İstatistik I

11. Hafta

25 Nisan 2022

11. Bazı Sürekli Dağılımlar ve Normal Dağılım

www.umitsarp.com

Ümit SARP, umit.sarp@ikcu.edu.tr

İZMİR KÂTİP ÇELEBİ ÜNİVERSİTESİ
2010

Giriş

- Bazı Sürekli Dağılımlar
  - Düzgün Dağılım
  - Üstel Dağılım
  - Gamma Dağılımı
  - Beta Dağılımı
- Normal Dağılım

Düzgün Dağılım

Düzgün Dağılım;

X ratgele olasılık yoğunluk fonksiyonu için

$$f(x) = \frac{1}{a-b}, \quad a \leq x \leq b$$

şeklindedir.

$$E(X) = \frac{a+b}{2} \qquad\qquad Var(X) = \frac{(b-a)^2}{12}$$

Üstel Dağılım

Üstel Dağılım;

Negatif olmayan değerler alan X rastgele değişken için olasılık yoğunluk fonksiyonu üstel dağılım için,

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & \text{aksi halde} \end{cases}$$

şeklindedir.

$$E(X) = \frac{1}{\lambda}, \qquad Var(X) = \frac{1}{\lambda^2}$$

Gamma Dağılımı

Gamma Dağılımı;

X, pozitif değerler alan sürekli rastgele değişken için, gamma dağılımı için olasılık yoğunluk fonksiyonu,

$$f(x) = \begin{cases} \frac{\alpha}{\Gamma(r)} \cdot (\alpha x)^{r-1} e^{-\alpha x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

şeklindedir.

$$E(X) = \frac{r}{\alpha}, \qquad Var(X) = \frac{r}{\alpha^2}$$

Beta Dağılımı

Beta Dağılımı;

X rastgele değşikeni beta dağılımı gösteriyorsa olasılık yoğunluk fonksiyonu,

$$f(x) = \begin{cases} \frac{1}{B(\alpha,\beta)} \cdot X^{\alpha-1} \cdot (1-x)^{\beta-1}, & 0 < x < 1 \\ 0, & \text{aksi halde} \end{cases}$$

şeklindedir.

$$E(X) = \frac{\alpha}{\alpha+\beta}, \quad Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2 \cdot (\alpha+\beta+1)}$$

Normal Dağılım

İstatistik teorisinde kullanılan temel olasılık dağılımlarından en önemlisi olan normal dağılımın günlük yaşamda pek çok uygulamasıyla karşılaşılır.

Örneğin, bir fabrikada üretilen ürünlerin raf ömürleri, şirket çalışanlarının aylık performans puanları, bir nakliye aracına yüklenen ürünlerin toplam ağırlığı vb. gibi sürekli değişkenlerin birçoğunun dağılımı normal dağılıma uyar.

Bunun birlikte, istatistiksel analiz tekniklerinin büyük bir kısmında, evren dağılımının normal dağılıma uyduğu varsayımı kullanılır.

Normal Dağılım

Normal Dağılım;

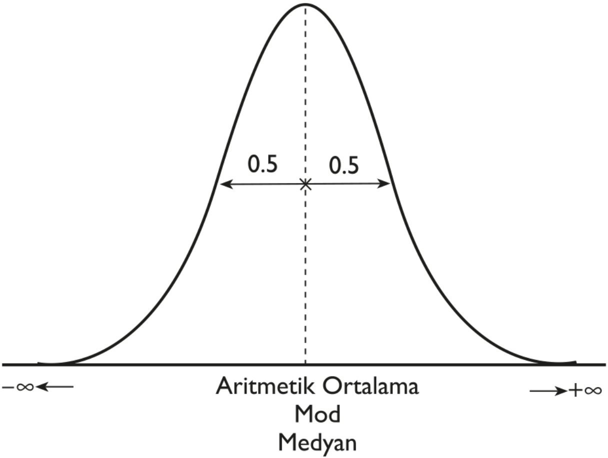
Normal dağılım eğrisinin fonksiyonu aşağıdaki şekildedir:

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

eşitliği ile verilir. Burada, X; sürekli rassal değişkenin aldığı değeri, μ; evren ortalamasını, σ; evren standart sapmasını, π; pi sayısı olan 3,14'ü ve e; doğal logaritma sisteminin tabanı olan 2,718 sayısını belirtmektedir. Dolayısıyla normal dağılımda, birincisi ortalama, ikincisi de standart sapma olmak üzere iki adet bilinmeyen parametre bulunur.

