6.3. Normal Dağılımdan Örneklem

Merkezi limit teoremine göre bazı koşullar altında, kitle dağılımı ne olursa olsun  örneklem hacmi yeterince büyükse, örneklem ortalaması dağılımda normal dağılıma yaklaşmaktadır. Bu kısımda,   dağılımından alınan  örnekleminin bazı özellikleri incelenecektir. Her bir  için  olup olasılık yoğunluk fonksiyonu,



şeklindedir.

**Teorem 6.3.1**  ler  dağılımlı bağımsız rasgele değişkenler olsun. Bu durumda,

 ve 

olmak üzere,

a)  b)  ile  bağımsız c) 

dır.

*İspat:* a)Bağımsız normal dağılıma sahip rasgele değişkenlerin toplamının da normal olduğunu biliyoruz. O halde,  nin beklenen değeri ile varyansını hesaplamak yeterlidir.  olduğundan  dir.

b) Bu ifadenin değişik kaynaklarda değişik ispatları vardır (Öztürk ve digerleri (2006) teoremin ispatını beş farklı yoldan vermiştir).  ve  iki rasgele değişken olmak üzere, ortak olasılık yoğunluk fonksiyonu marjinal olasılık yoğunluk fonksiyonlarının çarpımı olarak yazılabilirse,  ve  bağımsızdır. Örneklem ortalamasının tanımından



yazılır. Ayrıca,



eşitliğinden,  şeklinde yazılabilir. Yani,  sadece  lerin bir fonksiyonudur. Böylece,  ile  nin bağımsız olduğunu göstermek için,

 ile 

nin bağımsız olduğunu göstermek yeterlidir.  rasgele değişkenlerinin ortak olasılık yoğunluk fonksiyonu  ve  için,



şeklindedir. Şimdi,



dönüşümlerini tanımlayalım. Böylece,  ile  nin bağımsız olduğunu göstermek için  ile  rasgele değişkenlerinin bağımsız olduğunu göstermek yeterli olacaktır. Buradan



olduğundan ters dönüşümler



şeklindedir. Bu dönüşümlere ait Jacobien matrisi ve determinantı



olup  rasgele değişkenlerinin ortak olasılık yoğunluk fonksiyonu  için,





şeklindedir. Dolayısı ile,  ile  rasgele değişkenleri bağımsızdır. Yani,



rasgele değişkenleri bağımsızdır. Buradan da,



rasgele değişkenlerinin bağımsızlığı elde edilir. Dolayısı ile,



rasgele değişkenleri bağımsız olup,



rasgele değişkenlerinin bağımsız olduğu elde edilmiş olur.

c) Teoremin bu kısmının da değişik şekilde ispatlanabilir. Örneğin, çok değişkenli normal dağılımın özelliklerinden teoremin ispatı yapılabilir. Ayrıntılı bilgi için herhangi bir lineer model kitabında karesel formların dağılımına bakılabilir. Bunun için tümevarım tekniğini kullanalım. İşlemlerin kolay yürütülebilmesi için,  alalım ve iddianın  için doğru olduğunu gösterelim.  ise ,



olarak yazılabilir. Buradan da,



elde edilir. Yani, iddia  için doğrudur. Şimdi, iddia  için doğru olsun. Yani,  olsun ve iddianın  için doğru olduğunu gösterelim. Bunun için,



eşitliği (Bkz. Problem (6.5.1))  için



olarak yazılır. Varsayımdan  olup  ile  bağımsızdır. Ayrıca,



dir. Teoremin (b) kısmından  ile  bağımsız olup

 ile 

rasgele değişkenleri de bağımsızdır. Bağımsız ki-kare dağılımına sahip rasgele değişkenlerin toplamı yine *ki-kare* dağılımına sahip olduğundan



elde edilir. Böylece, her   için  dir

Bu teoremin bir sonucu olarak, istatistikte ve diğer alanlarda çok kullanılan ve *örneklem dağılımları* olarak bilinen ,,  ve  dağılımlarını inceleyelim. Aslında, ,  ve  dağılımları da normal dağılımla ilişkilidir. Hemen hemen bütün istatistiki sonuç çıkarımlar bu dört dağılıma dayanmaktadır. Bu dağılımlar kullanılarak, kitlenin parametreleri hakkında hipotez testleri yapılabilmekte, yine parametreler için güven aralıkları yazılabilmektedir. Bunların her birinde de verilerin normalliği ön plana çıkmaktadır. Normal olmayan veriler için istatistiki sonuç çıkarımlar ancak MLT nin geçerli olduğu durumlarda yapılabilir.

