Bilişimde İstatistiksel Analiz

Ders 5

Ali Mertcan KOSE Msc.
amertcankose@ticaret.edu.tr

İstanbul Ticaret Üniversitesi

Olasılık Dağılımları

- İstatistksel çözümlemelerde; değişkenlerin dağılma özellikleri, çözümleme yönteminin seçimi ve sonuçlarının yorumlanmasında önemlidir
- Dağılma özelliklerine **Olasılık Dağılımı** adı verilir.
- İstatistiksel çözümlemeler belirli çözümlemede kullanılan değişken(ler)in bu olasılık dağılımına uyması gerekir.

Raslantı değişkeni: Örneklem uzayındaki her sonuca ölçmeyi ya da saymayı gerektiren belli bir kuralı uygulayabiliriz. Tanımlanan kurala göre her sonuca bir sayı bağlanabilir. Örneğin; üç parayı birlikte atma denemesinde her atış için, tura sayısı değişik değerlerdir. Bu değerler rastlantı değişken olarak nitelendirilebilir. **Dağılım fonksiyonu:** X raslantı değişkeni ve x bir gerçel sayı ise, $(X \leq x)$ = s:$\{X_s \leq$x)} olayı tanımlanabilir. Bu olayın olasılığı $P(X \leq x)$'dir. $P(X \leq x)$ olasılığı x'e bağlıdır. Başka bir deyişle, x'in fonksiyonundur. Bu fonksiyona rastlantı değişkenin **dağılım fonksiyonu(birikimli dağılım fonksiyonu)** denir.

F(x)= $P(X \leq x)$.

Olasılık Dağılımları

- Herhangi olasılık dağılımı, y=f(x) biçiminde tanımlanan matematiksel bir fonksiyondur.
- y, x değerlerinin ortaya çıkma sıklığını gösterir.
- f(x), yoğunluk fonlsiyonu olarak da adlandırılır.
- f(x), x değişkeninin sürekli olması durumunda aşağıdaki özellikleri taşır.

$0\leq f(x)\leq 1 \quad -\infty\leq x\leq\infty$

$\left[\int_{-\infty}^{\infty} f(x)\, dx\right] = 1$

- f(x), x değişkeninin kesikli olması durumunda aşağıdaki özellikleri taşır.

$0 \leq f(x) \leq 1 \quad a \leq x \leq b$

$\sum_{x=a}^{b} f(x)$

X, kesikli raslantı değişkeni ise, dağılım fonksiyonu

F(x)= $\sum_{n \leq a} p(n)$ olur.

Beklenen Değeri(Aritmetik Ortalama): $\sum_{x=a}^{b} xf(x)$

Varyansı: $\sum_{x=a}^{b} x^2 f(x)$ - μ^2

X, sürekli raslantı değişkeni ise, dağılım fonksiyonu

$\sum_{n \leq a} p(n)$ olur.

X, sürekli raslantı değişkeni ise, dağılım fonksiyonu

F(x)= $\int_{-\infty}^{X} f(t)\, dt$

Beklenen Değeri(Aritmetik Ortalama): $[\int_{x=a}^{b} xf(x)]$

Varyansı: $[\int_{x=a}^{b} x^2 f(x)]$ - μ^2

Olasılık Dağılımları

Çok sayıda olasılık dağılımı bulunmaktadır. Bunlar arasından en sık kullanılanları:

- Binom Dağılımı
- Poisson Dağılımı
- Normal Dağılımı

Binom Dağılımı

Aynı koşullar altında, bir denemenin yinelendiğini her yinelemede bağımsız iki olaydan birinin ortaya çıktığını düşünelim. Bu olaylardan birinin olasılığı p ise, diğerinin q (1-p) dir. A ile gösterilirse, öteki $\bar{A}$ olur.

Denemenin iki kez yenilenmesinde AA, $A\bar{A}$, $\bar{A}A$, $\bar{A}\bar{A}$ sonuçlarından biri elde edilir.

Denemenin üç kez yinelenmesinde elde edilecek olası sonuçlar ve her sonucun çarpma ilkesine göre bulunan olasılığı aşağıda verilmiştir.

