Kenya Medical Training College
Department of Clinical Medicine
Year Two Semester One
Probability
10^{th} December 2020

Willis J. Opalla

Learning Objective

- To apply the principles of probability in healthcare related decision-making.

Learning Outcomes

- By the end of this session, you should be able to
  1. Define probability.
  2. Explain the types of probability.
  3. Explain the laws of probability.
  4. Explain the implication of probability with regard to the normal distribution curve.
  5. Apply the probability equations in making statistical inferences on relationships between variables.

Probability

- Probability is a measure of the likelihood that an event will occur in a random experiment.
- It is a number that reflects the chance or likelihood that a particular event will occur.
- It is quantified as a number between 0 and 1, where, 0 indicates impossibility and 1 indicates certainty.
- Probabilities can be expressed as proportions that range from 0 to 1, or as percentages ranging from 0% to 100%.

Probability

- A probability of 0 indicates that there is no chance that a particular event will occur.
- A probability of 1 indicates that an event is certain to occur.
- A probability of 0.45 (45%) indicates that there are 45 chances out of 100 of the event occurring.

Priori or Classical Probability

- If there are a total of *n* mutually exclusive and equally likely outcomes of a trial and if *nA* of these outcomes have an attribute *A*, then probability of *A* is fraction *nA/n*.
- This definition is used to find out the probability of the outcomes in the following cases:

1. Suppose a die is tossed, what is the probability of 4 coming up? Since there are six mutually exclusive and equally likely outcomes out of which 4 is only one, the probability of 4 coming up is 1/6

Priori or Classical Probability

2. Suppose 2 coins are tossed, there can be the following outcomes: HH, TH or HT and TT. The probability of HH or TH or HT or TT only = ¼.
3. Suppose 2 dice are cast and the probability of a total of 7 points is to be determined, a total of 7 can come in 6 ways (1-6, 2-5, 3-4, 4-3, 5-2, or 6-1). Since each die has 6 sides, the total number of equally likely mutually exclusive outcomes is 6 X 6 = 36. So the chance of getting a total of 7 when 2 dice are cast is 6/36.
4. Suppose there is a box containing 12 red beads and 8 green beads. The chances of getting a red bead is 12/20.

Posteriori or Frequency Probability

- The estimate of probability of a specified outcome based on a series of independent trials is given by

$$\text{Probability} = \frac{\text{The number of times the outcome occurred}}{\text{Total number of trials}}$$

- Sometimes this probability is referred to as *statistical probability*, frequency or empirical probability or a posterior probability i.e. after the event.

Posteriori or Frequency Probability

- Example:
  - To know the probability of success of a surgical procedure, a review of past experience of this surgical procedure under similar conditions will provide the data for estimating this probability.
  - The longer the series, the closer the estimate would be to the true value.

Laws of Probability for Independent Events

- Two important laws of probability which are useful in finding out probabilities in complex situations where the events concerned are independent.
  1. The Addition Law
  2. The Multiplication Law

The Addition Law

- If an event can occur in any one of several mutually exclusive ways, the probability of that event is the sum of the individual probabilities of the different ways in which it can occur.
  - For example, when a die is cast, what is the probability of getting 2 or 4 or 6?

The probability of 2 = 1/6

The probability of 4 = 1/6

The probability of 6 = 1/6

  - The probability of 2 or 4 or 6 = 1/6+1/6+1/6

    =3/6 =1/2

The Multiplication Law

- The probability of the simultaneous occurrence of 2 or more independent events is the product of the individual probabilities.
- For example, in tossing 2 coins

  Probability of head in one coin = ½

  Probability of head of another coin = ½

  Thus probability of head in both coins = ½ x ½

  $= \frac{1}{4}$

Probability

- The concept of probability can be illustrated in the context of a study of obesity in children 5-10 years of age who are seeking medical care at a certain pediatric facility.
- The population (sampling frame) includes all children who were seen in the facility in the past 12 months and is summarized in the table.

Unconditional Probability

	Age (Years)						
	5	6	7	8	9	10	Total
Boys	432	379	501	410	420	418	1560
Girls	408	513	412	436	461	500	2730
Total	840	892	913	846	881	918	5290

- A randomly selected child will have the equal probability of other children and it is 1/N, where N=the population size.
- Thus, the probability that any child is selected is 1/5,290 = 0.0002.