 bağımsız  dağılımlı rasgele değişkenlerin bir dizisi ise,  olduğunu biliyoruz. Ancak  bilinmediği zaman,  yerine tahmin edicisi  kullanılır. Böylece, istatistiki sonuç çıkarım için



istatistiğinin dağılımına ihtiyaç duyulur. Teorem (6.3.1c) ye göre,  olduğundan  olup  iken  dır. Dolayısı ile, Örnek (6.2.2) ye göre  iken  dir. Buradan da, Slutsky Teoremine göre  iken,  olur.

*dağılımı:*

 ile  bağımsız ve ,  olsun.  rasgele değişkeninin olasılık yoğunluk fonksiyonu (Bölüm 3),



dir. Bu olasılık yoğunluk fonksiyonuna sahip  rasgele değişkeni *serbestlik derecesi  olan dağılımına sahiptir* denir ve  ile gösterilir. Dağılımın oasılık yoğunluk fonksiyonunun grafiği Şekil (6.3.1) de verilmiştir. Grafikten de görüldüğü gibi, dağılımın şekli normal dağılıma benzer ve standart normal dağılımda olduğu gibi fonksiyon sıfır noktasına göre simetriktir. Dağılımın beklenen değer ve varyansı,

 için  ve  için 

dir.

 beklenen değeri , varyansı  olan normal dağılımdan bir örneklem olmak üzere,  istatistiğinin dağılımını bulalım. Kolayca görüleceği gibi  istatistiği,



şeklinde yazılabilir. Buradan,  ile  bağımsız ve dağılımları da

 ve 

dir (Teorem (6.3.1)). Buradan,  olmak üzere  olduğu görülür.

6

*Şekil: 6.3.1  için  dağılımının olasılık yoğunluk fonksiyonu*

Bu dağılımın bazı kullanım alanlarından (güven aralıkları, hipotez testleri gibi) dokuz ve onuncu bölümlerde bahsedilecektir.

*dağılımı:*

Uygulamada çok kullanılan dağılımlardan biri de  dağılımıdır. Bağımsız  ve  rasgele değişkenleri  ve  şeklinde ki-kare dağılımına sahip olsunlar. Buradan, üçüncü bölümde bahsedilen bilgiler kullanılarak  rasgele değişkeninin olasılık yoğunluk fonksiyonunun,



olduğu gösterilebilir. Olasılık yoğunluk fonksiyonu bu şekilde olan  rasgele değişkenine *serbestlik dereceleri  ve  olan  dağılımına sahiptir* denir ve  ile gösterilir. Fonksiyonun grafiği şekil (6.3.2) de verilmiştir.

6

*Şekil 6.3.2  dağılımının  ve  için olasılık yoğunluk fonksiyonu*

 olsun.  dağılımının beklenen değeri, ki-kare dağılımının beklenen değerinden bulunabilir.  ise bağımsız  ve  rasgele değişkenleri için  dir. Buradan,  ve  bağımsız olduğundan,  nin beklenen değeri,



dir(Bkz. Problem (5.5.7)).

 ve  şeklinde birbirinden bağımsız iki farklı örneklem olsun.  şeklinde varyansların oranını tahmin etmek için  örneklem varyanslarının oranını gözönüne alalım. Buradan,



nin dağılımı için,

 ve 

olduğınu biliyoruz. Ayrıca,  ile  bağımsız olduğundan



elde edilir. Buna göre  olduğu varsayımı altında,  olup bu istatistiğin değeri varyansların aynı olduğunu sınamak için kullanılır. Bu dağılımın da bazı kullanım alanlarından dokuz ve onuncu bölümlerde biraz bahsedilecektir.

 ve  dağılımları literatürde örneklem dağılımları olarak bilinir. Aynı normal dağılımda olduğu gibi, bu dağılımlar için de tablolar düzenlenmiştir. Verilerin normallik varsayımı bazen sağlanmayabilir. Böyle durumlarda, bazen merkezi limit teoremi bazen de dönüşümler ile normallik sağlanır.

