Continuous Probability Distribution

Prof. Dr. Kadir KIZILKAYA*

Normal Distribution

Let X be a random variable following *Normal* distribution with mean μ and standard deviation σ and it is denoted as $X \sim N(\mu, \sigma)$. The *Normal* distribution is a type of continuous probability distribution. The *Normal* probability density function of X variable is given in Equations 1 and 2 where $\pi=3.1415927$ and $e=2.7182818$.

$$\phi(X) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X-\mu)^2}{2\sigma^2}} \tag{1}$$

or

$$\phi(X) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{X-\mu}{\sigma}\right)^2} \tag{2}$$

The shape of the *Normal* distribution of X variable in Figure 1 is symmetric about mean and is called normal curve or bell shape.
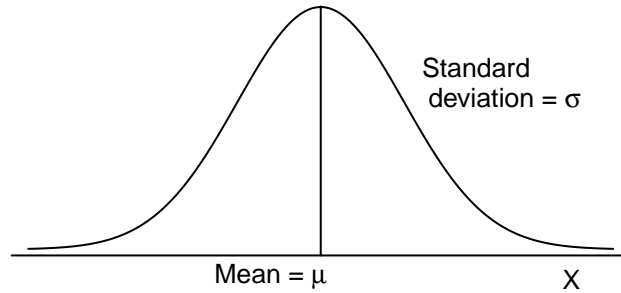

Figure 1: *Normal* distribution with mean μ and standard deviation σ

As seen in Figure 1 and in the formula of *Normal* probability density function (Equations 1 and 2), *Normal* distribution has two parameters, μ and σ which determine the shape of the distribtuion. The shape of the *Normal* distribution changes based on the parameters of μ and σ.

*Biometry and Genetics Unit, Department of Animal Science, Faculty of Agriculture, Aydın Adnan Menderes University

The mean of μ is the central tendency of the *Normal* distribution and defines the location of the peak of *Normal* distributions. As seen in Figure 1, data near the mean are more frequent in occurrence than data far from the mean. Figure 2 shows that when the mean changes, bell curve moves left or right on the X-axis.
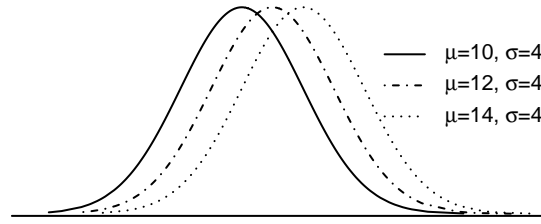

Figure 2: *Normal* distributions with different mean and same standard deviation

The standard deviation of σ is a measure of variability in data and as seen in Equation 3 it represents the distance between the observations (X_i) and the mean of μ.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}} \tag{3}$$

Larger (smaller) standard deviation indicates larger (smaller) distances between the observations (X_i) and the mean. The data points in the population with a low standart deviation tend to be very close to the mean, whereas the data in the population with a larger standart deviation is spread out over a large range of values (Figure 3 and 4).
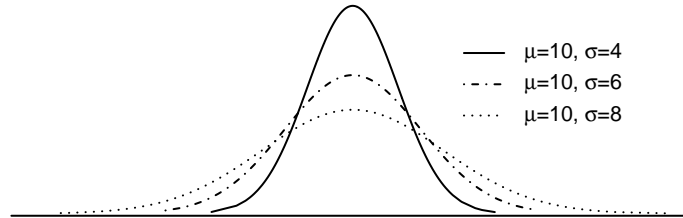

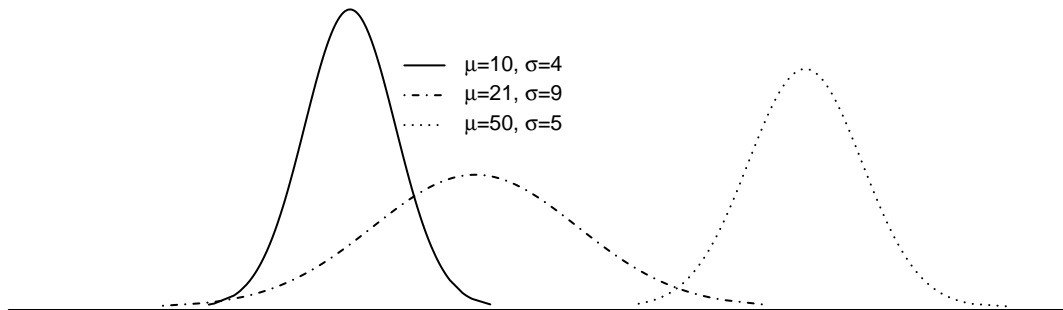Figure 3: *Normal* distributions with same mean and different standard deviation

Figure 4: *Normal* distributions with different mean and standard deviation

The figures given above indicate that the means and standard deviations of populations needs to be taken care of the comparison of populations.

Characteristics of *Normal* Distribution

Normal distribution has three characteristics.

1. Total area under *Normal* curve is 1 or 100% because of the *Normal* probability density function in Equations 1 and 2 (Figure 5).
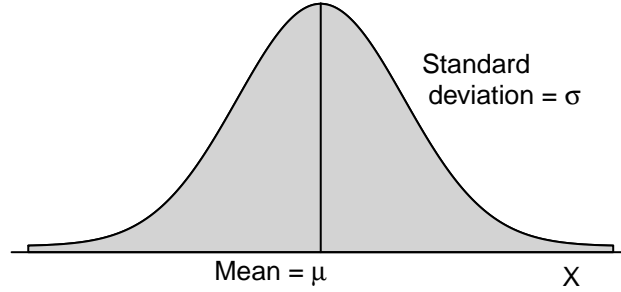

Figure 5: Total area under *Normal* curve is 1

Suppose X is a normally distributed random variable with mean of μ and standart deviation of σ and defined in the interval of a and b. The area between a and b under the *Normal* curve is calculated by evaluating the integral of *Normal* probabiliy function between a and b as below.

$$\phi(a < X < b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X-\mu)^2}{2\sigma^2}} dx = 1 \qquad (4)$$

The *Normal* cumulative function $(\Phi())$ of a random variable X is defines as

$$\phi(X < a) = \Phi(a) = \int_{-\inf}^a \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X-\mu)^2}{2\sigma^2}} dx \qquad (5)$$

and the *Normal* cumulative distribution function of variable X at a gives the probability that a random variable X takes a value less than or equal to a