Conditional Probability

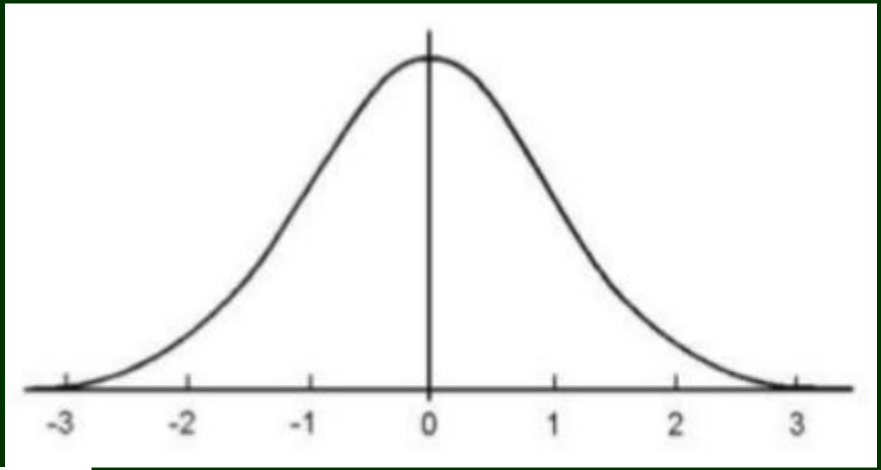
	Age (Years)						
	5	6	7	8	9	10	Total
Boys	432	379	501	410	420	418	1560
Girls	408	513	412	436	461	500	2730
Total	840	892	913	846	881	918	5290

- A purposeful selection of a population subset such as probability of 9 year old girls.
- This can be computed by the formula 461/2730 = 0.169 (16.9%).

Properties of Normal Probability Curve (Z Score)

- It is also called as normal distribution or Z score.
- It is based on the area or distribution of data.
- It is a bell shaped curve.
- Its mid-point is symmetrical if the Mean = Median = Mode,

  i.e. $(\bar{X}=M=Z)$



Normal Probability Curve (Z Score)

- When the Mean, Median and Mode are equal at the centre of the curve it is denoted as "μ" (mu).
- The line of the curve is extended to infinity at left side as well as right side.
- Total area of the normal curve is taken as "1".
- 1 is indicative of the maximum probability.

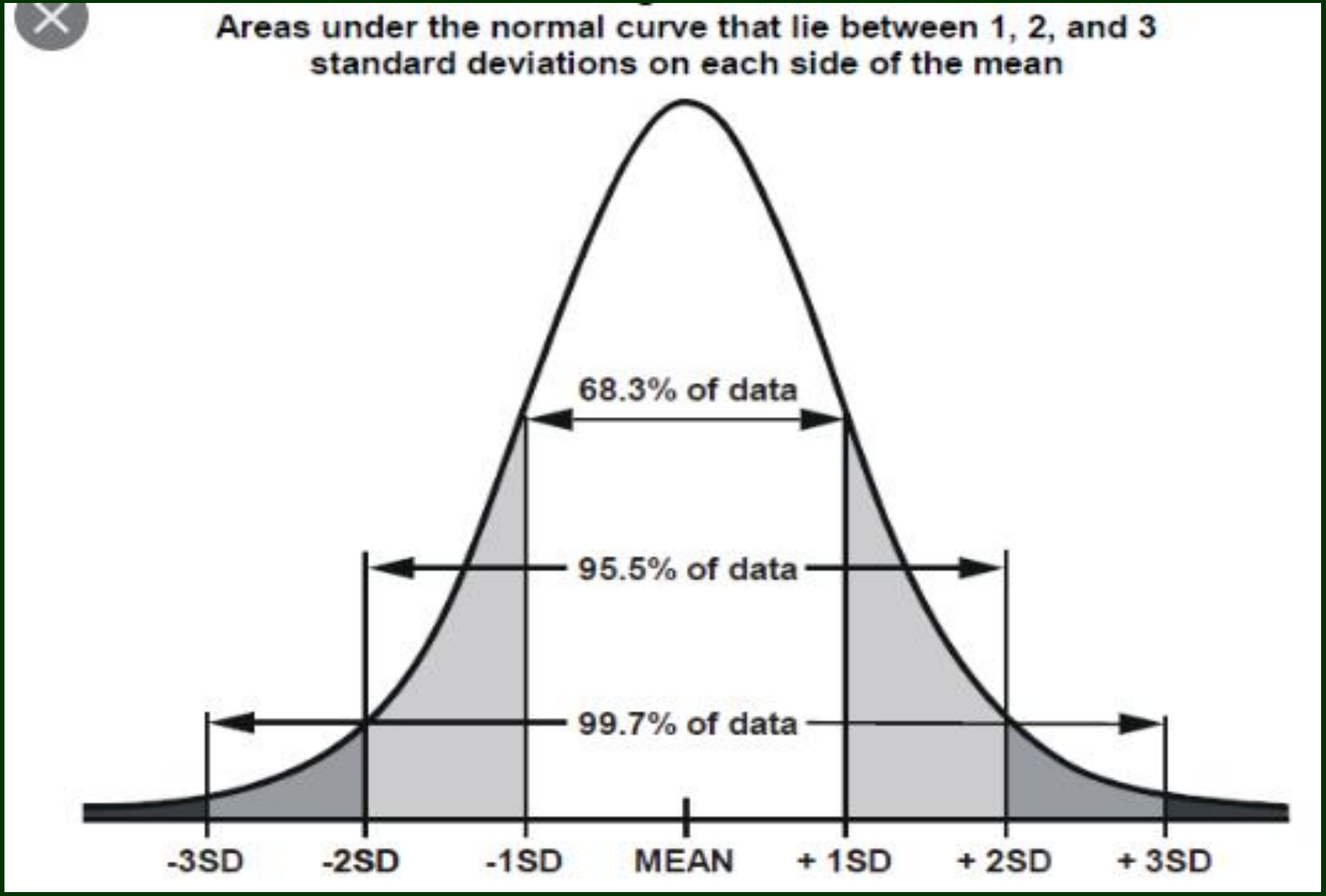
Normal Probability Curve (Z Score) Properties