**Örnek 6.3.1** a)  ise  rasgele değişkeninin olasılık yoğunluk fonksiyonunu bulalım.  rasgele değişkeninin olasılık yoğunluk fonksiyonu,



olup  nin olasılık yoğunluk fonksiyonu  dir. Burada,  denirse  olup  nin olasılık yoğunluk fonksiyonu,



şeklinde bulunur. Bu eşitlik biraz düzenlendiğinde,



olmak üzere  nin olasılık yoğunluk fonksiyonu,



şeklinde yazılabilir. Ayrıca  değeri yerine konulduğunda,



olup  rasgele değişkeninin olasılık yoğunluk fonksiyonu,



olarak elde edilir. Yani,  ise  dir.

b)  ise  dir.  ise dağılımın olasılık yoğunluk fonksiyonu,



dir.  denirse  ve  olup  nin olasılık yoğunluk fonksiyonunun  şeklinde olduğunu biliyoruz ( in olasılık yoğunluk fonksiyonu simetrik olduğundan  ile çarpılır).  olduğundan  nin olasılık yoğunluk fonksiyonu  için,





şeklinde bulunur. Bu da serbestlik dereceleri  ve  olan  dağılımının olasılık yoğunluk fonksiyonudur. Yani,  ise  dir.

c)  ise  olduğunu gösterelim.  nin değer kümesi  dir. Ters dönüşüm ve türevi

 ve 

olup  nin olasılık yoğunluk fonksiyonu için,





şeklinde yazılır. Bu da  dağılımının olasılık yoğunluk fonksiyonudur

6.4. Sıra İstatistikleri

İstatistiki veriler incelenirken genellikle dağılım hakkında herhangi bir bilgi yoktur. Veri analizine başlamadan önce dağılım hakkında görsel bazı bilgilere başvurulur. Bunlar genellikle *histogram*, *Box-Cox çiziti* ve *normal olasılık grafiği* dir. Bu grafiklerin oluşturulmasında verilerin sıralanmış halinden yararlanılır. Ayrıca, *mode*, *medyan*, *yüzdelikler*, *çeyreklikler* gibi değerler hesaplanırken de verilerin sıralanmış hali kullanılır.

 olasılık veya olasılık yoğunluk fonksiyonu  olan kitleden bir örneklem olsun. Bu rasgele değişkenler aynı  olasılık uzayı üzerinde tanımlıdır. Her  için  oluyorsa  rasgele değişkeni  rasgele değişkeninden küçük ya da eşittir denir ve  gösterimi kullanılır. Buna göre,

, ,…, 

olmak üzere, rasgele değişkenlerin  şeklinde bir sıralanmasından söz edilebilir. Buradaki,  lerin her biri örneklemin bir fonksiyonu olup birer tahmin edicidir. Bu istatistiklere *sıra istatistikleri* denir. Bu istatistikler için bazen  gösterimi de kullanılır.

 örneklemi için,  sıra istatistikleri göz önüne alındığında,





gibi bazı istatistikler de sıra istatistikleri ile tanımlanır.

**Örnek 6.4.1** Bir istatistik dersinden sınava giren öğrencilerin aldığı notlar aşağıdadır. Görüldüğü gibi dağılım hakkında herhangi bir bilgi yoktur.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 66 | 71 | 67 | 69 | 75 | 66 | 64 | 70 | 62 | 83 |
| 70 | 79 | 74 | 74 | 79 | 94 | 76 | 69 | 88 | 72 |
| 84 | 76 | 63 | 70 | 77 | 80 | 77 | 72 | 78 | 73 |
| 75 | 78 | 90 | 76 | 62 | 78 | 78 | 72 | 77 | 72 |
| 72 | 59 | 73 | 75 | 76 | 80 | 56 | 67 | 69 | 80 |

