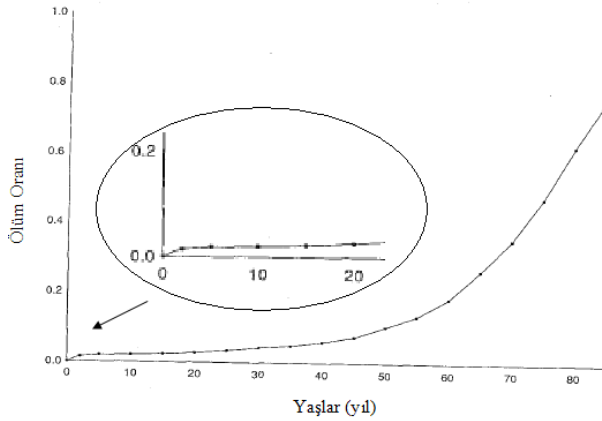
4. SAĞKALIM ANALİZİ

Sağkalım Analizi, belirli bir hastalığa yakalanan bir bireyin hastalığının tanısı üzerine, uygulanan bir girişimden (tıbbi tedavi, operasyon, kemoterapi vb.) sonra; daha ne kadar yaşayabileceğini, ve hastalığın ne kadar sürede nüksedebileceğini tahmin etmek, tedavi tiplerinin ve diğer faktörlerin sağkalım süresine etkilerini incelemek amacıyla geliştirilmiş yöntemleri içerir. (11, 34)

Bununla beraber, günümüzde Sağkalım Analizi'nin, üretim alanında bir aletin ömrünün analizi için kullanılabildiği gibi daha birçok alanda da kullanıldığını görmekteyiz. (<http://www.bmj.com/collections/statsbk/12.dtl> , Erişim tarihi: 19.06.2009)

4.1. Sağkalım Fonksiyonu (Sağkalım Eğrisi)

Bir değişkenin dağılımını tanımlamak için genellikle birikimli dağılım fonksiyonu ve histogram kullanılır. Sağkalım verileri de birikimli dağılım fonksiyonu ile gösterilebilir. Örneğin Şekil 21., 1974 yılında Amerika Birleşik Devletleri'nde belli bir yaştan önce ölme olasılığının bir tahminini, grafikte göstermektedir. (13)



Şekil 21. 1974'te Amerika Birleşik Devletleri'ndeki bireylere ait sağkalımın birikimli dağılımının bir tahminini gösteren grafik

Şekil 21., 1974 yılında Amerika Birleşik Devletleri'ndeki sağkalımın birikimli dağılımının bir tahminidir. Bu şekilden görüldüğü üzere ilk yıl boyunca ölüm oranında hızlı bir artış söz konusudur. İlk yıldan sonra, oran düşmekte ancak sonraki

yıllarda giderek artmaktadır. (Şeklin büyütülmüş kısmında ilk yıldaki ölüm oranının hızlı artışı daha açık görülmektedir.) Bu grafik bir bireyin belirlenen bir zamanda (t) veya bu zamandan önce ölme olasılığının bir tahminidir. Bu durum (4.1)'deki gibi gösterilebilir. (13)

$$F(x) = P [\text{Bireyin ölme zamanı} \leq t] \quad (4.1)$$

Topluluktaki tüm bireyler gözlemlendiğinde, bu değer birikimli dağılım fonksiyonunda tahmin edildiği gibi (4.2) formülüne göre tahmin edilir.

$$F(x) = \frac{\text{\textcolor{teal}{t} zamanı veya t zamanından önce ölen bireylerin sayısı}}{\text{\textcolor{teal}{Gözlemlenen tüm bireyler}}} \quad (4.2)$$

Sağkalım analizinde, ölümün birikimli olasılığı yerine buna eş bir fonksiyon olan *sağkalım fonksiyonu* veya *sağkalım eğrisi* kullanılmaktadır. *Sağkalım fonksiyonu*, belirli bir t zamanına kadar veya daha fazla yaşayan bireylerin oranı ya da yüzdesidir. Bu durumda *sağkalım eğrisi* x 'in bir fonksiyonu olmak üzere;

$S(x)$ = Belirli bir t zamanına kadar veya daha fazla yaşayan bireylerin yüzdesi

veya

$S(x)$ = Belirli bir t zamanına kadar veya daha fazla yaşayan bireylerin oranı'dır.

Sağkalımın birikimli dağılım fonksiyonu ile sağkalım eğrisi yakından ilişkilidir. İki eğri sürekli olduğunda (4.3) ya da (4.4) eşitlikleri ile birbirine bağlıdır.

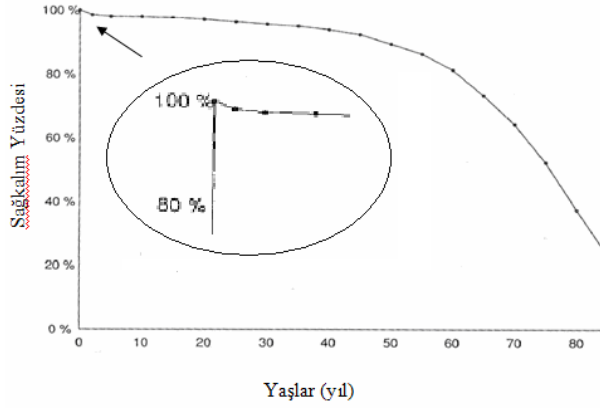
$$S(x) = 100(1 - F(x)) \quad (4.3)$$

veya

$$S(x) = 1 - F(x) \quad (4.4)$$

Şekil 22., daha önce birikimli dağılım eğrisi verilen Amerika Birleşik Devletleri'ndeki topluluğun sağkalım eğrisidir. Şekil 22.'den de görüldüğü gibi, sağkalım eğrisi, ölümün birikimli olasılığı eğrisinin ters dönmüş ve yüzdeliklerle

ifade edilmiş halidir. (13) (Şeklin büyütülmüş kısmında, ilk yıldaki sağkalım olasılığının hızlı düşüşü açıkça görülmektedir.)



Şekil 22. 1974'te Amerika Birleşik Devletleri'ndeki bireylere ait sağkalım eğrisi

Bir topluluktaki tüm bireyler gözlemlendiğinde sağkalım eğrisini tahmin etmek oldukça kolaydır. Bu durumda birikimli dağılım fonksiyonuna (4.2)'ye benzer olarak belirli bir t zamanına göre sağkalım tahmini (4.5)'teki gibi olacaktır.

$$S(x) = \frac{\text{Belirli bir } t \text{ zamanına kadar veya daha fazla yaşayan bireylerin sayısı}}{\text{Gözlemlenen tüm bireyler}} \quad (4.5)$$

Çoğu durumda bir topluluktaki tüm bireylerin tümünün ölene kadar ya da beklenen bir olay gerçekleşene kadar gözlemlenmesi mümkün olamamaktadır. Topluluktaki tüm bireylerin gözlemlenemediği durumda, sağkalım fonksiyonunu (eğrisini) tahmin etmek için bazı yöntemler geliştirilmiştir. Bunlardan biri, “Yaşam Tablosu Yöntemi”; diğeri ise, “Çarpım-Limit Yöntemi” veya “Kaplan-Meier Tahmini” dir.