- The curve is also called Gaussian or normal curve.
- The shape of the curve depends on mean and SD.
- If SD is large then width increases, but height decreases and vice versa.
- When the mean is 0 and SD is 1 the curve is said to be standard normal curve or Gaussian.

Characteristics of Normal Distributions

- Distributions that are normal or Gaussian have the following characteristics:
  1. Approximately 68% of the values fall between the mean and one standard deviation (in either direction)
  2. Approximately 95% of the values fall between the mean and two standard deviations (in either direction)
  3. Approximately 99.9% of the values fall between the mean and three standard deviations (in either direction).

Properties of Normal Probability Curve (Z Score)



Normal Probability Curve (Z Score)

- For a normally distributed variable with known population mean (μ) and standard deviation, the probability of particular values can be determined based on the equation for the normal probability model:

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

where

μ is the population mean,

σ is the population standard deviation,

π is a constant = 3.14159,

e is a constant = 2.71828.

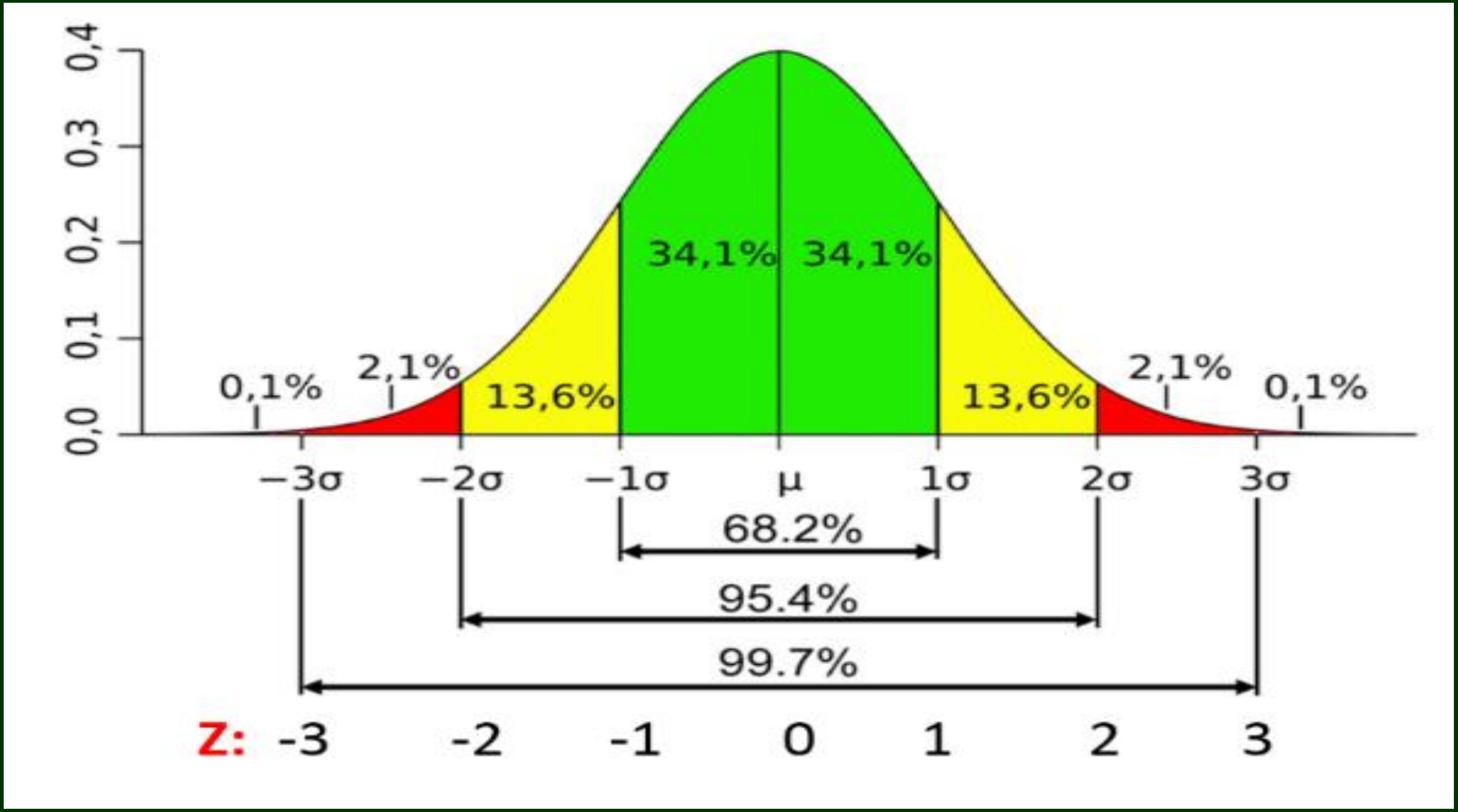
Gaussian

- The standard normal distribution is a special normal distribution that has a mean=0 and a standard deviation=1.
