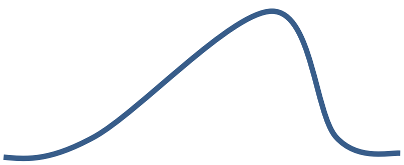
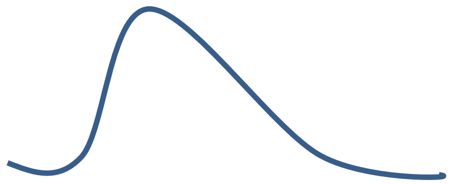
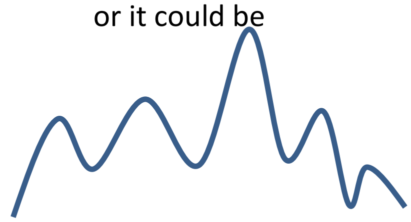
Normal Distribution – An Introduction

Let's say you collect some data from a large group of people about something like their height, weight age, or even the marks of the students in an exam. If you were to represent this data in a graph, what would the graph look like? There are several possibilities.

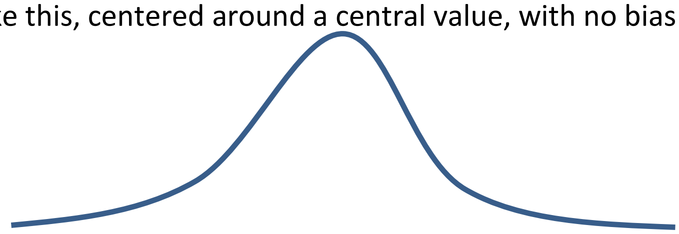
The graph could be skewed to the right or the left or it could be uneven







But, in most cases, if you have a large enough data to collect your data from, like for example, if you were to find the IQ of the entire human population, you would see that the graph would look something like this, centered around a central value, with no bias to the left or right.



This is what a Normal Distribution looks like. Its graph is a bell-shaped curve. There will be a central value, let's say the average IQ of a human being, and most people in the world will have an IQ of around that value. Therefore, you can see the curve bulging around the middle. There will be very few people in the world who are highly gifted with a very high IQ, and hence you see that the curve tapers towards the right. Similarly, there will be a low number of people with a very low IQ, and so the curve tapers towards the left as well. You'll see that heights, weights, blood pressure, income etc of the population will follow the same pattern. In a society, there are always few very rich people, few very poor, and most people will fall in the average middle class.

A Normal Distribution is a probability distribution of a continuous random variable. Let us take a look at the formal definition.

Prof. Anushri Tambe

Definition: A continuous random variable is said to have a normal or a Gaussian distribution with parameters μ (called mean) and σ (called standard deviation) if its p.d.f. is given by the probability law,

$$f(x,\mu,\sigma)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad , \begin{array}{l} -\infty < x < \infty \\ -\infty < \mu < \infty \\ \sigma > 0 \end{array}$$

We denote that a random variable x has a normal distribution by writing, $X \sim N(\mu,\sigma)$

Some properties of the normal distribution

- The graph of $f(x)$ is bell-shaped and is symmetrical about the mean μ
- Mean, median and mode of a normal distribution coincide
- $x=\mu\pm\sigma$ are the points of inflexion of the normal curve [i.e. where $f''(x)=0 \text{ and } f'''(x)\neq 0$]
- The mean deviation about the mean is given by $E(|X-\mu|)$

Normal Distribution is a very widely used concept, with many, many applications in various fields. If we're going to use the p.d.f. of a Normal Distribution to calculate the probabilities in a certain situation, you can see that the math would be very tedious, as the function denoting the p.d.f. is not exactly simple. Also, because it is so widely used, it makes more sense to develop a more standardized function for the p.d.f. This can be achieved by the substitution $Z=\frac{X-\mu}{\sigma}$.

Note that this is shift of origin and change of scale.

If $X \sim N(\mu,\sigma^2)$, then $Z=\frac{X-\mu}{\sigma}$ is a ***standard normal variate*** with E(Z) = 0 and Var(Z) = 1 and we write $Z \sim N(0,1)$. The p.d.f. of a standard normal variable is given by;

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

We shall now try out a few problems where we convert a random variable X that follows Normal Distribution to a Standard Normal Variate and vice-versa.