Sonuçlar: AAA, $AA\bar{A}$, $A\bar{A}A$, $\bar{A}AA$, $A\bar{A}\bar{A}$, $\bar{A}A\bar{A}$, $\bar{A}\bar{A}A$, $\bar{A}\bar{A}\bar{A}$

Olasılıklar: p^3, p^2q, p^2q, p^2q, pq^2, pq^2,pq^2, q^3

Binom Dağılımı

Yapılan üç yinelemede x=0, 1, 2, 3 kez A olayı elde edilebilir. A olayının sayısını gösteren raslantı değişkeni X olsun. X'in 0, 1, 2, 3 değerlerini alması olasılıkları

$P(X=0) = q^3$, $P(X=1) = 3pq^2$, $P(X=2) = 3p^2q$, $P(X=3) = p^3$

olur. Bu olasılıklar toplanırsa,

$q^3 + 3pq^2 + 3p^2q + p^3 = (p+q)^3$

elde edilir. Bu durum ikiterimlidir. Her olası sonucun olasılığı, bu iki terimlinin açılımındaki terimlerden biridir.

Yukarıdaki örneği geliştirebiliriz.

Olasılığı p(A)$=$ p olan A olayını ya da $\bar{A}$ olayını veren bir denemeyi düşünelim. Denemenin her yinelemesinde p olasılığı değişmiyor. Bu deneme n kez yineleniyor. n yinelemenin 0 tanesinde A, n tanesinde $\bar{A}$ ortaya çıkan sonuçların sayısı $\binom{n}{0}$; 1 tanesinde A, (n-1) tanesinde $\bar{A}$ ortaya çıkan sonuçların sayısı $\binom{n}{1}$; .... ; n tanesinde A ortaya çıkan sonuçların sayısı $\binom{n}{n}$'dir. Buna göre, örneklem uzayındaki tüm sonuçların sayısı,

$\binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{n} = 2^n$

olur. n bağımsız yinelenmenin x tanesinde A, n-x tanesinde $\bar{A}$ ortaya çıkan sonuçlarının sayısı $\frac{n!}{x!(n-x)!}$ dir.

Binom Dağılımı

Bu durumda binom dağılımı;

P(X=x) = p(x;n,p) = $\binom{n}{x}$ $p^x q^{n-x}$

yazılabilir. X raslantı değişkenin olasılık fonksiyonu, , parametreleri n ve p olan binom dağılımıdır.

p(x;n,p) = P(X=x) = p(x;n,p) = $\binom{n}{x}$ $p^x q^{n-x}$ x=0, 1, 2, . . . ,n

=0 öteki x değerleri için

Binom dağılımın beklenen değer np, Varyansı npq dir

Örnek

10 hasta için cerrahi tedavi uygulanmaktadır. Her bir hastanın cerrahi tedavisinin başarılı olma olasılı %70'dir bu durumda 5 ve 5 den az cerrahi tedavinin başarılı olma olasılığı nedir? $P(X \leq 5)$ Hastaların ortalama başarılı cerrahi tedavi sayısı ve varyansı nedir?

Poisson Dağılımı

0,1,2, . . . olası değerlerini alan X kesikli raslantı değişkenini göz önüne alalım. X'in olasılık fonksiyonu,

P(X=x) $= p(x; \lambda) = e^{-\lambda}\frac{\lambda^x}{x!}$, x= 0,1,2,3 . . . için $= 0$ diğer durumlar için

ise, X, Poisson dağılımına sahiptir denir. Burada $\lambda(\lambda > 0)$ parametredir.

Poisson dağılımın beklenen değer λ, Varyansı λ dir

Örnek

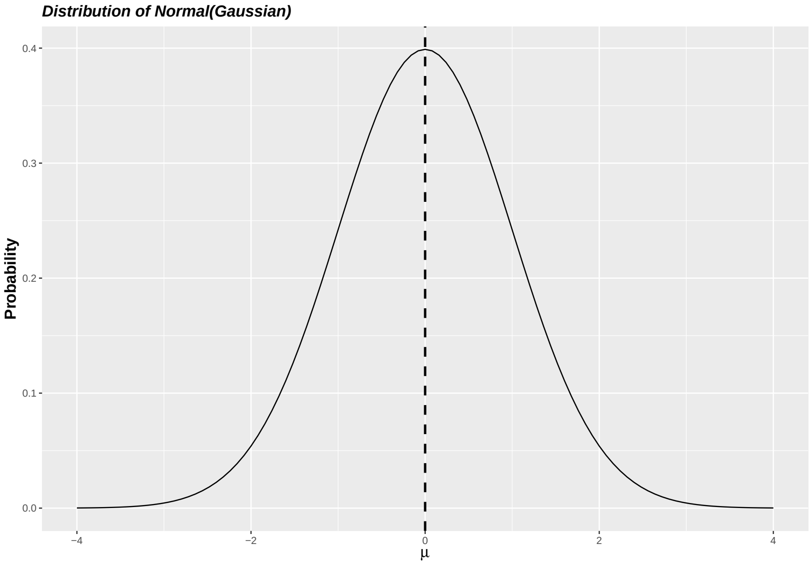
Bir hastanede meydana gelen doğumlar saatte 1.8 oranındadır. Hastanede bir saatte gözlenen 4 doğumun olma olasılığı nedir? ve bir saatte 2 ve 2 den fazla doğum olma olasılığı nedir?$P(X = 4)$ $P(X \geq 2)$ Hastane de meydana gelecek doğumların ortalama ve varyans değerleri nedir?

