NORMAL PROBABILITY CURVE (Z SCORE)

INTRODUCTION:

- If large values are collected for any character and a frequency table is prepared with small class interval, the frequency curve of this data will give a bell-shaped symmetrical curve, which is known as Gaussian or normal curve.

- The shape of this curve depends on the mean and SD of data. If the standard deviation (variation) is very high, the width of the curve is also more. If we move both sides from the midpoint (x = mean), the height of the curve decreases.

NORMAL DISTRIBUTION AND MEASURES OF RELATIONSHIP

- If the mean and standard deviation are 0 and 1, respectively, then normal curve called standard normal curve. If the total area under the curve is considered as unity, the normal curve is called normal probability curve The normal distribution is

Normal Distribution Formula

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

$\mu =$ mean of x

$\sigma =$ standard deviation of x

$\pi \approx 3.14159 \ldots$

$e \approx 2.71828 \ldots$

- In this formula: u = mean distribution (it also may be median and/or more) standard deviation, and its variance is therefore mean = Properties of Normal Probability Curve

- It is a normal bell-shaped curve.

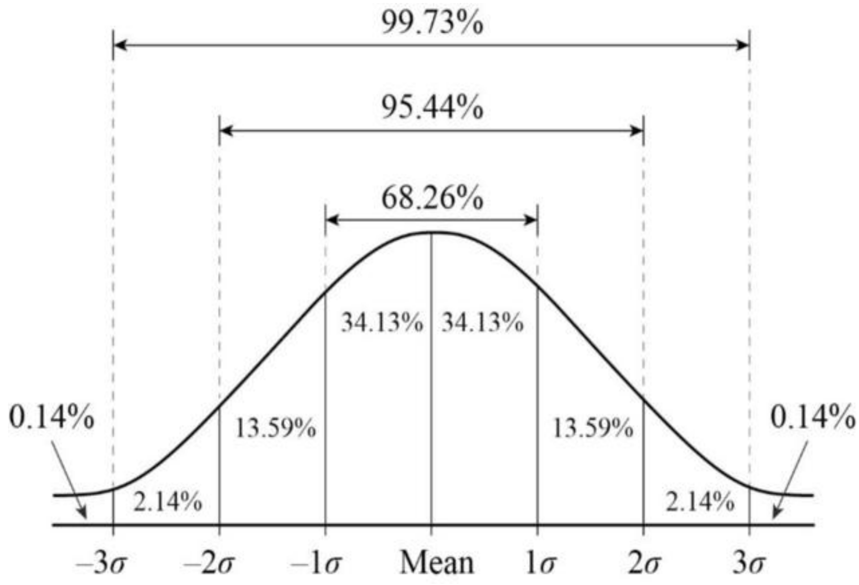
- Normal probability curve is continuous-type probability curve.

- The curve is symmetrical and asymptotic (i.e, touches at infinity).

- All the measures of central tendency are equal and stable on the highest peak axis, i.e. median mode.

- The total area under the curve is equal to unity. The quartiles Q,and Q are equidistant from the mean p and Qi = p -0.670, Qs = pi + 0.67o approximately.

- The normal curve has two parameters, i.e. mean (u) and standard deviation (o).
- . On the basis of mean (u) and standard deviation (o), the area of normal curve is distributed as:
- u = lo = 68.27%
- + 20 = 95.45%
- + 3o = 99.73%
- 0.670 = 50%
- u 1.960 = 95%
- 2.580 = 99%



- Normal curve is mesokurtic.

- Thus the total are aunder the curve above the x-axis is one square unit.
- The characteristic follows from the fact that the normal distribution is a probability distribution.
- Because of the symmetry already mentioned, 50% of the area is to the right of a perpendicular erected at the mean and 50% is to its left.

IMPORTANCE OF NORMAL PROBABILITY CURVE

- Normal distribution plays an important role in the sampling theory.
- It has been found that Data obtained from biological measurements approximately follow normal distribution Binomial and Poisson distribution can be approximated to normal distribution. For a large sample, any statistics (i.e. sample mean, sample standard deviation) approxi- mately follow normal distribution, and as such it can be studied with the help of a normal curve.
- Normal curve is used to find confidence limits of the population parameters.
- Normal distribution also forms the basis for the tests of significance.
- For example, the mean height of 500 students is 165 cm and the standard deviation is 5 cm.
- Assuming that the heights are normally distributed. Find how many students will have height between 153 and 180 cm.

SOLUTION

Here, mean (u) 165 cm, standard deviation (0) = 5 cm

We have $Z = \frac{X-\mu}{\sigma} = \frac{153-165}{5} = \frac{-12}{5} = -2.4$

$$= 180-165 = \frac{15}{5} = 3.0$$

Therefore, the proportion of students whose height is in between 153 and 180 cm

- Area under the standard normal curve between Z = -2.4 and 3.0

- (Area between Z = -2.4 and 0) + (Area between Z - 0 and 3.0)

-0.4918 +0.4987 (see Appendix V) 0.9905 Hence, the total number of students whose height is between 153 and 180 cm = 0.9905 +500 = 495 (approximately).

