Data Analytics

ECON 1008, Semester 1, 2019

Giulio Zanella
University of Adelaide
School of Economics

Chapter 8

Continuous probability distributions

Introduction

A **continuous random variable** is one that can assume an **uncountable** number of values in an interval [a,b].

Contrast this with the case of a discrete random variable, which takes a countable number of values

Now we cannot list all possible values because there is an infinite number of them in the range [a,b].

Because there is an infinite number of values their probabilities must sum up to 1, **the probability of each individual value is virtually 0**.

Point Probabilities are Zero

We can only determine the prob. of a ***range of values***

With a **discrete** random variable like tossing a die, it is meaningful to talk about P(X=5), say.

In a **continuous** setting, the probability the random variable of interest is infinitesimally small.

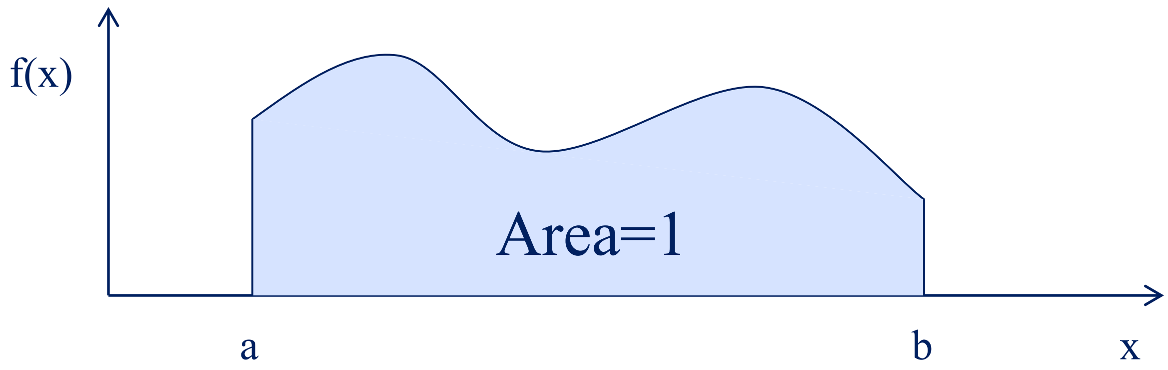
Example: X = time to complete your weekly assignment. The probability takes it **exactly** 5 minutes is infinitesimally small, hence P(X=5) = 0.

It is meaningful to talk about P(X ≤ 5). To determine these probabilities we employ a Probability Density Function.

Probability Density Function (PDF)

A function f(x) is called a **Probability Density Function** over the range **a ≤ x ≤ b** if it meets the following requirements:

1) f(x) ≥ 0 for all **x** between **a** and **b**, and

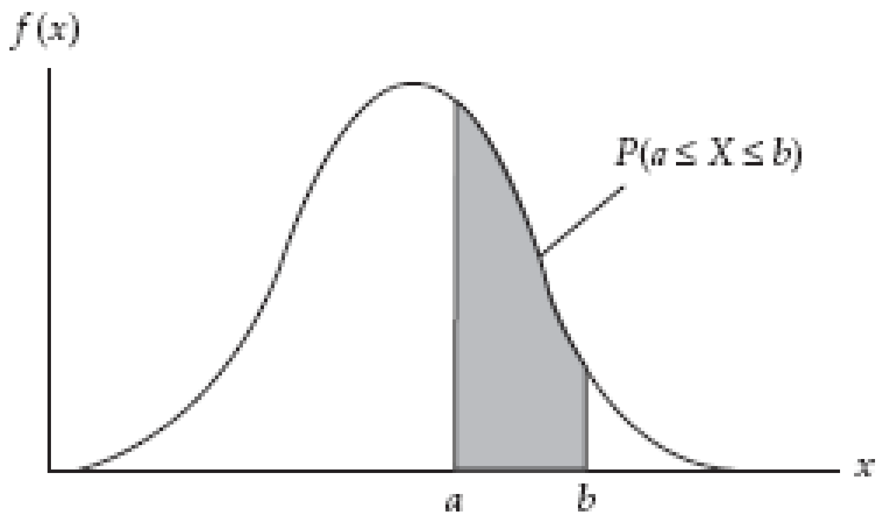


2) The total area under the curve between **a** and **b** is 1.

Probability density function…

The probability that x falls between *a* and *b* is found by calculating the area under graph of f(x) between *a* and b.

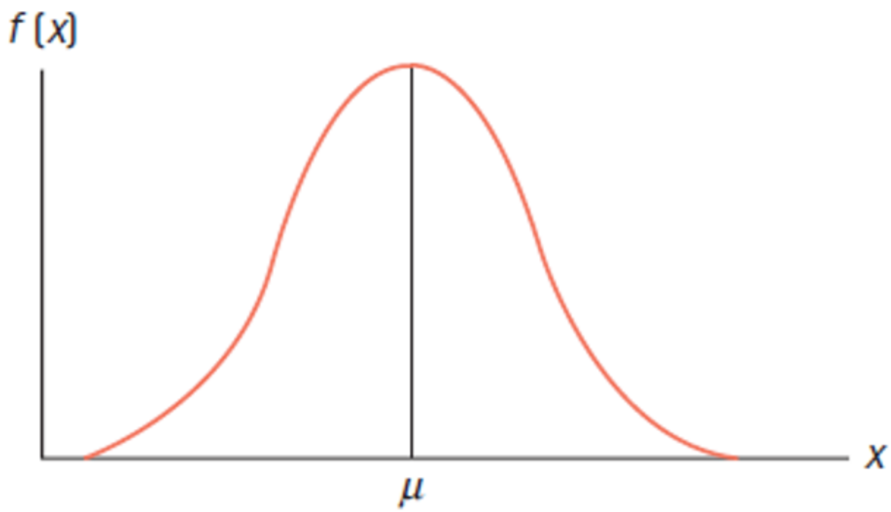
Shaded area is P(*a* ≤ X ≤ *b*)



Notice: area in the continuous case ("integral") is analogous to sum in the discrete case.

Normal distribution

The most important continuous distribution we study:
a bell-shaped distribution ranging from -infinity to +infinity, symmetrical around μ.



- Many random variables can be modelled as normal
- Many distributions can be approximated by a normal
- Cornerstone distribution in statistical inference.