Normal Dağılım

- İstatistik çözümlemelerde en çok yararlanılan olasılık dağılımıdır.
- μ, kitle ortalamasını ve σ^2 kitle varyansını göstermek üzere dağılım(yoğunluk fonksiyonu),

$$\mathrm{P(x)} = \frac{1}{\sqrt{2\pi\sigma}} \; \frac{e^{-(x-\mu)^2}}{2\sigma^2}$$

Normal Dağılım



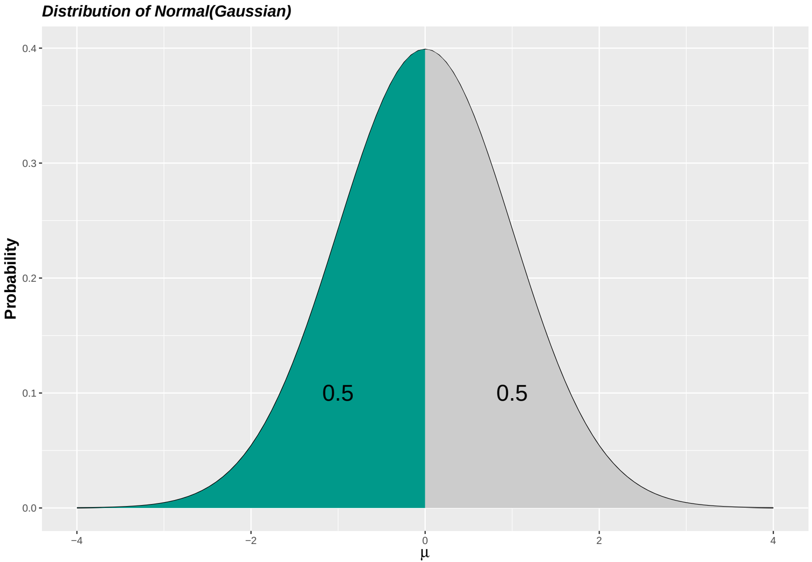
Normal Dağılım

- Dağılım ortalamaya simetriktir.
- Alanın %50'si ortalamadan geçen dikey çizginin sağına, %50'si soluna düşer.
- Eğri altında kalan toplam alan bir birim karedir.