STATISTICAL PACKAGES AND APPLICATIONS

INTRODUCTION:

- Traditionally, in the field of healthcare biostatistics, data has been handled manually; it involves laborious work, needing lots of paperwork.
- The results often become available too late and contain different types of errors. The advent of computers and their popularity in almost every sphere of life have significantly changed the way biostatistics is now applied to health sciences.
- Computer hardware and appropriate software have optimized the various processes involved in the application of biostatistics by making the data analysis and storage easier, faster, and errorless.
- On the other hand, it has also improved the comprehension of data through the use of sophisticated graphic presentation techniques
- . However, the use of computer in statistics has its share of disadvantages as well. Just as a car cannot be expected to drive itself sensibly under an incompetent driver, in the same way, the usefulness of the results from a computer depends critically upon the statistical competence of the user
- Several computer packages are available to be used for the statistical analysis; among them commonly used packages are Microsoft Excel, SPSS, SAS, Minitab, Stata, Systat, NCSS, etc.

APPLICATION OF STATISTICS IN HEALTH AND USE OF COMPUTER IN STATISTICS

• **Microsoft Excel**: Microsoft Excel is a very popular and useful spreadsheet program thar can be used for data entry.

- It has the capacity to generate random numbers and can be used for the computation of many standard statistical applications like computation of mean, range, standard deviation, etc. It can also be used for relatively sophisticated work like computing regression and ANOVA.

- However, statistical experts point out that there are problems in the algorithms of some versions of this programme, as a result of which the results could be erroneous. For example, in earlier version blank rows or columns were being treated as zero instead of being ignored in the calculations.
- The pseudorandom number generators are also known to be faulty. It is therefore better to export Excel data files to other packages such as SPSS or SAS before doing statistical analysis. However, Excel may be used for generating graphs such as bar charts, pie charts, scattered plots, etc.

SPSS: Originally known as Statistical Package for Social Sciences, SPSS was developed in 1960 at Stanford University to help solve problems in the social sciences. SPSS now stands for Statistical Product and Services Solutions and is among the most comprehensive and popular statistical packages.

- In addition to its provisioning for the standard procedures needed in descriptive statistics, it can also be provisioned for regression analysis and analysis of variance (ANOVA)
- . It is also capable of multivariate analysis involving sophis- ticated techniques, namely cluster analysis, time series analysis, etc. The product is also capable of determining sample size and estimating power of a statistical procedure.
- To know more about SPSS, visit the website www.spss.com
- **SAS:**

The Statistical Analysis System (SAS) is very comprehensive software developed by North Carolina State University.

- This software is divided into many modules and its licensing is flexible, based upon the need for functions. This system contains a very large variety of statistical methods and is the software of choice for many major businesses, including the entire pharmaceutical industry.

- SAS has also developed a PC SAS, which is compatible with the personal computer and has a user-friendly Windows interface. The more useful components of SAS are Base SAS, SAS/STAT, SAS/GRAPH, etc. An advan- tage of SAS is its capability to transport data files in various formats and convert them to SAS data set without much effort. To learn more visit their website www.asa.com.

- **Minitab:** This is another statistical package designed to facilitate the teaching of statistical methods by using computer.

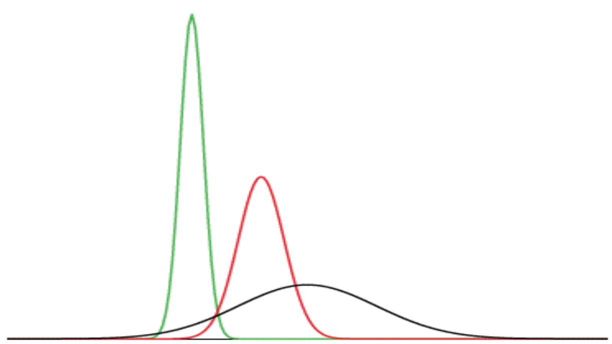
- The Minitab is also very user-friendly product with well- designed documentation facilities being widely used in educational institutions. To learn more about Minitab visit their website www.minitab.com.

The Normal Distribution

INTRODUCTION:

- It is sometimes called the "bell curve,". It is also called the "Gaussian curve" after the mathematician Karl Friedrich Gauss. Strictly speaking, it is not correct to talk about the normal distribution" since there are many normal distributions.

- Normal distributions can differ in their means and in their standard deviations shows three normal distributions. The green (left-most) distribution has a mean of -3 and a standard deviation of 0.5, the distribution in red (the middle distribution) has a mean of O and a standard deviation of 1, and the distribution in black (right-most) has a mean of 2 and a standard deviation of 3. These as well as all other normal distributions are symmetric with relatively more values at the center of the distribution and relatively few in the tails.

