3.6. Bazı Sürekli Dağılımlar

3.6.1 Normal Dağılım

Normal dağılım hem uygulamalı hem de teorik istatistikte kullanılan oldukça önemli bir dağılımdır. Normal dağılımın istatistikte önemli bir yerinin olmasının nedeni, yapılan birçok gözlem sonucunun, çan biçiminde bir dağılım vermesi ve çoğu dağılımın denek sayısı arttıkça normal dağılıma yaklaşmasıdır.

Sürekli bir X rasgele değişkeninin olasılık yoğunluk fonksiyonu,

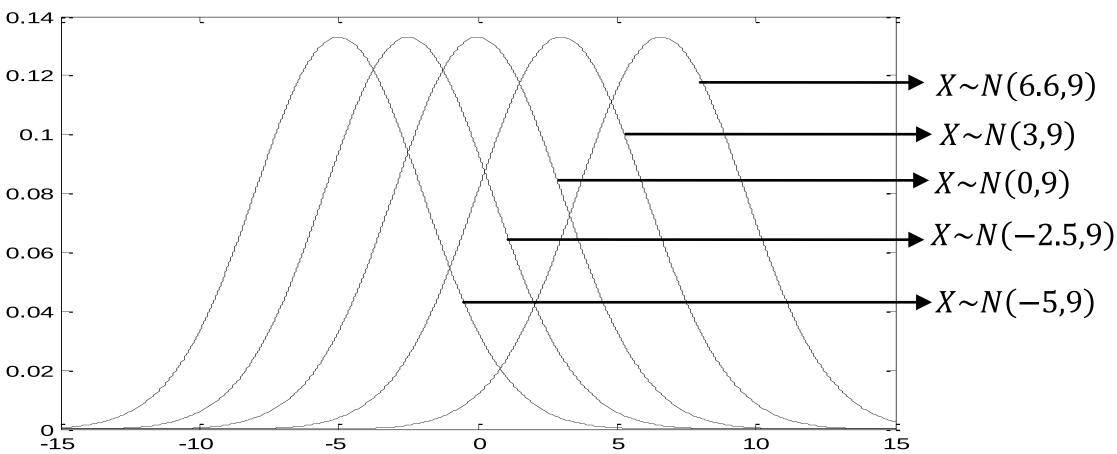
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0$$

biçiminde olduğunda, X rasgele değişkenine normal dağılıma sahiptir denir ve $X \sim N(\mu, \sigma^2)$ biçiminde gösterilir. Burada $\mu \epsilon R$ ve $\sigma^2 \in (0, \infty)$ dağılımın parametreleridir.

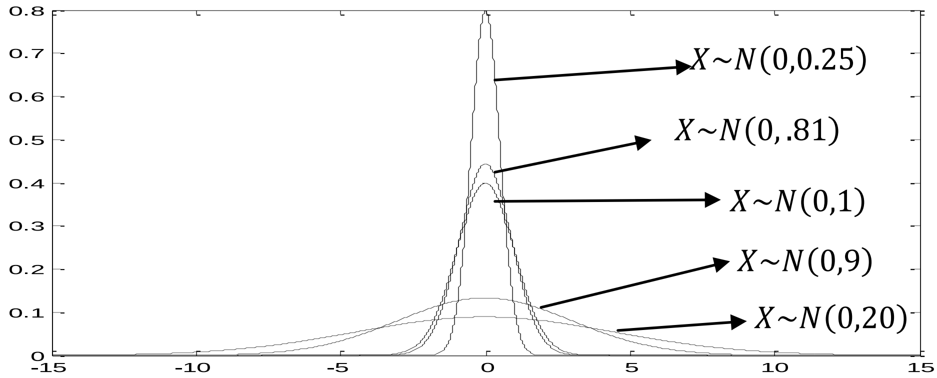
μ: kitle ortalaması
σ^2: kitle varyansı
$\mathrm{e} = 2.71825$
$\pi = 3.1416$

dır.