Normal distribution…

The normal distribution is fully defined by two parameters:

- its mean μ
- its variance σ^2

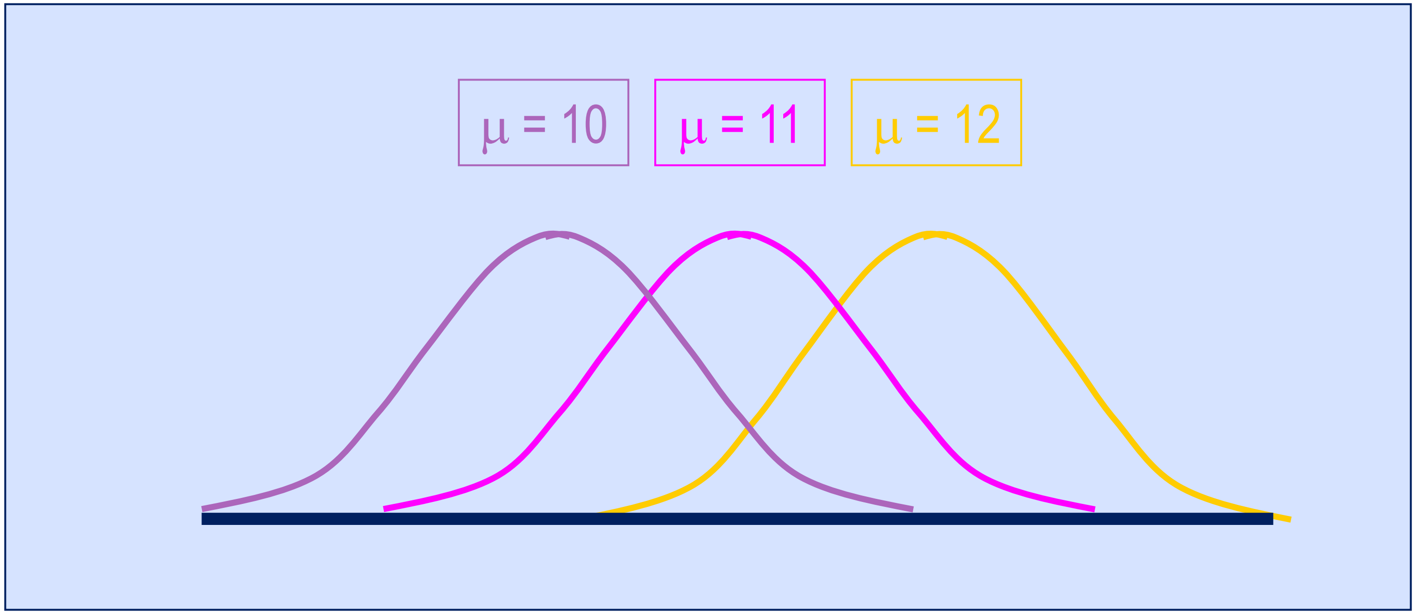
Mathematically, a random variable **X** with mean μ and variance σ^2 is normally distributed if its probability density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)[(x-\mu)/\sigma]^2} \qquad -\infty \le x \le \infty$$

where μ is the mean, σ is the standard deviation,

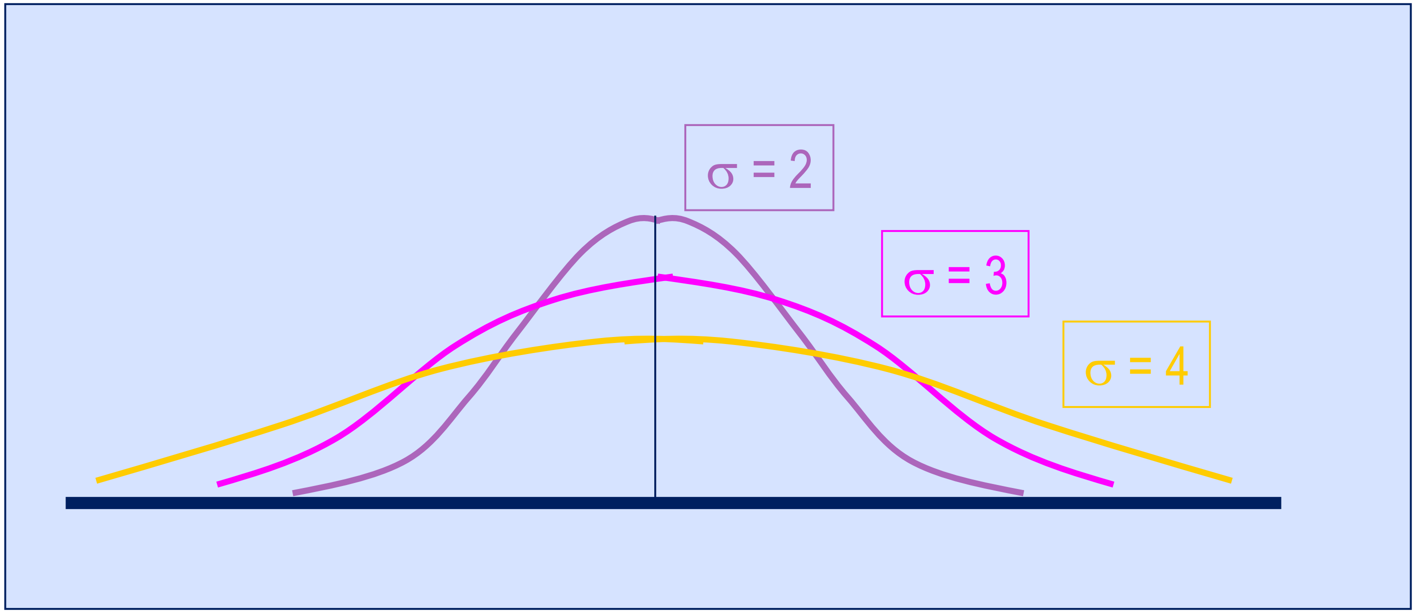
$\pi = 3.14159...$ and $e = 2.71828...$

How does the expected value affect the location of f(x)?



Increasing the mean shifts the curve to the right…

How does the standard deviation affect the location of f(x)?



Increasing the standard deviation 'flattens' the curve…

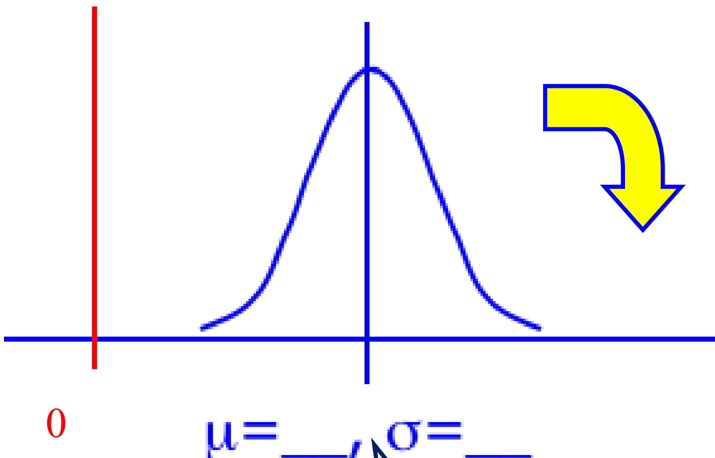
Calculating Normal Probabilities

Two facts help calculate normal probabilities:

- Normal distribution is symmetrical.
- Any normal distribution can be transformed into a special normal distribution, the **STANDARD NORMAL DISTRIBUTION.**

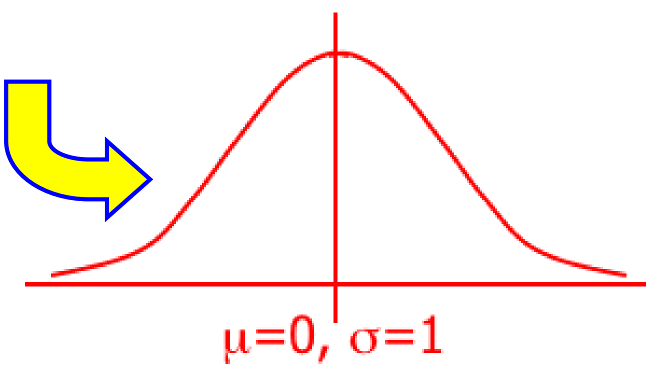
Calculating Normal Probabilities…

We can use the following function to convert **any** normal random variable to a **standard** normal random variable…



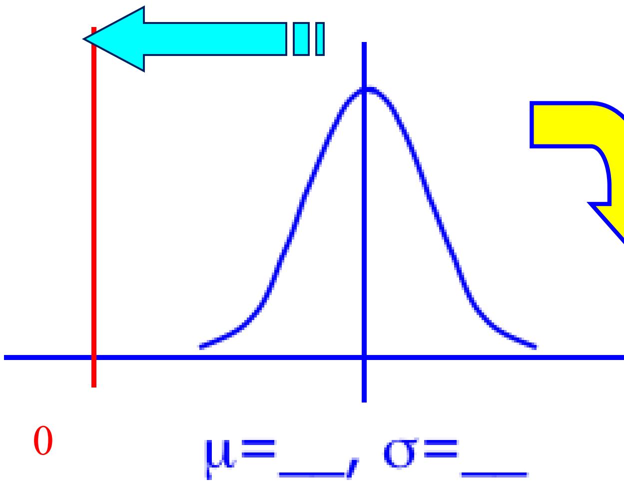
$$Z = \frac{X - \mu}{\sigma}$$

Some advice: always draw a picture!



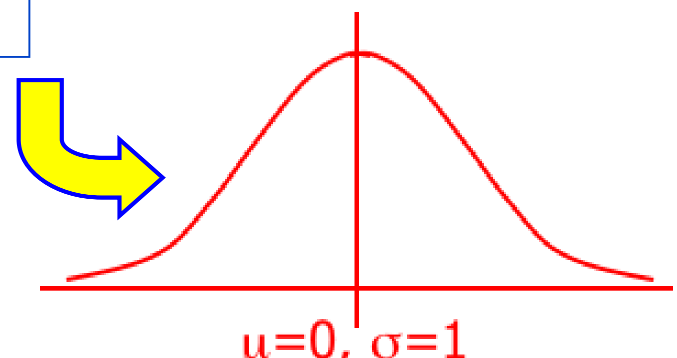
Calculating Normal Probabilities…

We can use the following function to convert any normal random variable to a **standard** normal random variable…



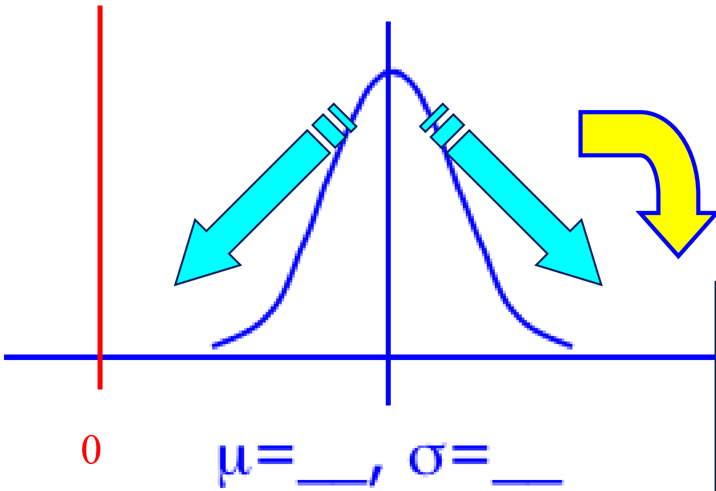
$$Z = \frac{X - \mu}{\sigma}$$

This shifts the mean of X to zero…



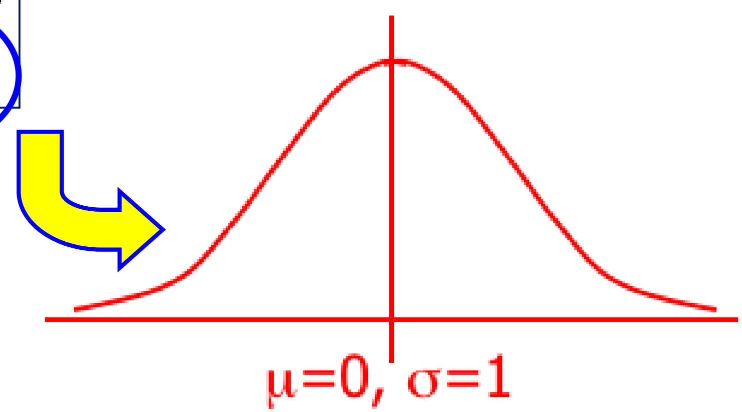
Calculating Normal Probabilities…

We can use the following function to convert any normal random variable to a **standard** normal random variable…



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

This changes the shape of the curve…



Calculating Normal Probabilities…

If we know the mean and standard deviation of a normally distributed random variable, we can always transform the probability statement about X into a probability statement about Z.

Consequently, to compute probabilities for normally distributed random variables we need only one table, Table 3 in Appendix B, the standard normal probability table.

Let’s familiarise with that table.