Normal Dağılım

Normal dağılım eğrisi, bir fonksiyon olarak kartezyen düzlemde aşağıdaki gibi çizilir:



Normal Dağılım

Şekilde görüldüğü gibi, normal olasılık dağılımının şekli çan eğrisi biçimindedir ve merkezinde tek bir tepe noktasına sahiptir. Dağılımın aritmetik ortalama, mod ve medyan değerleri eşittir ve bu değerler dağılımın merkezinde yer alırlar. Dolayısıyla, normal eğrinin altında kalan alanın yarısı bu noktanın sağında, diğer yarısı da bu noktanın solunda yer alır. Buna göre, dağılım aritmetik ortalama, mod ve medyana göre simetrik bir görünümdedir.

Dağılım eğrisi, merkez değerinden her iki yöne doğru düzgün bir şekilde azalır. Dolayısıyla, dağılım asimptotiktir. Bu da, eğrinin X-eksenine gitgide yaklaştığı ancak hiçbir zaman kesmediği anlamını taşır. Bir başka deyişle, eğrinin iki ucu teorik olarak X-ekseni ile sonsuzda kesişir. Fakat uygulamalarda hemen hemen tüm gözlem değerlerinin, aritmetik ortaladan $\pm\sigma 3$ uzaklığa kadar değer aldığı görülür ve bu aralığın dışında kalan bölge için eğri altında kalan alanın sıfır olduğu kabul edilmektedir.

Normal Dağılım

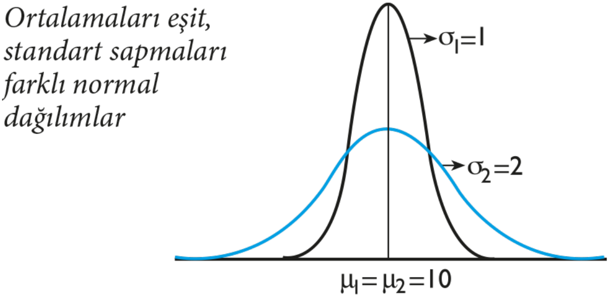
Sürekli olasılık dağılımlarında, dağılım eğrisinin altında kalan herhangi bir aralığın alanı, değişkenin bu aralıkta yer alma olasılığını verir. Ayrıca olasılık dağılımlarının özelliği gereği, olası tüm sonuçlar için normal dağılım eğrisinin altında kalan toplam alan 1'dir ve dağılımın aritmetik ortalamaya göre simetrik olması nedeniyle, aritmetik ortalamanın sağında ve solunda kalan yarı alanların büyüklüğü $0,5$'e eşittir.

Normal Dağılım

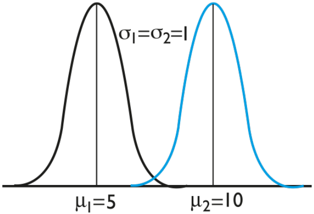
Normal dağılımın konumu, μ aritmetik ortalamasına, yayılması ise σ standart sapmasına göre belirlenir. Sadece tek bir eğri değil, birçok normal dağılım eğrisi çizilebilir.

Aşağıdaki şekilde aritmetik ortalamaları eşit ancak standart sapmaları farklı iki adet normal dağılım eğrisi görülmektedir. Burada verilen eğrilerin yukarıda sözü edilen özellikleri sağlamasının yanı sıra dağılımın standart sapma değeri küçüldükçe veriler aritmetik ortalama etrafında daha fazla yoğunlaşmakta ve eğri daha sivri bir şekil almaktadır.