- This is very useful for answering questions about probability because once how many standard deviations a particular result lies away from the mean is calculated, the probability of seeing a result greater or less than that can easily be determined.

Gaussian

- The following figure shows the percentage of observations that would lie within 1, 2, or 3 standard deviations from any mean in a data set that is more or less normally distributed.
- For a given value in the distribution, the Z score is the number of standard deviations above or below the mean.

Gaussian



0,4
0,3
0,2
0,1
0,0
34,1% 34,1%
13,6% 13,6%
2,1% 2,1%
0,1% 0,1%
−3σ −2σ −1σ μ 1σ 2σ 3σ
68.2%
95.4%
99.7%
Z: -3 -2 -1 0 1 2 3

Gaussian

- What is the probability of a value *less than the mean*? The obvious answer is 50%.
- What is the probability of a value less than I SD *below* the mean?

  P= 13.6+2.1+0.1=15.8%

- What is the probability of a value less than I SD *above* the mean?

  P= 34.1+34.1+13.6+2.1+0.1=84%

The Z Score

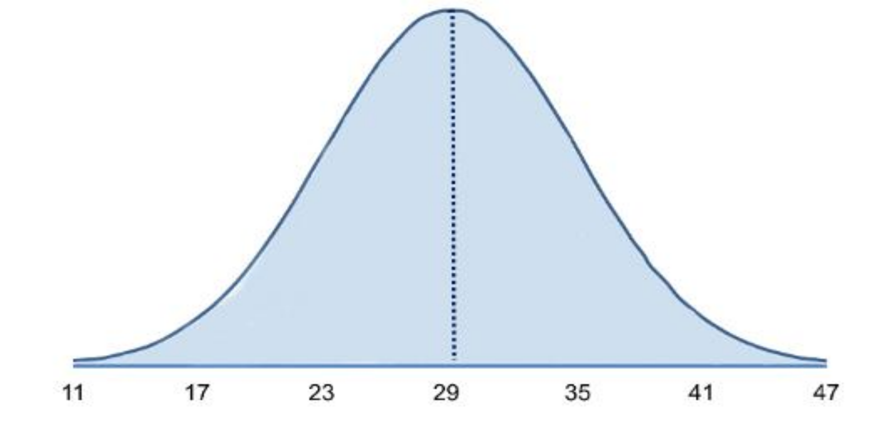
- Z score is the number of standard deviation units a given observation lies above or below the mean.
- There are tables and computer functions that can tell the probability of a value less than a given Z score.
- Example:
  - What is the probability of a Z score < 0?

    Answer: P= 34.1 + 13.6 + 2.1+ 0.1 = 50%
  - What is the probability of a Z score < +1?