Kaplan-Meier Tahmini'ne geçmeden önce, sağkalım analizinde tüm bireylerin beklenen bir olay (örneğin ölüm, hastalık, tekrarlama, iyileşme gibi araştırmacının ilgilendiği bir olay) gerçekleşene kadar gözlemlenemediği durumda ortaya çıkan “sansürlü veri” kavramından söz etmek doğru olacaktır. (10)

Sansürlü (censored) veriler: İleriye dönük (prospektif) bir analiz olan Sağkalım analizinde önceden belirlenen izlem süresi sonunda, ölüme ya da beklenen bir olguya

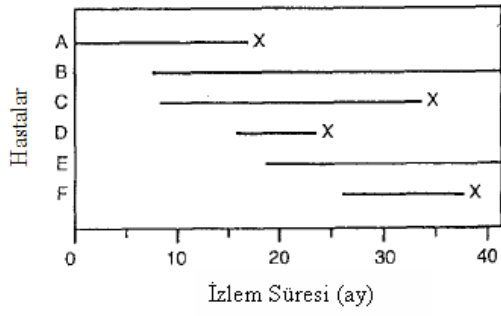
erişemeyen veriler (gözlemler) “sansürlü (censored) veriler” olarak adlandırılır. (3, 10) Analize dahil edilecek tüm hastaların ölüme ya da belirli bir olguya erişmesini beklemek, hangi tedavinin ya da yöntemin daha iyi sonuç verdiğini görme süresini uzatacağından, sansürlü veri kullanımına ihtiyaç duyulur.

Sansürlü veriler 3 farklı sebepten olabilir:

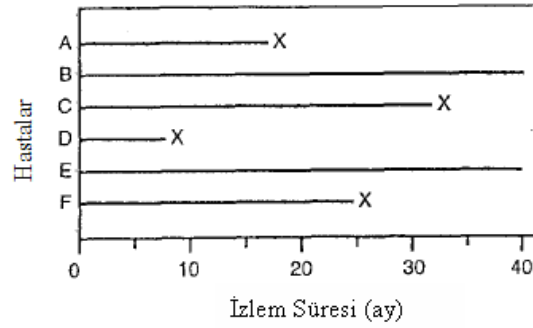
1. Çalışmanın bitiş noktasına kadar ölüm ya da beklenen olgunun gözlenmemesi (administrative censoring)
2. Çalışma bitmeden denekle ilgili bilgi alınamaması (lost to follow up)
3. Başka bir olgu (başka bir nedenden ölüm, ilaç reaksiyonu gibi) nedeni ile çekilme (withdrawing).

Sağkalım analizinde, hastaların izlenmeye başladığı tarih; tanı tarihi, tedaviye başladığı tarih veya cerrahi müdahale tarihi olabilir. Başlangıç noktasının tüm hastalar için aynı zamanda olması gerekmez. Veriler çalışmaya, çalışma süresi içinde farklı zamanlarda dahil olabilirler. Bu durumda gözlemler, “aşamalı (dereceli) sansürlü” olarak adlandırılır. Çalışma süresi (izlem süresi) bittiğinde aşamalı sansürlü olan gözlemlerin tümü, çalışmanın başlama noktasına çekilerek analiz edilirler.

Şekil 23.’te 40 aylık izlem süresine sahip, aşamalı sansürlü verileri daha açıkça görebiliriz. Burada, tüm hastaların izlenmeye başladıkları süreler farklıdır. “X” ile gösterilen hastalar ölen hastalar iken, B ve E hastaları sansürlü gözlemler olmaktadır. Şekil 24.’te ise görüldüğü üzere, farklı zamanlarda izlenmeye başlanan hastalar aynı noktaya çekilmişlerdir. (10, 13)



Şekil 23. Sağkalım analizi çalışmasına farklı zamanlarda giren hastaların çalışmaya girdikleri süreleri gösteren grafik



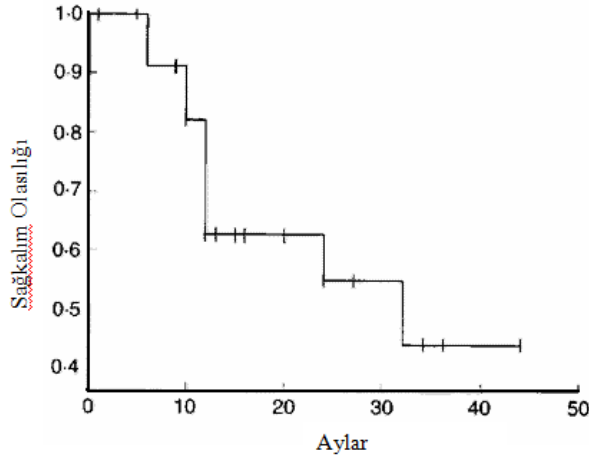
Şekil 24. Sağkalım analizi çalışmasına farklı zamanlarda giren hastaların çalışmaya girdikleri sürelerin çalışmanın başlama noktasına çekildiğini gösteren grafik

4.2. Sağkalım Fonksiyonu'nun Kaplan-Meier Tahmini

Çarpım-Limit Tahmini (Product Limit Estimator) olarak da bilinen Kaplan-Meier Tahmini, yaşam süresi verileri için sağkalım fonksiyonunu tahmin etmede kullanılan parametrik olmayan bir yöntemdir. (11, 34)

Yaşam tablosu yöntemi gibi diğer tahmin yöntemlerinde araştırmacının önceden belirlediği zaman aralıklarında gerçekleşen olgulara göre tahminler hesaplanmaktadır. Oysa Kaplan-Meier yönteminde, beklenen olgunun (ölümün) gerçekleştiği her noktada hesaplama yapılması söz konusudur. Bu özellik, Kaplan-Meier yönteminin parametrik yöntemlere göre üstün olan özelliğidir. Bu özellik doğrultusunda, Kaplan-Meier yöntemi ile sağkalım eğrisi oluşturulurken ölüm gözlenen zamandaki sıçramalar dışında eğri sabit olacaktır. (13) Şekil 25.'te de

görüldüğü üzere aylar ilerledikçe ölümün gerçekleştiği her bir zamanda sağkalım olasılığı düşmekte, ölüm görülmediği durumlarda (sansürlü gözlemler için) eğri sabit kalmaktadır. (<http://www.bmj.com/collections/statsbk/12.dtl> , Erişim tarihi: 19.06.2009)



Şekil 25. Sağkalım eğrisi

Kaplan-Meier tahmininde, sıfır anında sağkalım eğrisinin birikimli sağkalım olasılığı “1,00” yani %100 kabul edilir. (34) Sağkalım eğrisinin her yeni değerini bulmak için, ölüm (veya beklenen bir olguya erişilen) gerçekleşen belli bir zamandan önceki sağkalım olasılığı değerleri, bir önceki ölümün gerçekleştiği (veya beklenen bir olguya erişildiği) zamandaki sağkalım olasılığının tahmin değeri ile çarpılır. Bu çarpımdan dolayı Kaplan-Meier tahminine “Çarpım-Limit (Product-Limit) tahmini” denir. (13)