Aynı varyansa ($\sigma^2 = 9$) fakat farklı ortalamalara $\mu = -5, -2.5, 0, 3, 6.6$ sahip normal dağılımlı rasgele değişkenlere ilişkin olasılık yoğunluk fonksiyonlarının grafiği aşağıda verildiği gibidir,



Aynı kitle ortalaması ($\mu = 0$) ve farklı ($\sigma^2 = 0.25, 0.81, 1, 9, 20$) varyanslara sahip normal dağılımlı rasgele değişkenlere ilişkin olasılık yoğunluk fonksiyonlarının grafikleri aşağıdaki gibidir,



Normal Dağılımın Özellikleri;

a) Normal dağılım olasılık yoğunluk fonksiyonunun yani $f(x)$ in altında kalan alan 1 dir.

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1$$

b) Normal dağılım ortalamaya göre simetriktir. Yani,

$$\int_{-\infty}^{\mu} f(x)\,dx = \int_{\mu}^{+\infty} f(x) = \frac{1}{2}$$

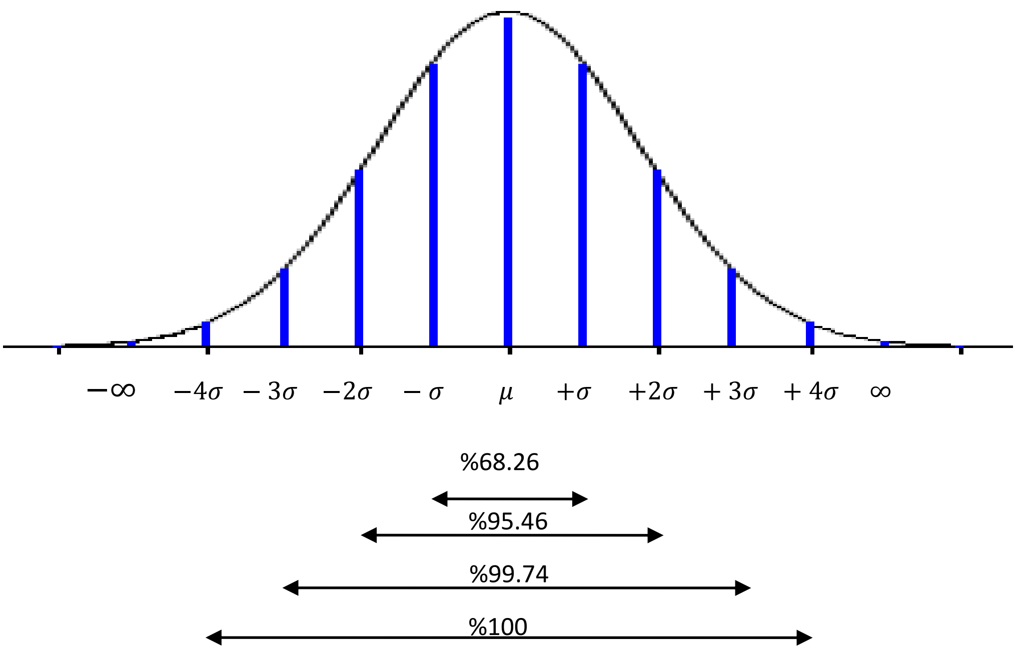
dir.

c) Normal dağılıma sahip bir rasgele değişkeninin aritmetik ortalaması, ortancası ve tepe değeri birbirlerine eşit ve μ' dür.

d) Deneklerin,

$\%68.26$' sı $\mu \pm 1\sigma$
$\%95.46$' sı $\mu \pm 2\sigma$
$\%99.74$' sı $\mu \pm 3\sigma$
$\%100$' ü $\mu \pm 4\sigma$

sınırları içinde yer alırlar.