Örneklem ortalaması ile örneklem varyansının değerleri

olarak hesaplanmıştır. Ayrıca, veriler küçükten büyüğe sıralandığında,

 ,  ,  ,  , 

değerleri gözlenmiştir. Bunlarla birlikte, bu verilere ait bazı özet bilgiler de

 , 

olarak elde edilmiştir. Verilerin küçükten büyüğe sıralanmış hali de aşağıdadır.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 56 | 58 | 60 | 62 | 62 | 62 | 64 | 66 | 66 | 66 |
| 68 | 68 | 68 | 70 | 70 | 70 | 70 | 72 | 72 | 72 |
| 72 | 72 | 72 | 72 | 74 | 74 | 74 | 74 | 76 | 76 |
| 76 | 76 | 76 | 76 | 76 | 78 | 78 | 78 | 78 | 78 |
| 78 | 80 | 80 | 80 | 82 | 84 | 84 | 88 | 90 | 94 |
| Tablo: Verilerin küçükten büyüğe doğru sıralanmış hali | | | | | | | | | |

Veriler küçükten büyüğe sıralandığında, %50 si medyan değerinden küçük ya da eşittir. %25’i bir  sayısından küçük ya da eşit kalıyorsa,  sayısına birinci çeyreklik, %75’i bir  sayısından küçük ya da eşit kalıyorsa  sayısına da üçüncü çeyreklik denir. İkinci çeyreklik medyandır. Benzer şekilde, verilerin %95’i bir  sayısından küçük ya da eşit kalıyorsa  sayısı dağılımın %95 lik kritik değeridir. Bu kritik değerlerden bazıları,

, , 

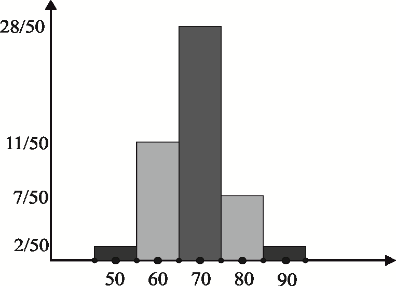
 ,  , 

olarak gözlenmiştir. Bu değerlere göre, verilerin %95’i  sayısından küçük ya da eşittir. Benzer şekilde, %90’ı  sayısından küçük ya da eşittir. Verilere ait diğer yüzdelikler de hesaplanabilir. Örneğin, verilerin %60’ı bir  değerinden küçük ya da eşit oluyorsa  sayısı %60 lık kritik değer olarak alınır. Buradaki  sayısı örneklem içinde bir sayı olmak zorunda değildir.

Veriler belli kategorilere ayrılarak da grafikler oluşturulur. Örneğin 90 ve üzerinde not alan öğrenciler birinci grup, 80 ile 90 arasında not alan öğrenciler ikinci grup olarak ele alınarak bir gruplandırma yapılabilir. Her gruptaki gözlem sayısına o grubun *sıklığı* (*frekansı*) denir. Her bir aralıktaki gözlem sayısı (frequency, sıklık), aralıktaki gözlem sayısının toplam gözlem sayısına oranı (relative frequency)  dir. Buna göre  olmak üzere aşağıdaki tablo oluşturulmuştur.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Grup | Aralık |  |  |  |
| 1 |  |  |  |  |
| 2 |  |  |  |  |
| 3 |  |  |  |  |
| 4 |  |  |  |  |
| 5 |  |  |  |  |

Bu tablo başka bir şekilde aşağıdaki gibi özetlenebilir. Bu özet şekline verilerin *Dal-Yaprak Grafiği* denir.



*Şekil 6.4.1 Dal Yaprak Grafiği(solda) ve Histogram(sağda)*

Bu oranlardan verilerin histogramı oluşturulur. Histogramdan yararlanarak dağılımın şekli hakkında görsel bilgiler elde edilir

Sıra istatistiklerinin olasılık fonksiyonları aşağıdaki teoremde özetlenmiştir. Teoremin ispatı burada verilmemiştir. Ayrıntılı bilgi için Casella ve Berger (2002) ye bakılabilir.