Daha anlaşılır olması için n bireyin gözlemlendiğini varsayalım. Bireylerin l tanesinde ölüm zamanı k farklı zamanda gözlenmiş olsun. $(t_1 < t_2 < \dots < t_k)$. m_i , t_i zamanında ölen bireylerin sayısı, $n-l$ birey sansürlü gözlem, n_i ise t_i zamanında ölüm riski bulunan bireylerin sayısı olsun; oran olarak ifade edilen sağkalım eğrisi Kaplan-Meier tahmini (4.6), (4.7), (4.8) ve (4.9) daki eşitliklerle hesaplanır:

$$S(t) = 1, \quad t < t_1, \quad (4.6)$$

$$S(t) = \prod_{j=1}^i \frac{n_i - m_i}{n_i}, \quad t_i \leq t < t_{i+1} \quad (i < k) \quad (4.7)$$

$$S(t) = 0, \quad t_k \leq t \text{ için, } m_k = n_k \text{ olduğunda}$$

$$(\text{yani kimse } t_k \text{ zamanında sağ kalmaz}) \quad (4.8)$$

$$S(t) = \prod_{j=1}^k \frac{n_j - m_j}{n_j}, \quad t_k \leq t \leq \text{en büyük sansürlü gözlem} \quad (4.9)$$

$m_k < n_k$ olduğunda, $S(t)$, “ $t > \text{en büyük sansürlü gözlem}$ ” için tanımsız olur.

Kaplan-Meier tahminine ait standart hata (SH) ise (4.10)’da verilmiştir:

$$SH(S(t)) = S(t) \sqrt{\sum_{j=1}^i \frac{m_j}{n_j(n_j - m_j)}} \quad t_i \leq t < t_{i+1} \quad (4.10)$$

5. YÖNTEM VE UYGULAMA

Bu tez çalışmada kullanılan veri seti (n=1499), İstanbul Üniversitesi Tıp Fakültesi Meme Cerrahi Kliniği'nde 13 yıllık izlem süresi (1993-2006 yılları arasında) sonunda elde edilen meme kanserli hastalara ait bir veri setidir.

Verilerin analizi için gerekli istatistikler EasyFit 5.0 (EasyFit 2008), GraphPad InStat 3.05 (GraphPad 2000), SPSS 17.0 (SPSS 2008) paket programları yardımıyla ve MS Excel'de oluşturulan formül yazılımları ile elde edilmiştir.

5.1. Tanımlayıcı İstatistikler

Hastaların izlenme süresi sona erdiğinde, 1499 hastanın 1406'sının (%93.8) hayatta kaldığı, 93'ünün (%6.2) öldüğü gözlenmiştir. Çalışmada yer alan bazı önemli değişkenlerin dağılımı Tablo 1.'deki gibidir:

Tablo 1. Çalışmada kullanılan veri setine ait bazı önemli değişkenlerin dağılımı

<u>Değişken</u>	<u>Hasta Sayısı</u>	<u>N%</u>	<u>Geçerli Yüzde</u>	<u>Ölen Hasta Sayısı</u>	<u>Sağ kalan Hasta Sayısı</u>
<u>Yaş</u>					
≤50	799	53.3	53.8	61	738
>50	686	45.8	46.2	32	654
Kayıp	14	0.9			
<u>Evre</u>					
Evre0	41	2.7	2.7	6	35
Evre1	465	31.0	31.0	22	443
Evre2A	625	41.7	41.7	35	590
Evre2B	368	24.5	24.5	30	338
<u>Tümör Çapı</u>					
T1a	253	16.9	17.2	21	232
T1b	101	6.7	6.9	5	96
T1c	332	22.1	22.6	9	323
T2	782	52.2	53.3	52	730
Kayıp	31	2.1			

Tablo1. (devam) Çalışmada kullanılan veri setine ait bazı önemli değişkenlerin dağılımı

<u>Değişken</u>	<u>Hasta Sayısı</u>	<u>Yüzde</u>	<u>Geçerli yüzde</u>	<u>Ölen Hasta Sayısı</u>	<u>Sağkalan Hasta Sayısı</u>
<u>Patolojik Çap</u>					
≤23	814	54.3	58.8	29	785
>23	571	38.1	41.2	51	520
Kayıp	114	7.6			
<u>Histolojik Seviye</u>					
Seviye1	37	2.5	2.8	0	37
Seviye2	651	43.5	50.0	43	608
Seviye3	614	41.0	47.2	41	573
Kayıp	197	13.1			
<u>Lenfovasküler İnvazyon</u>					
LVK (-) *	1043	69.6	69.6	72	971
LVK (+) *	456	30.4	30.4	21	435
<u>Cerrahi Tedavi</u>					
MKC *	620	41.4	41.5	36	584
MRM *	875	58.4	58.5	57	818
Kayıp	4	0.2			
<u>Aksiler Küraj</u>					
Hayır	101	6.7	7.7	2	99
Evet	1208	80.6	92.3	58	1150
Kayıp	190	12.7			

* LVK (-): Lenfovasküler invazyon (yayılım) yok, LVK (+): Lenfovasküler invazyon var, MKC: Meme koruyucu cerrahi, MRM: Modifiye Radikal mastektomi

Çalışmada kullanılan meme kanserli hastalara ait veri tabanında yer alan yaş değişkeninin ortalaması $\pm$ standart sapması, 50.99 ± 12.744 ve patolojik çap değişkeninin ortalaması $\pm$ standart sapması, 22.70 ± 10.447 olduğundan veriler Tablo 1.'de bu değerlere göre nitel veri şeklinde gösterilmiştir.

Tablo 1. doğrultusunda, meme kanserli hastaların %53.8'i 50 yaş altında iken, %46.2'si 50 yaş üstündedir. Hastaların %58.5'ine *modifiye radikal mastektomi* (MRM) uygulanırken, %41.5'ine *meme koruyucu cerrahi* (MKC) uygulanmıştır. Evre, tümör çapı, patolojik çap, histolojik seviye, lenfovasküler invazyon ve aksiler küraj gibi diğer bazı önemli değişkenlerin dağılımı Tablo 1.'de yer almaktadır.

5.2. Ters Normal Dağılım ve Sağkalım Analizi Sonuçları (Bulguları)

Ters Normal dağılım belli bir olguya erişilinceye (bir ürünün son kullanma tarihine kadar olan raf ömrü veya bireyin ölümü gibi) kadar geçen sürenin dağılımını için kullanıldığından bu tez çalışmasında, meme kanserli hastalara ait veri setinde ölümü gerçekleşmiş hastaların, tanı konduktan sonra ölümüne kadar geçen süreleri ele alınmıştır. Başka bir deyişle, ölen hastaların izlem sürelerinin dağılımı incelenmiştir.

Meme kanseri çalışmasında, tanı konduktan sonra ölen hastaların ölümüne kadar geçen süreler (ölen hastaların izlem süreleri) Tablo 2.'de verilmiştir (n=93);

Tablo 2. Tanı konduktan sonra ölen meme kanserli hastaların ölümüne kadar geçen süreler

Takip Süresi (Ay)					
27	24	108	72	59	23
33	24	48	62	76	48
46	48	39	48	19	57
15	74	60	94	20	61
59	60	84	60	140	84
34	36	24	80	14	28
40	48	12	37	49	30
43	72	24	32	112	61
45	48	60	56	19	35
48	60	48	21	50	30
48	27	36	48	28	20
24	45	60	89	26	40
35	48	36	32	21	52
51	36	24	14	40	
36	43	82	55	36	
30	102	48	15	30	

Öncelikle Tablo 2.'de yer alan verilerin, EasyFit 5.0 paket programı (EasyFit 2008) ve Ters Normal çarpık basıklık katsayıları yaklaşımı ile Ters Normal dağılım

gösterip göstermediği incelenmiştir. Bu analizler için test hipotezleri aşağıdaki gibidir:

H_0 : Tanı konduktan sonra ölen hastaların, öldükleri zamana kadar geçen süreleri (izlem süresi) ters normal dağılmaktadır.