e) Normal dağılım için dağılım fonksiyonu;

$$F(x) = \Phi(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

olarak tanımlanır. Bu fonksiyon belli bir x değerinden daha küçük olma olasılıklarını verir. Dağılım fonksiyonu kullanılarak,

$$P(X < \infty) = 1$$

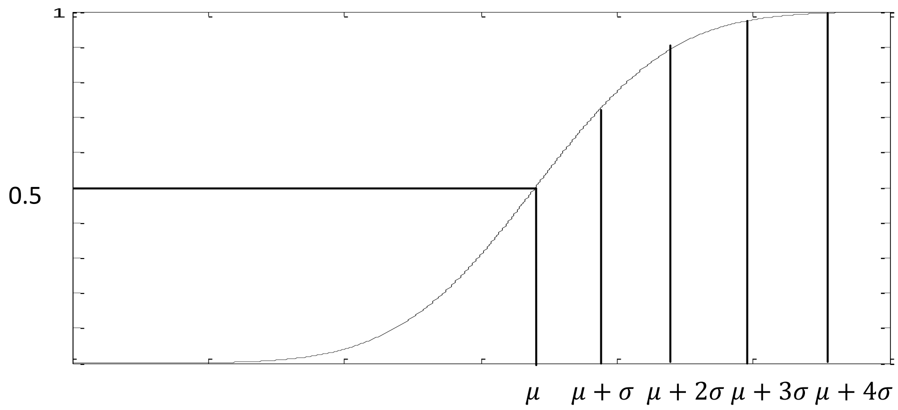
$$P(X < \mu) = 0.5$$

$$P(X < \mu + \sigma) = 0.8413$$

$$P(X < \mu + 2\sigma) = 0.9772$$

$$P(X < \mu + 3\sigma) = 0.9987$$

olasılıkları hesaplanabilir. Dağılım fonksiyonunun şekli aşağıdaki gibidir.



$\mu = 0$ ve $\sigma^2 = 1$ olan normal dağılıma **standart normal dağılım** denir. Standart normal dağılıma sahip rasgele değişken genellikle Z harfi ile gösterilir. $Z \sim N(0,1)$ rasgele değişkeninin olasılık yoğunluk fonksiyonu,

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, -\infty < z < \infty$$

dır.

$X \sim N(\mu, \sigma^2)$ normal dağılıma sahip bir rasgele değişken olmak üzere, $z = \frac{x-\mu}{\sigma}$ rasgele değişkeni standart normal dağılıma sahip bir rasgele değişken olur ve bu durumda,

$$P(X \leq x) = \phi\left(\frac{x-\mu}{\sigma}\right) = \phi(z) = P(Z \leq z)$$

dır. Buna standartlaştırma işlemi denir. Simetri özelliğinden,

$$\phi(-z) = 1 - \phi(z) = 1 - P(Z \leq z)$$

biçiminde yazılabilir. Standart normal dağılım için dağılım fonksiyonundan bulunan olasılıkları veren tablolar düzenlenmiştir. Bu tablolar kullanılarak belli bir z değerine karşılık gelen olasılık bulunabildiği gibi, belli bir olasılığa karşılık gelen z değeri de bulunabilir.

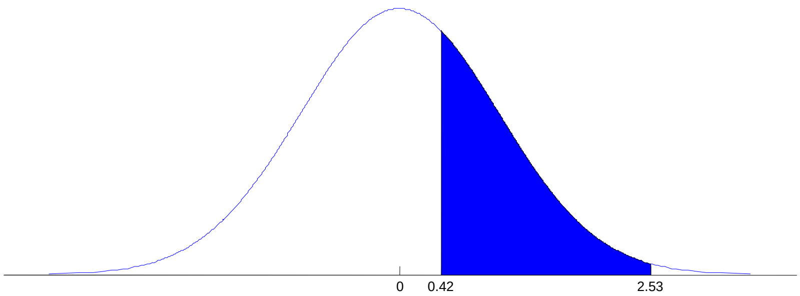
Örnek 3.23. Standart normal dağılıma sahip Z değişkeni için aşağıda istenilen olasılıkları hesaplayınız.

a) $P(0.42 < Z < 2.53)$ =?

b) $P(Z > 1)$ =?

c) $P(-1.56 < Z < 2.53)$ =?

a)