2. Normal curve is symemetric about the mean of μ and the areas below and above the mean of μ are 0.5 or 50% (Figure 6).

The area below the mean of μ $\Phi(\mu) = \phi(X < \mu)$ is calculated as:

$$\phi(X < \mu) = \Phi(\mu) = \int_{-\inf}^\mu \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X-\mu)^2}{2\sigma^2}} dx \qquad (6)$$
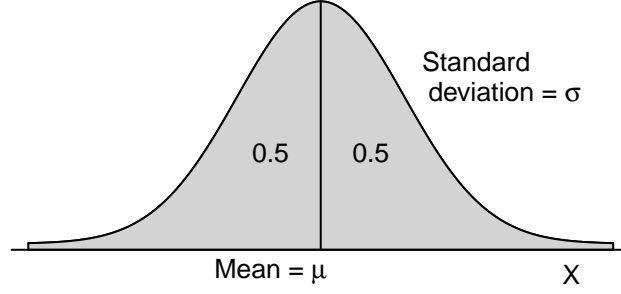
Figure 6: The areas below and above the mean of μ are equal

and the area above the mean of μ $\phi(X > \mu)$ is calculated as:

$$\phi(X > \mu) = 1 - \Phi(\mu) = 1 - \phi(X < \mu) \tag{7}$$

3. Tails of *Normal* distirbution go to infinity without touching X-axis. Therefore, the areas below and above the 3 standard deviation far from mean of μ ($\mu \pm 3\sigma$) are assumed to be zero (Figure 7).
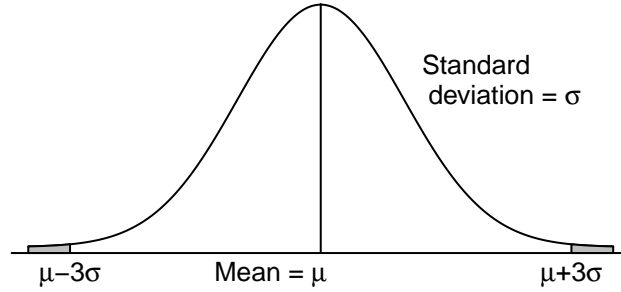
Figure 7: Tails of *Normal* distirbution goes to infinity without touching X-axis

The area below the 3 standard deviation far from mean of μ ($\mu - 3\sigma$) is calculated as:

$$\phi(X < \mu - 3\sigma) = \Phi(\mu - 3\sigma) = \int_{-\inf}^{\mu - 3\sigma} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X-\mu)^2}{2\sigma^2}} dx \cong 0 \tag{8}$$

Example

Let X be a continuous random variable that is normally distributed with a mean of μ=3.2 and a standard deviation of σ=0.3.

a. Find whole area under the *Normal* curve

$$\phi(-\inf < X < \inf) = \int_{-\inf}^{\inf} \frac{1}{\sqrt{2\pi} \times 0.3} e^{-\frac{(X-3.2)^2}{2 \times 0.3^2}} dx \tag{9}$$

4

Example

Figure 8: X normally distributed variable with a mean of $\mu = 3.2$ and a standard deviation of $\sigma = 0.3$

```
####################################################################################
## R-Code
####################################################################################
set.seed=1234
mean.x=3.2
sd.x=0.3
x = rnorm(5000, mean.x, sd.x)


f <- function(x) {1/(sqrt(2*pi)*sd.x)*exp(-(x-mean.x)^2/(2*sd.x^2))}
aa = integrate(f, lower = min(x), upper = max(x))
aa

## 0.9992926 with absolute error < 8.6e-06

cat("The shaded area under normal curve: ", aa$value,"~", round(aa$value,2))

## The shaded area under normal curve:  0.9992926 ~ 1

####################################################################################
```

> The probability that X assumes a value within the defination space is 1.0.

b. Find the probability that X assumes a value less than 3.2

$$\phi(X < 3.2) = \Phi(3.2) = \int_{-\inf}^{3.2} \frac{1}{\sqrt{2\pi} \times 0.3} e^{-\frac{(X-3.2)^2}{2 \times 0.3^2}} \, dx \tag{10}$$

```
####################################################################################
## R-Code
####################################################################################
```
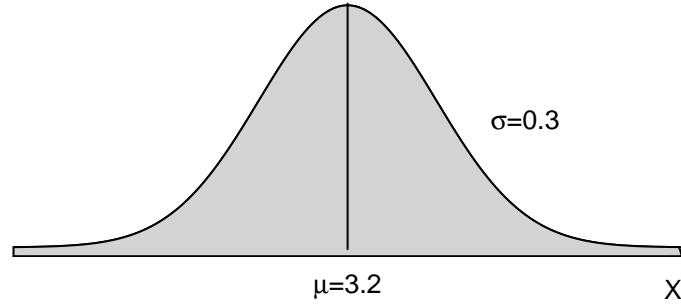
Example

Figure 9: X normally distributed variable with a mean of $\mu = 3.2$ and a standard deviation of $\sigma = 0.3$

```r
set.seed=1234
mean.x=3.2
sd.x=0.3
x = rnorm(5000, mean.x, sd.x)


a=min(x)
b=3.2


f <- function(x) {1/(sqrt(2*pi)*sd.x)*exp(-(x-mean.x)^2/(2*sd.x^2))}
aa = integrate(f, lower = min(x), upper = 3.2)
aa
```

```
## 0.4999229 with absolute error < 1.8e-13
```

```r
cat("The shaded area under normal curve: ", aa$value,"~", round(aa$value,2))
```

```
## The shaded area under normal curve:  0.4999229 ~ 0.5
```

```
################################################################################
```

> The probability that X assumes a value less than 3.2 is
>
> $$\phi(Z < 3.2) = \Phi(3.2) = 0.5$$

c. Find the probability that X assumes a value between 2.8 and 3.6

> $$\phi(2.8 < X < 3.6) = \Phi(3.6) - \Phi(2.8) = \int_{2.8}^{3.6} \frac{1}{\sqrt{2\pi} \times 0.3} e^{-\frac{(X-3.2)^2}{2 \times 0.3^2}} dx \qquad (11)$$

```
################################################################################
## R-Code
```
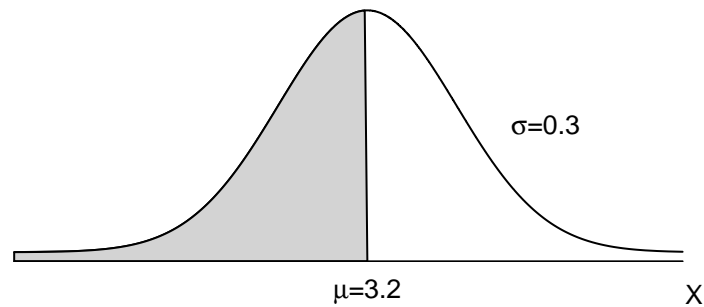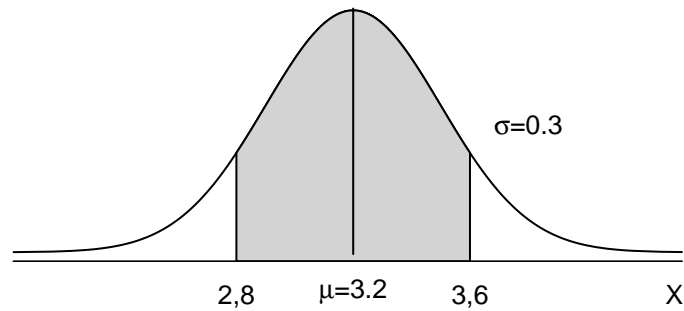
Figure 10: X normally distributed variable with a mean of $\mu = 3.2$ and a standard deviation of $\sigma=0.3$

```
#######################################################################################
set.seed=1234
mean.x=3.2
sd.x=0.3
x = rnorm(5000, mean.x, sd.x)


f <- function(x) {1/(sqrt(2*pi)*sd.x)*exp(-(x-mean.x)^2/(2*sd.x^2))}
aa = integrate(f, lower = 2.8, upper = 3.6)
aa

## 0.8175776 with absolute error < 9.1e-15

cat("The shaded area under normal curve: ", aa$value,"~", round(aa$value,2))

## The shaded area under normal curve:  0.8175776 ~ 0.82

#######################################################################################
```