H_1 : Tanı konduktan sonra ölen hastaların, öldükleri zamana kadar geçen süreleri (izlem süresi) ters normal dağılmamaktadır.

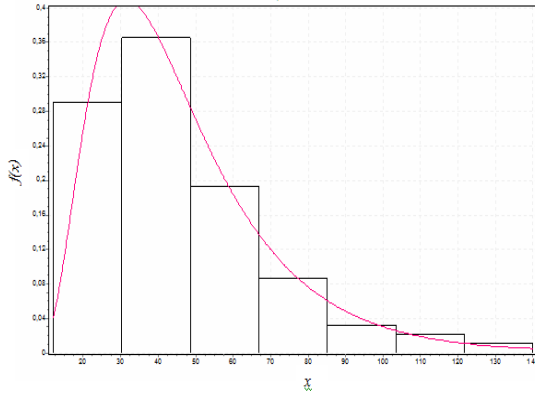
Tüm hesaplamalarda $\alpha = 0.05$ 'tir.

EasyFit 5.0 paket programı sürekli dağılımlar için Kolmogorov-Smirnov testi ile Anderson-Darling uyum iyiliği testlerini kullanmaktadır. Tablo 3.'ten görüleceği üzere, EasyFit 5.0 paket programı uyum iyiliği testleri sonucu meme kanserli hastalara ait bu veri setinde ölen hastalara ait izlem süreleri Ters Normal dağılıma uygundur.

Tablo 3. EasyFit paket programı Ters Normal dağılım uyum iyiliği test sonuçları

Kolmogorov-Smirnov					
Örnek Büyüklüğü	93				
Test İstatistiği	0.09424				
p değeri	0.35822				
α	0.2	0.1	0.05	0.02	0.01
Kritik değer	0.10947	0.12506	0.13891	0.15533	0.16666
Red-Kabul?	Kabul	Kabul	Kabul	Kabul	Kabul
Anderson-Darling					
Örnek Büyüklüğü	93				
Test İstatistiği	0.50012				
α	0.2	0.1	0.05	0.02	0.01
Kritik değer	1.3749	1.9286	2.5018	3.2892	3.9074
Red-Kabul?	Kabul	Kabul	Kabul	Kabul	Kabul

Ölen hastaların izlem süresi dağılımına ait olasılık yoğunluk fonksiyonu grafiği Şekil 26.'da verilmiştir. Bu grafiğe göre verilerin, Ters Normal dağılımın özelliği olan sağa çarpık olduğu gözlenmektedir. (Grafik EasyFit 5.0 paket programı ile elde edilmiştir.) (EasyFit 2008)



Şekil 26. Ölen meme kanserli hastaların izlem süreleri dağılımının olasılık yoğunluk fonksiyonu grafiği

Çarpıklık ve basıklık yaklaşımı ile yapılan testlerin hesaplamaları, gerekli veri tabanı ve tezde bahsi geçen formüller kullanılarak MS Excel’de oluşturulan bir yazılım ile yapılmıştır. Bu testlere ait sonuçlar aşağıdaki gibidir:

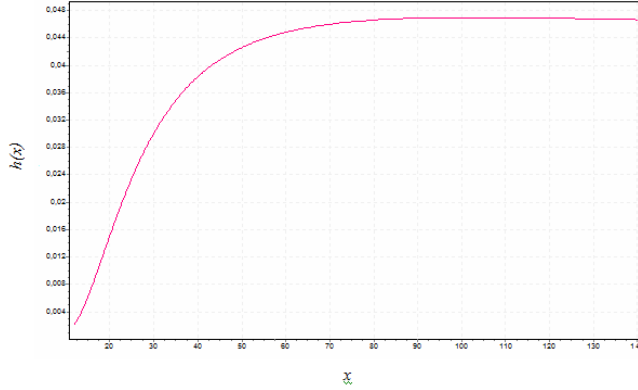
Dağılımın parametreleri, (3.14) ve (3.17) eşitlikleri kullanılarak $\hat{\mu} = 46.5054$ ve $\hat{\lambda} = 157.7974$ olarak tahmin edilir.

(3.33) numaralı formülden $\hat{\theta} = 3.3931$ olarak hesaplanır. (3.29) eşitliğinden çarpıklık için Jackknife test istatistiği $\tilde{d}_1 = -0.1849$ olarak bulunur. (3.31) ve (3.32) eşitliklerinden $m(\tilde{d}_1) = -0.0058$ ve $s(\tilde{d}_1) = 0.2864$ elde edilir. (3.35) ve (3.36) eşitlikleri kullanılarak $Z \approx -0.63$ ve bir $p = 0.528$ değeri hesaplanır, bunun sonucunda verilerin ters normal dağılım gösterdiği kabul edilmiş olur. ($p > 0.05$)

Benzer şekilde, basıklık katsayısı için, (3.40) eşitliğinden Jackknife test istatistiği $\tilde{T}_2 = 0.7513$ elde edilir. (3.43) eşitliği kullanılarak $s(\tilde{T}_2) = 0.2620$ hesaplanır. $m(\tilde{T}_2) = \ln 2 = 0.6931$ ortalama ile (3.44) eşitliği kullanılarak $Z \approx 0.222$ olarak bulunur. (3.36) ile $p = 0.825$ elde edilir. Bu durumda basıklık testi doğrultusunda da Ters Normal dağılım kabul edilmiş olur. ($p > 0.05$)

Ters Normal çarpıklık ve basıklık katsayıları yaklaşımı ile yapılan testlerde, verilerin Ters Normal dağılıma uygunluk gösterdiği saptanmıştır.

Ters Normal dağılıma ait, yaşam süresi verilerinin dağılımının teşhisi için incelenmesi gereken önemli bir fonksiyonda risk (hazard) fonksiyonudur. Risk (hazard) fonksiyonunun grafiği ise Şekil 27.'de verilmiştir. (Grafik EasyFit 5.0 paket programı ile elde edilmiştir.) (EasyFit 2008)



Şekil 27. Ölen meme kanserli hastaların izlem süreleri Ters Normal dağılımının risk (hazard) fonksiyonu grafiği

EasyFit 5.0 ve çarpıklık-basıklık yaklaşımı ile yapılan test sonuçlarına göre ilgilenilen verilerin Ters Normal dağıldığı kabul edilmiş olsa da, dağılıma ait risk (hazard) grafiği, Ters Normal dağılımın gerektirdiği gibi (Bakınız Şekil 20 $\mu = 1$ ve $\mu = 2$ için.) ilk zamanlarda yükselen ve bir maksimumda doruk yaparak hızla azalan bir yapı göstermemektedir.