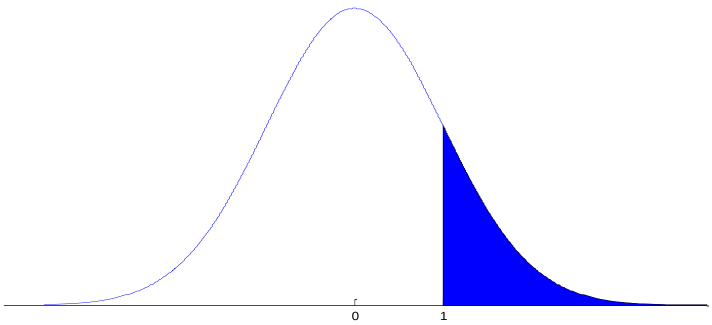


$$P(0.42 < Z < 2.53) = P(Z < 2.53) - P(Z < 0.42)$$

$$= 0.9943 - 0.6628$$

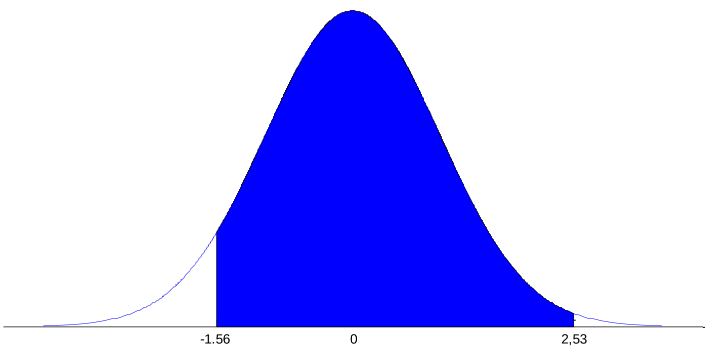
$$= 0.3315$$

b)



$$P(Z > 1) = 1 - P(Z < 1) = 1 - 0.8413 = 0.1587$$

c)



$$P(-1.56 < Z < 2.53) = P(Z < 2.53) - P(Z < -1.56)$$

$$= P(Z < 2.53) - (1 - P(Z < 1.56))$$

$$= 0.9943 - (1 - 0.9406) = 0.9349$$

Örnek 3.24 Bir hastanede belli bir hastalıkla ilgili bulunan hastaların tansiyonlarının ortalaması 15 ve varyansı 9 olan normal dağılıma sahip oldukları bilinmektedir. Bu hastalar içinden rasgele seçilen bir hastanın tansiyonunun,

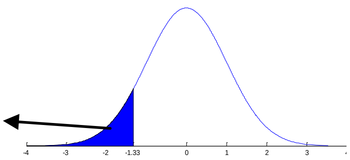
a) 11 den küçük

b) 12 den büyük

c) 9 ile 16 arasında

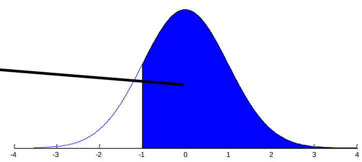
olması olasılıklarını hesaplayınız.

$X \sim N(15,9)$ olduğu biliniyor. Buna göre,



a) $P(X \leq x) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = P\left(Z \leq \frac{11-15}{3}\right) = P(Z \leq -1.33)$

$= 1 - P(Z \leq 1.33) = 1 - 0.9082 = 0.0918$



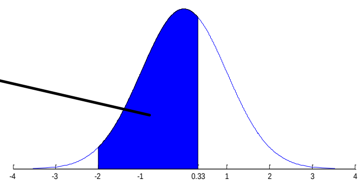
b) $P(X > x) = P\left(Z > \frac{x-\mu}{\sigma}\right) = P\left(Z > \frac{12-15}{3}\right) = P(Z > -1)$

$= 1 - P(Z < -1) = 1 - \left(1 - P(Z < 1)\right)$

$= P(Z < 1) = 0.8413$

c) $P(9 < X < 16) = P\left(\frac{9-15}{3} < Z < \frac{16-15}{3}\right)$

$= P(-2 < Z < 0.33)$



$= P(Z < 0.33) - P(Z < -2)$

$= P(Z < 0.33) - \left(1 - P(Z < 2)\right)$

$= 0.6293 - (1 - 0.9772)$

$= 0.6065$

Örnek 3.25. Bir doğumevinde doğan bebeklerin ağırlıklarının ortalamasının 3.2 kg ve standart sapmasının ise 0.3 kg olan bir normal dağılım gösterdiği bilinmektedir. Buna göre;

a) Bu doğumevinde doğan bebeklerin 3.2 kg ile 3.9 kg arasında olması olasılığı nedir?