> The probability that X assumes a value between 2.8 and 3.6 is
>
> $$\phi(2.8 < X < 3.6) = \Phi(3.6) - \Phi(2.8) = 0.82$$
>
> .

Standard Normal Distribution

Standard Normal distribution with mean equal to zero (0) and standard deviation equal to 1 is a special member of *Normal* distibution family. The random variable used for *Standard Normal* distribution is denoted by Z and Z is called a *standard normal random variable*. Probability density function of Z is

$$\phi(Z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad -\infty < z < \infty \tag{12}$$

and cumulative density function of z, $\Phi(z)$, is

$$\Phi(z) = \phi(Z \leq z) = \int_{-\inf}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dx \qquad -\infty < Z < \infty \tag{13}$$

Distribution of the *standard normal random variable (Z)* is in Figure 11. The tails of *Standard Normal* distribution go to minus infinity $(-\infty)$ and plus infinify (∞) and the area under the curve is equal to 1. As ssen in Figure 11, the *standard normal random variable (Z)* is ranged between -4 and 4 since the area under the curve almost equals to 1 within this range.
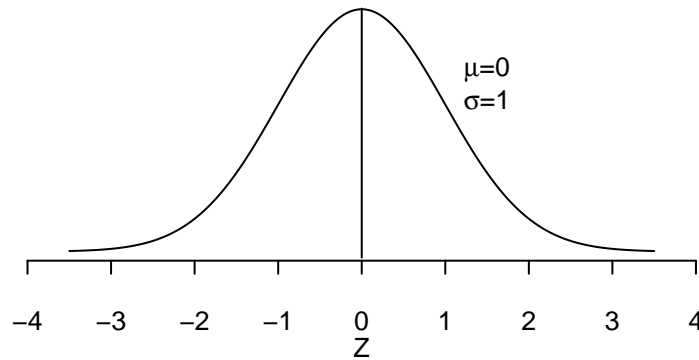
Figure 11: *Standard Normal* distribution curve

Z values *Standard Normal* distribution are called standard units or standard scores. As seen in Figure 11, the Z values on the left side of the mean (μ=0) are negative and the Z values on the right side of the mean (μ=0) are positive. Any Z value on the horizontal axis of *Standard Normal* distribution indicates the distance between the mean (μ=0) and location of Z in terms od standard deviation. For example, the locations with Z=-1 on the left and Z=1 on the right side of the mean give 1 standard deviation distance to the left and right of the mean, respectively. Also, the locations with Z=-2 and -3 on the left and Z=2 and 3 on the right side of the mean give 2 and 3 standard deviations distance to the left and right of the mean, respectively.

Example

   a. Find the area under the *Standard Normal* curve from Z=0 to Z=1.

$$\Phi(1) - \Phi(0) = \phi(0 < Z < 1) = \int_{0}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dx \tag{14}$$
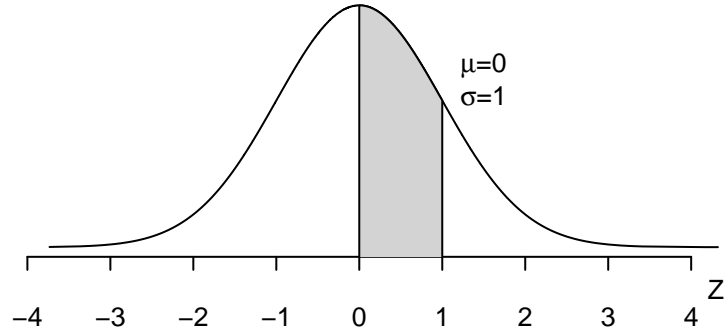
Figure 12: Shaded area between $Z=0$ and $Z=1$ under *Standard Normal* curve

```
################################################################################
## R-Code
################################################################################
set.seed=1234
mean.z=0
sd.z=1
z = rnorm(5000, mean.z, sd.z)


f <- function(z) {1/(sqrt(2*pi))*exp(-z^2/2)}
aa = integrate(f, lower = 0, upper = 1.0)
aa

## 0.3413447 with absolute error < 3.8e-15

cat("The shaded area under Standard Normal curve: ", aa$value,"~", round(aa$value,4),"tür.")

## The shaded area under Standard Normal curve:  0.3413447 ~ 0.3413 tür.

################################################################################
```

> The shaded area gives the probability that Z assumes a value between $Z=0$ and $Z=1$ is
>
> $$\phi(0 < Z < 1) = \Phi(1) - \Phi(0) = 0.3413$$
>
> or gives that 34.13% of Z values are between $Z=0$ and $Z=1$.

b. Find the area under the *Standard Normal* curve from $Z=-1$ to $Z=0$.

$$\Phi(0) - \Phi(-1) = \phi(-1 < Z < 0) = \int_{-1}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, dx \tag{15}$$

Example

Figure 13: Shaded area between Z=-1 and Z=0 under *Standard Normal* curve

```
###############################################################################
## R-Code
###############################################################################
set.seed=1234
mean.z=0
sd.z=1
z = rnorm(5000, mean.z, sd.z)


f <- function(z) {1/(sqrt(2*pi))*exp(-z^2/2)}
aa = integrate(f, lower = -1, upper = 0)
aa
```

0.3413447 with absolute error < 3.8e-15

```
cat("The shaded area under Standard Normal curve: ", aa$value,"~", round(aa$value,4),"tür.")
```

The shaded area under Standard Normal curve: 0.3413447 ~ 0.3413 tür.

```
###############################################################################
```