Bu sonuçlar bize dağılımın teşhisi ve diğer dağılımlardan ayırt edilmesi için hazard fonksiyonun ne kadar önemli olduğunu göstermektedir. Tüm bu elde edilen sonuçlar, tanı konduktan sonra ölen meme kanserli hastalara ait yaşam süresi verilerinin Ters Normal dağılıma oldukça yakın olan Lognormal dağılım gösterdiği fikrini doğurmaktadır.

Bu kez verilerin Lognormal dağılım gösterip göstermediğini anlayabilmek üzere, ölen hastaların izlem süresi verilerinin logaritması alınarak normallik testi yapılmıştır. (Çünkü Lognormal dağılım, verilere uygulanan logaritmik dönüşümün, verilerin yaklaşık olarak simetrik dağılmasını gerektirir.) Dönüşüm sonucu verilerin normallik testi sonuçları Tablo 4.'teki gibidir. (Hesaplama GraphPad InStat paket programı versiyon 3.05 ile yapılmıştır.) (GraphPad 2000)

Tablo 4. Logaritması alınan verilerin Kolmogorov-Smirnov normallik testi

	Kolmogorov-Smirnov test istatistiği	p (çift yönlü)
Ln takip süresi	0.09	0.389

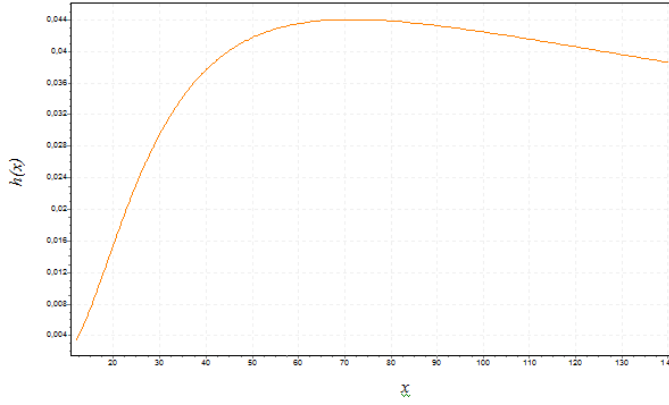
Burada, Kolmogorov-Smirnov testi, logaritmik dönüşüm yapılan verilerin normal dağılıma uygun olup olmadığını test etmek için kullanılmıştır. Elde edilen p değeri 0.05'ten büyük olduğu için normallik üzerine kurulan hipotez kabul edilmiş ve logaritması alınan verilerin normal dağılım gösterdiği, dolayısıyla veri setindeki ölen hastalara ait verilerin Lognormal dağıldığı kanıtlanmıştır.

Tablo 5.'te, Lognormal dağılım için yapılan EasyFit 5.0 uyum iyiliği testleri sonucunda da verilerin Lognormal dağılıma uygun olduğu görülmektedir.

Tablo 5. EasyFit paket programı Lognormal dağılım uyum iyiliği test sonuçları

Kolmogorov-Smirnov					
Örnek Büyüklüğü	93				
Test İstatistiği	0.09427				
p değeri	0.35787				
α	0.2	0.1	0.05	0.02	0.01
Kritik değer	0.10947	0.12506	0.13891	0.15533	0.16666
Red-Kabul?	Kabul	Kabul	Kabul	Kabul	Kabul
Anderson-Darling					
Örnek Büyüklüğü	93				
Test İstatistiği	0.39996				
α	0.2	0.1	0.05	0.02	0.01
Kritik değer	1.3749	1.9286	2.5018	3.2892	3.9074
Red-Kabul?	Kabul	Kabul	Kabul	Kabul	Kabul

Ayrıca Şekil 28., ölen hastaların izlem süresi verilerine ait Lognormal risk (hazard) fonksiyonu grafiği incelendiğinde Lognormal risk (hazard) fonksiyonun gerektirdiği gibi (Bakınız Şekil 16.) olduğu görülmektedir. (Grafik EasyFit 5.0 paket programı ile elde edilmiştir.) (EasyFit 2008)



Şekil 28. Ölen meme kanserli hastaların izlem sürelerinin Lognormal dağılımının risk (hazard) fonksiyonu grafiği

Sonuçlar doğrultusunda, ele alınan meme kanserli hastaların risk (hazard) fonksiyonunun Ters Normal dağılımdan çok Lognormal dağıldığına uygun olduğu söylenebilir. Bunun sonucunda da, meme kanserli hastalara ait verilerin tanı konduktan sonra ölene kadar geçen süreleri Ters Normal dağılım yerine daha çok Lognormal dağılmaktadır.

Lognormal dağılım gösteren bu verilerin ortalaması ve takiben anti logaritması alınarak hastalar için ortalama bir risk (hazard, ölüm) süresi elde edilir. Bu süre 41.059 aydır. Ayrıca bu sonuç, Sağkalım analizi Kaplan-Meier sağkalım olasılık tahmin sonuçları ile birleştirilmiştir.

Sağkalım analizi Kaplan-Meier tahmin sonuçları Tablo 6.'da verilmiştir. Tabloda Lognormal dağılımın ortalaması olan 41.059'uncu ayda birikimli sağkalım olasılığı yaklaşık 0.96'dır. (Kaplan-Meier sağkalım olasılık tahminleri SPSS 17.0 paket programı ile hesaplanmıştır.) (SPSS 2008)

Tablo 6. İzlem sürelerine göre hastaların sağkalım olasılıklarının bir kısmı

İzlem Süresi	Durum	Birikimli Sağkalım Olasılığı		Birikimli vaka (ölüm) sayısı
		Kaplan Meier Tahmini	Standart Hata	
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
40	Kayıp	0.964	0.005	44
40	Hayatta	.	.	44
40	Hayatta	.	.	44
40	Hayatta	.	.	44
40	Hayatta	.	.	44
40	Hayatta	.	.	44
.	"	.	.	.
.	"	.	.	.
41	Hayatta	.	.	44
41	Hayatta	.	.	44
41	Hayatta	.	.	44
.	"	.	.	.
.	"	.	.	.
42	Hayatta	.	.	44
42	Hayatta	.	.	44
42	Hayatta	.	.	44
42	Hayatta	.	.	44
42	Hayatta	.	.	44
.	"	.	.	.
.	"	.	.	.
43	Kayıp	.	.	45
43	Kayıp	0.962	0.005	46
.	.	.	.	.
.	.	.	.	.
46	Kayıp	0.959	0.006	49
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

Not: Tüm veri setine ait sağkalım süresi, 1,249 standart hata ile ortalama 153,107 aydır.

Ele alınan verilerin Ters Normal dağılım gösterdiği düşünüldüğünde meme kanserli hastaların risk altında kaldığı süre ortalama 46.505 aydır. Tablo 6.'dan 46 ay için birikimli sağkalım olasılığı değerine bakıldığında Lognormal dağılım ile elde edilen sonuca benzer olarak yaklaşık 0.96'dır.

6. TARTIŞMA VE ÖNERİLER

Yaşam ve insana ait veriler genellikle pozitif ölçümlerden oluştuğundan bu verilerin üstel dağılım gösterdiği varsayılır. Yaşam süresi de insan yaşamına ait veriler arasındadır. Üstel dağılım, Gama dağılımı, Weibull dağılımı, Lognormal dağılım ve Ters Normal dağılım üstel dağılım ailesine ait başlıca yaşam süresi dağılımlardır.