Normal Dağılım

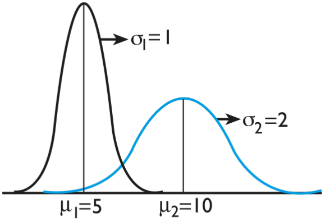


Normal Dağılım



Ortalamaları farklı, standart sapmaları eşit normal dağılımlar

Bu kez standart sapmaları aynı ancak aritmetik ortalamaları farklı,



Ortalamaları ve standart sapmaları farklı normal dağılımlar

Hem aritmetik ortalamaları hem de standart sapmaları farklı iki adet normal dağılım eğrisi görülmektedir

Normal Dağılım

Aritmetik ortalama ve standart sapma değerlerine ilişkin olarak sonsuz sayıda farklı normal dağılım eğrisi çizilebileceğinden, tüm eğriler için olasılık hesabında kullanılacak tabloların oluşturulması mümkün değildir. Bu nedenle, olasılıkların belirlenmesinde standart normal dağılımdan yararlanılır.

Normal Dağılım

Normal dağılımın özel bir durumu olan standart normal dağılım, elde edilebilecek tüm normal dağılımlar için olasılık belirlemede kullanılabilmektedir.

Aritmetik ortalaması 0 ve standart sapması 1 olan normal dağılıma standart normal dağılım adı verilir.

Normal Dağılım

Herhangi bir normal dağılım, gözlem değerlerinden aritmetik ortalama değeri çıkartıldıktan sonra, bu farkın standart sapmaya bölünmesi yoluyla standart normal dağılıma dönüştürülebilir. Bu işlem sonucu elde edilen değerlere z değerleri ya da standart normal değerler adı verilir. z değerleri standart değerler olduğu için birimleri yoktur.

X gözlem değerleri aşağıdaki eşitlik yardımıyla z değerine dönüştürülür:

$$z = \frac{X - \mu}{\sigma}$$

Normal Dağılım

Burada, X; Herhangi bir gözlem ya da ölçüm değerini, µ; dağılımın aritmetik ortalamasını, σ ise dağılımın standart sapmasını belirtir.

μ aritmetik ortalama ve standart sapması ile normal dağılıma sahip X değerleri standartlaştırıldığında elde edilen z değerlerinin dağılımı, aritmetik ortalaması 0 ve standart sapması 1 olan normal dağılım, dolayısıyla standart normal dağılımdır.

z değerleri, X gözlem değerlerinin aritmetik ortalamadan kaç standart sapma uzaklıkta olduğunu belirten değerlerdir.

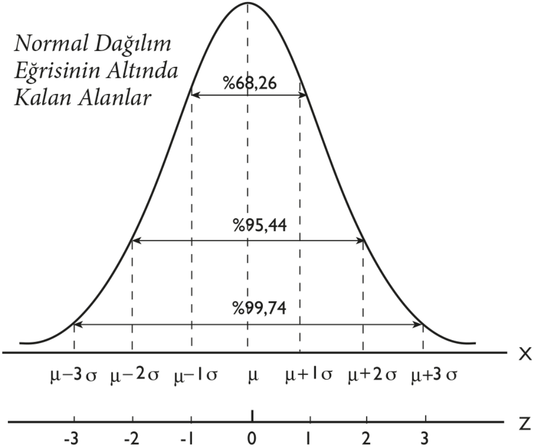
Normal Dağılım

Aritmetik ortalaması ve standart sapması belli bir normal dağılımda μ ve değerleri arasında değişmez ilişkiler vardır. Bu ilişki yardımıyla aritmetik ortalamadan ne kadar uzaklaşılırsa hangi büyüklükte bir alan elde edileceği belirlenebilir. Örneğin, aritmetik ortalamadan 1μ kadar sağ ve sol tarafa ilerlenirse bu noktalarda sınırlanan yarı alanlar % 34,13 olup bu iki alanın toplamı % 68,26 olur. Bunun gibi, aritmetik ortalamadan 2σ ve 3σ kadar sağ ve sol tarafa ilerlendiğinde elde edilecek değerler Tabloda görülmektedir.

