**Practical 5**

**Probability Distribution& Normal distribution**

Knowledge of the probability distribution of the variables provides the clinicians and researchers with a powerful tool for describing a set of data and for reaching conclusion about population on the basis of a sample drawn from that population.

A probability distribution defines the relationship between the outcomes and their likelihood of occurrence. We have several types of distribution in statistics:

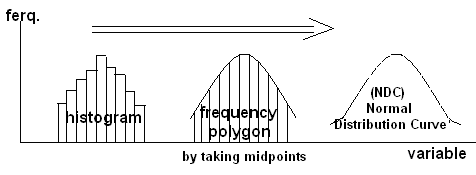
I- For discrete variables we have

1. Binomial Distribution: dichotomous outcomes (A-B, heads-tails, yes-no, on-off, is-is not, right-wrong, etc.)
2. Poisson Distribution Useful for studying rare random events.

II- But the “Normal distribution" "Gaussian distribution", for continues variables is the most important one. This is because:

* Many human variables naturally have a “bell shaped” distribution.
* The distributions are tied to probabilities, and it is the probability which will be of interest to us

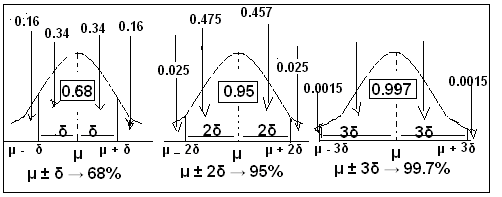
If we have a group of continuous variables with certain class interval, we can represent them by histogram and frequency polygon. But suppose we have a group of variables which is huge and the class interval is very small so the frequency polygon will take a shape of very smooth curve and that curve is called "normal distribution curve"

****

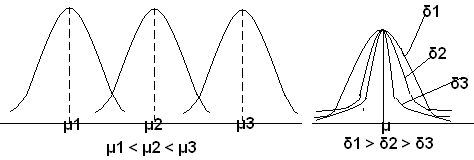
NORMAL DISTRIBUTION (Gaussian distribution): it is the most important distribution in statistics and mostly used. Parameters are population mean (μ) the measure of central tendency and standard deviation (δ) the measure of dispersion.

**Characteristics of normal distribution curve include:**

* It is used for continuous variables only.
* Symmetrical around its mean i.e. the right side is equal to the left.
* The mean, mode and median are equal.
* Total probability under the curve (area under the curve - AUC -) equals to one.
* 50% of AUC lies to the right of the mean & 50% to the left.
* Probability Limits around the mean: If you move by one standard deviation (δ) away from the mean on each side; the AUC limited by ± 1δ equals to 68% of total AUC, & so: μ ± 1δ → 68%, μ ± 2δ → 95%, μ ± 3δ → 99.7%, and as 99.7% ≈ 1(or100%) so AUC ≈ 6δ (3 on each side of μ).

****

Different values of μ and δ shift the graph of distribution along X & Y axes. If we change μ while keeping δ constant, the curve will shift to the right on increasing μ & to the left on decreasing μ. On changing δ and keeping μ constant; the curve will become more flat on increasing δ and narrower on decreasing δ without any shifting the curve to any side. μ1 < μ2 < μ3

****

* Since we know the shape of the curve, we can (using calculus) calculate the area under the curve
* The percentage of that area can be used to determine the probability that a given value could be pulled from a given distribution.
* Each normal distribution with its own values of m and s (unit) would need its own calculation of the area under various points on the curve