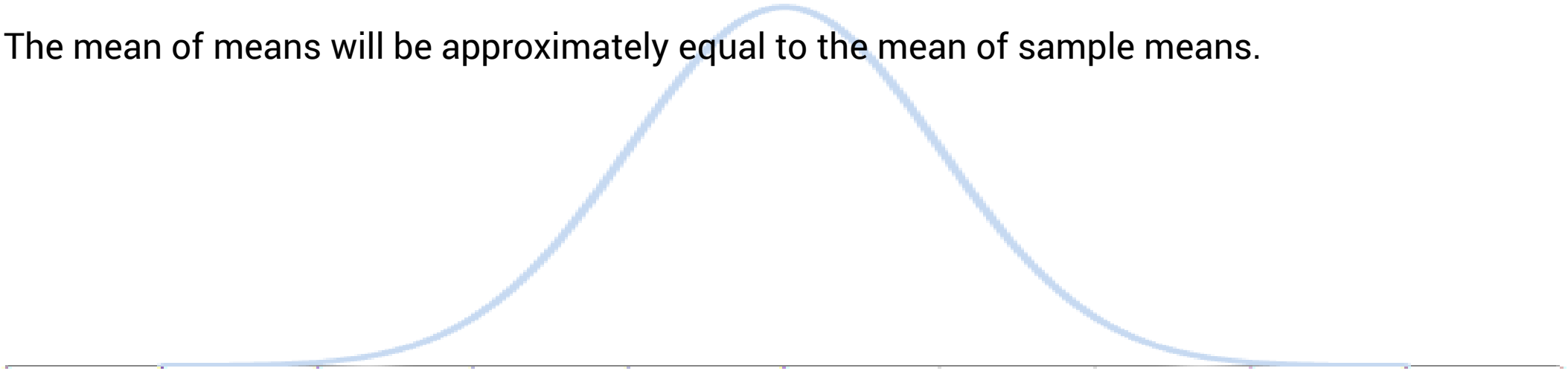
Introduction to Data Science

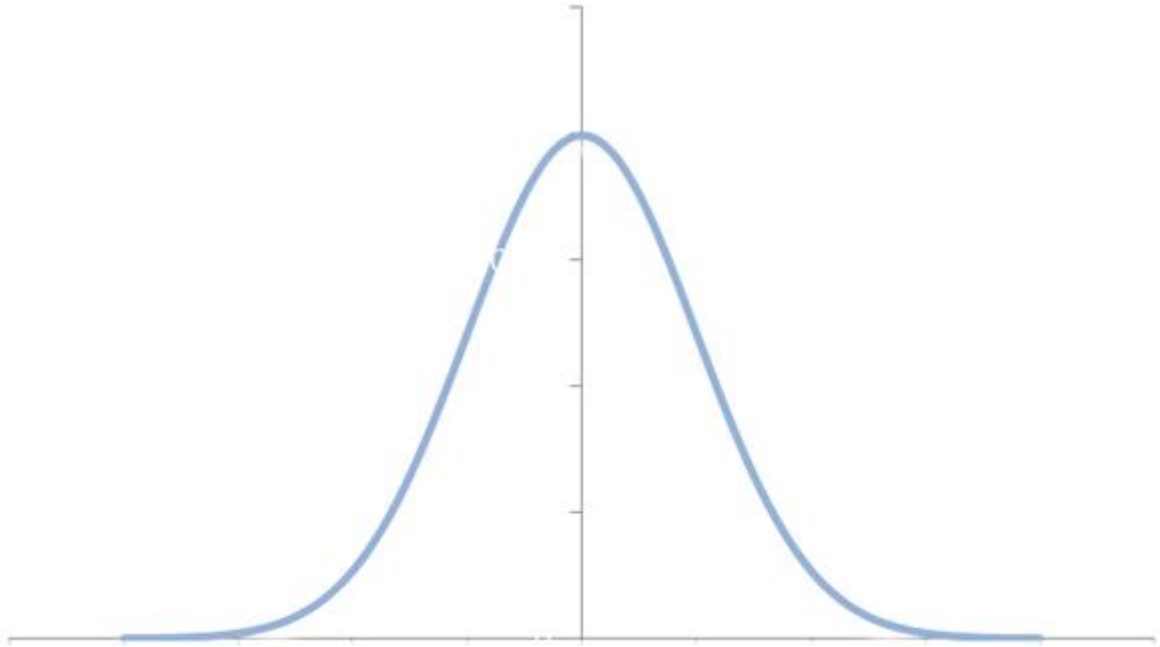
The Central Limit Theorem

- If we take means of random samples from a distribution and we plot the means, the graph approaches to a normal distribution when we have taken sufficiently large number of such samples.
- The mean of means will be approximately equal to the mean of sample means.