b) Bu doğumevinde doğan bebeklerin 3.2 kg'dan daha hafif olması olasılığı nedir?

c) Bu doğumevinde doğan bebeklerin 2.8 kg ile 3.6 kg arasında olması olasılığı nedir?

d) Bir günde ortalama 200 bebeğin doğduğu kabul edilirse bu bebeklerden kaç tanesinin ağırlığı 4 kg' dan daha fazladır?

e) Bebeklerden en ağır %2.5' i hangi ağırlığın üzerindedir?

f) Bebeklerden en hafif %5' i hangi ağırlıktan daha düşüktür?

a) $P(3.2 < X < 3.9) = P\left(\frac{3.2-3.2}{0.3} < Z < \frac{3.9-3.2}{0.3}\right) = P(0 < Z < 2.33)$

$$= P(Z < 2.33) - P(Z < 0)$$

$$= 0.9901 - 0.5 = 0.4901$$

Bu doğumevinde doğan bebeklerin %49.01' inin ağırlığı 3.2 kg ile 3.9 kg arasındadır.

b) $P(X < 3.2) = P\left(Z < \dfrac{3.2-3.2}{0.3}\right) = P(Z < 0) = 0.5$

Bu doğumevinde doğan bebeklerin 3.2 kg' dan daha hafif olması olasılığı %50'dir.

c) $P(2.8 < X < 3.6) = P\left(\frac{2.8-3.2}{0.3} < Z < \frac{3.6-3.2}{0.3}\right) = P(-1.33 < Z < 1.33)$

$$= P(Z < 1.33) - P(Z < -1.33)$$

$$= P(Z < 1.33) - (1 - P(Z < 1.33))$$

$$= 2P(Z < 1.33) - 1$$

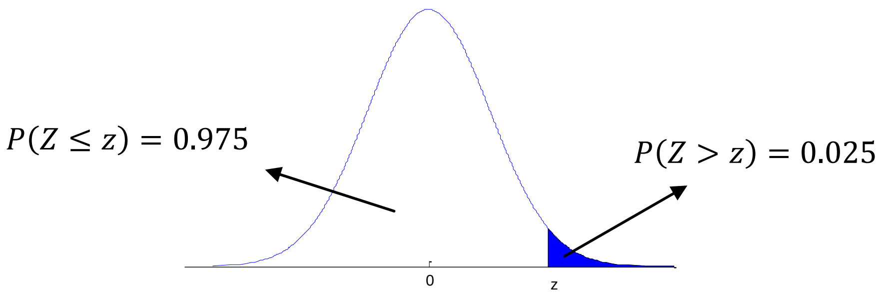
$$= 2(0.9082) - 1 = 0.8164$$

Bu doğumevinde doğan bebeklerin %81.64' ünün doğum ağırlığı 2.8 kg ile 3.6 kg arasındadır.

d) $P(X > 4) = P\left(Z > \frac{4-3.2}{0.3}\right) = P(Z > 2.67) = 1 - P(Z < 2.67) = 1 - 0.9962 = 0.0038$

Yani, bir günde doğan 200 bebeğin %0.38' inin ağırlığı 4 kg' dan daha fazladır. Bu durumda doğum ağırlığı 4 kg' dan daha fazla olan bebek sayısı 200×0.0038=0.76, yaklaşık olarak 1' dir.

e) Bu soruda standart normal dağılım tablosuna tersten bakmak gerekir. Yani, elimizde belli bir değerden büyük olması olasılığı varken bu değerin tablodan belirlenmesi gerekir. Bu değere z diyelim,



ve tablodan 0.975 olasılığına karşılık gelen nokta arandığında z=1.96 olarak bulunur.

