Continuous probability distributions
Normal Distribution

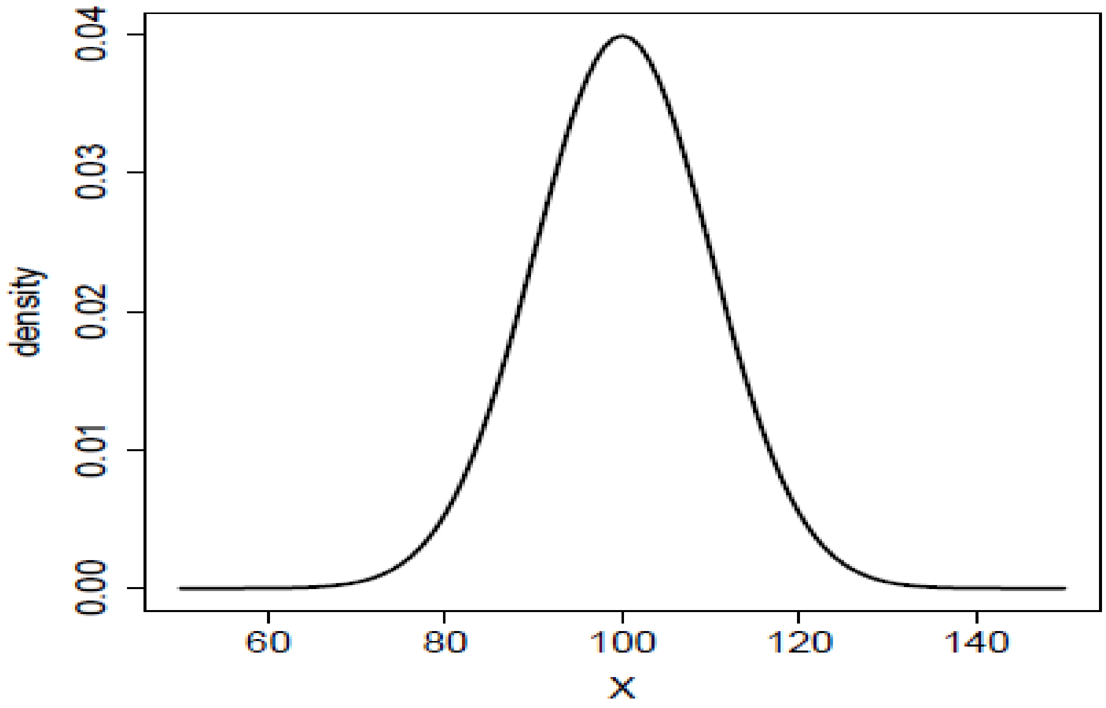
When we considered the Binomial and Poisson distributions. we saw that the probability distributions were characterized by a formula for the probability of each possible discrete value.

All of the probabilities together sum up to 1.

For continuous data we don't have equally spaced discrete values so instead we use a curve or function that describes the probability *density* over the range of the distribution.

The curve is chosen so that the area under the curve is equal to 1.

A continuous probability distribution



The Normal Distribution

There will be many, possible probability density functions over a continuous range of values.

