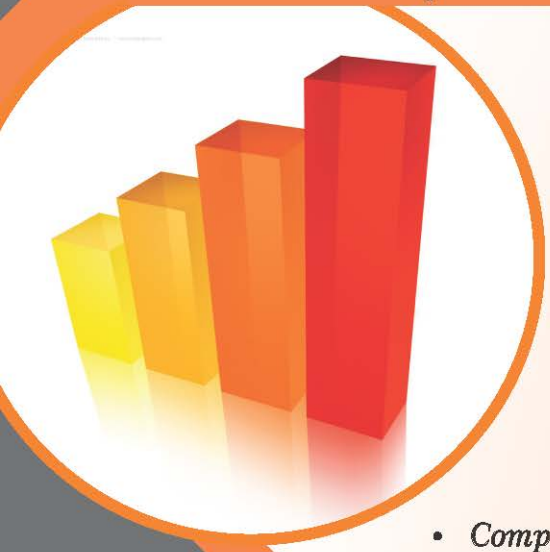
BIOSTATS

Simplified for PGMEE & FMGE

Dr. Mukhmohit Singh
Dr. Shveta Saini



- *Complete review of Biostatistics for PG Entrance and FMG Exams*
- *MCQs based on pattern from recent PG Entrance exams from AIIMS, AIPG, PGI, JIPMER, NEET, DNB and other central exams*
- *Numericals solved on Sample size calculations, Standard dist^n and screening of disease*
- *Concepts of every chapter discussed in quick review format*
- *Illustrations and images in color*

Contribution : Dr. Prince Goyal

ALTIS VORTEX

Normal distribution curve, inferential statistics, concepts of p-value

3

Probability and Probability Distribution

The probability distribution curve

- Also known as normal distribution curve

Features of the normal distribution curve

1. Bilateral symmetrical curve
2. The two ends never touch the baseline
3. Mean = Median = Mode; all coincide at '0' point of the NDC
4. Area Under Curve (AUC) = 1
5. Standard deviation = 1, Variance =1

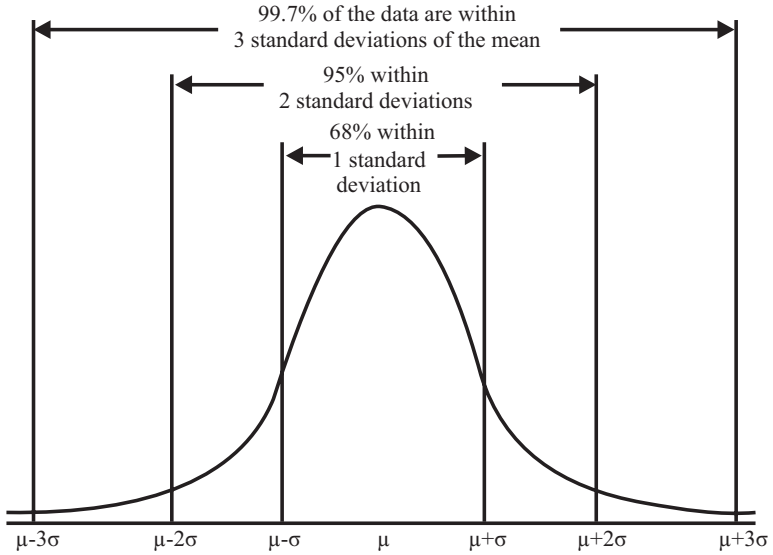
Derivation of the NDC

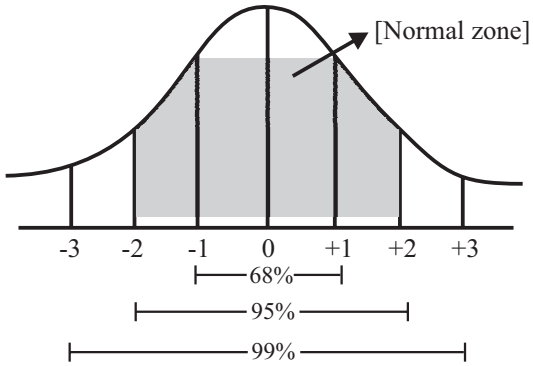
± 1 SD comprises of 68.23% (~68%) of the total population