Tüm istatistiksel dağılımlarda olduğu gibi yaşam süresi dağılımlarını tanımlamada da dağılımların bazı özelliklerinden faydalanılır. Dağılımın olasılık yoğunluk fonksiyonu, birikimli dağılım fonksiyonu, sağkalım fonksiyonu ve özellikle risk (hazard) fonksiyonu bu tanımlamalar arasındadır.

Diğer yaşam süresi dağılımları üstel dağılım üzerinden türemiş olduğu için üstel dağılım önemli bir yaşam süresi dağılımıdır. Üstel dağılımın önemli bir başka özelliği de risk (hazard) oranının sabit olmasıdır ve bu durum matematiksel uygulamada kolaylık sağlar. Ancak uygulamalarda sabit risk (hazard) oranına çok sık rastlanılmadığından farklı dağılımların türetilmesine gerek duyulmuştur.

Üstel dağılımın genelleştirilmiş bir şekli olan Gama dağılımı, monoton değişen risk (hazard) oranına uygunluk gösteren daha esnek bir dağılımdır. Sabit risk (hazard) oranı nedeni ile zamana karşı dayanma (yaşam süresi) modeli olarak yetersiz kalan üstel dağılıma seçenek olarak risk oranının zamana bağlı değiştiği ve Gama dağılımına göre de daha genel ve daha esnek bir dağılım olan Weibull dağılımı ortaya çıkmıştır. Weibull dağılımında dikkat edilmesi gereken nokta risk oranı değişiminin monoton olduğu durumlarda dağılımın uygunluk göstermesidir. (Monoton azalan, monoton artan veya sabit). Weibull dağılımı, endüstride bir ürünün güvenilirliği için oldukça uygun bir model sağlarken, insan ömrü için her zaman geçerli olamamaktadır. (özellikle insan ölümcül bir hastalığa yakalandığında risk oranı (ölüm hızı) monotonluk göstermez).

Lognormal dağılım, olayların genellikle dağılımın sol kuyruğunda yoğunlaştığı, verilere uygulanan logaritmik dönüşüm sonucu verilerin yaklaşık olarak simetrik dağılmasını sağlayan bir yaşam süresi dağılımıdır. Bahsedilen diğer üstel dağılımlar

arasında Ters Normal dağılıma en çok benzerlik gösteren dağılımdır. Her ikisi de, sağa çarpık ve pozitif verinin analizinde kullanılırlar ve monoton olmayan risk (hazard) oranı gösterirler. Ancak Ters Normal dağılım, fonksiyonun doruk noktasını ve uç değerlerini daha iyi tanımlayan bir dağılımdır. (6) Ayrıca, çok yönlü ve esnek bir model elde etmeyi sağlar. Ters Normal dağılımın Lognormal dağılıma göre bir üstünlüğü de x sonsuza giderken $(x \rightarrow \infty)$ Lognormal dağılımın risk (hazard) fonksiyonunun 0'a yaklaşmasıdır ki bu da fiziksel olgular için kabul edilebilir olmaz. (insan ömrü gibi, çünkü insan ömründe risk (hazard) hiçbir zaman 0 olmaz, her insan ölümlüdür) Oysa x sonsuza giderken $(x \rightarrow \infty)$ Ters Normal dağılımın hazard

fonksiyonu 0'a değil $\frac{\lambda}{2\mu^2}$ değerine yakınsar. (6)

Sağkalım Analizi, birçok konuda risk faktörlerini, prognostik öğeleri, tedavi başarılarını vb ortaya koyabilmesi bakımından önem taşır ve tıbbın çeşitli alanlarında (bireyin sağ kalma süresi, protezin dayanma süresi gibi) kullanılır.

Sağkalım, belli bir başlangıç noktasından (ameliyat sonrası, hastalığın teşhisinin konulduğu tarih) itibaren, bir izleme süresi içindeki olgunun, araştırılan özel bir konuma (ölüm gibi) erişmesine kadar geçen sürenin tanımı için kullanılabilir. Sağkalım olasılığını tahmin etmek için kullanılabilecek yöntemler geliştirilmiştir. Parametrik olmayan bir yöntem olan “Kaplan-Meier (Çarpım-Limit)” yöntemi yaygın olarak kullanılan bir yöntemdir. Kaplan-Meier tahmin yönteminin önemli özelliklerinden biri, sağkalım olasılığı hesaplanırken ölümün gerçekleştiği her noktada sağ kalan tüm bireyler için hesaplama yapmasıdır.

Lloyd D. Fisher ve Gerald van Belle “Biostatistics, A Methodology for the Health Science” isimli kitaplarında Sağkalım analizine bu ismin verilme nedenini şöyle vurgulamaktadırlar: “Ölümden bahsetmek üzücü ancak yaşamdan (sağkalımdan) söz etmek daha memnun edicidir”. (13) Bu tez çalışmasında konuya, ölümcül bir hastalığa yakalanmış hasta veya hasta yakını tarafından yaklaşılarak, hastanın hastalığa yakalandıktan sonra ölene kadar geçen sürenin dağılımı ele alınmış ve ortalama ne kadar ömrünün kaldığı belirlenmeye çalışılmıştır.

Ters Normal dağılıma ait olasılık yoğunluk fonksiyonu grafiği dikkate alındığında ölüm olayı ya da beklenen bir olguya kadar geçen süre için fonksiyon değeri ilk zamanda hızlı bir artış gösterirken bu değer, ilerleyen ilk zamanlarda göreceli olarak hızla düşmektedir ve daha sonra (grafik'in kuyruk kısmında) bu düşüş yavaşlamaktadır. (Bakınız Şekil 17.) Meme kanserli hastalara ait veri setinin kullanıldığı bu çalışmada, yaşam süresi verileri dağılımının (bir başka deyişle tanı konduktan sonra ölen hastalara ait izlem süresi, dolayısıyla tanıdan sonra yaşam süresi verilerinin) olasılık yoğunluk fonksiyonuna ait grafikten (sağa çarpık) yola çıkarak verilerin Ters Normal dağıldığı varsayımında bulunulmuştur.

Yapılan analizlerin bulguları incelendiğinde, meme kanserli hastalara ait veri setinde ölen hastaların izlem sürelerinin (tanı konduktan sonra yaşam sürelerinin) Ters Normal dağılıma uygun olduğu görülmüştür. Yaşam süresi dağılımının teşhisinde önemli yer tutan risk (hazard) fonksiyonu bu veri seti için incelenmiştir. Verilerin risk (hazard) fonksiyonunu grafiğinin Ters Normal dağılımın risk (hazard) fonksiyonu grafiğine bezediği görülmüştür. Bunun sonucunda verilerin hangi dağılıma ait olabileceği sorusu yeniden gündeme gelmiştir.

Meme kanserli hastalara tanı konduktan sonra ölene kadar geçen sürenin, olasılık yoğunluk fonksiyonu bakımından Ters Normal dağılıma en yakın ve Ters Normal dağılım gibi ilk zamanlarda riskte önemli bir artış gösteren bir dağılım olan Lognormal dağılıma uygunluk gösterebileceği varsayılarak gerekli testler yapılmıştır. Yapılan testler sonucunda verilerin Lognormal dağılıma uygunluk gösterdiği kanıtlanmıştır. Ayrıca Şekil 28.'de verildiği üzere, dağılıma ait risk (hazard) fonksiyonu grafiği incelendiğinde Lognormal'e uygun olduğu görülmüştür.