Table 3: Standard Normal Distribution

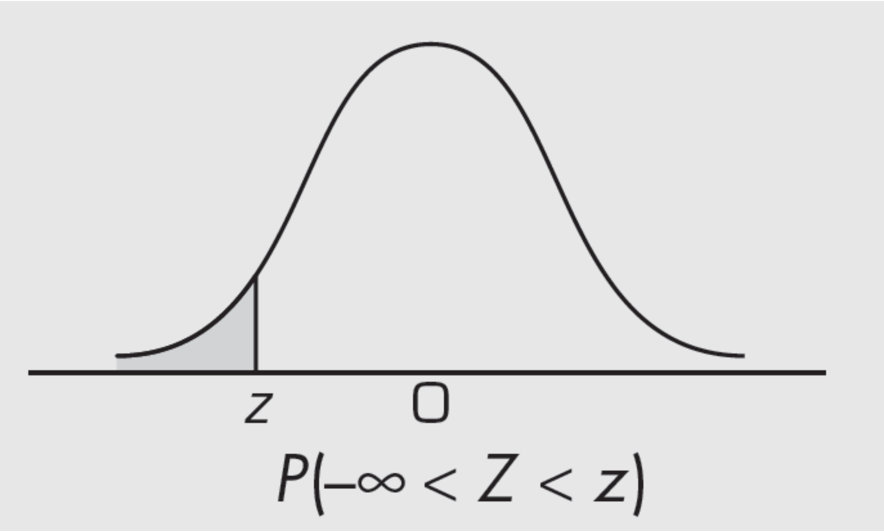
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
–3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
–2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
–2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
–2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
–2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
–2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
–2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
–2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
–2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
–2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
–2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
–1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
–1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
–1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
–1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
–1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
–1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681

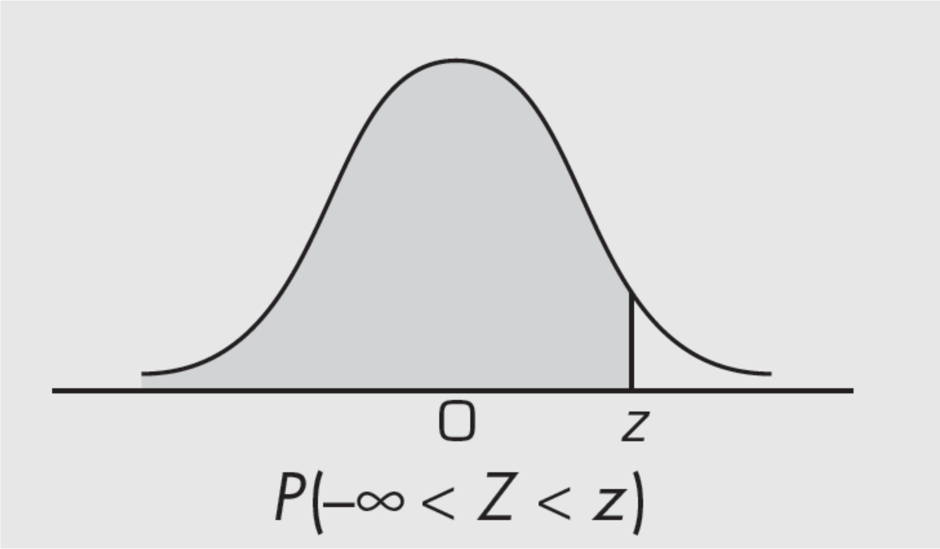
Table 3: Standard Normal Distribution

The standard normal table lists cumulative probabilities

P(Z < z) or equivalently **P(-∞ < Z < z)**

for values of z ranging from −3.09 to +3.09



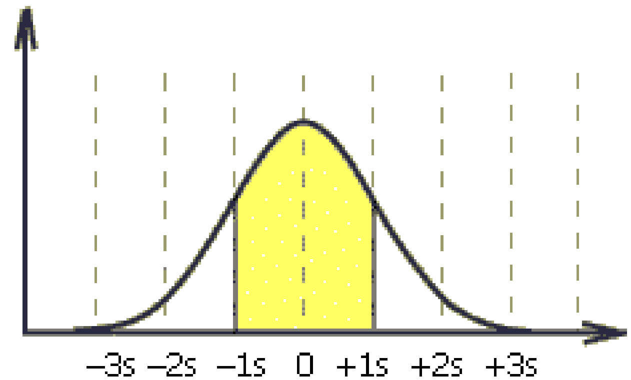


Exam-like question (why are we limiting to [-3.09 , 3.09]?)

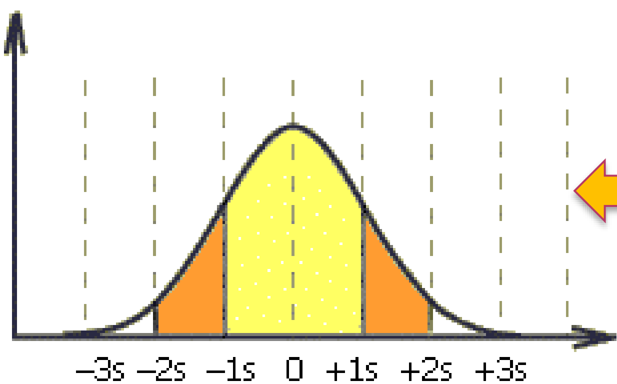
What fraction of realisations of the random variables will be between -3.09 and 3.09?

Empirical rule…

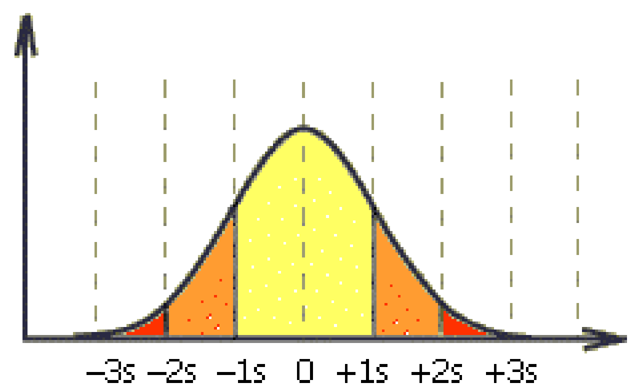
Approximately 68% of all observations fall within **one** standard deviation of the mean.



Approximately 95% of all observations fall within **two** standard deviations of the mean.



Approximately 99.7% of all observations fall within **three** standard deviations of the mean.