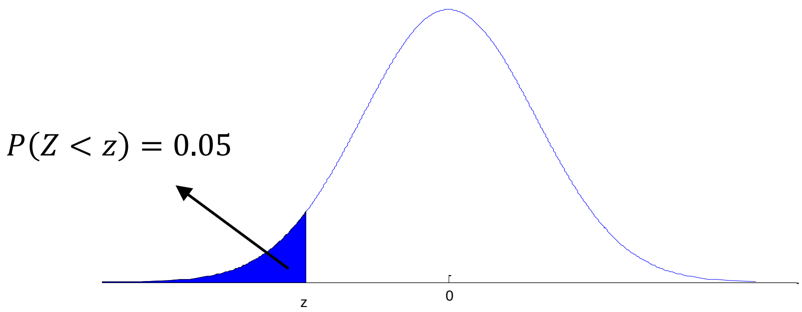
$$z = \frac{x-\mu}{\sigma} \text{ olduğundan,}$$

$$1.96 = \frac{(x-3.2)}{0.3} \text{ olup}$$

$$x = 1.96 \times 0.3 + 3.2 = 3.788$$

dır. Yani bu doğumevinde doğan bebeklerin %2.5' inin ağırlığı 3.788 kilonun üzerindedir.

f) Aynı e şıkkı gibi burada da olasılıktan noktaya geçmek gerekir.



Belli bir z değerinden küçük kalma olasılığı 0.05 olduğundan aranılan değer tablonun negatif kısmındadır. Bu değere $-z$ denilirse, simetriklik özelliğinden,

$$P(Z < -z) = 1 - P(Z < z) = 0.05$$

olup,

$$P(Z < z) = 0.95$$

dir. Tablodan 0.95 olasılığına karşılık gelen nokta arandığında en yakın 0.9495 olasılığı ile 1.64 noktası ve 0.9505 olasılığı ile 1.65 noktası bulunur. 0.95 olasılığına karşılık gelen noktayı bulmak için 1.64 ile 1.65 noktalarının ortalamasını almak gerekir. Buna göre,

$$z = \frac{1.64 + 1.65}{2} = 1.645$$

dir. Aranılan nokta $-z$ idi. Buna göre,

$$-1.645 = \frac{x - 3.2}{0.3}$$

olup,

$$x = (-1.645) \times 0.3 + 3.2 = 2.7065$$

dir. Yani, bu doğumevinde doğan bebeklerin %5' inin ağırlığı 2.7065 kilonun altındadır.

Binom Dağılımına Normal Dağılım Yaklaşımı

X rasgele değişkeni binom dağılımına sahip ise, bu değişkenini olasılık fonksiyonu,

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0,1, \dots, n$$

biçiminde idi. p çok küçük iken deney sayısı yani n arttığında binom dağılımının poisson dağılımına yaklaştığı daha önce ifade edilmişti. p çok küçük değilse ve n de büyükse binom dağılımdan yararlanarak olasılıkları hesaplamak güçleşir. $n \to \infty$ iken binom olasılıkları normal dağılıma yaklaştırılabilir. Bu dağılımın ortalaması ve varyansı, binom dağılımından

$\mu = E(X) = np$

$\sigma^2 = Var(X) = npq$

biçiminde hesaplanır. Bu durumda $X \sim N(np, npq\ )$ olur. Bu nedenle p çok küçük değilse ve n de büyükse binom dağılımına sahip X değişkeninin belli bir değerden küçük ya da büyük olması ve ya herhangi iki değer arasında yer alması olasılıkları standart normal dağılım tablosu kullanılarak hesaplanabilir. Fakat binom dağılımı kesikli ve normal dağılım sürekli olduğundan standart normale dönüştürme yapılırken düzeltme işlemi yapılması gerekir. Bu düzeltme işlemi ± 0.5 değeri kullanılarak yapılır. Olasılık hesaplarında yapılacak bu süreklilik düzeltmesi aşağıdaki tabloda verilmiştir.