Aritmetik Ortalama		Standart Sapma	Kapsadığı Alan
μ	∓	1σ	0,3413+0,3413=0,6826
μ	∓	2σ	0,4772+0,4772=0,9544
μ	∓	3σ	0,4987+0,4987=0,9974

Normal Dağılım

Şekilde, aritmetik ortalamadan $\pm 1\sigma$'dan $\pm 3\sigma$'ya kadar uzaklaşıldığında ortaya çıkan alanlar görülmektedir.



Normal Dağılım

Burada, x değerleri z değerlerine dönüştürüldüğünde, ölçek de değişmektedir. Dönüşüm sonucu elde edilecek z değerleri de grafiğin alt kısmında görülmektedir. Örneğin, μ değerine karşılık gelen z değeri 0'dır, $\mu + 1\sigma$ değerine karşılık gelen z değeri ise $+1$'dir.

Normal Dağılım

Örnek;

X rassal değişkeni, 100 aritmetik ortalaması ve 10 standart sapması ile normal dağılıma sahiptir. X'in 100 ile 123,3 arasında olma olasılığını bulunuz.

Normal Dağılım

Örnek;

X rassal değişkeni, 100 aritmetik ortalaması ve 10 standart sapması ile normal dağılıma sahiptir. X'in 100 ile 123,3 arasında olma olasılığını bulunuz.

Çözüm: Burada, $P(100 < x < 123,3)$ olasılığı sorulmaktadır. Bu olasılığın bulunabilmesi için, ilk olarak alt sınır 100 ve üst sınır 123,3 değerlerinin, 25 no'lu eşitlik kullanılarak z değerlerine dönüştürülmesi gerekir.

Normal Dağılım

Örnek;

$x = 100$ değerine karşılık gelen z değeri, $z = \frac{100-100}{10} = 0$'dır. Yukarıda belirtildiği üzere, aritmetik ortalamaya karşılık gelen z değeri daima 0'dır. $x = 123,3$ değerine karşılık gelen z değeri ise, $z = \frac{123,3-100}{10} = 2,33$, olarak elde edilir. Ardından standart normal dağılım tablosuna bakılır. Tabloda, bu tablonun bir bölümü görülmektedir. $z = 0$ ile $z = 2,33$ arasında kalan bölgenin alanını bulabilmek için tabloda z harfi ile belirtilen ilk sütunda $2,3$ değerini bulana kadar aşağıya doğru ilerlenir. Daha sonra, bu satırda sağa doğru ilerlenir ve $0,03$ ile başlayan sütunun altındaki alan değeri $0,4901$ olarak bulunur. Bu da, $0,00$ ile $2,33$ arasında eğrinin altında kalan alanın $0,4901$ olduğu anlamına gelir ve bu değer X'in 100 ile $123,3$ arasında yer alma olasılığıdır.

Normal Dağılım

Örnek;

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	...
⋮								
2,0	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	
2,1	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	
2,2	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	
2,3	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	
2,4	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	
2,5	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	
2,6	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	
⋮								

Normal Dağılım

Örnek;

Otomobil lastiği üretimi yapan bir fabrikada, üretilen lastiklere uygulanan bir dayanıklılık testi sonucunda lastiklerin teste dayanma sürelerinin, aritmetik ortalaması 242 sn. ve standart sapması 36 sn. ile normal dağılıma sahip olduğu bilinmektedir. Bu fabrikada üretilen ve teste tabi tutulan lastikler arasından rassal olarak seçilen bir lastiğin teste dayanma süresinin,

1. 300 sn.'den fazla olma,
2. 220 ile 290 sn. arasında olma,
3. 256 ile 308 sn. arasında olma olasılıklarını bulunuz.
4. Toplam 150 lastiğe test uygulandığı bilindiğine göre, bu teste 300 sn.'den daha fazla dayanabilen lastik sayısı kaçtır?