The Normal distribution describes a special class of

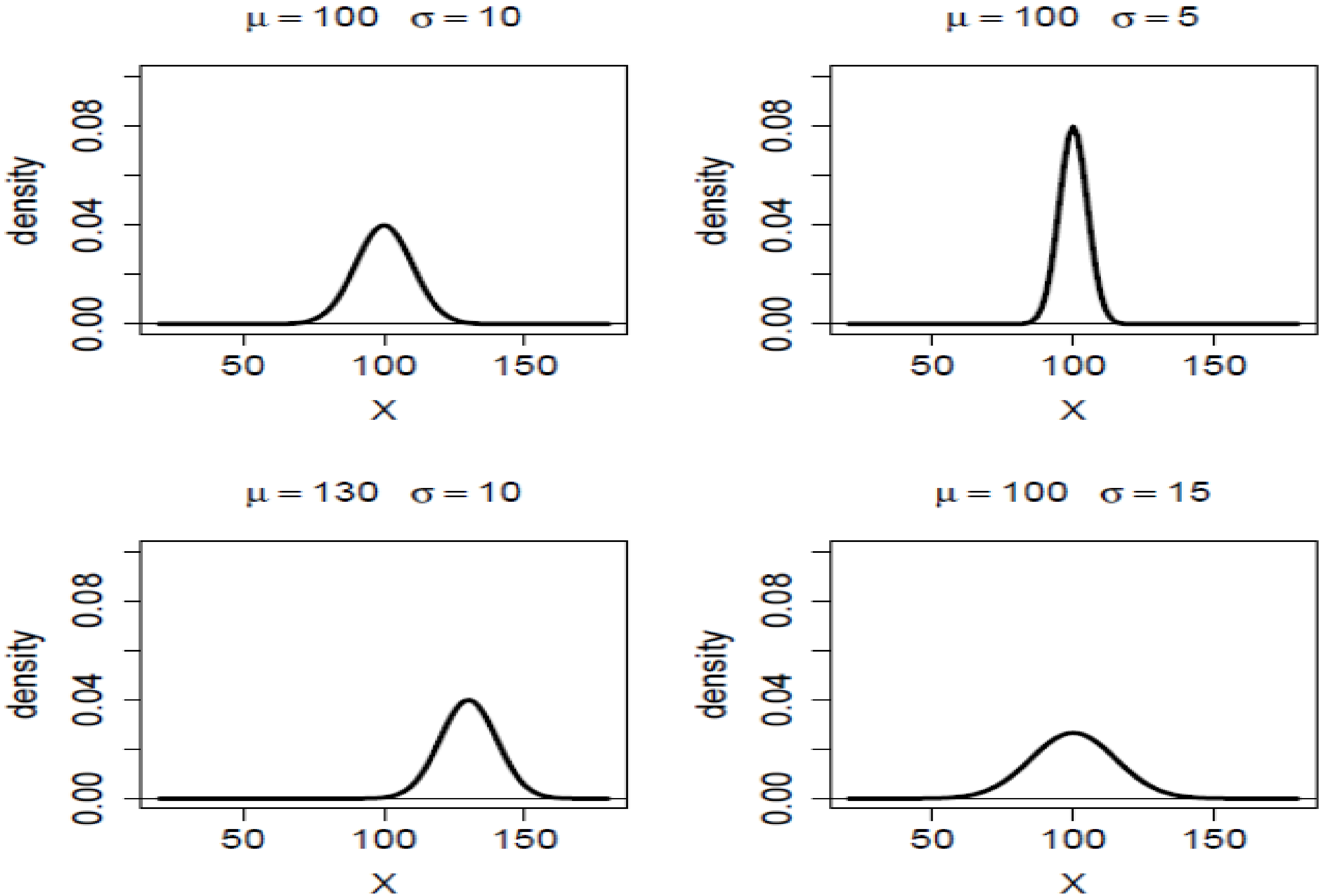
such distributions that are symmetric and can be described by two parameters

(i) μ = The mean of the distribution

(ii) σ = The standard deviation of the distribution

Changing the values of and alter the positions and

shapes of the distributions.


μ = 100 σ = 10
μ = 100 σ = 5
μ = 130 σ = 10
μ = 100 σ = 15
density
0.00
0.04
0.08
50
100
150
X

If X is Normally distributed with mean and standard deviation , we write

$$X \sim N(\mu, \sigma^2)$$

μ and σ are the **parameters** of the distribution.

The probability density of the Normal distribution is given by

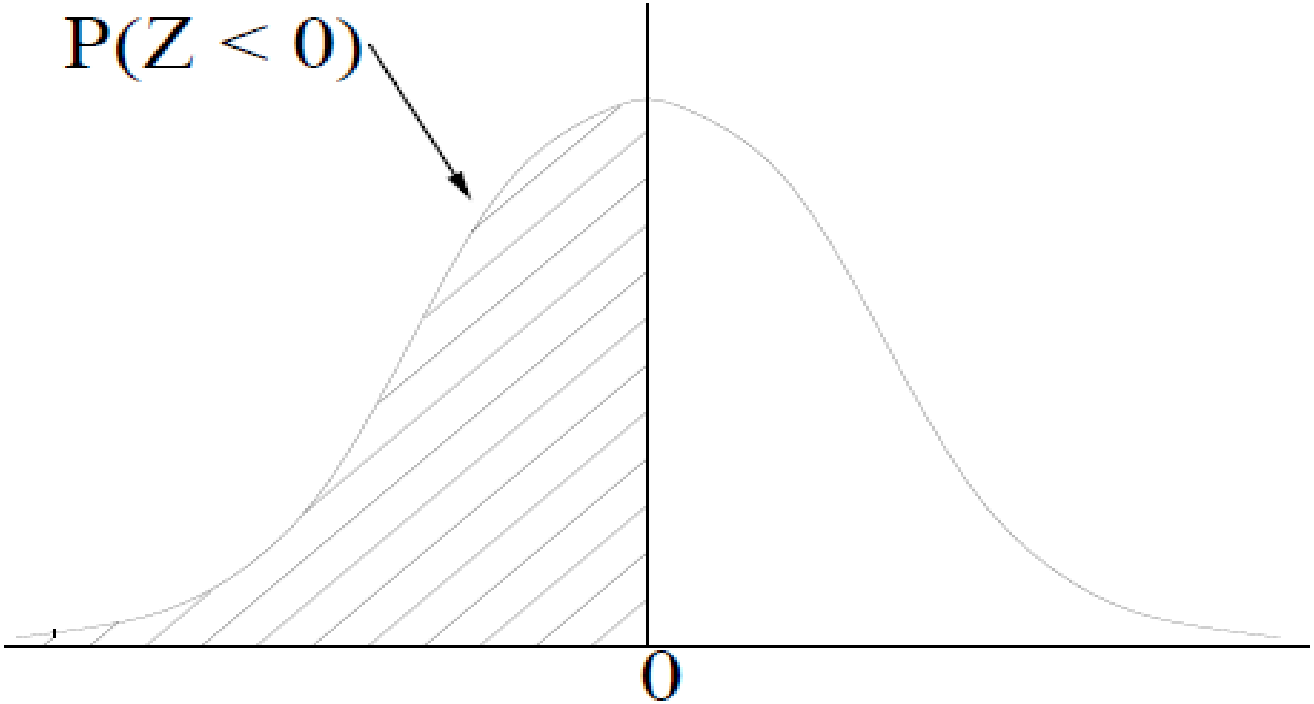
$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} * e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