Solved Example

1. On a final exam in Mathematics, the mean was 75 and the s.d, 12. Determine the standard scores of students receiving the marks (a) 60 (b)78 and (c) 98

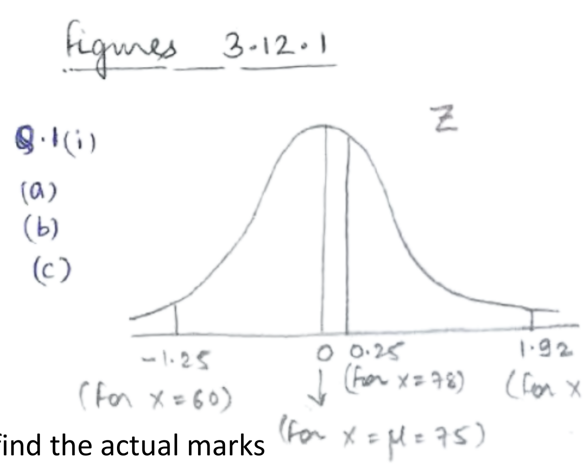
Also find the marks corresponding to the standard scores (a)-2 (b)0 (c)1.6

Assume normal distribution for marks

Solution: Let X be the random variable denoting the marks

Given: $X \sim N(\mu = 75, \sigma = 12)$ The standard scores are given by $Z = \dfrac{X-\mu}{\sigma}$

(Check Figures 3.12.1)



(a) $X = 60 \Rightarrow Z = \dfrac{60-75}{12} = -1.25$

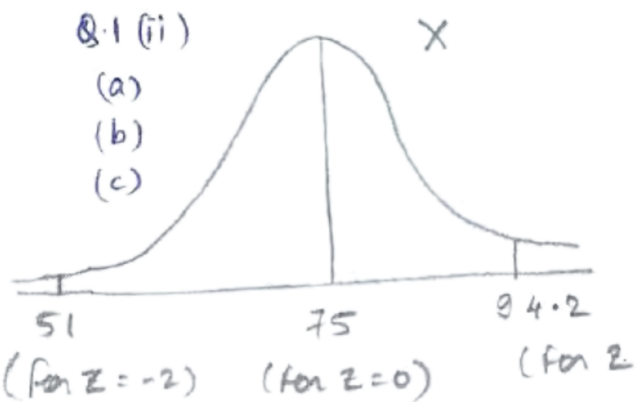
(b) $X = 78 \Rightarrow Z = \dfrac{78-75}{12} = 0.25$

(c) $X = 98 \Rightarrow Z = \dfrac{98-75}{12} = 1.9167$

Given the standard scores, we are required to find the actual marks

Now $Z = \dfrac{X-\mu}{\sigma} \Rightarrow X = \sigma Z + \mu$

(Check Figures 3.12.1)



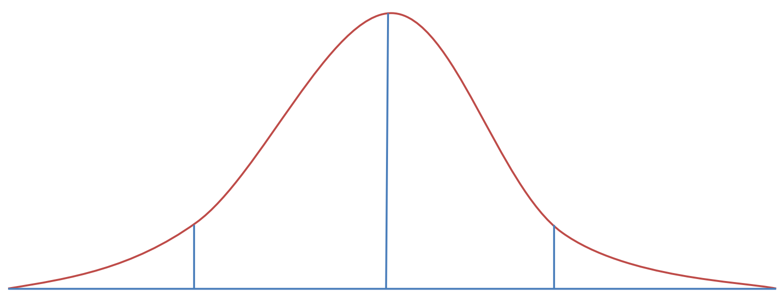
(a) $Z = -2 \Rightarrow X = 12(-2) + 75 = 51$

(b) $Z = 0 \Rightarrow X = 12(0) + 75 = 75$

(c) $Z = 1.6 \Rightarrow X = 12(1.6) + 75 = 94.2$

Prof. Anushri Tambe

The curve and the area under it:



The area under the bell-shaped curve of normal distribution denotes probability, and hence the total area under the curve is 1. The line that splits it in the middle is the line for z=0, and is the mean, median and the mode for the standard normal variate. Exactly half the area lies on either side of this line. Hence the area to the left of this line = the area to the right of this line = 0.5. The standard deviation is 1. Recall that a property of the normal curve is that the points of inflexion are $x = \mu \pm \sigma$. The smaller lines you see on either side denote the numbers +1 and -1. 68% of the total area under the curve lies between -1 and +1, that is, within the first standard deviation. If you were to go upto the second standard deviation, that is +2 and -2, you would see that over 95% of the area lies between these two numbers. And within the third standard deviation lies 99.74% of the data (area).

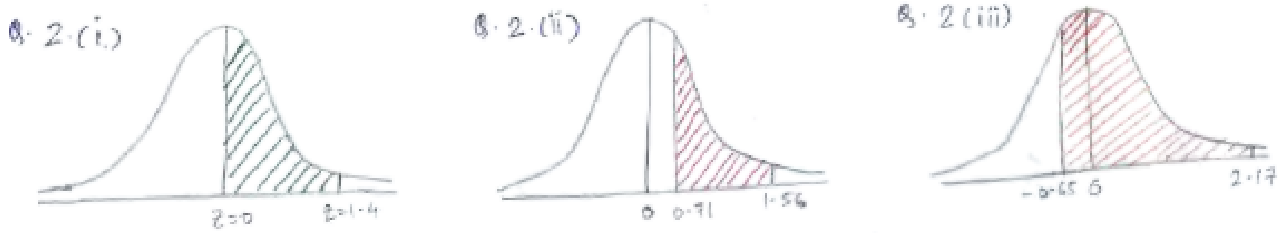
Typically in a normal distribution problem, you may have to find some part of the area under this curve. This task is made easy with the help of the Normal Distribution table.

How to read the table:

(The Normal Distribution table is uploaded separately.)

The numbers in the first column, i.e. the numbers indicating the rows are the numbers you will be looking for, which will be on the horizontal bottom of the graph. The numbers on top give the second decimal place. For example, look up the number 0.87. Scroll down to 0.8 in the rows and look for 0.07 in the columns. The number that is in the intersection of this row and column is 0.3078. This means, that the area between z=0 and z=0.87 is 0.3078. So for any given number z, the table gives you the area under the curve between 0 and z.

Prof. Anushri Tambe



Take a look at a few of these simple examples.

2. Find the area under the normal curve in each of the following cases:
   (Check Figures 3.12.2)

   (i) Between $Z = 0$ *and* $Z = 1.4$

   Simply look up 1.40 in the tables, as explained before. We get the required area to be 0.4192

   (ii) Between $Z = 0.71$ *and* $Z = 1.56$

   In order to find this area, from the tables, we can look up the value for 1.56 and it is 0.4406, i.e. the area between 1.56 and 0. The value for 0.71 is 0.2611, which is the area between 0.71 and 0. We want the area between these two numbers which is given by 0.4406-0.2611=0.1795, which is the required area.

   (iii) Between $Z = -0.65$ *and* $Z = 2.17$

   From the tables, the area from -0.65 to 0 is given by 0.2422. Since the curve is symmetric, this area is the same as between 0.65 and 0. The area between 2.17 and 0 is given by 0.4850. For the total area, we add the two. We get the required area to be 0.7272.

3. Determine the values of Z in each of the following cases:

   (Check Figures 3.12.3)

   (i) The area between 0 and Z is 0.4345

   This time, look for the number 0.4345 in the numbers in the table. We get $z = \pm 1.51$

   (ii) The area to the left of Z is 0.8621

   You will see that 0.8621 is not a number present in the table. This area is more than 0.5 and hence it must include more than half the area under the curve. Since the area to the left of z is given, z must lie to the right of the central line, so that more than 05 of the area is to the left of it. But, to read the table, we are only interested in the area bwtween z and 0. This will be given by 0.8621-0.5=0.3621. Now, from the tables we get $z = 1.09$

figures 3.12.3