**Teorem 6.4.1** Dağılım fonksiyonu , olasılık veya olasılık yoğunluk fonksiyonu da  olan kitleden bir örneklem  olsun. Bu örnekleme ilişkin sıra istatistikleri  olmak üzere,

a)  rasgele değişkeninin olasılık yoğunluk fonksiyonu  için

,

b)  ve  rasgele değişkenlerinin ortak olasılık yoğunluk fonksiyonu  için

,

c)  rasgele değişkenlerinin ortak olasılık yoğunluk fonksiyonu da,



dir (Casella ve Berger, 2002, sayfa 229-230)

**Örnek 6.4.2** ,  dağılımından bir örneklem olsun. Yani  lerin olasılık yoğunluk fonksiyonu ve dağılım fonksiyonu sırası ile

 , 

olarak verilmiş olsun.

a)  sıra istatistiğinin olasılık yoğunluk fonksiyonu Teorem (6.4.1a) dan,  için



eşitliğinden



olarak yazılır.  sıra istatistiğinin olasılık yoğunluk fonksiyonu dağılım fonksiyonu yardımı ile de bulunabilir. ,  rasgele değişkeninin dağılım fonksiyonunu göstermek üzere,  için  ve ve  için  olup  için,



dir. Buradan  nin olasılık yoğunluk fonksiyonu dağılım fonksiyonunun türevinden (dağılım fonksiyonunun türevlenebildiği yerlerde),



şeklinde elde edilir.  nin ilk iki momenti ve varyansı,

, 

ve



şeklinde hesaplanmıştır. Buradan da,  rasgele değişkeninin beklenen değeri ile varyansı da

 ve 

olur.

b)  örneklem genişliği istatistiğinin olasılık yoğunluk fonksiyonunu bulalım. Önce,  ve  nin ortak olasılık yoğunluk fonksiyonu Teorem (6.4.1b) den



şeklindedir.  nin olasılık yoğunluk fonksiyonu için  yardımcı dönüşümünü tanımlayalım. Ters dönüşümler ,  şeklinde olup Jacobien matrisi ve determinantı



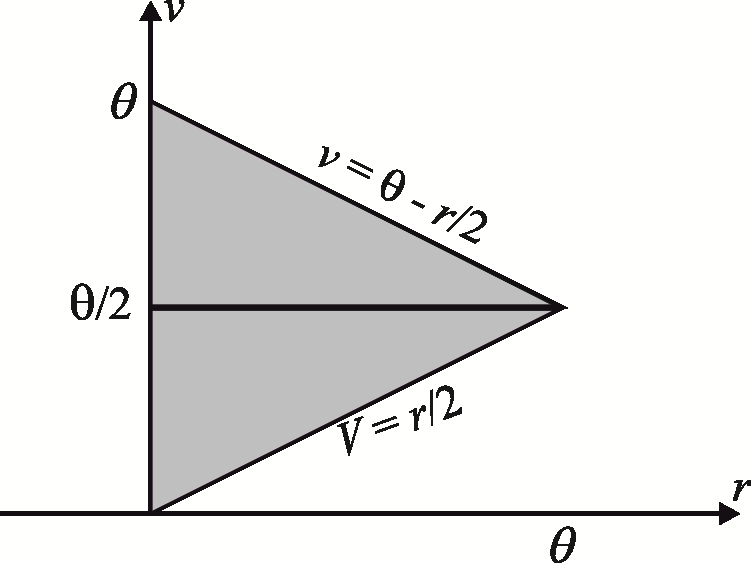
olarak bulunmuştur. Burada, rasgele değişkenlerin sınırlarının belirlenmesi gerekir.  ve  rasgele değişkenlerinin her ikisinin değer kümesi  ve  olduğundan  olduğu açıktır.  nin üst sınırı için  nin alacağı en büyük değerden  rasgele değişkeninin alacağı en küçük değer çıkartıldığında  nin üst sınırının  olduğu görülür. Diğer taraftan,  olduğundan,  ise  dir. Ayrıca,  olduğundan  olup  ve  nin ortak olasılık yoğunluk fonksiyonu  ve  için



şeklinde olur. Yani,  ve  nin ortak olasılık yoğunluk fonksiyonu,



dir. Ortak olasılık yoğunluk fonksiyonunun değer kümesi Şekil (6.4.2) de gösterilmiştir.