For the purposes of this course we do not need to use this expression. It is included here for future reference.

Calculating probabilities from the Normal distribution

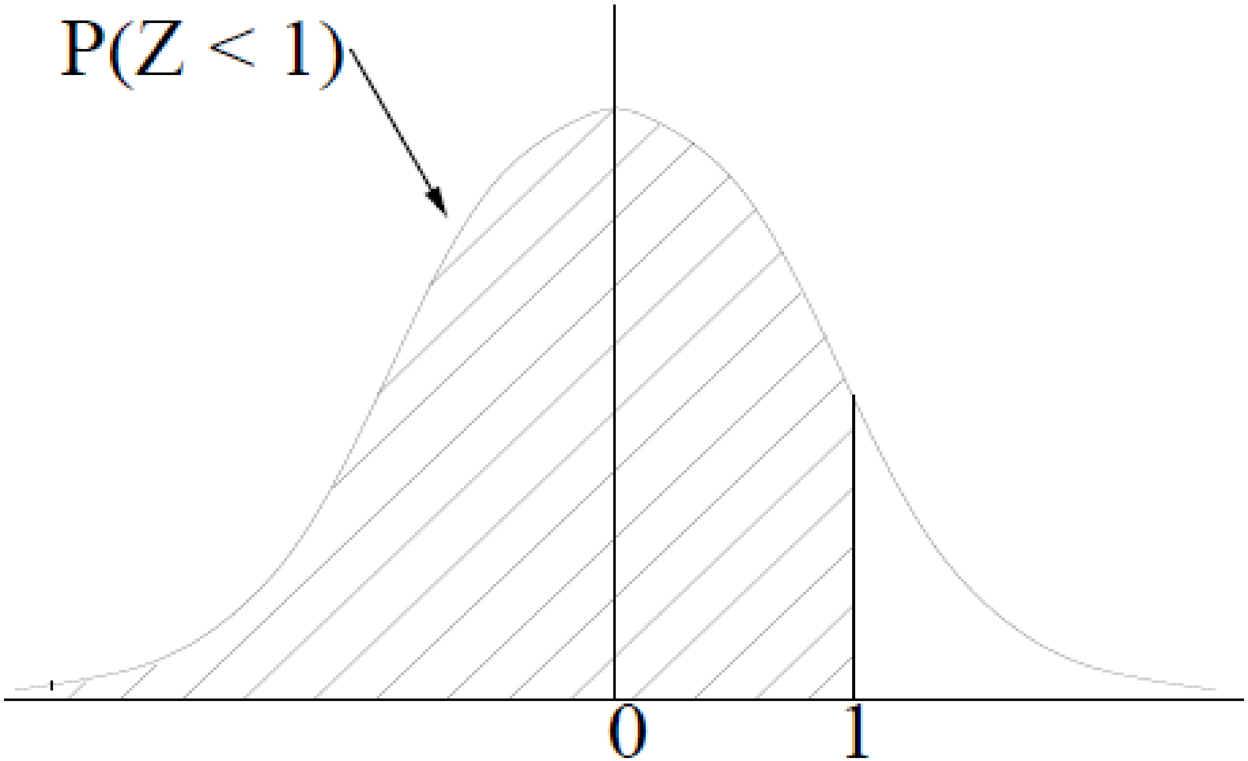
- For a discrete probability distribution we calculate the probability of being less than some value x, i.e.$P(X < x)$, by simply summing up the probabilities of the values less than x.
- For a continuous probability distribution we calculate the probability of being less than some value x, i.e. $P(X < x)$, by calculating the area under the curve to the left of x.