$\int_{x=a}^{b} f(x)dx = 1$

- Aritmetik Ortalama = Medyan = Mod

Normal Dağılım



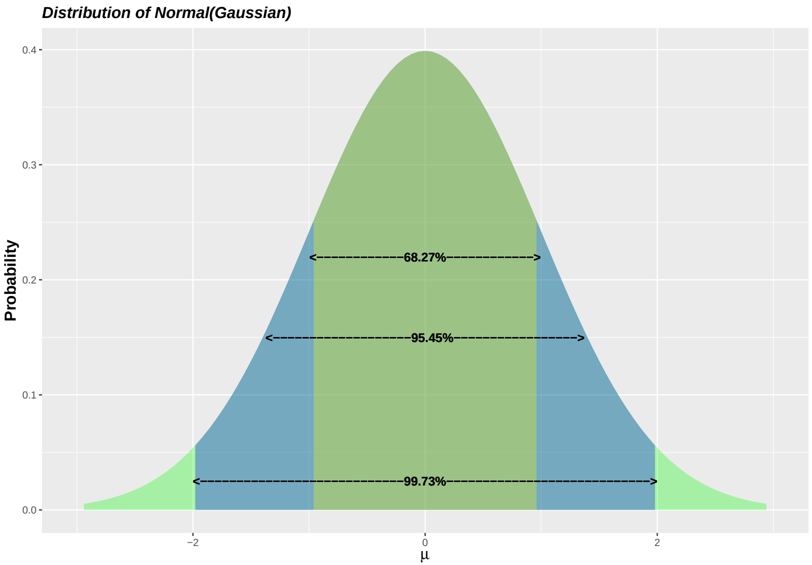
Normal Dağılım

$$P(\mu - \sigma \leq x \leq \mu + \sigma) = 0.6830$$

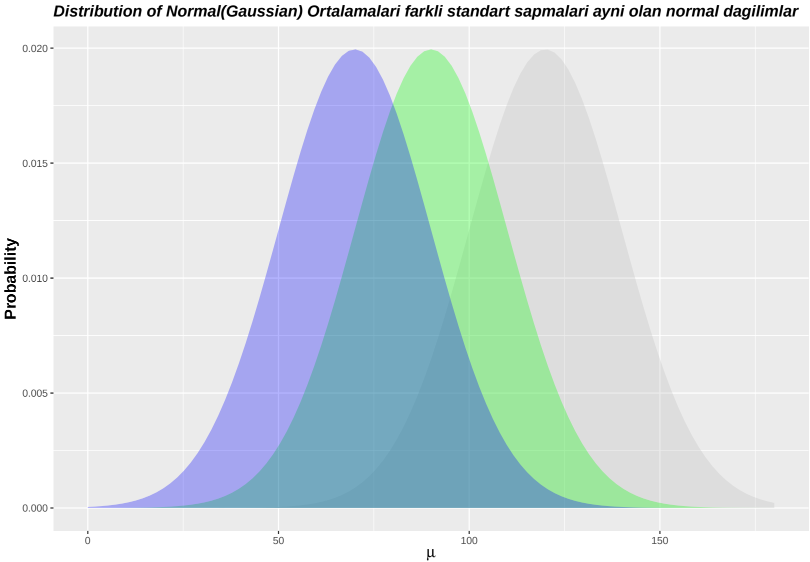
$$P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = 0.9540$$

$$P(\mu - 3\sigma \leq x \leq \mu + 3\sigma) = 0.9970$$

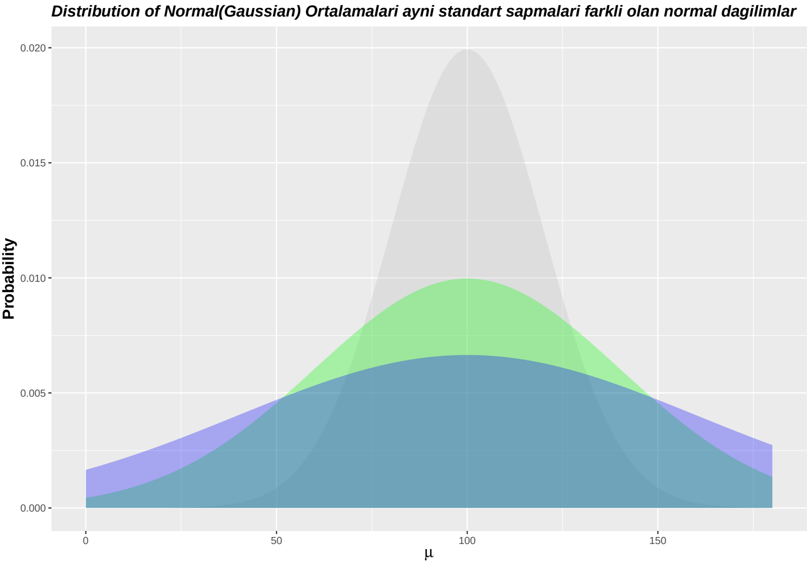
Normal Dağılım



Normal Dağılım



Normal Dağılım



Normal Dağılım

Normal dağılımda birikimli olasılıklar,

$P(x \leq b) = \int_{-\infty}^{b} f(x)\, dx$ işlemi ile

herhangi [a,b] aralığına ilişkin olasılık

$P(a \leq x \leq b) = \int_{a}^{b} f(x)\, dx$ işlemi ile bulunabilir.

Yukarıdaki hesaplamaları yapmak kolay olmadığından; bu hesaplamalar için standart normal dağılım yaklaşımından yararlanılır.