    Answer: P= 34.1+34.1+13.6+2.1+0.1=84%

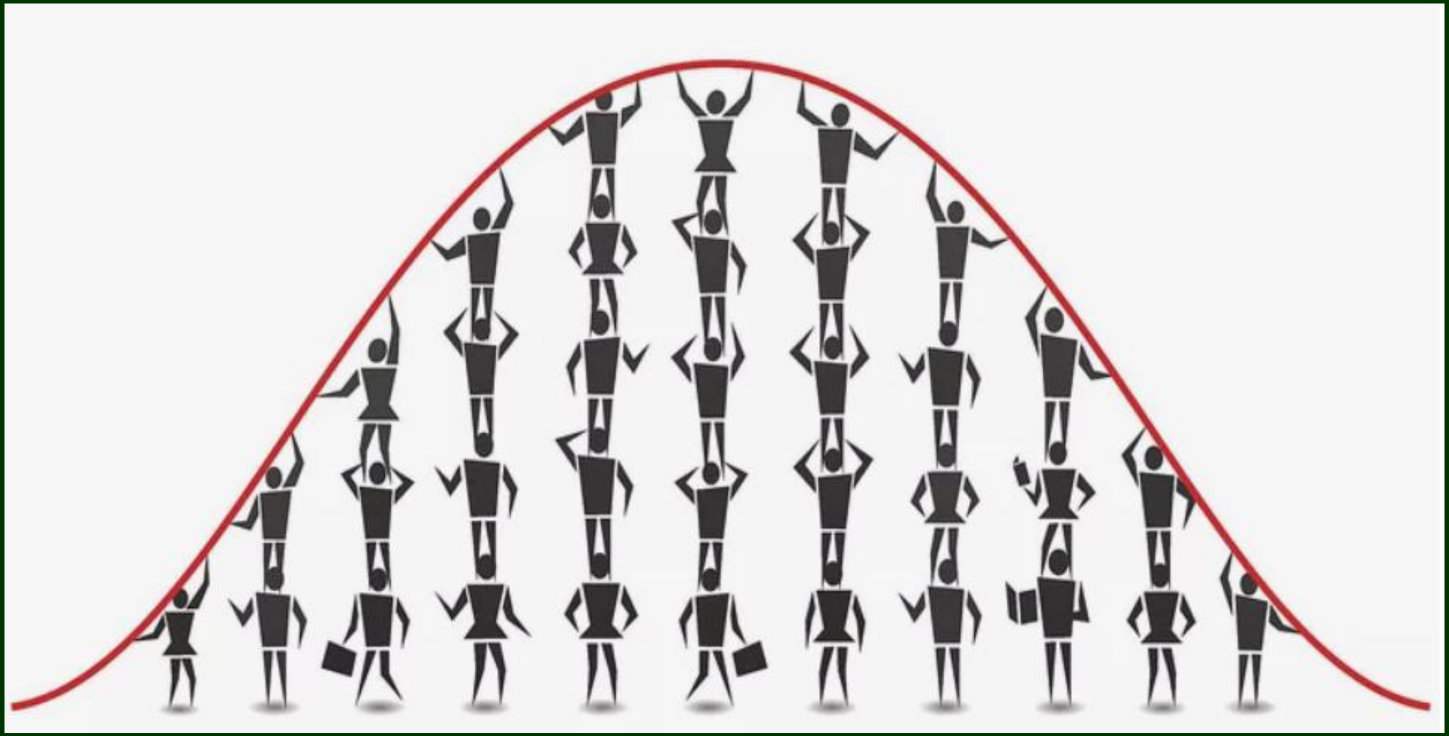
Properties of Normal Probability Curve (Z Score)

- Worked Example
  - Consider BMI in a population of 60 year old males in whom BMI is normally distributed and has a mean value = 29 and a standard deviation = 6.
  - The standard deviation gives a measure of how spread out the observations are.



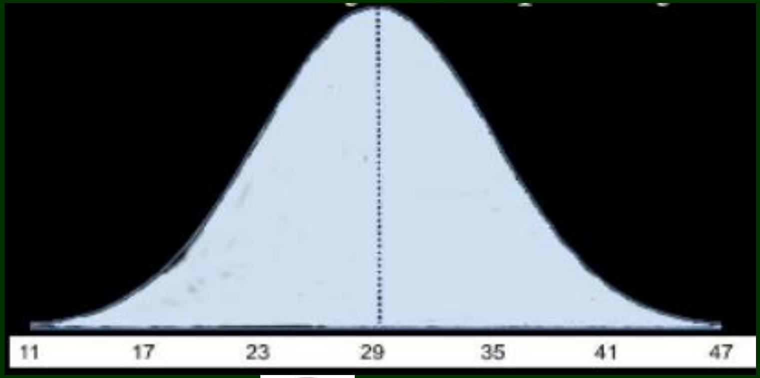
Properties of Normal Probability Curve (Z Score)

- Worked Example



Worked Example

- The mean ($\mu = 29$) is in the centre of the distribution and the horizontal axis is scaled in increments of the standard deviation ($\sigma = 6$) and the distribution essentially ranges from $\mu - 3\sigma$ to $\mu + 3\sigma$.
- It is possible to have BMI values below 11 or above 47, but such extreme values occur very infrequently.