*Şekil 6.4.2 Örnek (5.2.4b) Olasılık yoğunluk fonksiyonunun değer kümesi*

Buna göre,  nin olasılık yoğunluk fonksiyonu için  ortak olasılık yoğunluk fonksiyonunun  nin değer kümesi üzerinden integrali alınır. Bu integral,



olup  nin olasılık yoğunluk fonksiyonu,



olarak bulunmuştur.  nin olasılık yoğunluk fonksiyonu için grafikten de görüldüğü gibi integralin iki ayrı şekilde hesaplanması gerekir. İntegral değerleri  için,

,

 için,



şeklindedir. Buna göre  nin olasılık yoğunluk fonksiyonu,



olarak elde edilmiştir.  için  olduğu açıktır.

c)  beklenen değeri  olan üstel dağılımdan bir örneklem olsun.  ve  sıra istatistiklerinin ortak olasılık yoğunluk fonksiyonu Teorem (6.4.1b) den ( ve )  için,



şeklindedir. Şimdi,  örneklem uzunluğu istatistiğinin olasılık yoğunluk fonksiyonunu bulalım. Bunun için  yardımcı dönüşümü ile ters dönüşümler  ve  olup Jacobien matrisinin determinantı  dir. Buna göre,  ve  istatistiklerinin ortak olasılık yoğunluk fonksiyonu  için



olarak yazılabilir. Bu fonksiyonun  nin değer kümesi () üzerinden integrali



olup  nin olasılık yoğunluk fonksiyonu,



şeklinde bulunmuştur

Sıra istastitikleri ile rasgele değişkenlerin yüzdelikleri de tahmin edilir.  rasgele değişkeninin dağılım fonksiyonu  aralığında sürekli olsun.  in dağılım fonksiyonu  olmak üzere  in  yüzdeliği (quantile)  için  olarak verilir. Buna göre,  rasgele değişkeninin medyanı,  dir.

 bir örneklem,  de bu örnekleme karşılık gelen sıra istatistikleri olsun.  diyelim. Burada,  sayısı  sayısının tam kısmını göstermektedir.  in olasılık yoğunluk fonksiyonu  ise  in  ye kadar integrali  dır. Buna göre,  denirse,  nin beklenen değeri,



olur (Hogg, Mckean ve Craig, 2005). Buradan,  olduğunda  sıra istatistiği  nin bir tahmin edicisi olarak alınabilir.

Herhangi bir istatistiksel analiz için  istatistik (five number summary) değeri ön plana çıkmaktadır. Bunlar

, 

birinci çeyreklik , üçüncü çeyreklik  ve

medyan (ortanca) 

dır.

**Örnek 6.4.3**  birimlik örneklem değerleri sıralanmış olarak aşağıda verilmiştir.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 56 | 70 | 89 | 94 | 96 | 101 | 102 | 102 | 102 | 105 | 106 | 108 | 110 | 113 | 116 |

Bu verilere göre,  olup  özet istatistiğin değeri

 ,  ,  , , 

olarak elde edilir

Veri analizinde, aykırı değerlerin (outliers) belirlenmesi önemlidir. Bu aykırı değerlerin belirlenmesinde de sıra istatistiklerinden yararlanılır. Bunun için alt ve üst çitler (lower fence, upper fence) belirlenir.  olmak üzere,  ve  şeklinde alt ve üst çitler tanımlanır.  aralığının dışında kalanlar aykırı değerlerdir. Örnek (6.4.3) de  olup alt ve üst çitler

 ve 

olduğundan  aralığının dışında kalan gözlemler aykırı gözlemlerdir. Buradan,  ve  bu aralığın dışında olduğundan, aykırı gözlemlerdir.

 sürekli rasgele değişkeninin dağılım fonksiyonu  şeklinde olsun. Burada,  dağılım fonksiyonu biliniyor olmasına rağmen,  ve  bilinmiyor olabilir.  ise  nin dağılım fonksiyonu  olur.  için  değeri  rasgele değişkeninin  yüzdeliğini göstersin.  de  nin  yüzdeliğini göstersin.  biliniyorsa  de biliniyor demektir.  olduğundan,  şeklinde bir ilişki elde edilir. Pratikte,  yüzdelikleri bilinmeyen parametrelerdir. Bu parametrelerin verilen  örneklemine göre tahmin edilmesi gerekir.  örneklemi için  sıra istatistikleri olmak üzere  için  diyelim. Böylece,  sıra istatistiği  nin tahmin edicisidir.  olmak üzere,  lerin  lere karşı grafikleri çizilir. Bu grafikler de  *çizitleridir*. Bu grafik doğrusal ise,  rasgele değişkeninin dağılımının  yapısına uygun olduğu söylenebilir. Uygulamada normallik çok kullanıldığından bu grafiğe bazen normal olasılık grafiği de denir.