Calculating Standard Normal Probabilities

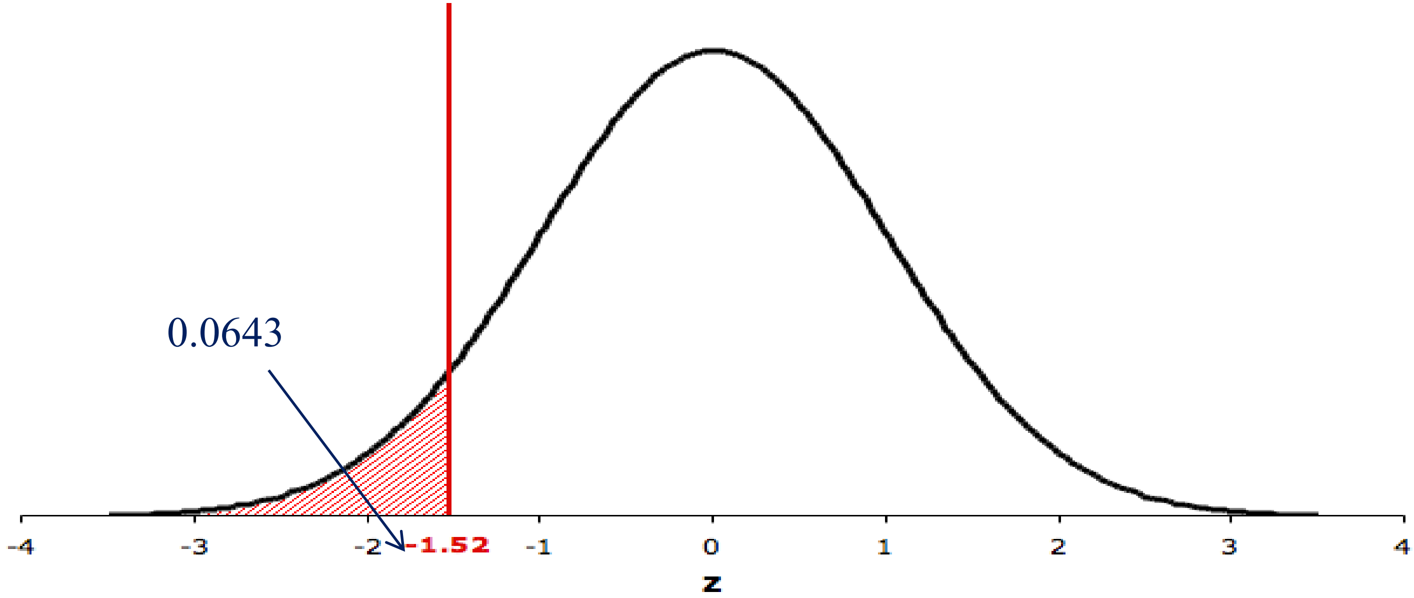
Suppose we want to determine the following probability: **$P(Z < -1.52)$**

We first find **−1.5** in the left margin. We then move along this row until we find the probability under the **.02** column heading. Thus,

$$P(Z < -1.52) = 0.0643$$

Calculating Standard Normal Probabilities

$P(Z < -1.52) = 0.0643$



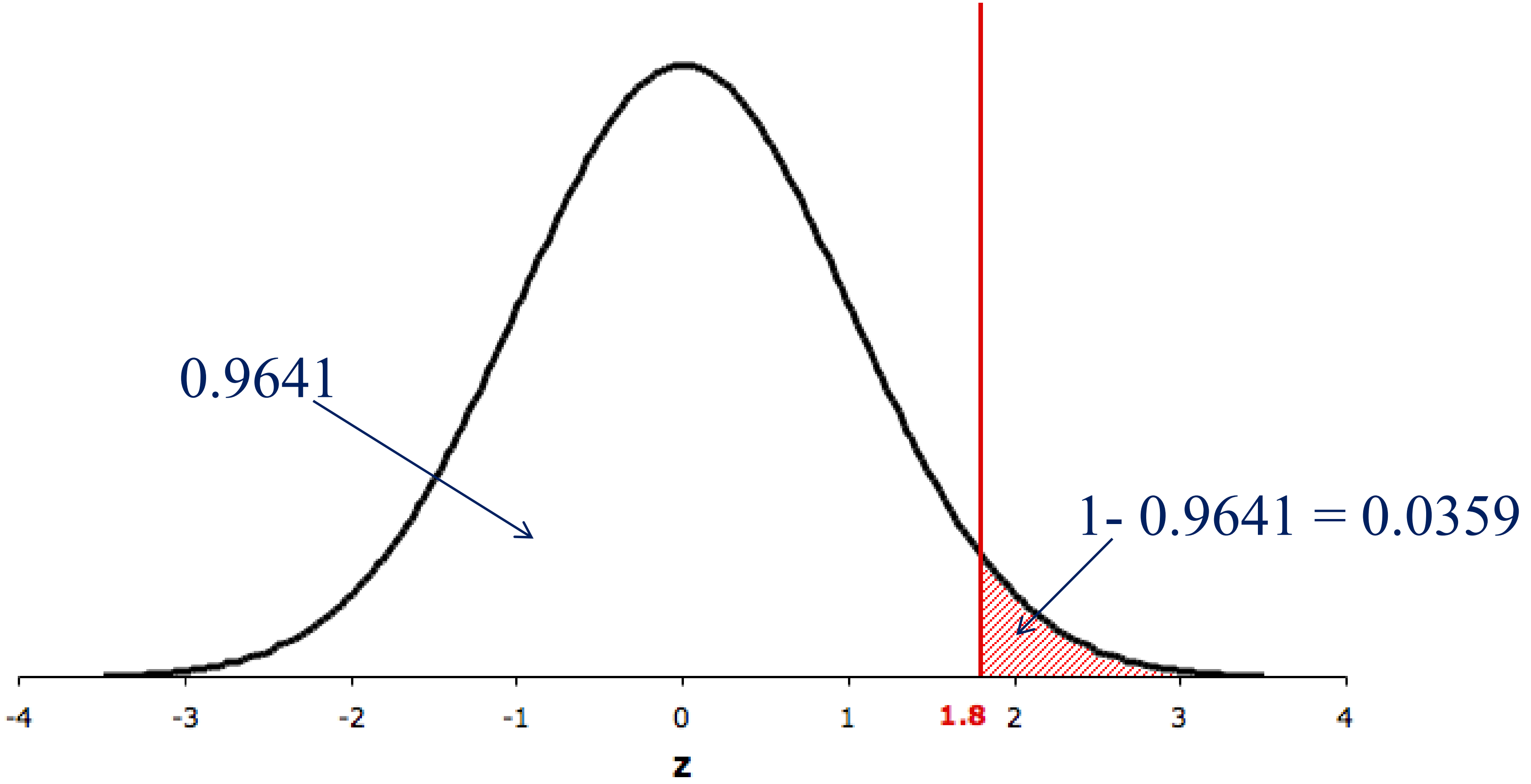
Calculating Standard Normal Probabilities

We can find the probability that Z is greater than 1.80, **$P(Z > 1.80)$**.