İstenen Olasılık	**Süreklilik Düzeltmesi**
$P(X = x)$	$P(x - 0.5 \leq X \leq x + 0.5)$
$P(X \leq x)$	$P(X \leq x + 0.5)$
$P(X < x)$	$P(X \leq x - 0.5)$
$P(X \geq x)$	$P(X \geq x - 0.5)$
$P(X > x)$	$P(X \geq x + 0.5)$
$P(a \leq X \leq b)$	$P(a - 0.5 \leq X \leq b + 0.5)$

Bu durumda standartlaştırma işlemi, $X \sim N(np, npq\ )$ olmak üzere,

$$\frac{x \pm 0.5 - np}{\sqrt{npq}}$$

şeklinde olacaktır.

Örnek 3.26. Alkol bağımlılığı ile ilgili yapılan araştırmalarda alkolik anne babadan doğan çocuklarda alkol bağımlılığı oranının %80 olduğu saptanmıştır. 23-45 yaş grupları arasından 200 kişi seçilmiştir. Buna göre

a) $P(X \geq 150)$

b) $P(X = 178)$

c) $P(158 < X < 161)$

d) $P(X < 190)$

olasılıklarını bulunuz.

p: 0.80

$np = 200 \times 0.80 = 160$

$npq = 200 \times 0.80 \times 0.20 = 32$

Buna göre, $X \sim N(160{,}32)$

a) $P(X \geq 150)$= $P(X = 150) + P(X = 151) + \cdots + P(X = 200)$

Olasılığının çözümü binom formülü kullanılarak,

$$P(X \geq 150) = \sum_{x=150}^{200} \binom{200}{x} (0.80)^x (0.20)^{200-x}$$

eşitliğinden hesaplanabilir. Görüldüğü gibi bu işlem oldukça uzundur. Oysaki normal dağılıma yaklaşım ile standart normal dağılım tablosu kullanılarak daha az işlemle sonuca ulaşılabilir. Buna göre,

$$\begin{aligned} P(X \geq 150) &= P(X \geq 149.5) \\ &= 1 - P(X < 149.5) \\ &= 1 - P\left(Z < \frac{149.5 - 160}{\sqrt{32}}\right) \\ &= 1 - P(Z < -1.84) \\ &= 1 - \big(1 - P(z < 1.84)\big) \\ &= 1 - (1 - 0.9671) \\ &= 0.9671 \end{aligned}$$

b) $P(X = 178) = P(177.5 \leq X \leq 178.5)$

$$\begin{aligned} &= P\left(\tfrac{177.5-160}{5.7} \leq Z \leq \tfrac{178.5-160}{5.7}\right) \\ &= P(3.07 \leq Z \leq 3.25) \end{aligned}$$

$= P(Z \leq 3.25) - P(Z \leq 3.07)$

$= 0.9994 - 0.9989$

$= 0.0005$

c) $P(158 < X < 161) = P(158.5 \leq X \leq 160.5)$

$$= P(\frac{158.5 - 160}{5.7} \leq Z \leq \frac{160.5 - 160}{5.7})$$

$= P(-0.26 \leq Z \leq 0.09)$

$= P(Z \leq 0.09) - P(Z \leq -0.26)$

$= P(Z \leq 0.09) - (1 - P(Z \leq 0.26))$

$= 0.5359 - (1 - 0.6026)$

$= 0.1385$

d) $P(X < 190) = P(X \leq 189.5)$

$$= P\left(Z \leq \frac{189.5 - 160}{5.7}\right)$$

$= P(Z \leq 5.16)$

$= 1$

Örnek 3.27 Bir ilaç hastaların %60' ında etkili olmaktadır. 30 hasta bu ilaçla tedavi edildiğinde iyileşmeyen hasta sayısının 6' dan az olması olasılığı nedir?

Burada yine, binom dağılımına normal dağılım yaklaşımı kullanılacak. Buna göre,

X: İyileşmeyen hasta sayısı

p: 0.40

$np = 30 \times 0.40 = 12$

$npq = 30 \times 0.40 \times 0.60 = 7.2$

olmak üzere $X \sim N(12, 7.2)$ olacaktır.

İstenen olasılık,

$$P(X < 6) = P\left(Z < \frac{(6 - 0.5) - 12}{\sqrt{7.2}}\right) = P(Z < -2.42) = 1 - P(Z < 2.42)$$

$$= 1 - 0.9922 = 0.0078$$