Elde edilen bilgilerin ışığı altında, herhangi bir dağılıma uygun olup olmadığı analiz edilen bir veri setinin tek özellik açısından incelenmesinin yeterli olmadığı, verilere ait dağılımın çok yönlü olarak irdelenmesi sonucu ortaya çıkmıştır.

Sonuç olarak, Lognormal dağılıma uygun olan verilerin ortalaması hesaplanmış ve meme kanseri tanısı konmuş bir kişinin ölümüne kadar geçen (bir başka deyişle ölüm riski altında bulunan) sürenin ortalama 41.059 ay olduğu hesaplanmıştır. Kaplan-Meier yöntemi ile ise, 41.059 ay için birikimli sağkalım olasılığı yaklaşık

0.96 olarak bulunmuştur. 41.059 ay (yaklaşık üç buçuk yıl) sonunda meme kanserli bir kişinin sağkalım olasılığının çok yüksek olduğu görülmüştür.

Lognormal için hesaplanan risk süresi, aynı veri seti için Ters Normal dağılım varsayılarak hesaplandığında bu süre ortalama 46.505 ay bulunmuştur. 46.505 ay için Kaplan-Meier sağkalım tahmini yine yaklaşık 0.96 olmaktadır. Bu durumda, riskli sürede (yaklaşık 41 - 46 ay) dahi sağkalım olasılığının yüksek olduğu görülmektedir.

Ters Normal dağılım için, verilere ait risk (hazard) fonksiyonu grafiğinde doruk noktasına erişilmesine kadar geçen süre uzun olmaktadır (46.505 ay yaklaşık dört yıl) ve doruktan sonra da hızlı bir düşüş görülmemektedir.

Meme kanseri tanısı konan ve sadece ölüm olayı gerçekleşmiş bu hastalarda, tanı konduktan sonra ölüme kadar geçen sürenin risk (hazard) fonksiyonunun Ters Normal Dağılıma uygunluk göstermemesinin nedeni, hastalığın tedavisinde oldukça yol kat edilmiş olması şeklinde açıklanabilir. Ancak, hızla ilerleyen (hastalığın teşhisinden sonraki ilk zamanlarda ölüm riskinin yüksek olduğu) kanser türleri için (akciğer, karaciğer, serviks kanseri gibi) Ters Normal dağılımın daha uygun olacağını düşünmekteyiz. Buna rağmen, ülkemizde adı geçen kanser türlerinde, Ters Normal dağılımı iyi derecede ortaya koyabilecek organize veri toplama çalışmaları ve doğru veri tabanlarına erişim yetersizdir. Bu bakımdan, bu çalışmada kullanılan, uzun yıllar izlenen (on üç yıl) ve 1499 kişiden oluşan geniş veri seti (tabanı) oldukça değerlidir.

Ölümcul bir hastalığa yakalandığında, risk altında kalınan ortalama sürenin belirlenebilmesi ve ele alınan konuda faydalı çıkarımlar yapılabilmesi için bu gibi konularda organize veri toplama ve sağlıklı veri tabanları oluşturmaya yönelik teşviklerde bulunulabilir.

İlk dönemlerde ölüm riski yüksek olan kanser türlerine (ya da başka bir hastalığa) ait doğru (organize) veriler elde edildiğinde, “Ne kadar ömrüm kaldı?” sorusuna yanıt olarak hastalığa yakalanan kişilerin ölüm sürelerinin dağılımı için mutlaka Ters Normal dağılımın irdelenmesi önerilmektedir.

KAYNAKLAR

1. Bacanlı S, Demirhan YP. (2006). Ters Normal dağılımda ortalamanın testi. 5. İstatistik Günleri Sempozyumu Bildiri Özetleri Kitabı, Ankara Üniversitesi. Ankara. s. 125.
2. Bacanlı S, Demirhan YP. (2008). A group sequential test for the Inverse Gaussian mean. *Statistical Papers*, 49:377-386.
3. Bekiroğlu N. (1999). İki veya ikiden fazla sağkalım olasılıkları karşılaştırılmasının bir uygulamayla irdelenmesi. 4. Ulusal Biyoistatistik Kongresi Bildirileri Kitabı, Ankara Üniversitesi. Ankara. s. 23-34.
4. Bekiroğlu N. (1998) . Açıklamalı Biyoistatistik Terimleri Sözlüğü. 1. baskı. Nobel Tıp Kitabevleri, İstanbul.
5. Blischke WR, Murthy DNP. (2000). Reliability: Modelling, Prediction and Optimization, Wiley-IEEE, New York, p. 135-162
6. Chhikara RS, Folks JL. (1977). The Inverse Gaussian distribution as a lifetime model. *Technometrics*, 19(4):461-468.
7. Chhikara RS, Folks JL. (1998). The Inverse Gaussian Distribution, Theory, Methodology and Applications. In: *The Statistics: Textbooks and Monographs*. Ed: Spall JC. Marcel Dekker, New York, p.1-15.
8. Evans M, Hastings N, Peacock B. (2000). Statistical Distributions. Eds: Cressie NAC, Fisher NI, Johnstone IM, Kadane JB, Scott DW, Silverman BW, Smith AFM, Teugels JL, Barnett V, Bradley RA, Hunter JS, Kendall DG., 3rd ed, John Wiley&Sons Inc, New York, p. 12-203.
9. Davis AS. (1980). Use of the likelihood ratio test on the Inverse Gaussian distribution. *The American Statistician*, 34(2):108-110.
10. Dawson B, Trap RG. (2001). Basic&Clinical Biostatistics. 3rd ed, McGraw Hill, US. p. 211-225.
11. Dişçi R. (2008). Temel ve Klinik Biyoistatistik. İstanbul Medikal Yayıncılık Ltd. Şti., İstanbul, s.221-235.

12. Elitok Ö. (2006). Weibull Dağılımı ve Uygulamaları. K.Ü. Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Kırıkkale, (Danışman: Doç. Dr. H Erbay).
13. Fisher LD, Belle Gv. (1993). Biostatistics, A Methodology for the Health Sciences. John Wiley&Sons Inc, New York, p. 786-803.
14. Folks JL, Chhikara RS. (1978). The Inverse Gaussian distribution and its statistical application-a review. *Journal of the Royal Statistical Society*, B40(3):263-289.
15. GraphPad Software Inc. (2000). GraphPad InStat for Windows. Versiyon 3.05, California.
16. Günak N. (2007). İleri Teknik Analiz Uygulamaları ve Bu Uygulamaların İMKB’de Test Edilmesi. İ.Ü. Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, İstanbul, (Danışman: Doç. Dr. A Köse).
17. Johnson NL, Kotz S. (1970). Continuous Univariate Distributions-1. Houghton Mifflin Company, Boston, p. 137-271.
18. Kalbfleisch JD, Prentice RL. (2002). Failure Time Models. In: *The Statistical Analysis of Failure Time Data*. Eds: Balding DJ, Bloomfield P, Cressie NAC, Fisher NI, Johnstone IM, Kadane JB, Ryan LM, Scott DW, Smith AFM, Teugels JL, Barnett V, Hunter JS, Kendall DG., 2nd ed, John Wiley&Sons Inc, New York, p. 31.39.
19. Leiva V, Hernandez H, Sanhueza A. (2008). An R package for a general class of Inverse Gaussian distributions. *Journal of Statistical Software*, 26(4):1-21.
20. Lüküslü D. (2005). Modelleme Sürecinde Weibull Dağılımının Kullanılması. M.Ü. Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul, (Danışman: Prof. Dr. M Tez)
21. Marshall AW, Olkin I. (2007). Life Distributions: Structure of Nonparametric, Semiparametric and Parametric Families. Springer, New York, p. 291-361, 451-471.
22. MathWave Technologies. (2008). EasyFit Professional for Windows. Versiyon 5.0, Waltham.
23. Mudholkar GS, Srivastava DK, Freimer M. (1995). The exponentiated Weibull family: a reanalysis of the bus-motor-failure data. *Technometrics*, 37(4):436-445.

