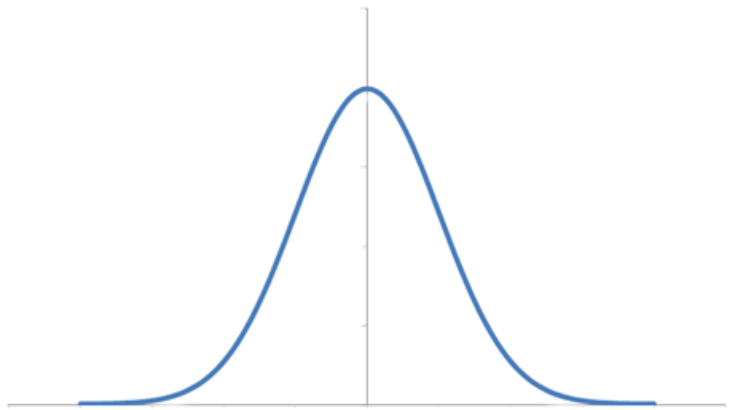
Introduction to Data Science

In the previous video

- In a normal distribution 68% of the data falls within one standard deviation of the mean
- 95% of the data falls within two standard deviations around the mean of the normal distribution
- Area under the curve would give the probability of any value falling in that range



$\mu = 150$

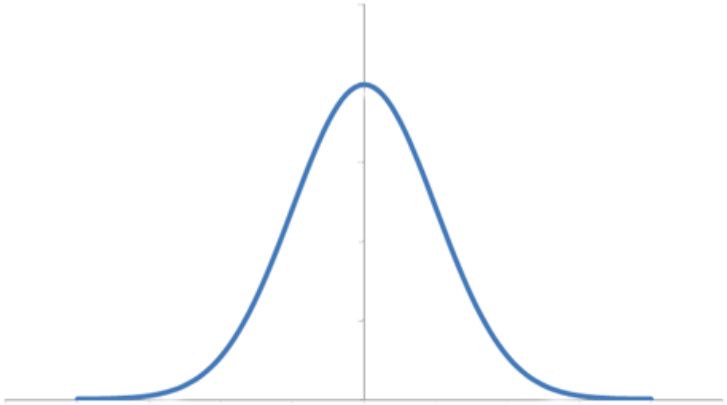
$\sigma = 15$

Analytics Vidhya

Learn everything about analytics

The Problem Statement

- Let's see where 172 falls on the distribution. How many people have cholesterol less than 172.

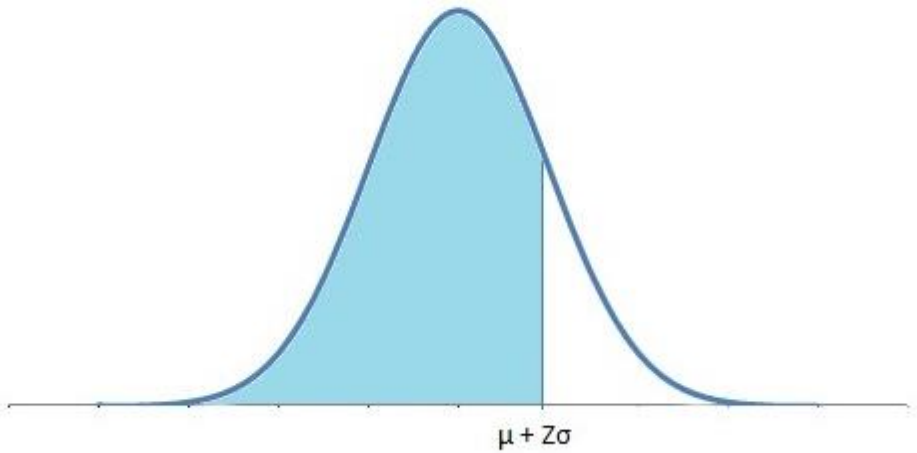


$\mu = 150$

$\sigma = 15$

Z scores

- The distance in terms of number of standard deviations, the observed value is away from the mean, is the standard score or the Z score.
- A positive Z score indicates that the observed value is Z standard deviations above the mean. Negative Z score indicates that the value is below the mean.
- Observed value = μ+zσ [μ is the mean and σ is the standard deviation]



Z Table

- For a particular z score, we can look into the table to find the probability for values to fall less than that particular z value
- For z = 1.47, let's calculate the probability-

Analytics Vidhya
Learn everything about analytics

Problem Statement

- The average IQ is 100, with a standard deviation of 15. What percentage of the population would you expect to have an IQ more than 120

Analytics Vidhya
Learn everything about analytics