This probability can be obtained by determining the probability that Z is less than 1.80 and subtracting that value from 1.

$$P(Z > 1.80) = 1 - (P(Z < 1.80)$$

Calculating Standard Normal Probabilities



Calculating Standard Normal Probabilities

Applying the complement rule we get

$$P(Z > 1.80) = 1 - P(Z < 1.80)$$
$$= 1 - 0.9641 = 0.0359$$

Alternatively, using symmetry,

$$P(Z > 1.80) = P(Z < -1.80) = 0.0359$$

Calculating Standard Normal Probabilities

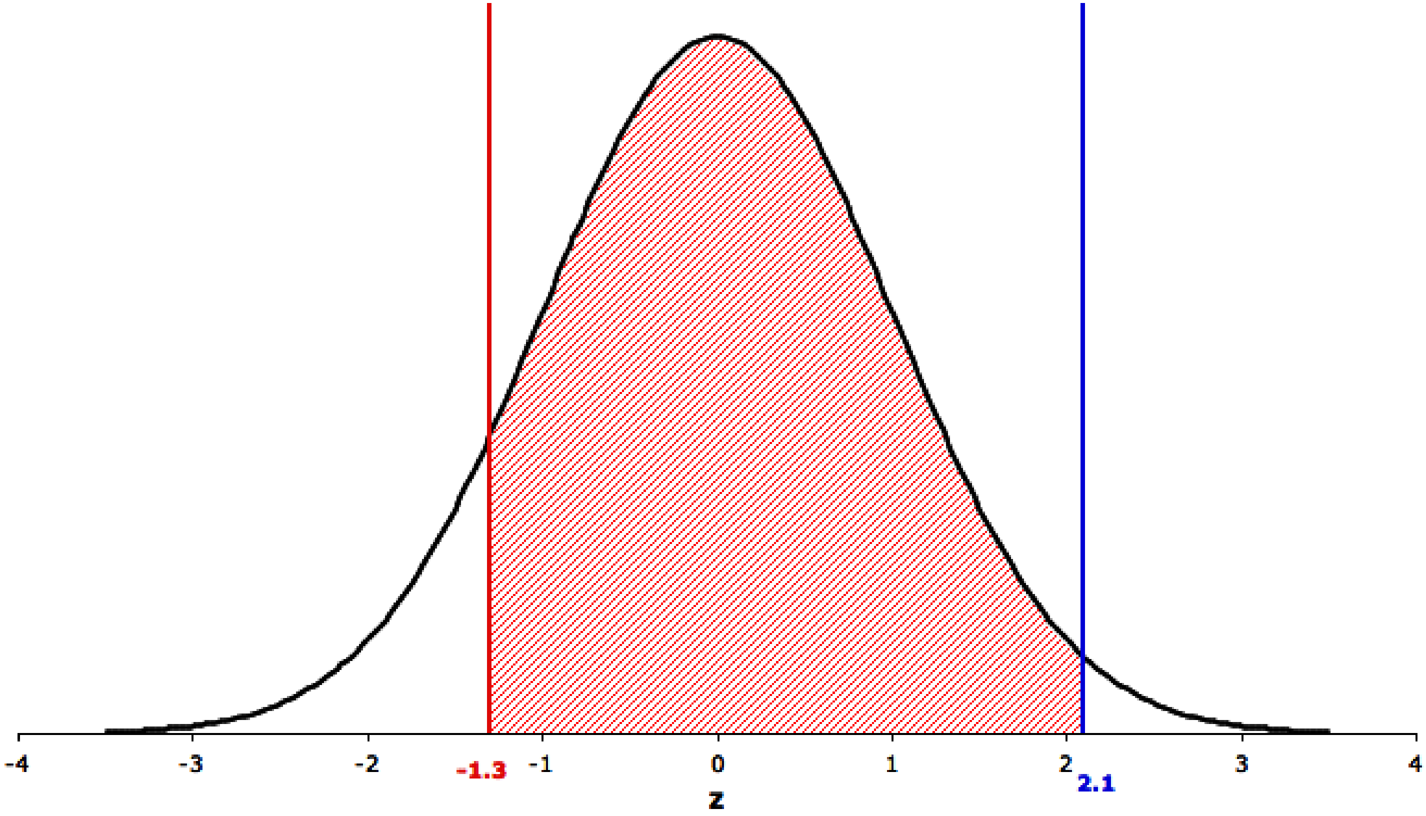
We can also easily determine the probability that a standard normal random variable lies between two values of z. For example, we want to find the probability

$$P(-1.30 < Z < 2.10).$$

This can be obtained by finding the two cumulative probabilities and calculating their difference,

$$P(-1.30 < Z < 2.10) = P(Z < 2.10) - P(Z < -1.30)$$

Calculating Standard Normal Probabilities



Calculating Standard Normal Probabilities

From the z table,

$$P(Z < -1.30) = 0.0968$$

and $$P(Z < 2.10) = 0.9821$$

Hence,

$$\begin{aligned} P(-1.30 < Z < 2.10) &= P(Z < 2.10) - P(Z < -1.30) \\ &= 0.9821 - 0.0968 \\ &= 0.8853 \end{aligned}$$

Table 3: Standard Normal Distribution

Notice that the largest value of z in the table is 3.09, and that $P(Z < 3.09) = 0.9990$. This means that

$$P(Z > 3.09) = 1 - 0.9990 = 0.0010$$

However, because the table lists no values beyond 3.09, we approximate any area beyond 3.10 as 0. That is,

$$P(Z > 3.10) = P(Z < -3.10) \approx 0$$

Example 1

Suppose that at a country town petrol station, the daily demand for regular petrol is normally distributed with a mean of 1,000 litres and a standard deviation of 100 litres.

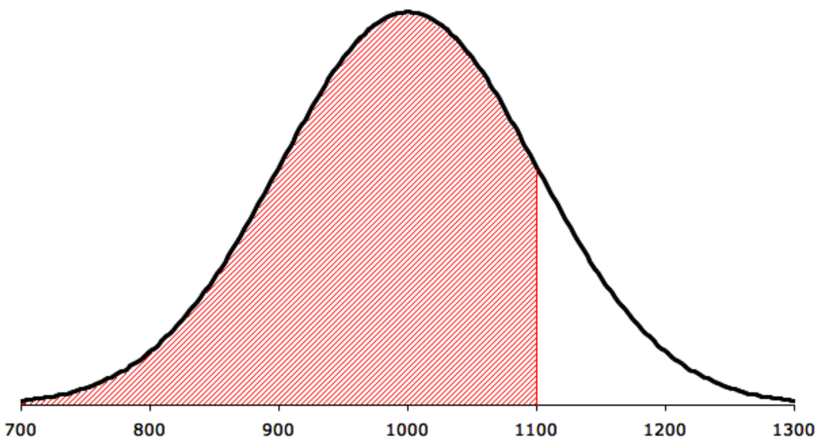
The station manager has just opened the station for business and notes that there is exactly 1,100 litres of regular petrol in storage. The next delivery is scheduled later today at the close of business. The manager would like to know the probability that he will have enough regular petrol to satisfy today's demands.