Suppose $Z \sim \mathrm{N}(0,\, 1)$, what is $P(Z < 0)$?



Symmetry $\Rightarrow$ $P(Z < 0) = 0.5$

What about $P(Z < 1.0)$?

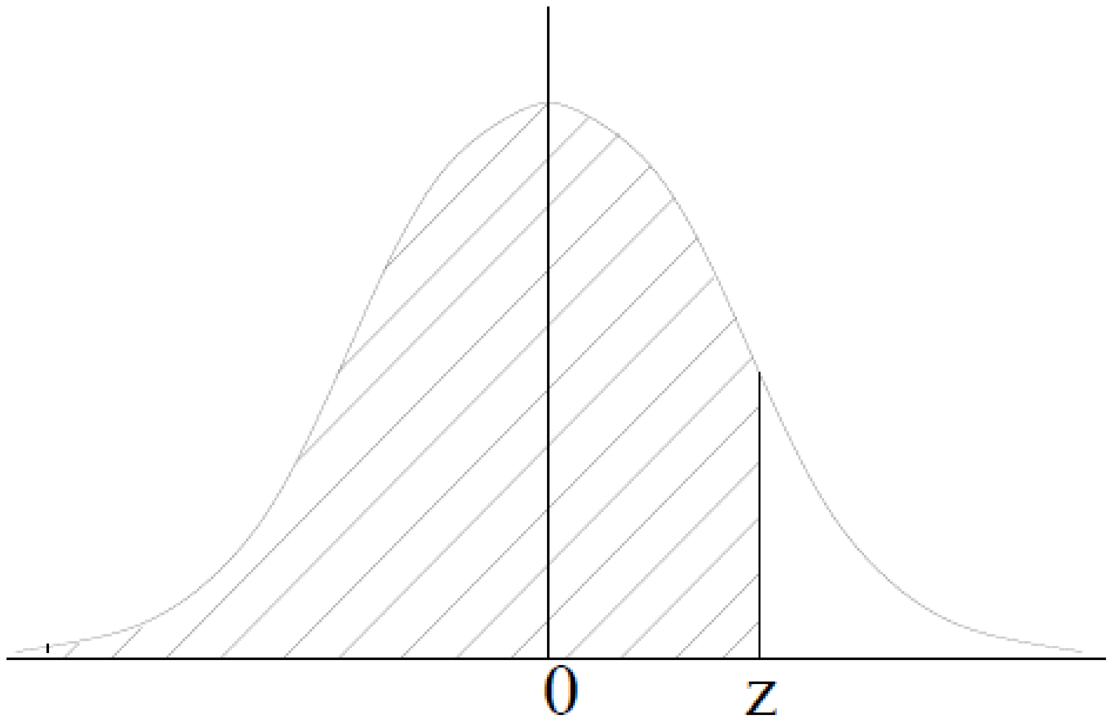


Calculating this area is not easy and so we use probability tables. Probability tables are tables of probabilities that have been calculated on a computer. All we have to do is identify the right probability in the table and copy it down!

Only one special Normal distribution, $N(0,1)$ has been tabulated.

- The $N(0,1)$ distribution is called the **standard Normal distribution**.

The tables allow us to read off probabilities of the form $P(Z < z)$.