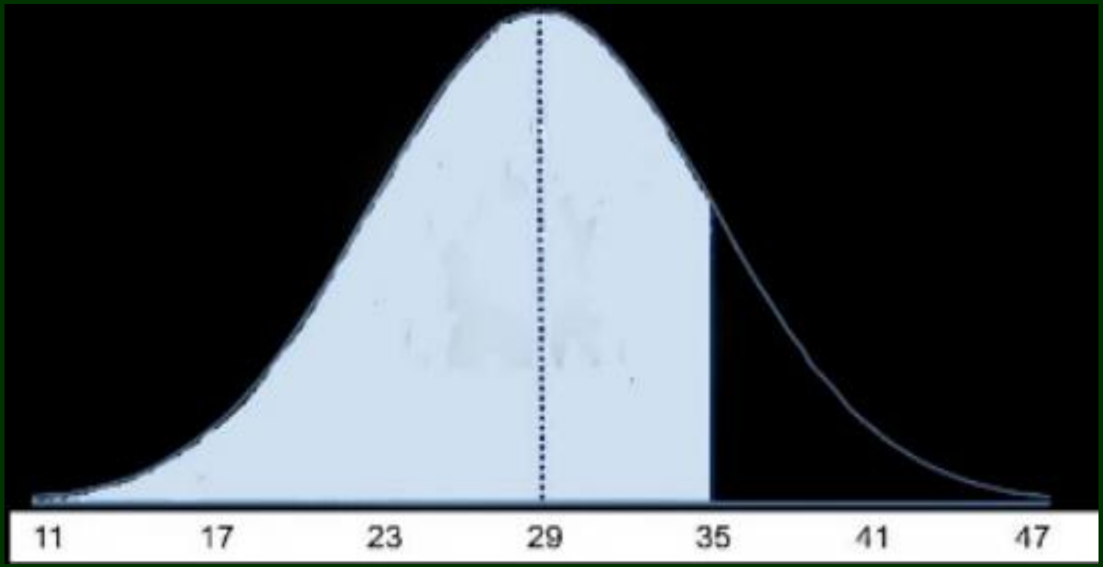


Worked Example

- Areas under the curve are computed to compare probabilities from normal distributions.
  - The total area under the curve is 1.
  - Here the *mean* is equal to *median*, so half (50%) of the area under the curve is above the mean and half is below, so Probability of BMI $< 29 = 0.50$.
  - Consequently, if a man is selected at random from this population, what is the probability that his BMI is less than 29?
  - The answer is 0.50 or 50%, since 50% of the area under the curve is below the value BMI = 29.

