

PROBABILITIES, RANDOM VARIABLES AND THEIR DISTRIBUTIONS

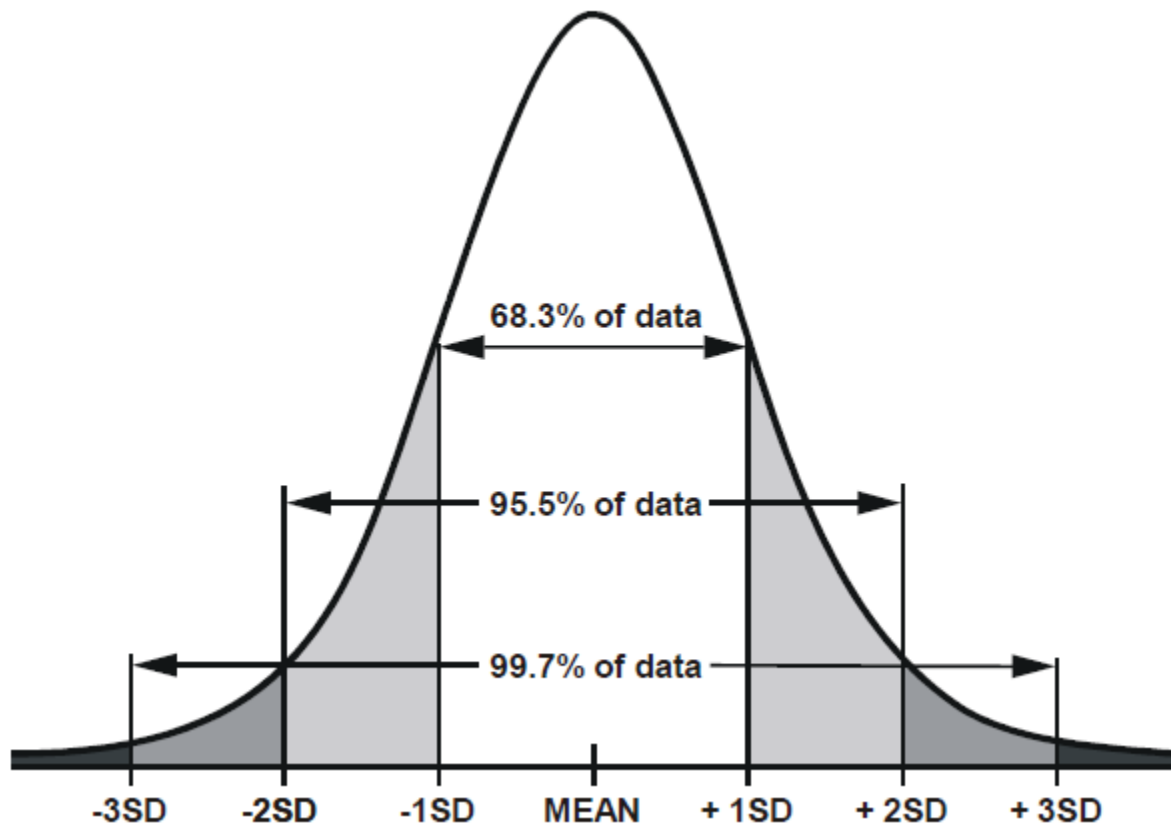
3.4 Continuous Random Variable Distributions

Continuous random variables take on an infinite number of possible values i.e. interval data e.g. a value between 1 and 2 like 1.3333899967. Height is an example of a continuous random variable. Continuous random variables are modelled using their curves and the height of the curve on the y-axis is called the **probability density function (pdf)** represented by $f(x)$, values the random variable takes on are on the x-axis.

3.4.1 Normal Distribution

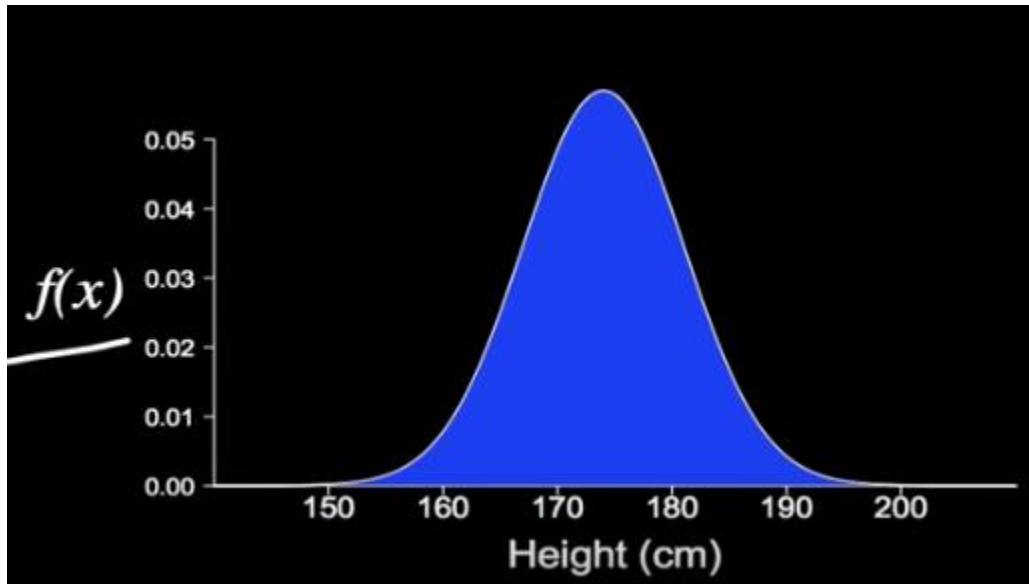
- This is the most important distribution in statistical applications because most measurements have an approximate normal distribution. Symmetrical continuous data are modelled using a normal distribution
- A normal distribution is bell-shaped and partitioned in standard deviations (sd) as seen below.

Areas under the normal curve that lie between 1, 2, and 3 standard deviations on each side of the mean



Suppose you let X be a random variable representing “height of AGG 3842 female students measured in cm”. Assume that the mean or average height on the x-axis is 175cm.

- Values where the curve is high are more likely to occur than where the curve is low. This means most students will be around 175cm. A few taller ones will be on the right end and some shorter ones on the left end.



Normal distribution curves may have different spreads but they share common attributes. The area under the normal curve represents the **probability** of an event happening between two points. The common attributes are: (refer to first distribution curve above)

- Within 1 standard deviation (sd), there is a 68.3% probability of an event occurring here. This can also be understood as 68.3% of the area lies within 1 sd.
- Within 2 sd, there is a 95.5% probability of an event occurring here.
- Within 3 sd, there is a 99.7% probability of an event occurring here.

The probability of a random variable X between two points a and b is represented by

$P(a < X < b)$ and represents the **area under the curve** between points a and b .

Properties of a normal distribution

- 1) The curve is symmetrical around the mean (μ) and it is bell shaped
- 2) The mean (μ) is at the middle and divides the area into two equal halves
- 3) Mean, median and mode for a normal distribution are all equal

- 4) The distribution is denser at the center and less dense in the tails
- 5) The total area under the curve is 1. (Remember the probability of the normal distribution is the area under the curve).

The two parameters important for a normal distribution are the **mean (μ)** and **variance (σ^2)**. A random variable X is said to be normally distributed with mean (μ) and variance (σ^2), this is represented as:

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