Normal Dağılım

Standart Normal Dağılım

- Normal Dağılımın özel bir biçimidir. Normal dağılıma dayalı hesaplamalarda kullanıcılara kolaylık sağlar.
- $\mu = 0$ ve $\sigma = 1$ dir.
- Yoğunluk fonksiyonu aşağıdaki gibidir.

$$\mathrm{P(z)} = \frac{1}{\sqrt{2\pi}} \; \frac{e^{-(z)^2}}{2}$$

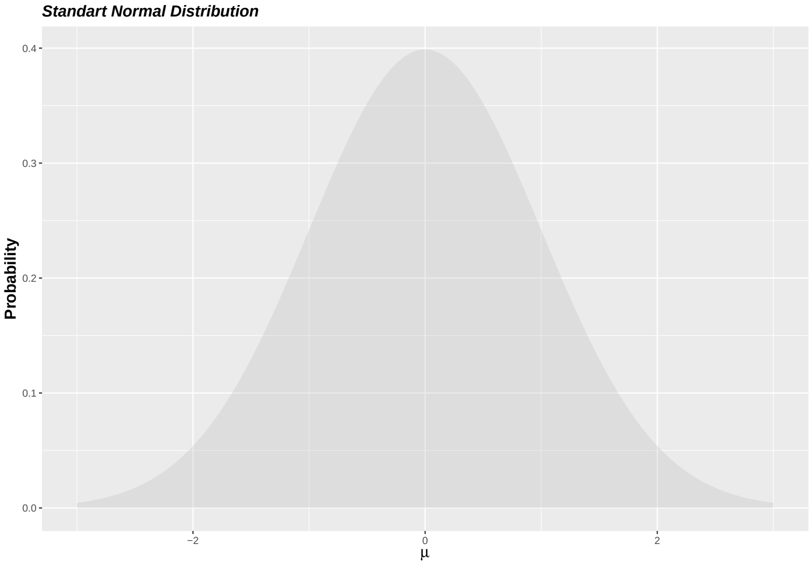
Normal Dağılım

Eğer bir x değişkeninin normal dağıldığı biliniyorsa

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

eşitliği ile elde edilen z değerleri ortalaması 0 ve varyansı 1 olan standart normal dağılıma uyar.

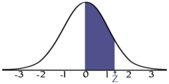
Normal Dağılım



Normal Dağılım

- Bu özellik, ortalama ve standart sapmanın değerine bağlı değildir.
- Ortalama ve standart sapma ne olursa olsun x değişkenin normal dağılması bu özelliğin geçerliği için yeterlidir.
- Çeşitli z değerleri için 0 ile z arasında kalan alanı gösteren z tablosu geliştirilmiştir. Bu tablodan yararlanarak normal dağılıma dayalı hesaplamalar yapılabilir.

Normal Dağılım



STANDARD NORMAL TABLE (Z)

Entries in the table give the area under the curve between the mean and z standard deviations above the mean. For example, for $z = 1.25$ the area under the curve between the mean (0) and z is 0.3944.

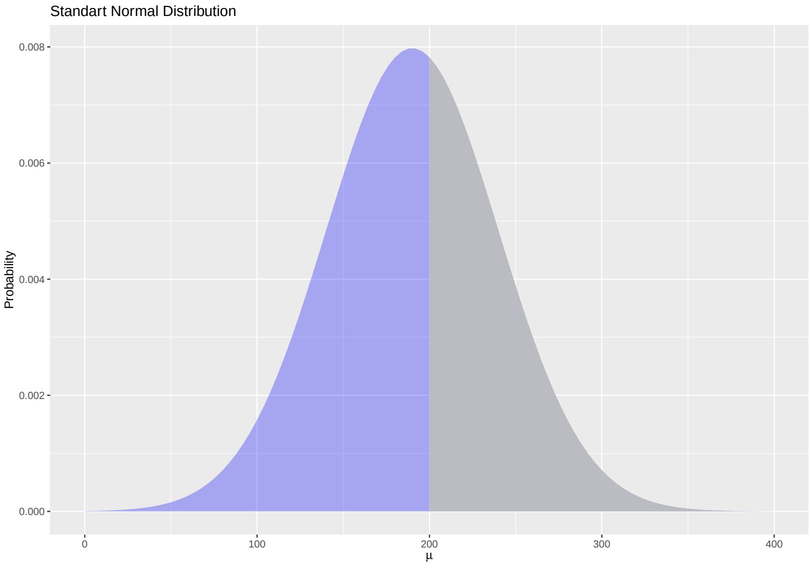
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998

Figure 1: Standard Normal Dağılım.