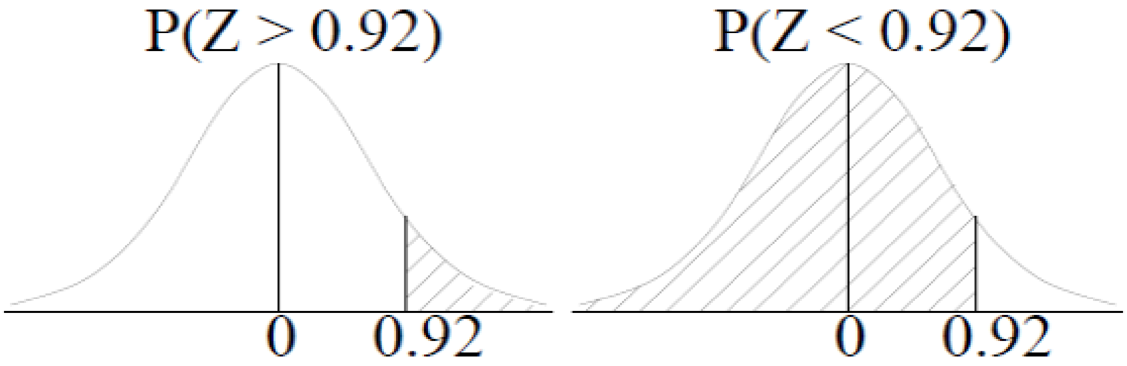


z	0.0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	5040	5080	5120	5160	5199	5239	5279	5319	5359
0.1	0.5398	5438	5478	5517	5557	5596	5636	5675	5714	5753
0.2	0.5793	5832	5871	5910	5948	5987	6026	6064	6103	6141
0.3	0.6179	6217	6255	6293	6331	6368	6406	6443	6480	6517
0.4	0.6554	6591	6628	6664	6700	6736	6772	6808	6844	6879
0.5	0.6915	6950	6985	7019	7054	7088	7123	7157	7190	7224
0.6	0.7257	7291	7324	7357	7389	7422	7454	7486	7517	7549
0.7	0.7580	7611	7642	7673	7704	7734	7764	7794	7823	7852
0.8	0.7881	7910	7939	7967	7995	8023	8051	8078	8106	8133
0.9	0.8159	8186	8212	8238	8264	8289	8315	8340	8365	8389
1.0	$\boxed{0.8413}$	8438	8461	8485	8508	8531	8554	8577	8599	8621
1.1	0.8643	8665	8686	8708	8729	8749	8770	8790	8810	8830

From this table we can identify that $P(Z < 1.0) = 0.8413$

Example 1

If $Z \sim$ N(0, 1) what is $P(Z > 0.92)$?

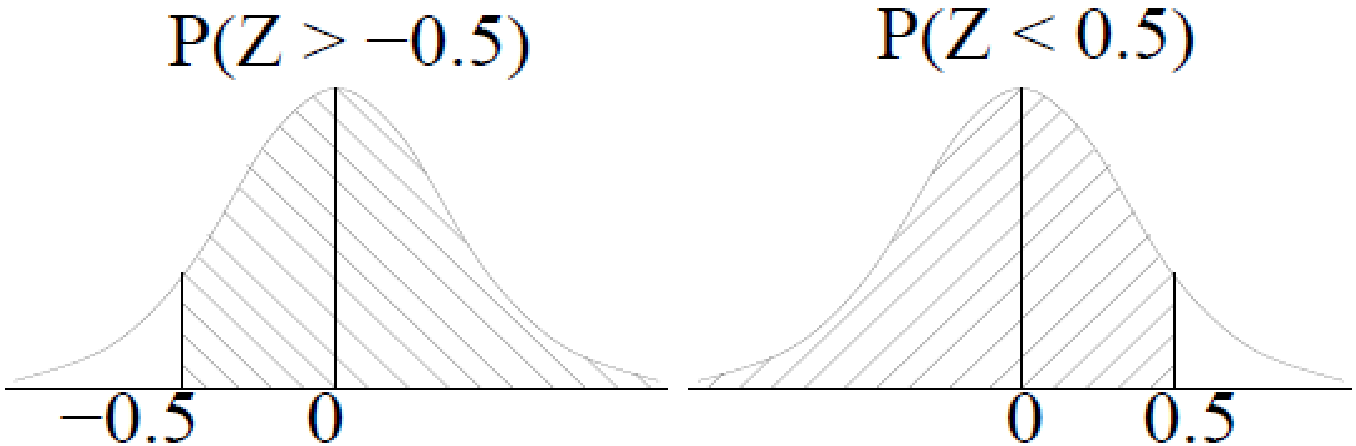


We know that $P(Z > 0.92) = 1 - P(Z < 0.92)$ and we can calculate $P(Z < 0.92)$ from the tables.