Example 1: Solution

The demand for petrol is normally distributed with mean $\mu = 1{,}000$ and standard deviation $\sigma = 100$. We want to find the probability

$P(X < 1{,}100)$

Graphically we want to calculate the shaded area:



Example 1: Solution…

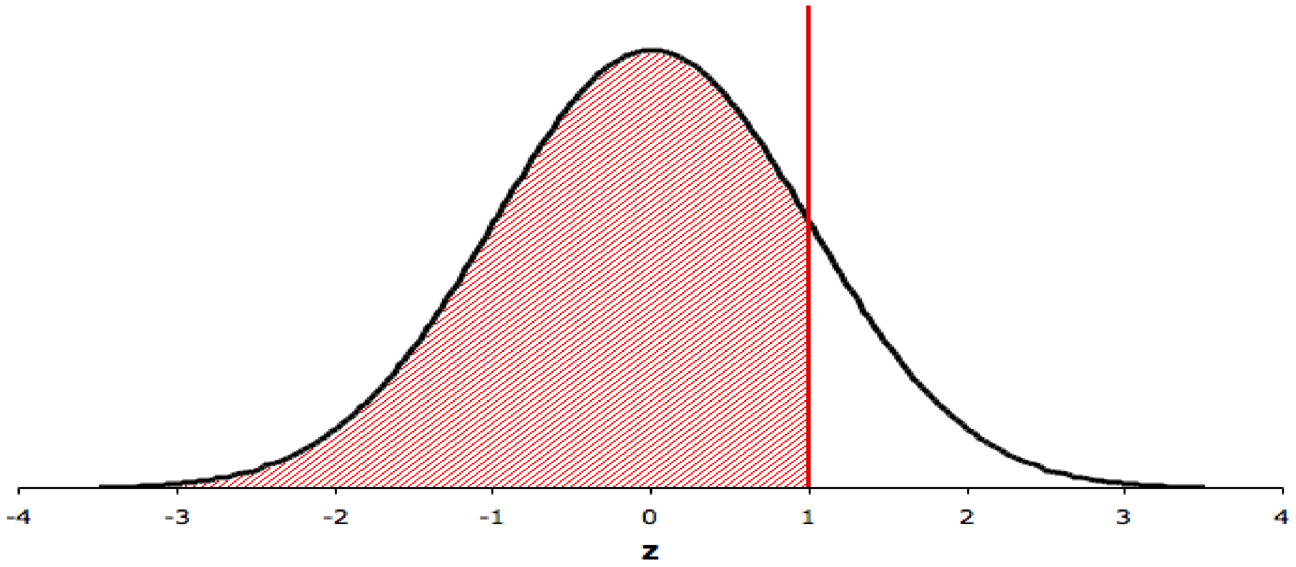
The first step is to standardize X. However, if we perform any operations on X, we must perform the same operations on 1,100. Thus,

$$P(X < 1{,}100) = P\left(\frac{X-\mu}{\sigma} < \frac{1100-1000}{100}\right)$$

$$= P(Z < 1.00)$$

Example 1: Solution…

The figure below graphically depicts the probability we seek.



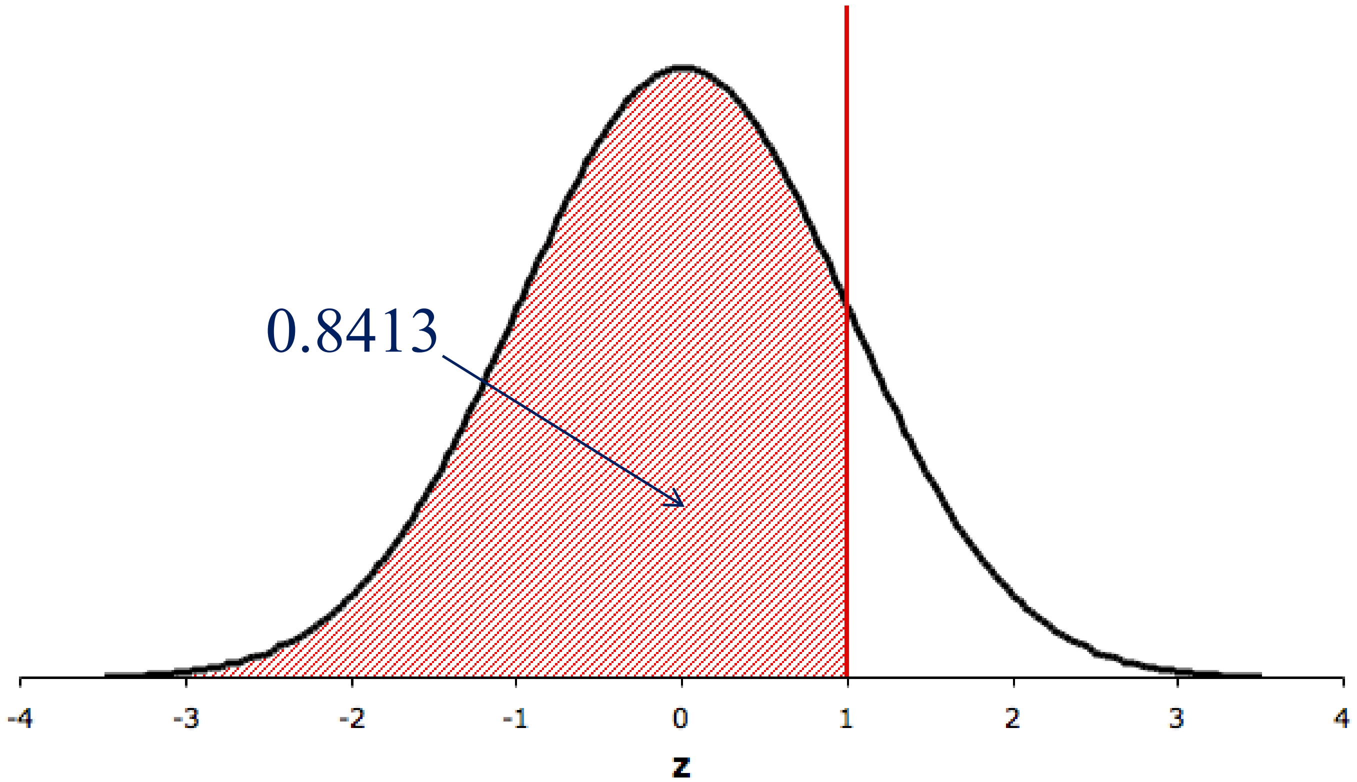
Example 1: Solution…

The values of Z specify the location of the corresponding value of X. A value of Z = 1 corresponds to a value of X that is 1 standard deviation above the mean. Notice as well that the mean of Z, which is 0 corresponds to the mean of X.

The probability that we seek is

$$P(X < 1{,}100) = P(Z < 1.00) = 0.8413$$

Example 1: Solution…



Example 2

Consider an investment whose return is normally distributed with a mean of 10% and a standard deviation of 5%.

a. Determine the probability of losing money.

b. Find the probability of losing money when the standard deviation is equal to 10%.