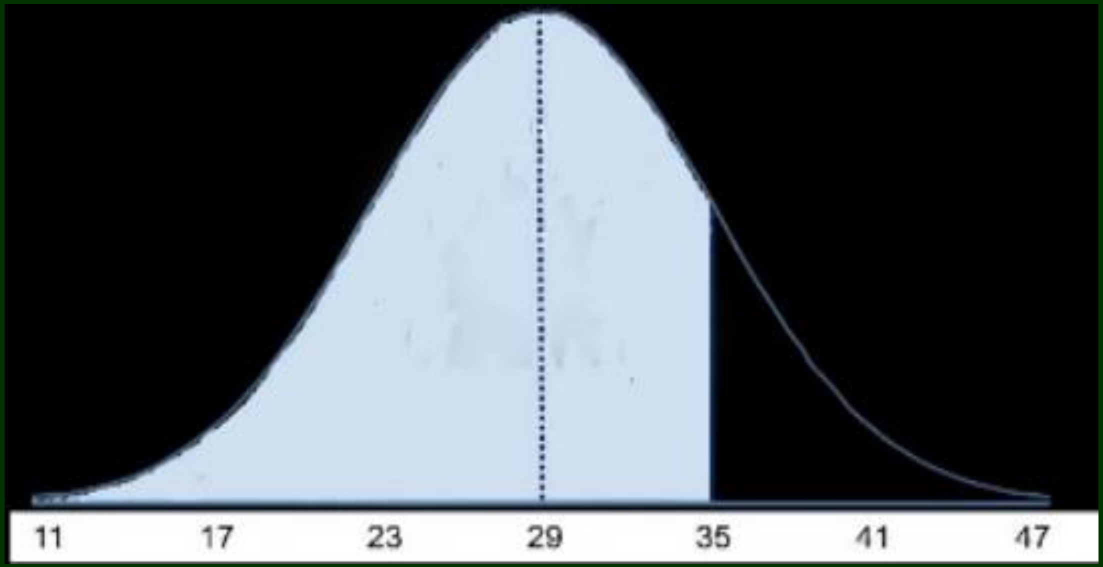
Worked Example

- What is the probability that a 60 year old male has BMI less than 35?
- The probability is displayed graphically and represented by the area under the curve to the left of the value 35 in the figure below.



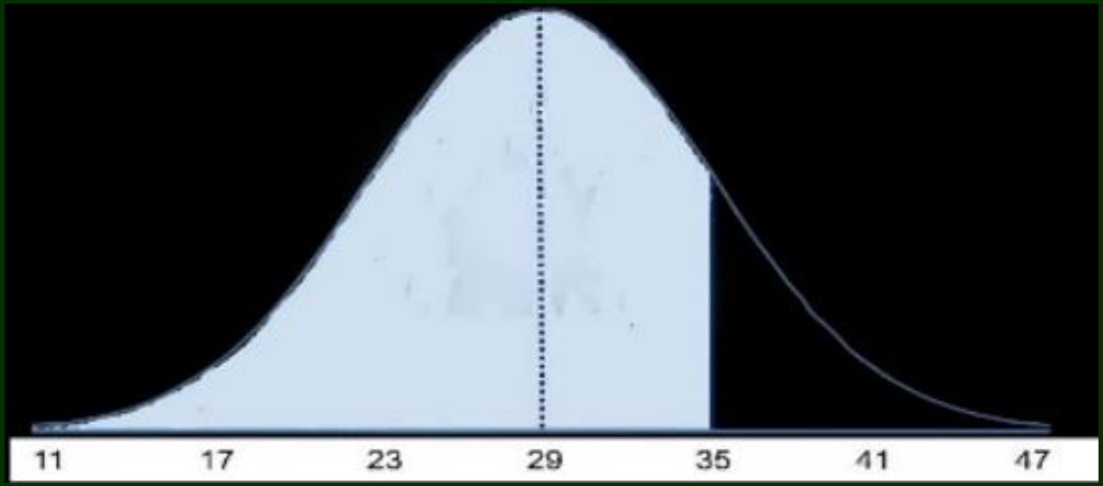
Worked Example

- Note that BMI = 35 is 1 standard deviation above the mean.
- For the normal distribution, approximately 68% of the area under the curve lies between the mean plus or minus one standard deviation.



Worked Example

- 68% of the area under the curve therefore lies between 23 and 35.
- Normal distribution is symmetric about the mean, therefore $P(29 < X < 35) = P(23 < X < 29) = 0.34$.
- Consequently, $P(X < 35) = 0.5 + 0.34 = 0.84$ or 84%.



Worked Example

- This can also be calculated using the formula:

  $Z = X - \mu / \sigma$.

  where

  μ is the mean,

  σ is the standard deviation of the variable X.

- In order to compute $P(X < 30)$, the $X=30$ is converted to its corresponding Z score.

  $Z = (30-29)/6$

  $= 1/6 = 0.17$ (refer the Z table for corresponding value i.e 0.0675)

  $= 0.0675 + 0.5 = 0.5675 = 56.75\%$

Worked Example 2

- The mean height of 500 students is 165 cm and the SD is 6. Assuming that heights are normally distributed, how many students will be between 155 and 175cm.

  $Z = (X - \mu) / \sigma$.

  $Z = (155-165)/6 = -10/6 = -1.67$

  $Z = (175 - 165)/6 = 10/6 = 1.67$

  Area under the standard normal curve is btwn Z = -1.67 and 1.67.

  = (area btwn Z = -1.67 and 0) + area btwn Z = 0 and 1.67.

  = (0.9525 – 0.5 = 0.4525) + 0.4525 = 0.9050

  = 90.5%

  = (0.9050x500) = 452.5 = 452 students are having height between 155cm to 175cm.

Importance of Normal Probability Curve

- Data obtained from biological measurements approximately follow normal distribution.
- Binominal and Poisson distribution can be approximated to normal distribution.
- Binominal distribution is a fixed trial with limited probability.
- It can have only two results, e.g. tossing a coin.