Örnek 1

10000 yetişkin üzerinde yapılan kolesterol tarama testi sonucunda kolesterol değerlerinin 190 ortalama ve 50 standart sapma ile normal dağıldığı görülmüştür. Kolesterol normal sınırlarının 150-200 olduğu bilindiğine göre kaç kişinin kolesterolü yüksektir?

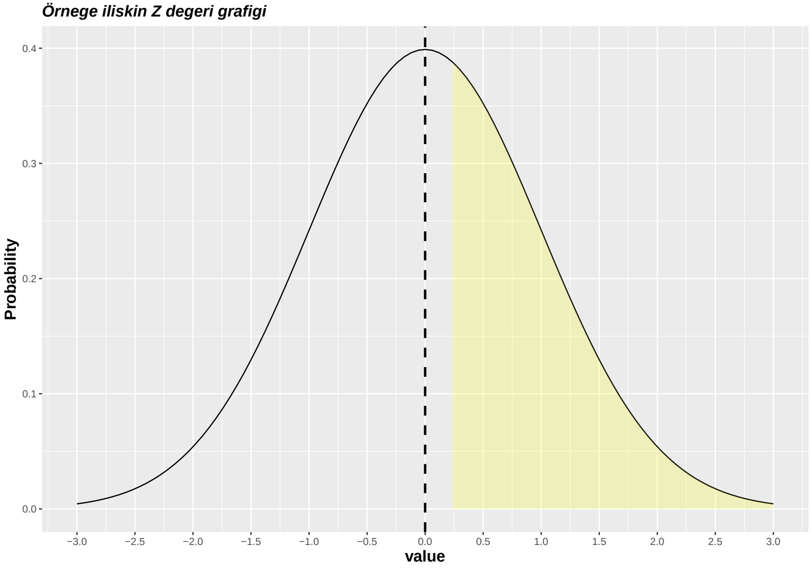
Örnek 1


Standart Normal Distribution
Probability
0.008
0.006
0.004
0.002
0.000
0
100
200
300
400
μ

Örnek 1

Standart Normal dağılımı yaklaşımını ve z$= \frac{x-\mu}{\sigma}$ eşitliğini kullanarak z$= \frac{200-190}{50}$ $=$0.2 olarak bulunur.

Örnek 1



Örnege iliskin Z degeri grafigi
Probability
0.4
0.3
0.2
0.1
0.0
−3.0
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
value

Örnek 1

Tablo Değeri

P(z$>$0.2)$=$ 0.5-0.0793 $=$ 0.4207

10000*0.4207 $=$ 4207 kişinin kolesterolü yüksektir.

Örnek 2

Biyoistatisitik dersinin not ortalamasının 70.07 standart sapmasının 10.27 olan normal dağılıma sahip olduğu bilinmektedir. Bu öğrencilerin % kaçının notu 70.07 ile 85 arasındadır? $P(70.07 \leq x \leq 85)$

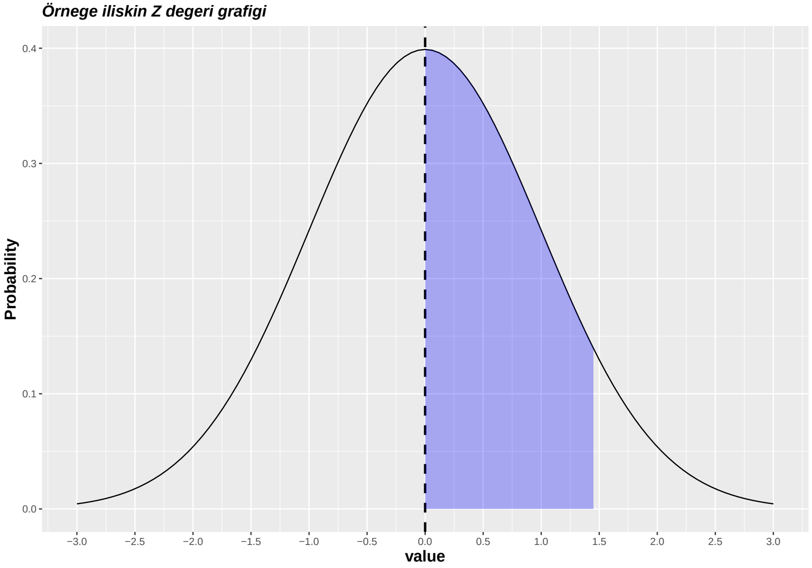
Standart Normal dağılımı yaklaşımını ve $z = \frac{x-\mu}{\sigma}$ eşitliğini kullanarak