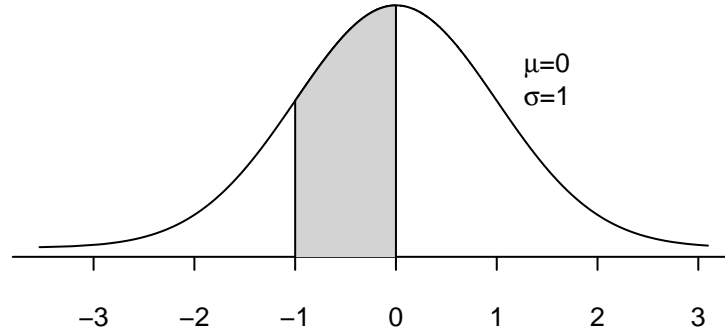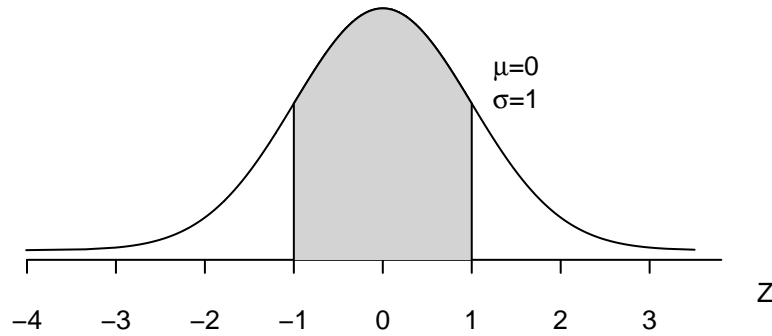The shaded area gives the probability that Z assumes a value between Z=-1 and Z=0 is

$$\phi(-1 < Z < 0) = \Phi(0) - \Phi(-1) = 0.3413$$

or gives that 34.13% of Z values are between Z=-1 and Z=0.

c. What is the total area within one standard deviation on the mean?

$$\Phi(1) - \Phi(-1) = \phi(-1 < Z < 1) = \int_{-1}^{1} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dx \qquad (16)$$

10

Figure 14: Shaded area between Z=-1 and Z=1 under *Standard Normal* curve

```
#################################################################################
## R-Code
#################################################################################
set.seed=1234
mean.z=0
sd.z=1
z = rnorm(5000, mean.z, sd.z)


f <- function(z) {1/(sqrt(2*pi))*exp(-z^2/2)}
aa = integrate(f, lower = -1, upper = 1)
aa
```

```
## 0.6826895 with absolute error < 7.6e-15
```

```
cat("The shaded area under Standard Normal curve: ", aa$value,"~", round(aa$value,4),"tür.")
```

```
## The shaded area under Standard Normal curve:  0.6826895 ~ 0.6827 tür.
```

```
#################################################################################
```

> This area is given by the sum of the areas between Z=-1 and Z=0 and Z=0 and Z=1. As shown in (a) and (b), each of these two areas is 0.3413. Consequently, the total area between Z=-1 and Z=1 is 0.6826 or 68.26%.
>
> The shaded area gives the probability that Z assumes a value between Z=-1 and Z=1 is
>
> $$\phi(-1 < Z < 1) = \Phi(1) - \Phi(-1) = 0.6826$$
>
> or gives that 68.26% of Z values (observations) are between Z=-1 and Z=1.

d. What is the total area within two standard deviations on the mean?

Example

$$\Phi(2) - \Phi(-2) = \phi(-2 < Z < 2) = \int_{-2}^{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, dx \qquad (17)$$
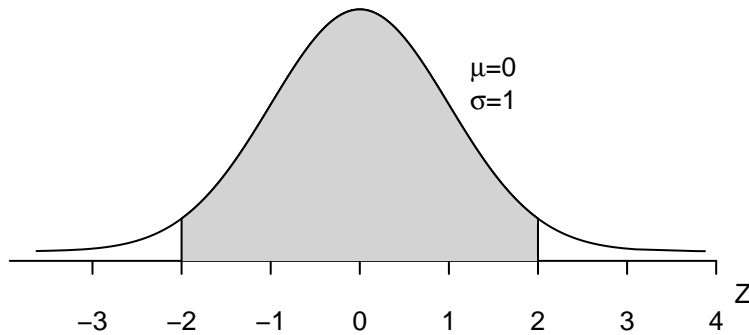
Figure 15: Shaded area between Z=-2 and Z=2 under *Standard Normal* curve

```
###########################################################################
## R-Code
###########################################################################
set.seed=1234
mean.z=0
sd.z=1
z = rnorm(5000, mean.z, sd.z)


f <- function(z) {1/(sqrt(2*pi))*exp(-z^2/2)}
aa = integrate(f, lower = -2, upper = 2)
aa
```

0.9544997 with absolute error < 1.8e-11

```
cat("The shaded area under Standard Normal curve: ", aa$value,"~", round(aa$value,4),"tür.")
```

The shaded area under Standard Normal curve: 0.9544997 ~ 0.9545 tür.

```
###########################################################################
```

The shaded area gives the probability that Z assumes a value between Z=-2 and Z=2 is

$$\phi(-2 < Z < 2) = \Phi(2) - \Phi(-2) = 0.9545$$

or gives that 95.45% of Z values (observations) are between Z=-2 and Z=2.

e. What is the total area within three standard deviations on the mean?

$$\Phi(3) - \Phi(-3) = \phi(-3 < Z < 3) = \int_{-3}^{3} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, dx \tag{18}$$
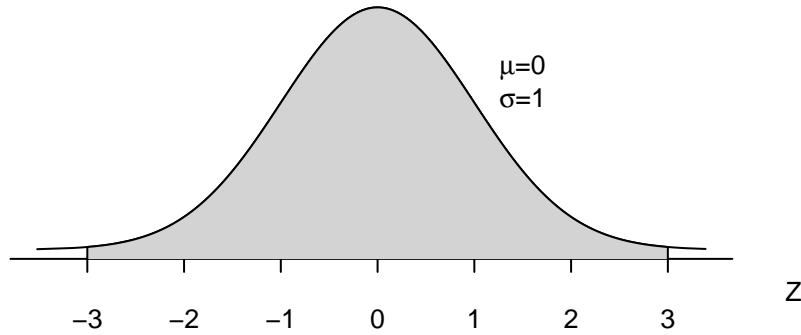
Figure 16: Shaded area between Z=-3 and Z=3 under *Standard Normal* curve

```
####################################################################################
## R-Code
####################################################################################
set.seed=1234
mean.z=0
sd.z=1
z = rnorm(5000, mean.z, sd.z)


f <- function(z) {1/(sqrt(2*pi))*exp(-z^2/2)}
aa = integrate(f, lower = -3, upper = 3)
aa
```

0.9973002 with absolute error < 9.3e-07

```
cat("The shaded area under Standard Normal curve: ", aa$value,"~", round(aa$value,4),"tür.")
```

The shaded area under Standard Normal curve: 0.9973002 ~ 0.9973 tür.

```
####################################################################################
```