24. Mudholkar GS, Hutson AD. (1996). The exponentiated Weibull family: some properties and a flood data application. *Commun. Statist. Theory Meth*, 25(12):3059-3083.
25. Mudholkar GS, Srivastava DK, Kollia GD. (1996). A generalization of the Weibull distribution with application to the analysis of survival data. *Journal of the American Statistical Association*, 91(436):1575-1583.
26. Mudholkar GS, Natarajan R. (1999). Approximations for the Inverse Gaussian probabilities and percentiles. *Commun Statist, Simula*, 28(4):1051-1071.
27. Mudholkar GS, Natarajan R, Chaubey YP. (2001). A goodness-of-fit tests for the Inverse Gaussian distribution using its independence characterization. *The Indian Journal of Statistics*, B63(3):362-374.
28. Mudholkar GS, Tian L. (2002). An entropy characterization of the Inverse Gaussian distribution and related goodness-of-fit test. *Journal of Statistical Planning and Inference*, 102:211-221.
29. Mudholkar GS, Natarajan R. (2002). The Inverse Gaussian models: analogues of symmetry, skewness and kurtosis. *Annals of the Institute of Statistical Mathematics*, 54(1):138-154.
30. Natarajan R, Mudholkar GS. (2004). Moment-based goodness-of-fit tests for the Inverse Gaussian distribution. *Technometrics*, 46(3):339-347.
31. Shadri V. (1993). The Inverse Gaussian Distribution, A Case Study in Exponential Families. Oxford University Press Inc, New York.
32. Seshadri V. (1999). The Inverse Gaussian Distribution, Statistical Theory and Applications. In: *Lecture Notes in Statistics*. Eds: Bickel P, Diggle P, Fienberg S, Krickeberg K, Olkin I, Wermuth N, Zeger S. Springer, New York.
33. SPSS Inc. (2008). SPSS for Windows. Version 17.00, Chicago.
34. Şenocak M. (1998). Biyoistatistik. 2. baskı, Cerrahpaşa Tıp Fakültesi Yayını, İstanbul, p. 187-219

35. Türkan AH. (2007). Güvenilirlik Analizinde Kullanılan İstatistiksel Dağılım Modelleri. Ç.Ü. Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Adana, (Danışman: Prof. Dr. H Erol).
36. Tweedie MC. (1957). Statistical properties of Inverse Gaussian distributions. *The Annals of Mathematical Statistics*, 28(2):362-377.
37. Varan SE. (2007). Sistem Güvenirliği Analizi ve Bir Uygulama. Y.T.Ü. Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul, (Danışman: Yrd. Doç. Dr. A Evren).
38. Whitmore Ga. (1979). An Inverse Gaussian model for labour turnover. *Journal of the Royal Statistical Society*, A142(4):468-478.

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı	Pınar	Soyadı	Günel
Doğum Yeri	İstanbul	Doğum Tarihi	13.10.1982
Uyruğu	T.C.	TC Kimlik No	31522634916
E-mail	gunelpinar@yahoo.com	Tel	0505 670 82 43

Eğitim Düzeyi

	Mezun Olduğu Kurumun Adı	Mezuniyet Yılı
Doktora/Uzmanlık	-	
Yüksek Lisans	-	
Lisans	Yıldız Teknik Üniversitesi	2004
Lise	Kabataş Erkek Lisesi	2000

İş Deneyimi (Sondan geçmişe doğru sıralayın)

	Görevi	Kurum	Süre (Yıl - Yıl)
1.	Proje Yöneticisi	Promart Yönetim Danışmanlık ve Eğitim Hizmetleri Ltd. Şti.	2005-2008
2.	Anket Editörlüğü ve Kodlama	GFK Pazarlama Araştırmaları	2005
3.	Customer Developer Assistance	Delta Socks	2004

Yabancı Dilleri	Okuduğunu Anlama*	Konuşma*	Yazma*
İngilizce	Çok iyi	Orta	İyi
Fransızca	Zayıf	Zayıf	Zayıf

* Çok iyi, iyi, orta, zayıf olarak değerlendirin

Yabancı Dil Sınav Notu #								
KPDS	ÜDS	IELTS	TOEFL IBT	TOEFL PBT	TOEFL CBT	FCE	CAE	CPE
68	71.25							

	Sayısal	Eşit Ağırlık	Sözel
ALES Puanı	88.629	86.445	82.994
(Diğer) Puanı			

Bilgisayar Bilgisi

Program	Kullanma becerisi
MS Office	Çok iyi
NCSS, SPSS, SAS	İyi
Turbo C programlama	Zayıf

*Çok iyi, iyi, orta, zayıf olarak değerlendirin

Uluslararası ve Ulusal Yayınları/Bildirileri/Sertifika/Ödülleri/Diğer

1- P.Günel, N.Bekiroğlu, F.Yılmaz, Ş.Aktaş, R.Konyalıoğlu. “Rasch Analysis: Validity, Reliability of Turkish Version of The Questionnaire ECOS-16 (Assesment of Health Related Quality Of Life in Osteoporosis) in Post-menopausal Women With Osteoporosis”, (Oral presentation)

5th Conference of the Eastern Mediterranean Region of the International Biometric Society, 10-14 May 2009, İstanbul, Türkiye.

2- Nural Bekiroğlu, Pınar Günel, Burak Aksu. Türkiye Klinikleri, Deneye Dayalı Çalışmalarda Bazı İstatistiksel Kavramlar

Türkiye Klinikleri Tıp Bilimleri Dergisi., (Makale kabul edildi.)

3- Filiz Ersoz, Nural Bekiroglu, Safak Aktas, Pinar Günel, Yeliz Sevimli, Aytekin Yamac. “Clustering Cities in Turkey with using Cluster and Discriminant Analysis for Some Demographic, Educational and Health Indicators”, (Oral presentation).

HDM-2008 International Conference on Multivariate Statistical Modelling & High Dimensional Data Mining, 19-23 June 2008, Erciyes Üniversitesi, Kayseri, TURKEY.

4- Pınar Günel, Şafak Aktaş, Yeliz Sevimli, Rana Konyalıoğlu, Nural Bekiroğlu. Ters Normal Dağılım- Sağkalım Analizi, (Sözlü sunum).

XI. Ulusal Biyoistatistik Kongresi, 27-30 Mayıs 2008, İnönü Üniversitesi, Malatya, TÜRKİYE