Thus, $P(Z > 0.92) = 1 - 0.8212 = 0.1788$

Example 2

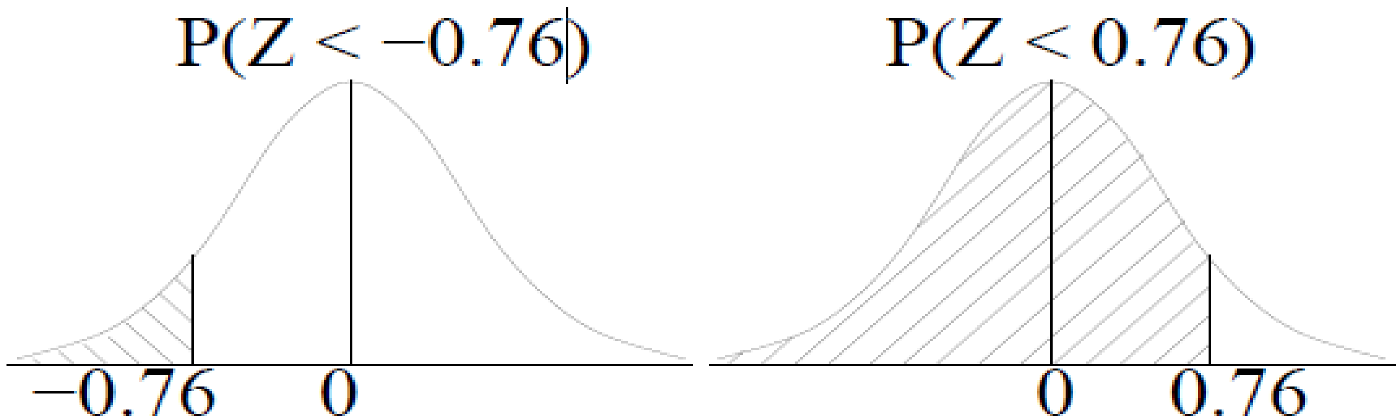
If $Z \sim \mathrm{N}(0, 1)$ what is $P(Z > -0.5)$?



The Normal distribution is symmetric so we know that
$P(Z > -0.5) = P(Z < 0.5) = 0.6915$

Example 3

If $Z \sim$ N(0, 1) what is $P(Z < -0.76)$?