$z_1 = \frac{70.07-70.07}{10.27} = 0$

$z_2 = \frac{85-70.07}{10.27} = 1,45$

olarak bulunur.

Örnek 2


Örnege iliskin Z degeri grafigi
Probability
0.0
0.1
0.2
0.3
0.4
value
−3.0
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0

Örnek 2

$P(0 \leq z \leq 1.45) = 0.4265$

100*0.4265 = %42,65 kişinin sınav notu 70.07 ile 85 arasında yer almaktadır.

Örnek 3

Bir hastanede belli bir hastalıkla ilgili bulunan hastaların tansiyonlarının ortalaması 15 ve varyansı 9 olan normal dağılıma sahip oldukları bilinmektedit. Bu hastalar ,çinden rasgele seçilen bir hastanın tansiyonun,

a) 11'den küçük
b) 12 den büyük
c) 9 ile 16 arasında olma olasılıklarını hesaplayınız.

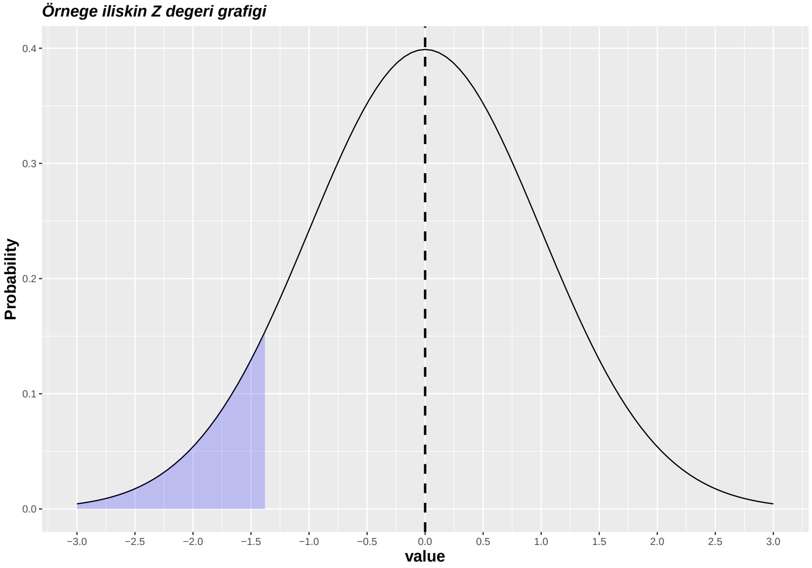
Örnek 3

Standart Normal dağılımı yaklaşımını ve z$= \frac{x-\mu}{\sigma}$ eşitliğini kullanarak

$z = \frac{11-15}{3} =$ -1.33

p(z$<$-1.33)

Örnek 3


Örnege iliskin Z degeri grafigi
Probability
0.4
0.3
0.2
0.1
0.0
−3.0
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
value

Örnek 3

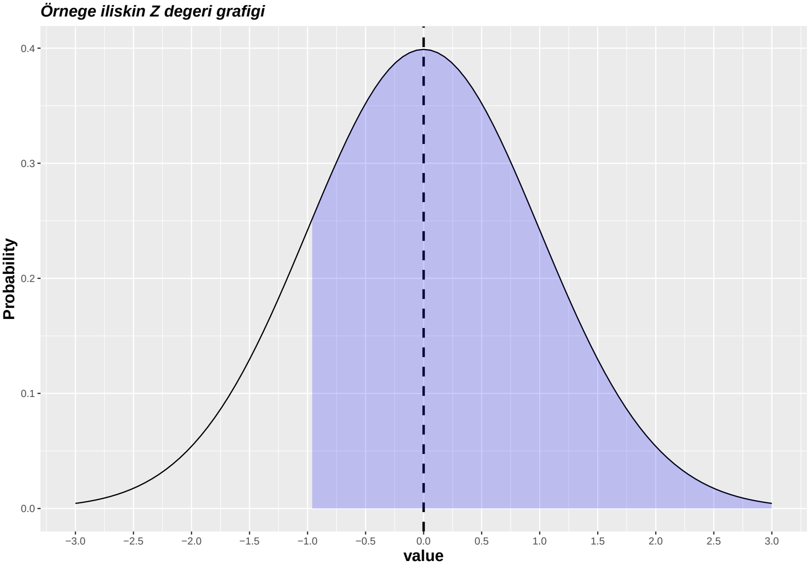
$P(z< -1.33) = 0.5\text{-}0.4082 = 0.0918$