The shaded area gives the probability that Z assumes a value between Z=-3 and Z=3 is

$$\phi(-3 < Z < 3) = \Phi(3) - \Phi(-3) = 0.9973$$

or gives that 99.73% of Z values (observations) are between Z=-3 and Z=3.
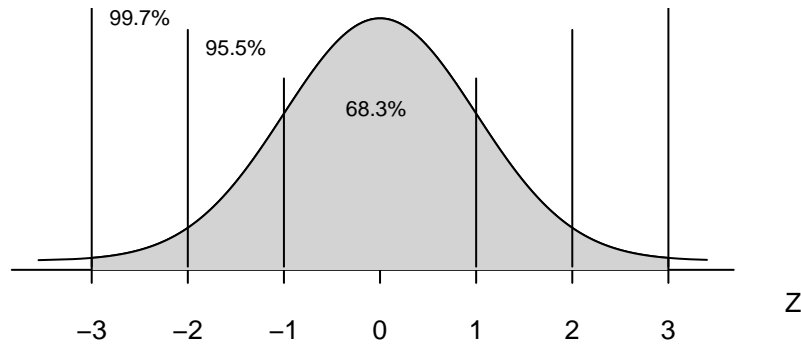
Figure 17: Percentage of observations based on standard deviation

For a (approximately) *Normally* distributed random variable,

1. Roughly 68.3% of the values (observations) are within ± 1 standard deviation of the mean.

2. Roughly 95.5% of the values (observations) are within ± 2 standard deviations of the mean.

3. Roughly 99.7% of the values (observations) are within ± 3 standard deviations of the mean.

Z-Table

Z-Table includes the probability that Z assumes a value less than Z=z, $\Phi(z) = \phi(Z < z)$. This probability can be found by calculating the defined area under *Standard Normal* curve based on the integral of the probability density function of Z random variable:

$$\Phi(z) = \phi(Z < z) = \int_{-3.5}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dx \tag{19}$$
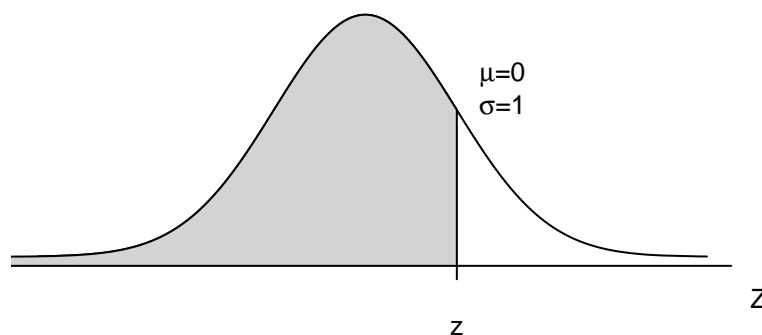
Figure 18: Shaded area less than $Z = z$ under *Standard Normal* curve

```
################################################################################
## R-Code
################################################################################
```

```r
Ztable = matrix(0,71,11)
n=0
#rownames(Ztable)=round(seq(-3.5, 3.5, by=0.1), 2)
Ztable[,1] = round(seq(-3.5, 3.5, by=0.1), 2)
for(i in seq(-3.5, 3.5, by=0.1)){
  n=n+1
  m=1
  for(j in seq(0, 0.09, by=0.01)){
    m=m+1
    z=i+j
    f=function(z) {1/(sqrt(2*pi))*exp(-z^2/2)}
    result=integrate(f, lower=-3.5, upper=z)
    Ztable[n,m]=round(result$value,3)
  }
}
library(kableExtra)
knitr::kable(
  Ztable[,1:11], longtable=T,
  caption="Standat Normal Dağılım Tablosu $\\Phi(z)=\\phi(Z < z)$",
  col.names=c("z","0.00","0.01","0.02","0.03","0.04","0.05","0.06","0.07","0.08","0.09"),
  align=c("c"),
  booktabs=T, linesep="") %>%
  kable_styling(bootstrap_options="striped", font_size=9)
```

Table 1: Standat Normal Dağılım Tablosu $\Phi(z) = \phi(Z < z)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.5	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-3.4	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-3.3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-3.2	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
-3.1	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
-3.0	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.002
-2.9	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
-2.8	0.002	0.002	0.002	0.003	0.003	0.003	0.003	0.003	0.003	0.003
-2.7	0.003	0.003	0.003	0.004	0.004	0.004	0.004	0.004	0.004	0.004
-2.6	0.004	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.006	0.006

-2.5	0.006	0.006	0.006	0.007	0.007	0.007	0.007	0.007	0.008	0.008
-2.4	0.008	0.008	0.008	0.009	0.009	0.009	0.009	0.010	0.010	0.010
-2.3	0.010	0.011	0.011	0.011	0.012	0.012	0.012	0.013	0.013	0.013
-2.2	0.014	0.014	0.014	0.015	0.015	0.016	0.016	0.016	0.017	0.017
-2.1	0.018	0.018	0.019	0.019	0.019	0.020	0.020	0.021	0.021	0.022
-2.0	0.023	0.023	0.024	0.024	0.025	0.025	0.026	0.027	0.027	0.028
-1.9	0.028	0.029	0.030	0.031	0.031	0.032	0.033	0.033	0.034	0.035
-1.8	0.036	0.036	0.037	0.038	0.039	0.040	0.041	0.042	0.042	0.043
-1.7	0.044	0.045	0.046	0.047	0.048	0.049	0.050	0.051	0.052	0.053
-1.6	0.055	0.056	0.057	0.058	0.059	0.060	0.062	0.063	0.064	0.065
-1.5	0.067	0.068	0.069	0.071	0.072	0.073	0.075	0.076	0.078	0.079
-1.4	0.081	0.082	0.084	0.085	0.087	0.088	0.090	0.092	0.093	0.095
-1.3	0.097	0.098	0.100	0.102	0.104	0.105	0.107	0.109	0.111	0.113
-1.2	0.115	0.117	0.119	0.121	0.123	0.125	0.127	0.129	0.131	0.133
-1.1	0.135	0.138	0.140	0.142	0.144	0.147	0.149	0.151	0.154	0.156
-1.0	0.158	0.161	0.163	0.166	0.168	0.171	0.173	0.176	0.179	0.181
-0.9	0.184	0.187	0.189	0.192	0.195	0.197	0.200	0.203	0.206	0.209
-0.8	0.212	0.215	0.217	0.220	0.223	0.226	0.229	0.232	0.236	0.239
-0.7	0.242	0.245	0.248	0.251	0.254	0.258	0.261	0.264	0.267	0.271
-0.6	0.274	0.277	0.281	0.284	0.288	0.291	0.294	0.298	0.301	0.305
-0.5	0.308	0.312	0.315	0.319	0.323	0.326	0.330	0.333	0.337	0.341
-0.4	0.344	0.348	0.352	0.355	0.359	0.363	0.367	0.370	0.374	0.378
-0.3	0.382	0.386	0.390	0.393	0.397	0.401	0.405	0.409	0.413	0.417
-0.2	0.421	0.424	0.428	0.432	0.436	0.440	0.444	0.448	0.452	0.456
-0.1	0.460	0.464	0.468	0.472	0.476	0.480	0.484	0.488	0.492	0.496
0.0	0.500	0.504	0.508	0.512	0.516	0.520	0.524	0.528	0.532	0.536
0.1	0.540	0.544	0.548	0.551	0.555	0.559	0.563	0.567	0.571	0.575
0.2	0.579	0.583	0.587	0.591	0.595	0.598	0.602	0.606	0.610	0.614
0.3	0.618	0.621	0.625	0.629	0.633	0.637	0.640	0.644	0.648	0.651
0.4	0.655	0.659	0.663	0.666	0.670	0.673	0.677	0.681	0.684	0.688
0.5	0.691	0.695	0.698	0.702	0.705	0.709	0.712	0.715	0.719	0.722
0.6	0.726	0.729	0.732	0.735	0.739	0.742	0.745	0.748	0.752	0.755
0.7	0.758	0.761	0.764	0.767	0.770	0.773	0.776	0.779	0.782	0.785
0.8	0.788	0.791	0.794	0.796	0.799	0.802	0.805	0.808	0.810	0.813
0.9	0.816	0.818	0.821	0.824	0.826	0.829	0.831	0.834	0.836	0.839
1.0	0.841	0.844	0.846	0.848	0.851	0.853	0.855	0.857	0.860	0.862
1.1	0.864	0.866	0.868	0.871	0.873	0.875	0.877	0.879	0.881	0.883
1.2	0.885	0.887	0.889	0.890	0.892	0.894	0.896	0.898	0.899	0.901
1.3	0.903	0.905	0.906	0.908	0.910	0.911	0.913	0.914	0.916	0.918