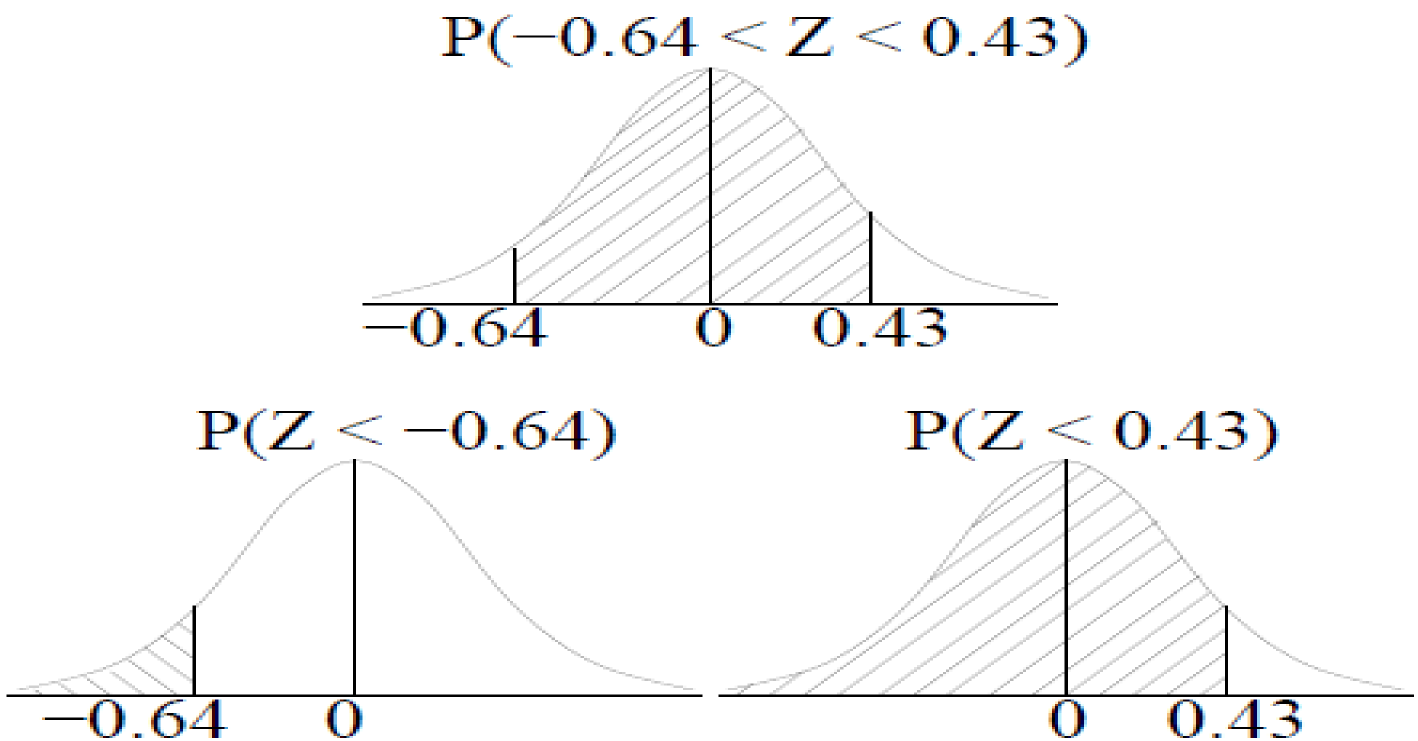


By symmetry

$$\begin{aligned} P(Z < -0.76) = P(Z > 0.76) &= 1 - P(Z < 0.76) \\ &= 1 - 0.7764 \\ &= 0.2236 \end{aligned}$$

Example 4

If $Z \sim \mathrm{N}(0,\, 1)$ what is $P(-0.64 < Z < 0.43)$?



We can calculate this probability as

$$\begin{aligned}
P(-0.64 < Z < 0.43) &= P(Z < 0.43) - P(Z < -0.64) \\
&= 0.6664 - (1 - 0.7389) \\
&= 0.4053
\end{aligned}$$

Example 5

Consider $P(Z < 0.567)$?

From tables we know that $P(Z < 0.56) = 0.7123$ and $P(Z < 0.57) = 0.7157$
To calculate $P(Z < 0.567)$ we *interpolate* between these two values