Example 2: Solution

a. The investment loses money when the return is negative. Thus we wish to determine

P(X < 0)

The first step is to standardize both X and 0 in the probability statement.

$$P(X < 0) = P\left(\frac{X-\mu}{\sigma} < \frac{0-10}{5}\right) = P(Z < -2.00)$$

Example 2: Solution…

From Table 3 we find

$$P(Z < -2.00) = 0.0228$$

Therefore, the probability of losing money is 0.0228.

Example 2: Solution…

b. If we increase the standard deviation to 10% the probability of suffering a loss becomes

$$P(X < 0) \quad = \quad P\left(\frac{X-\mu}{\sigma} < \frac{0-10}{10}\right)$$

$$= P(Z < -1.00)$$

$$= 0.1587$$

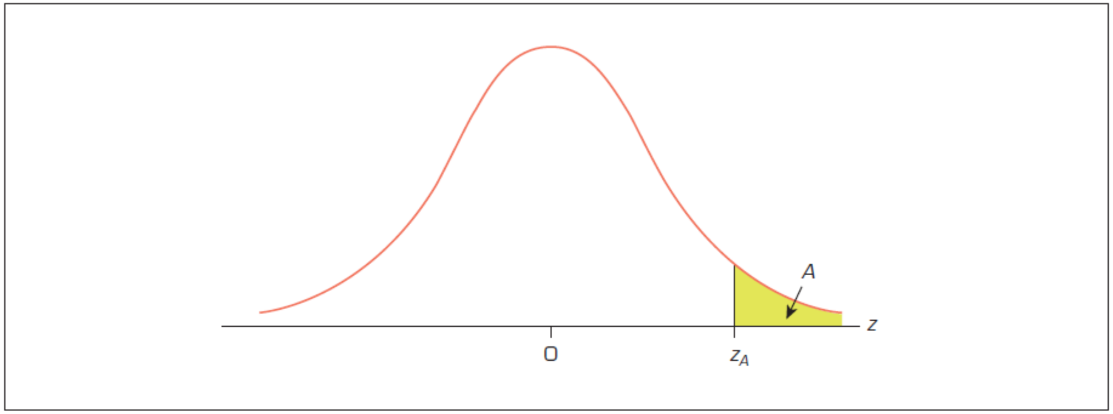
Finding the Value of z

Often we're asked to find some value of z for a given probability, i.e. given an area (A) under the curve, what is the corresponding value of z (z_A) on the horizontal axis that gives us this area? That is:

For given A, find z_A such that

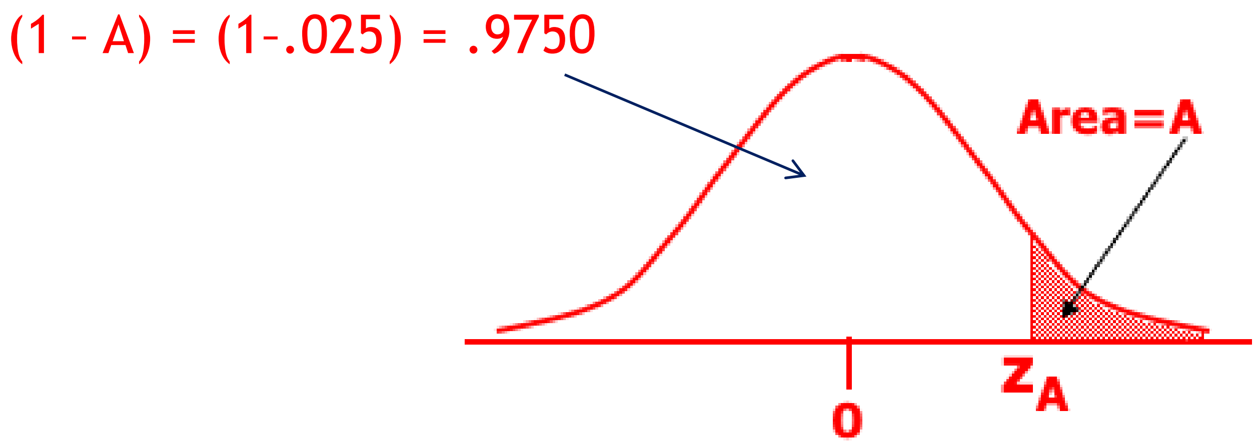
$$P(Z > z_A) = A$$

Figure 8.17 $P(Z > z_A) = A$



Finding the Value of z…

What value of z corresponds to an area under the curve of 2.5%? That is, what is z_A such that $P(Z > z_A) = 0.025$?



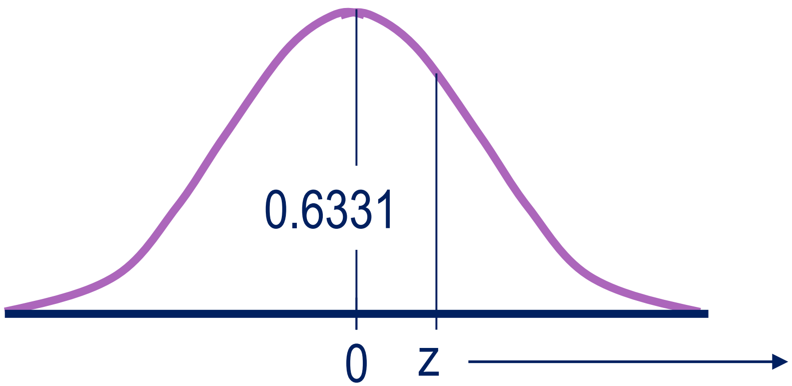
If you do a 'reverse look-up' on Table 3 for 0.9750, you will get the corresponding $z_A = 1.96$.

Since $P(z > 1.96) = 0.025$, we say: $z_{.025} = 1.96$.

Example 4

If Z is a standard normal variable, determine the value z for which $P(Z < z) = 0.6331$.

Solution:



z = 0.34

[Using Table 3, Appendix B]

Example 4

The weekly incomes of shift foremen in the glass industry follow the normal probability distribution with a **mean of $1,000** and a **standard deviation of $100**.

What is the ***z* value** for the income, let's call it *x*, of a foreman who earns **$1,100** per week? For a foreman who earns **$900** per week?

Example 4: solution

The weekly incomes of shift foremen in the glass industry follow the normal probability distribution with a **mean of $1,000** and a **standard deviation of $100**.

What is the ***z* value** for the income, let's call it *x*, of a foreman who earns **$1,100** per week? For a foreman who earns **$900** per week?

For $x = \$1{,}100$:

$$z = \frac{x - \mu}{\sigma} = \frac{\$1{,}100 - \$1{,}000}{\$100} = 1.00$$

For $x = \$900$:

$$z = \frac{x - \mu}{\sigma} = \frac{\$900 - \$1{,}000}{\$100} = -1.00$$

INTERPRETATION: the income of the first is 1 standard deviation above the mean, while the income of the second worker is 1 standard deviation below the mean.