Normal Dağılım

Örnek;

X rassal değişkeni, "Lastiğin dayanıklılık testine dayanma süresi" olarak tanımlandığında, soruda verilen bilgilere göre bu X rassal değişkeninin dağılımı, ortalaması 242 sn. ve standart sapması 36 sn. ile normal dağılımdır.

Normal Dağılım

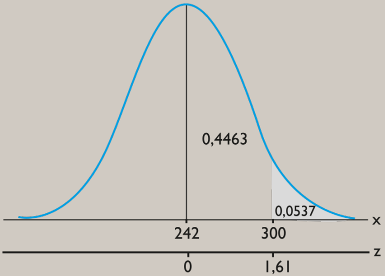
Örnek;

1-) Seçilen lastiğin teste dayanma süresinin 300 sn.'den fazla olma olasılığını bulabilmek için öncelikle bu X değerine karşılık gelen z değeri hesaplanmalıdır. Diğer bir deyişle, z dönüşümü uygulanmalıdır.

$x = 300$ noktasına karşılık gelen z değeri, $z = \frac{x-\mu}{\sigma} = \frac{300-242}{36} = 1,61$ olarak elde edilir. Buna göre soruda istenen olasılık; $z = 1,61$ değerinin sağında kalan bölgenin alanı olacaktır. Dikkat edilirse aritmetik ortalamanın (ya da $z = 0$ noktasının) sağında ve normal dağılım eğrisinin altında kalan bölgenin alanı, toplam alanın yarısı olan $0,5$'tir. Standart normal dağılım tablosuna göre $z = 0$ ile $z = 1,61$ arasındaki bölgenin alanı $0,4463$ olarak bulunur. Buna göre, soruda istenen olasılığı bulabilmek için $0,5$ değerinden, $z = 1,6$ değerine karşılık gelen alanı, yani $0,4463$ değerini çıkartmak gerekir.

Normal Dağılım

Örnek;



Buna göre, seçilen lastiğin teste dayanma süresinin 300 sn.'den fazla olma olasılığı;
$P(x > 300) = P(z > 1,61) = 0,5 - P(0 < z < 1,61) = 0,5 - 0,4463 = 0,0537$ olarak bulunur.

Ödevler

Yukarıdaki son örnekte yer alan 2., 3. ve 4. maddelerin çözümleri yapınız.

12-13-14 Hafta: Momentler, Hipotez Testi, Regrasyon, Korelasyon

Kaynaklar I

[1] K. Mert Çubukçu,
"Planlamada ve Coğrafyada Temel İstatistik ve Mekansal İstatistik",
Ankara: Nobel Akademik Yayıncılık Eği. Dan. Tic. Ltd. Şti., (2015).

[2] A. Özmen, F. Er, M. Atlas, E. Şıklar,
"İstatistik (AÖF)",
Eskişehir: Anadolu Üniversitesi Yayınları, (2012).

[3] L. İşbilen Yücel,
"İstatistik Maliye Uzaktan Eğitim",
İstanbul Üniversitesi Açık ve Uzaktan Eğitim Fakültesi Ders Notu.

[4] Ö. Serper,
"Uygulamalı İstatistik 1",
Bursa: Ezgi Kitapevi, (2004).

Kaynaklar II

[5] Murat Komisyon,
"İstatistik",
Murat Açıköğretim Yayınları, (2004).

[6] N. Gürsakal, A. Oğuzlar,
"Betimsel İstatistik",
Dora Yayıncılık, (2019).

[7] Y. Baykul, C. O. Güzeller,
"Sosyal Bilimler için İstatistik Uygulamaları",
Ankara: Pegem Akademi, (2014).

[8] Ankara Üniversitesi Açık Ders Sunumları,
"AKT102 İSTATİSTİK, BÖLÜM 3 OLASILIK",
https://acikders.ankara.edu.tr