$$P(Z < 0.567) = 0.3 \times 0.7123 + 0.7 \times 0.7157 = 0.71468$$

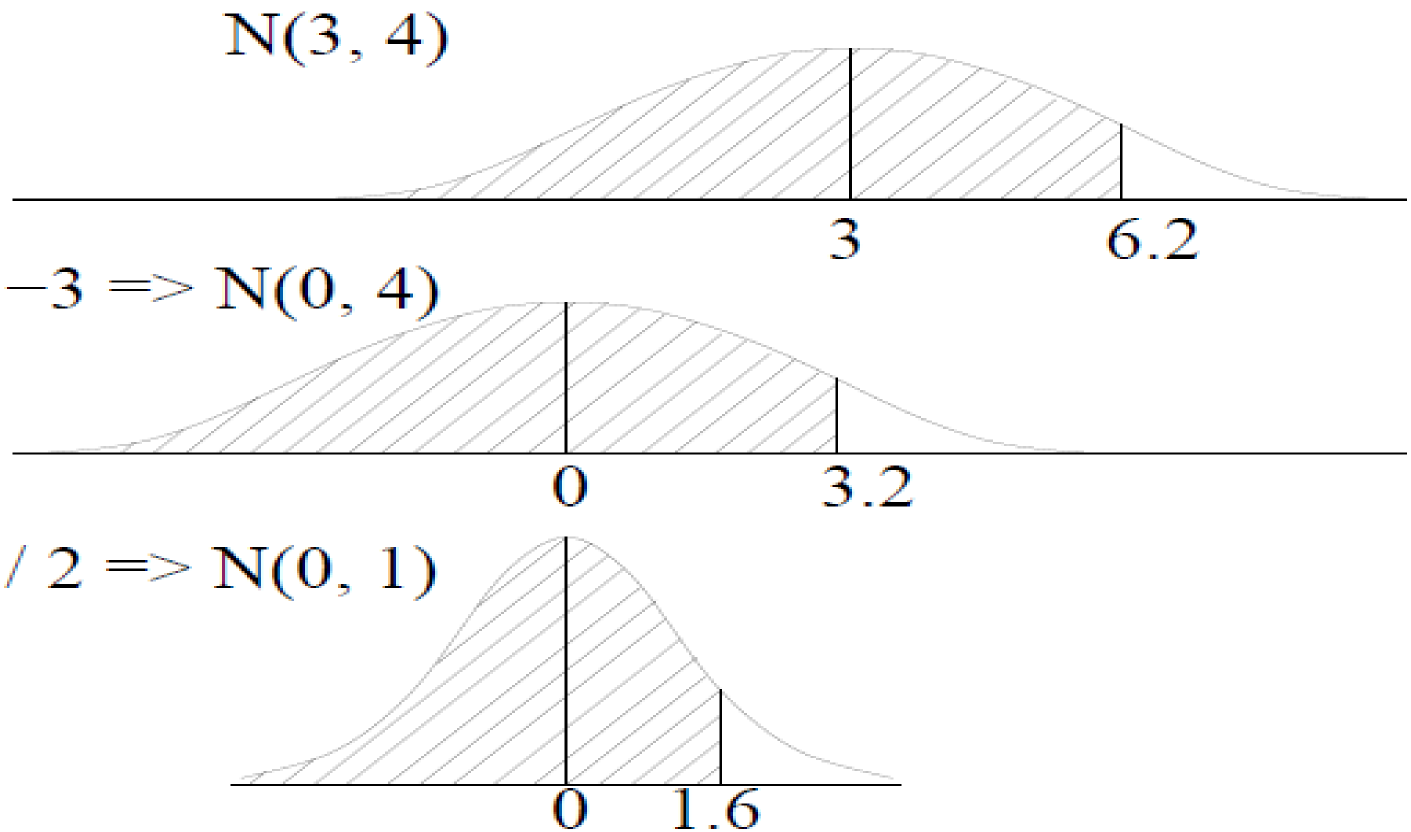
Standardization

All of the probabilities above were calculated for the standard Normal distribution $N(0,1)$. If we

want to calculate probabilities from different Normal distributions we convert the probability to one involving the standard Normal distribution.

This process is called **standardization**.

Suppose $X \sim$ N(3, 4), what is $P(X < 6.2)$?



We convert this probability to one involving the $N(0,1)$ distribution by:

(i) Subtracting the mean

(ii) Dividing by the standard deviation

Subtracting the mean re-centers the distribution on zero.

Dividing by the standard deviation re-scales the distribution so it has standard deviation 1 . If we Also

transform the boundary point of the area we wish to calculate we obtain the equivalent boundary point for the $N(0,1)$ distribution.

$P(X < 6.2) = P(Z < 1.6) = 0.9452$ Where $Z \sim N(0,1)$

This process can be described by the following rule

$$\boxed{\begin{array}{l}\text{If } X \sim \mathrm{N}(\mu,\ \sigma^2) \text{ and } Z = \frac{X-\mu}{\sigma} \\ \text{then} \\ \qquad\qquad Z \sim \mathrm{N}(0,\ 1)\end{array}}$$

Example 6

Suppose we know that the birth weight of babies is Normally distributed with mean 3500g and standard deviation 500g. What is the probability that a baby is born that weighs less than 3100g?

That is $X \sim \mathrm{N}(3500,\ 500^2)$ and we want to calculate $P(X < 3100)$?

We can calculate the probability through the process of standardization.

Drawing a rough diagram helps

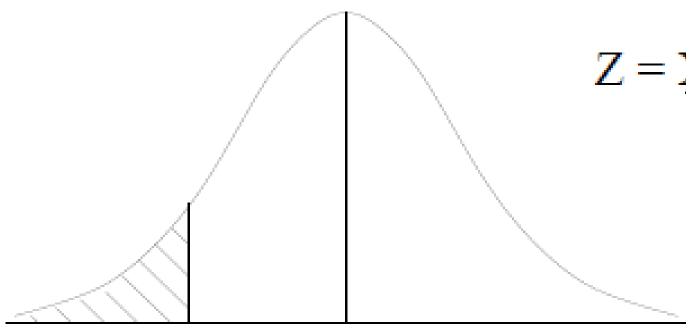
$X \sim N(3500, 500^2)$

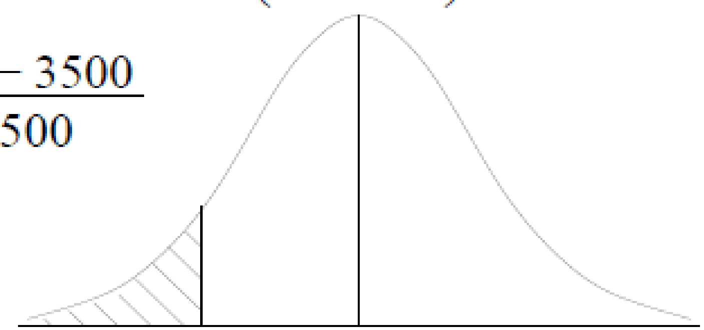
$Z \sim N(0, 1)$

$P(X < 3100)$

$P(Z < -0.8)$

$$Z = \frac{X - 3500}{500}$$





$$
\begin{aligned}
P(X < 3100) &= P\left(\frac{X - 3500}{500} < \frac{3100 - 3500}{500}\right) \\
&= P(Z < -0.8) \qquad \text{where } Z \sim \mathrm{N}(0,1) \\
&= 1 - P(Z < 0.8) \\
&= 1 - 0.7881 \\
&= 0.2119
\end{aligned}
$$