1.4	0.919	0.920	0.922	0.923	0.925	0.926	0.928	0.929	0.930	0.932
1.5	0.933	0.934	0.936	0.937	0.938	0.939	0.940	0.942	0.943	0.944
1.6	0.945	0.946	0.947	0.948	0.949	0.950	0.951	0.952	0.953	0.954
1.7	0.955	0.956	0.957	0.958	0.959	0.960	0.961	0.961	0.962	0.963
1.8	0.964	0.965	0.965	0.966	0.967	0.968	0.968	0.969	0.970	0.970
1.9	0.971	0.972	0.972	0.973	0.974	0.974	0.975	0.975	0.976	0.976
2.0	0.977	0.978	0.978	0.979	0.979	0.980	0.980	0.981	0.981	0.981
2.1	0.982	0.982	0.983	0.983	0.984	0.984	0.984	0.985	0.985	0.986
2.2	0.986	0.986	0.987	0.987	0.987	0.988	0.988	0.988	0.988	0.989
2.3	0.989	0.989	0.990	0.990	0.990	0.990	0.991	0.991	0.991	0.991
2.4	0.992	0.992	0.992	0.992	0.992	0.993	0.993	0.993	0.993	0.993
2.5	0.994	0.994	0.994	0.994	0.994	0.994	0.995	0.995	0.995	0.995
2.6	0.995	0.995	0.995	0.995	0.996	0.996	0.996	0.996	0.996	0.996
2.7	0.996	0.996	0.997	0.997	0.997	0.997	0.997	0.997	0.997	0.997
2.8	0.997	0.997	0.997	0.997	0.998	0.998	0.998	0.998	0.998	0.998
2.9	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998	0.998
3.0	0.998	0.998	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
3.1	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
3.2	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
3.3	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999	0.999
3.4	0.999	0.999	0.999	0.999	0.999	0.999	0.999	1.000	1.000	1.000
3.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Z-Transformation

A *Normal* random variable $X \sim N(\mu, \sigma)$ with mean of μ and standard deviation of σ is transformed to the *Standard Normal* random variable (Z) by using the formula below:

$$Z = \frac{X - \mu}{\sigma} \tag{20}$$

Z-transformation for *Normal* X random variable is a one-to-one transformation and results in at most $z \in Z$ for every $x \in X$ random variable. Therefore, any area defined by $\phi(a < X < b)$ under the *Normal* curve of X random variable is calculated easily by using *Standard Normal* distribuiton after **Z-transformation** of *Normal* X random variable.

Let X be a continuous random variable that is normally distributed with a mean of μ=3.2 and a standard deviation of σ=0.3.

The area between 2.9 and 3.5 under the *Normal* curve of X random variable is calculated by using **Z-**

transformation and **Z-table** as below:

$$\phi(2.9 < X < 3.5) = \phi(\frac{2.9 - 3.2}{0.3} < \frac{X - \mu}{\sigma} < \frac{3.5 - 3.2}{0.3}) = \phi(-1.0 < Z < 1.0) \tag{21}$$

then $\Phi(Z < 1.00) = 0.841$ and $\Phi(Z < -1.00) = 0.158$ are from **Z-table**

$$\phi(2.9 < X < 3.5) = \phi(-1.00 < Z < 1.00) = \Phi(Z < 1.00) - \Phi(Z < -1.00) = 0.841 - 0.158 = 0.683. \tag{22}$$
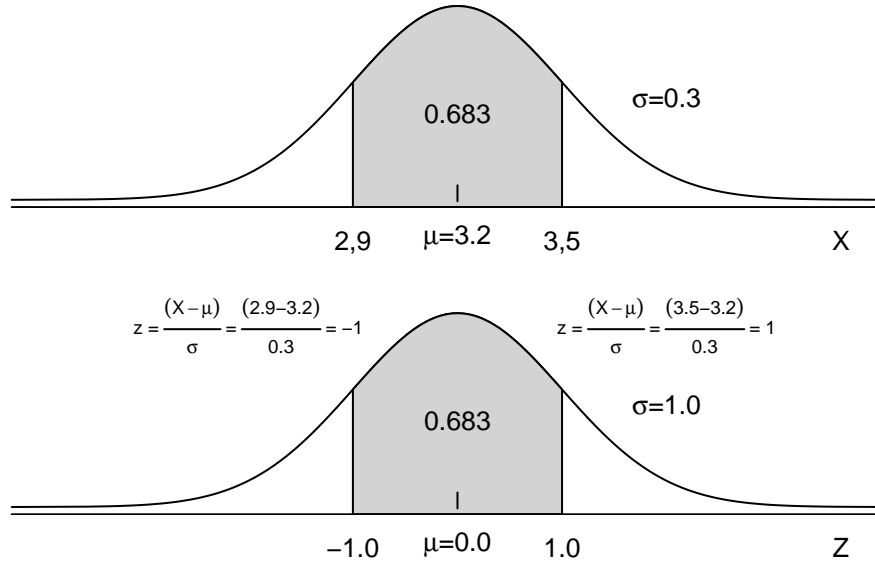

Figure 19: Areas of $\phi(2.9 < X < 3.5)$ and $\phi(-1.00 < Z < 1.00)$ under *Normal* and *Standard Normal* curves