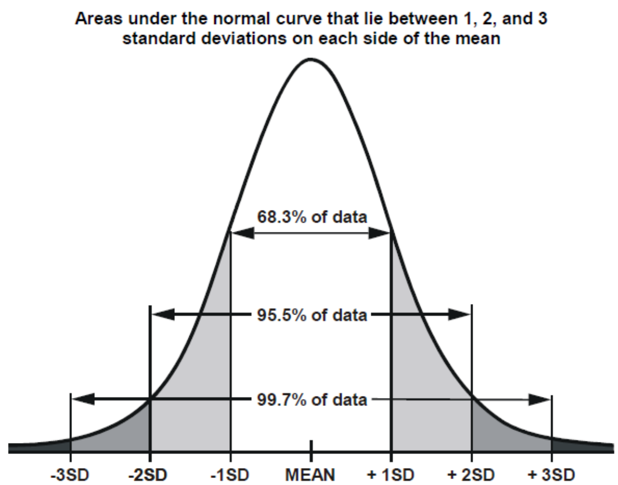


FEATURES OF NORMAL DISTRIBUTION

Normal distributions are symmetric around their mean. The mean, median, and mode of a normal distribution are equal. The area under the normal curve is equal to 1.0.

Normal distributions are denser in the center and less dense in the tails. Normal distributions are defined by two parameters, the mean (u) and the standard deviation (o). 68% of the area of a normal distribution is within one standard deviation of the mean. Approximately 95% of the area of a normal distribution is within two standard deviations of the mean.

The Normal Distribution:

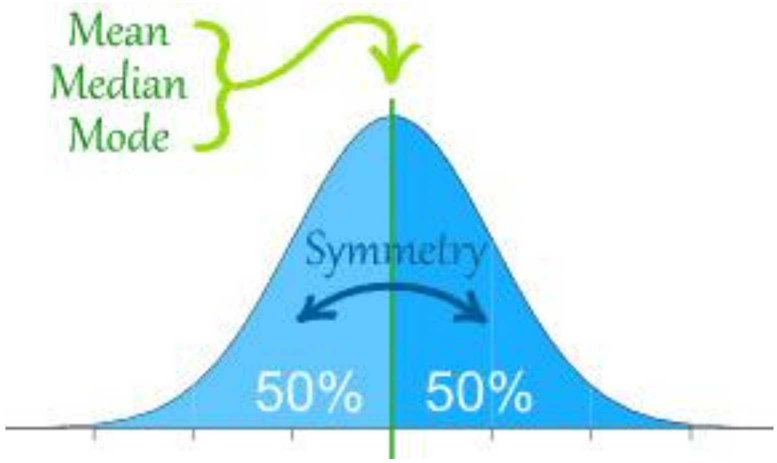


- It is often called a "Bell Curve"because it looks like a bell.

  Many things closely follow a Normal Distribution:

- heights of people
- size of things produced by machines
- .errors in measurements
- blood pressure
- marks on a test

The Normal Distribution has:



- mean = median = mode
- symmetry about the center
- 50% of values less than the mean and 50% greater
- than the mean.
- Distributions of data and probability distributions are not all the same shape.

Some are asymmetric and skewed to the left or to the right. Other distributions are bimodal and have two peaks. Another feature to consider when talking about a distribution is the shape of the tails of the distribution on the far left and the far right. Kurtosis is the measure of the thickness or heaviness of the tails of a distribution.

The kurtosis of a distributions is in one of three categories of classification:

- Mesokurtic
- Leptokurtic
- Platykurtic

1-MESOKURTIC

Kurtosis is typically measured with respect to the normal distribution. A distribution that has tails shaped in roughly the same way as any normal distribution, not just the standard normal distribution, is said to be meso kurtic. The kurtosis of a mesokurtic distribution is neither high nor low, rather it is considered to be a baseline for the two other classifications. Besides normal distributions, binomial distributions for which p is close to 1/2 are considered to be meso kurtic.

2-LEPTOKURTIC

A leptokurtic distribution is one that has kurtosis greater than a mesokurtic distribution. Leptokurtic distributions are sometimes identified by peaks that are thin and tall. The tails of these distributions, to both the right and the left, are thick and heavy. Leptokurtic distributions are named by the prefix "lepto" meaning "skinny." There many examples of leptokurtic distributions. One of the most well known leptokurtic distributions is Student's t distribution.

3-PLATYKURTIC

The third classification for kurtosis is platykurtic. Platykurtic distributions are those that have slender tails. Many times they possess a peak lower than a mesokurtic distribution. The name of these types of distributions come from the meaning of the prefix "platy" meaning "broad." All uniform distributions are platykurtic. In addition to this the discrete probability distribution from a single flip of a coin is platykurtic.