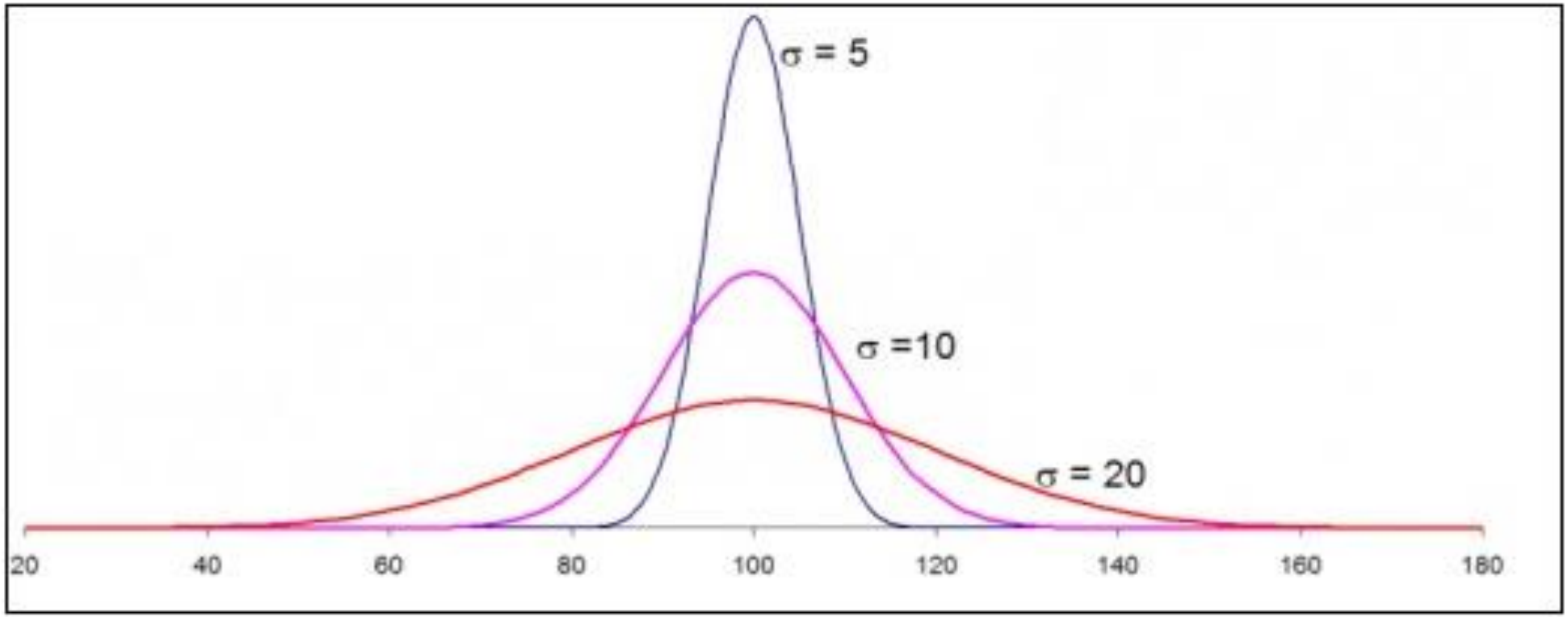


Properties of the normal distribution

- The distribution is symmetric about the mean.

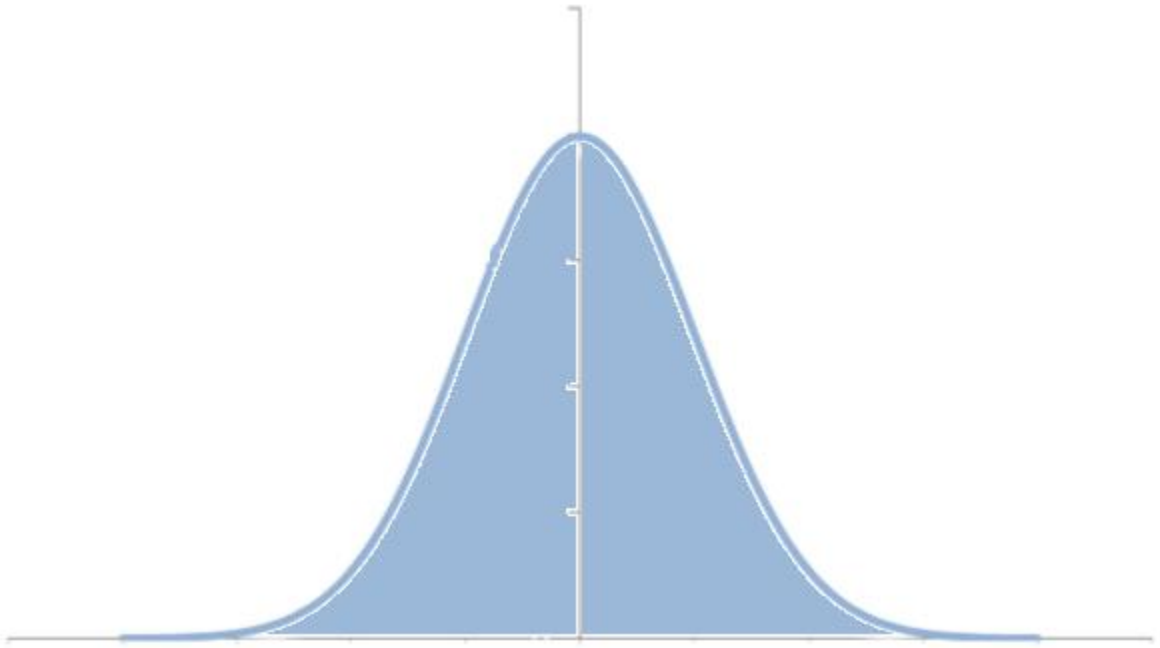


Properties of the normal distribution



Standard normal distribution

- Replace Frequencies with probabilities
- Area under the curve would be equal to 1



The Normal Distribution

- The Equation of the normal distribution is-

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{(2\sigma^2\pi)}}\left(e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)$$

- Here μ is the mean of the data while σ is the standard deviation of the data

Summarizing normal distribution

- Area under a probability density function gives the probability for the random variable to be in that range.
- If I have a population data and I take random samples of equal size from the data, the sample means are approximately normally distributed
- There is large probability for the means to be around the actual mean of the data, than to be farther away
- Normal distributions for higher standard deviations are flatter as compared to those for lower standard deviations