The area greater than 3.788 under the *Normal* curve of X random variable is calculated by using **Z-transformation** and **Z-table** as below:

$$\phi(X > 3.788) = \phi(\frac{X - \mu}{\sigma} > \frac{3.788 - 3.2}{0.3}) = \phi(Z > 1.96) \tag{23}$$

then $\Phi(Z < 1.96) = 0.975$ is from **Z-table**

$$\phi(X > 3.788) = \phi(Z > 1.96) = 1 - \Phi(Z < 1.96) = 1 - 0.975 = 0.025. \tag{24}$$

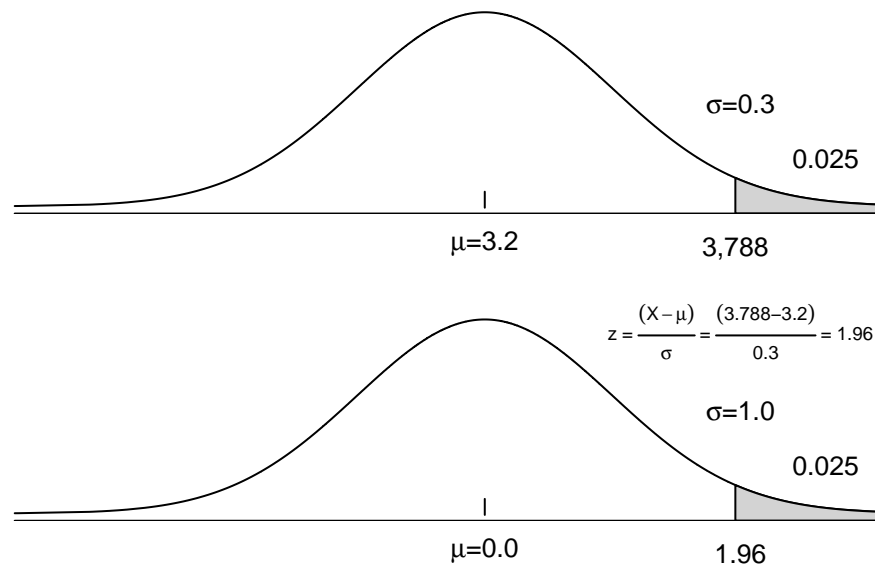$$z = \frac{(X-\mu)}{\sigma} = \frac{(3.788-3.2)}{0.3} = 1.96$$

Figure 20: Areas of $\phi(X > 3.788)$ and $\phi(Z > 1.96)$ under *Normal* and *Standard Normal* curves

Probability Distributions in R

Probability distributions in statistics are important to model random phenomena (events), to calculate the probability of a random event and to test the hypothesis. There are two types of probability distributions:

- Discrete probability distributions for discrete variables
- Continuous probability distributions for continuous variables

R-project includes probability mass/density functions for discrete/probability distributions in statistics. Probability mass/density functions in R-project are listed below:

```r
#################################################################################
## R-Code
#################################################################################
library(kableExtra)
DistName = read.table("DistNames.txt", header=F)
knitr::kable(
  DistName[,1:5], longtable=T,
  caption="Probability Distributions in R-Project",
  col.names=c("Distribution Name","R Root Name","Parameter-1","Parameter-2","Parameter-3"),
  align=c("l"),
  booktabs=T, linesep="") %>%
  kable_styling(bootstrap_options="striped", font_size=9)
```

Table 2: Probability Distributions in R-Project

Distribution Name	R Root Name	Parameter-1	Parameter-2	Parameter-3
beta	beta	shape1	shape2	.
binomial	binom	size	prob	.
Cauchy	cauchy	location	scale	.
chi-squared	chisq	df	.	.
exponential	exp	rate	.	.
F	f	df1	df2	.
gamma	gamma	shape	rate	.
geometric	geom	p	.	.
hypergeometric	hyper	m	n	k
log-normal	lnorm	meanlog	sdlog	.
logistic	logis	location	scale	.
negative-binomial	nbinom	size	prob	.
normal	norm	mean	sd	.
Poisson	pois	lambda	.	.
Student-t	t	df	.	.
uniform	unif	min	max	.
Weibull	weibull	shape	scale	.

There are four functions for every distribution in R-Project which are defined by the letters below:

- p : **p**robability, (the cumulative distribution function)
- q : **q**uantile, (the inverse of the cumulative distribution function)
- d : **d**ensity, (the density function)
- r : **r**andom, (a random variable)

These functions for each distribution are carried out by using the letters as a prefix for R root name of distributions.

Normal Distribution

The use of these functions for **Normal** distribution is shown as follows:

```
dnorm(x, mean, sd)
pnorm(x, mean, sd)
qnorm(p, mean, sd)
rnorm(n, mean, sd)
```

where

- `mean` is the mean of *Normal* distribution (default `mean=0`),

- `sd` is the standard deviation of *Normal* distribution (default `sd=1`),

- `x` is a number or a vector from the defined *Normal* distribution,

- `p` is a value of probability or a vector of probability and

- `n` is the number of observations generated from the defined *Normal* distribution

Let X be a continuous random variable that is normally distributed with a mean of $\mu=3.2$ and a standard deviation of $\sigma=0.3$.

a. Find the area between 2.9 and 3.5 under the *Normal* curve of X random variable:

$$\phi(2.9 < X < 3.5) = \Phi(X < 3.5) - \Phi(X < 2.9) = 0.8413447 \text{ - } 0.1586553 = 0.6826895$$

```
################################################################################
## R-Code
################################################################################
 pnorm(3.5, mean=3.2, sd=0.3)
```

```
## [1] 0.8413447
```

```
 pnorm(2.9, mean=3.2, sd=0.3)
```

```
## [1] 0.1586553
```

```
 pnorm(3.5, mean=3.2, sd=0.3) - pnorm(2.9, mean=3.2, sd=0.3)
```

```
## [1] 0.6826895
```

```
################################################################################
```


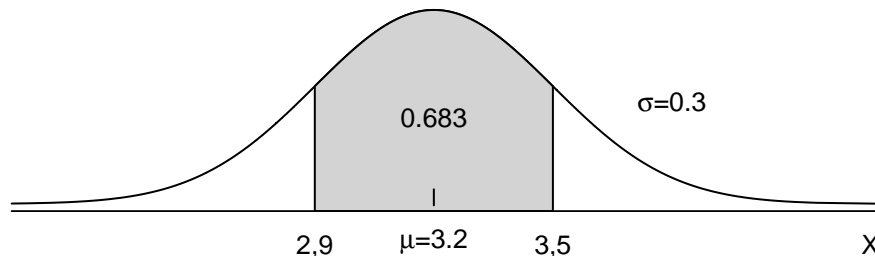Figure 21: Area of $\phi(2.9 < X < 3.5)$ under *Normal* curve

$\Phi(X < 3.5)=0.8413447$ is from 'pnorm(3.5, mean=3.2, sd=0.3)'

```
#####################################################################################
## R-Code
#####################################################################################
 pnorm(3.5, mean=3.2, sd=0.3)
```