Binomial Distribution

- Knowing the proportion of individuals in a population who possess a particular character may be of interest.
- An estimate of this proportion is calculated on the basis of a suitably drawn sample from this population and the corresponding sampling distribution.
- The sampling distribution is given by a theoretical frequency distribution known as *Binomial distribution.*

Binomial Distribution

- Binomial distribution is the theoretical frequency distribution that is used to determine the sampling distribution.
- It gives knowledge about the proportion of a population that possesses a particular characteristic.

  e.g. Healthy and Sick people in a population.

Binomial Distribution

H S

H S H S

H S H S H S H S

H S H S H S H S H S H S H S H S

Binomial Distribution

- H: healthy person and S: the sick person.
- The probabilities of various compositions are obtained by using addition and multiplication laws on probability.
- The probability of a sick person being chosen is 0.4 and the probability of a healthy person being chosen is 0.6.
- One person being sick SHHH, HSHH, HHSH, HHHS: 0.4 x 0.6 x 0.6 x 0.6 + 0.6 x 0.4 x 0.6 x 0.6 + 0.6 x 0.6 x 0.4 x 0.6 + 0.6 x 0.6 x 0.6 x 0.4

  = 0.3456

Poisson Distribution

- Poisson distribution is an infinite trial with multiple outcome of results, e.g. printing mistakes of a book.
- In case of large samples it can be used to study the descriptive statistics such as mean, SD etc.
- It is used to find confidence limits of the population parameters.
- It is the basis of test of significance.
- Useful for measuring those events that did not occur.

Poisson Distribution

- There are situations in which the number of times an event occurs can be counted but the number of times the event did not occur cannot be counted.
- The amount of radioactive fallouts during a certain duration, following a nuclear explosion can be measured but not amount of nuclear dust that has not fallen during this period.

Poisson Distribution

- The probability are generated by the mathematical function

$$\frac{e^{-m}\, m^{x}}{x!}$$

- Where *x* is the number of times the event occurs, *m* is the mean of the distribution and *x!* =1 x 2 x 3 x … x *X* and *e* is the natural logarithm.
- The Binomial and Poisson distribution address the occurrence of distinct events such as the number of sick, number of accidents, number of cells, number of coins etc.

Sample Study Questions

1. A coin is thrown 3 times .what is the probability that at least one head is obtained?
2. There are 5 green 7 red balls. Two balls are selected one by one without replacement. Find the probability that first is green and second is red.
3. Three dice are rolled together. What is the probability as getting at least one '4'?
4. A problem is given to three persons P, Q, R whose respective chances of solving it are 2/7, 4/7, 4/9 respectively. What is the probability that the problem is solved?
5. Fifteen people sit around a circular table. What are odds against two particular people sitting together?
6. Three bags contain 3 red, 7 black; 8 red, 2 black, and 4 red & 6 black balls respectively. 1 of the bags is selected at random and a ball is drawn from it. If the ball drawn is red, find the probability that it is drawn from the third b

Sample Study Questions

7. In a class, 40% of the students study math and science. 60% of the students study math. What is the probability of a student studying science given he/she is already studying math?
8. Five men and five women are ranked according to their scores on an exam. Assume that no two scores are the same and all possible rankings are equally likely. Let the random variable X be the highest ranking achieved by a women. What is the probability mass function of X?

Sample Study Questions

9. Most graduate schools of business require applicants for admission to take the Graduate Management Admission Council's GMAT examination. Scores on the GMAT are roughly normally distributed with a mean of 527 and a standard deviation of 112. What is the probability of an individual scoring above 500 on the GMAT?
10. From Qn 9, how high must an individual score on the GMAT in order to score in the highest 5%?
11. The length of human pregnancies from conception to birth approximates a normal distribution with a mean of 266 days and a standard deviation of 16 days. What proportion of all pregnancies will last between 240 and 270 days (roughly between 8 and 9 months)?
12. What length of time marks the shortest 70% of all pregnancies?

Summary

- Probability is a measure of the likelihood that an event will occur in a random experiment.
- Binomial distribution gives knowledge about the proportion of a population that possesses a particular characteristic.
- Poisson distribution is used to find confidence limits of the population parameters, is the basis of significance test and is useful for measuring those events that did not occur.

References

- Joseph, J. K. (n.d) *Measures of Relationship*, [Online] Available: https://www.slideshare.net/JohnykuttyJoseph/measures-of-relationship, (Retrieved 26.11.2020)
- Kothari, C. R., (2004) *Research Methodology, Methods and Techniques*, 2nd ed., New Age International Publishers, New Delhi.