Standart Normal dağılımı yaklaşımını ve z= $\frac{x-\mu}{\sigma}$ eşitliğini kullanarak

$z = \frac{12-15}{3} = -1$

$P(z>-1)$

Örnek 3



Örnege iliskin Z degeri grafigi
Probability
0.4
0.3
0.2
0.1
0.0
−3.0
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
value

Örnek 3

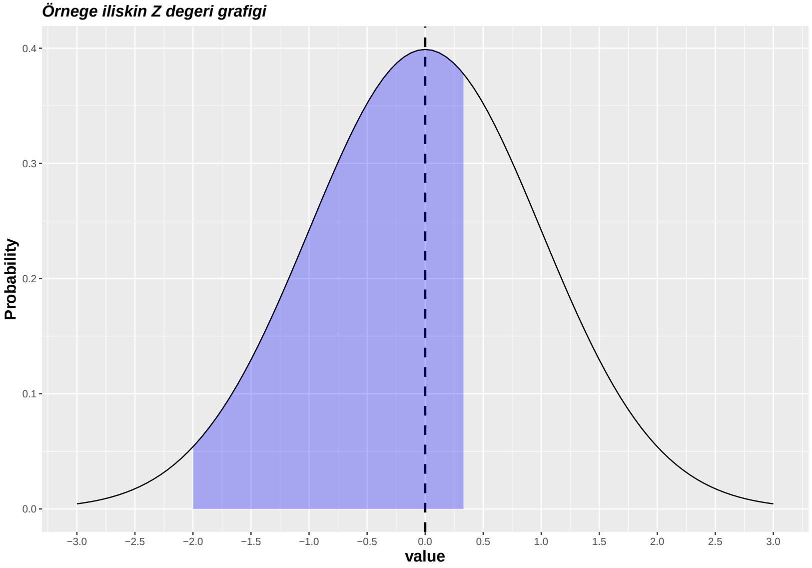
$\mathbf{P(z>\text{-}1) = 0.5+0.3413 = 0.8413}$

$z_1 = \frac{9-15}{3} =$ -2

$z_2 = \frac{16-15}{3} = 0.33$

p(-2 $\leq$ z $\leq$ 0.33)

Örnek 3



Örnege iliskin Z degeri grafigi
Probability
0.0
0.1
0.2
0.3
0.4
−3.0
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
value

Örnek 3

$$\mathbf{p(\text{-}2 \leq z \leq 0.33) = 0.4772 + 0.1293 = 0{,}6065}$$

ÖDEV 1

Z standart normal dağılıma sahip raslantı değişkeni olmak üzere $P(Z<1.23)= 0.8907$ ve $P(Z<3.06) = 0.9889$ ise $P(-3.06<Z<1.23)$ olasılığı kaçtır?

ÖDEV 2

Hastaların uyku düzeni için elde edilen PSQI(Pittsburgh Sleep Quality Index/Uyku Kalitesi İndeksi) değerlerinden bir inceleme yapılmak amaçlanmıştır. Bu doğrultuda kontrol grubu üzerinden yapılan çalışmada hastaların PSQI değerleri ortalaması 13.45 standart sapması 3.61 olarak elde edilmiştir. Bu durumda hastaların en az 8 ve en fazla 21 arasında değer alma olasılığı nedir?

ÖDEV 3

Belli bir ameliyatın başarılı sonuçlanması olasılığı %80'dir. Ameliyat edilen 10 hastadan 6 sının iyileşmesi olasılığı nedir?

ÖDEV 4

Bir şehirde ender rastlanan bir hastalıktan, bir hafta içinde ortalama ölen kişi sayısı 4'dür. Belli bir hafta içinde bu hastalıktan, hiç kimsenin ölmemesi olasılığı nedir?