[1] 0.8413447

```
#####################################################################################
```
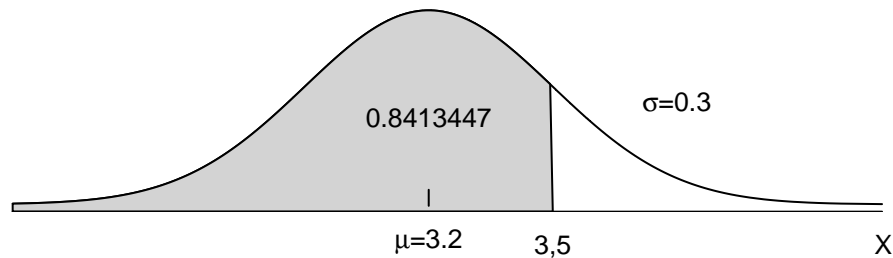

Figure 22: Area of $\phi(X < 3.5)$ under *Normal* curve

and

$\Phi(X < 2.9)$=0.1586553 is from 'pnorm(2.9, mean=3.2, sd=0.3)'

```
#####################################################################################
## R-Code
#####################################################################################
 pnorm(2.9, mean=3.2, sd=0.3)
```

[1] 0.1586553

```
#####################################################################################
```


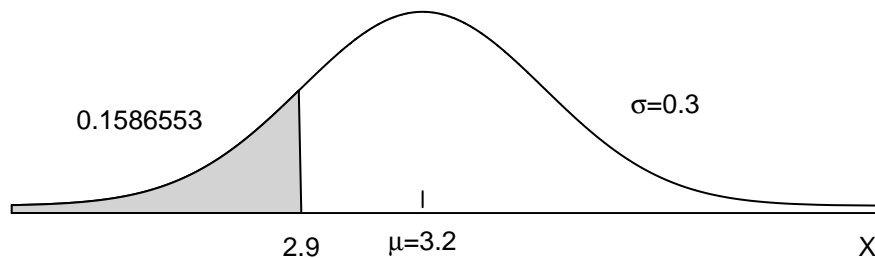Figure 23: Area of $\phi(X < 2.9)$ under *Normal* curve

b. Generate 10 random number form $X \sim (\mu = 3.2, \sigma = 0.3)$

```
###################################################################################
## R-Code
###################################################################################
 rnorm(n=10, mean=3.2, sd=0.3)
```

```
##  [1] 3.067112 3.348455 3.082890 3.301901 3.330237 3.494076 3.187586 3.333402
##  [9] 3.132174 3.678173
```

```
###################################################################################
```

b. Generate 10000 random number form $X \sim (\mu = 3.2, \sigma = 0.3)$ and find minimum, maximum, mean, median, variance, standard deviation and coefficient of variation of X variable.

```
###################################################################################
## R-Code
###################################################################################
X = rnorm(n=10000, mean=3.2, sd=0.3)
cat("Minimum value of X=", min(X),"\n")
```

```
## Minimum value of X= 2.068803
```

```
cat("Maximum value of X=", max(X),"\n")
```

```
## Maximum value of X= 4.505925
```

```
cat("Mean of X=", mean(X),"\n")
```

```
## Mean of X= 3.201811
```

```
cat("Median of X=", median(X),"\n")
```

```
## Median of X= 3.202036
```

```
cat("Variance of X=", var(X),"\n")
```

```
## Variance of X= 0.09186935
```

```
cat("Standard deviation of X=", sd(X),"\n")
```
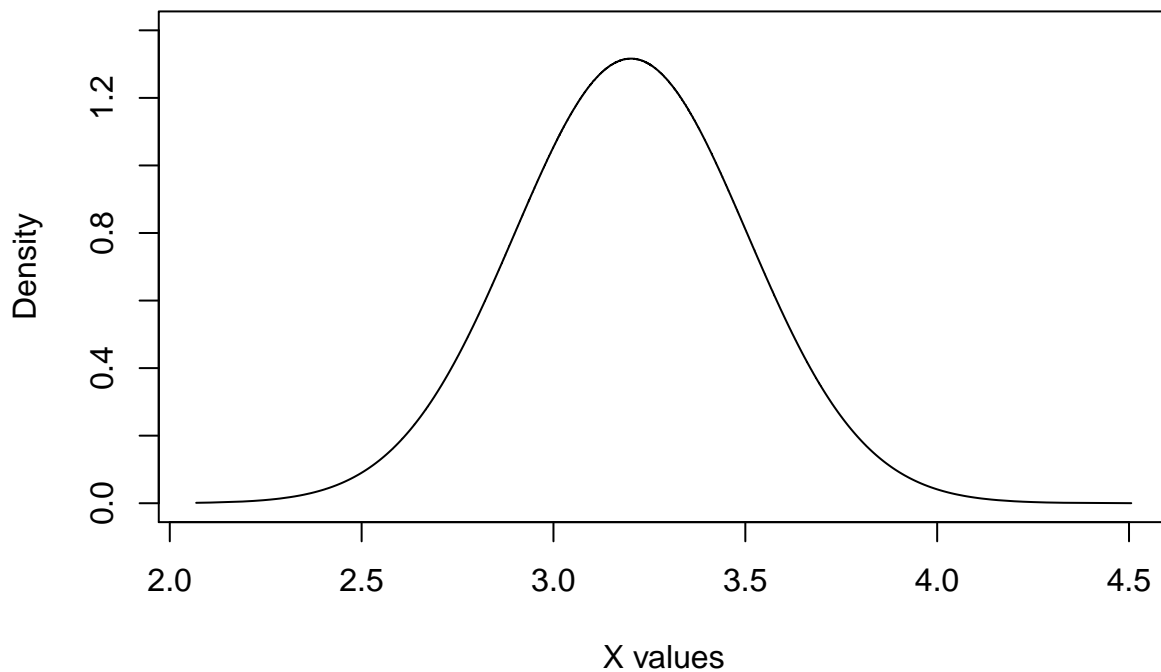
```
## Standard deviation of X= 0.3030996
```

```
cat("Coefficient of variation of X=", sd(X)/mean(X)*100,"%","\n")
```

```
## Coefficient of variation of X= 9.466505 %
```

```
plot(sort(X), dnorm(sort(X),mean(X), sd(X)),
     main="Distribution of 10000 observations from Normal distribution", ylim=c(0, 1.4),
     xlab="X values", ylab="Density",
     axes=T, type="l", lty=1, lwd=1, pch=19, cex=0.7)
```

Distribution of 10000 observations from Normal distribution


```
dnorm(3.2, mean=3.2, sd=0.3) # density of X value
```

```
## [1] 1.329808
```

```
qnorm(0.5, mean=3.2, sd=0.3) # the X value providing the probability of 0.5
```

```
## [1] 3.2
```

```
############################################################################
```