± 2 SD comprises of 95.45% (~95%) of the total population

± 3 SD comprises of 99.7% (~99%) of the total population

Concept of confidence levels:





Assumption we take 95% confidence limits as the normal zone. Hence any value falling within the 95% confidence limits is known as normal and any value located outside the 95% confidence limits is known as 'outside normal zone' (abnormal zone)

Now, the limits on the either side can be taken as the confidence limits and the 95% Confidence limits are taken as the zone of normalcy. Hence by this notion, any value which lies more than 95% Confidence limits would also lie outside the normal zone. This measurement of location on an entity could be derived mathematically using a standard deviate or 'Z Score'

Z Score is given by: $\frac{\text{Observed value} - \text{Expected value}}{\text{S tandard Deviation}}$

Z score is thus a measure of the location of the variable and the result is in terms of the standard deviations away from the mean or the expected value.

If the Value of the Z score is more than 2, implies that the value is lying more than 2 SD's away from the mean (or the expected value).

As we know that, 2 SD's approximately correspond to 95.45% of population, thus 95% of population is in 1.96 SD. Hence the limits for Score at 95% Confidence limits correspond to 1.96 SD.

Concept of p-value:

If we plot the p-values, on a normal distribution curve. The p-value of 1 corresponds to the mean or the central point and the p-value of 0.05 corresponds to the 95% confidence limits or 1.96 SD (~2SD).

Hence we can summarize that the p-value less than 0.05 lies in the abnormal zone and the variation observed in the data is a significant variation which is due to some external

factors.

Concept of Alpha value or alpha level:

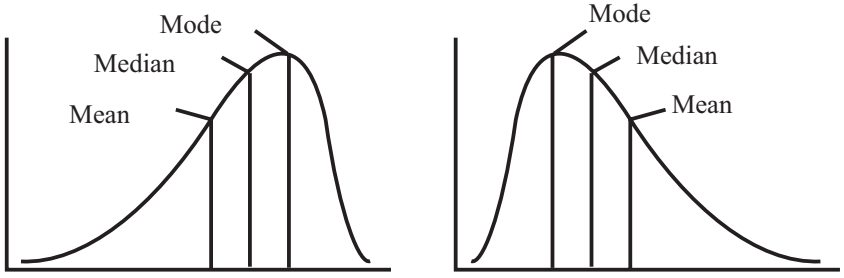
If the p-values are plotted on a normal distribution curve:

2 SD corresponds to 0.05 p value and an alpha level of 5%. This implies that if the results obtained after comparing the variables shows an alpha level of 5% or p-value lesser than 0.05. the null hypothesis must be rejected, in the sense that there is less than 5% chance of error in the results obtained and the observed variation or difference is the real variation.

SKEW

Skew means a deviation from the normal shape and structure.

In some cases, the data is not showing a normal distribution, in the sense, that it does not have the mean coinciding with the median and the mode.



Left Skew Right Skew

Remember:

In the Right skew: Mean > Median > Mode

In the Left Skew: Mode > Median > Mean

Solved Multiple Choice Questions

1. Which of the following is true about the standard normal distribution?

a. Is skewed to the left
b. Has mean = 1.0
c. Has standard deviation 0.0
d. Has variance 1.0

Ans: (d)

- It is a bell-shaped smooth curve
- It has two tails and it is symmetrical
- Both the ends do not touch the base line
- The measures of central tendency - mean, median and mode coincide
- Variance 1 because variance = $(SD)^2$ and SD is taken as 1

2. The distribution of random blood glucose measurements from 30individuals was found to be 120 ± 10 mg / dL.Which of the following is a correct statement about the shape of the distribution of random blood glucose in these first year medical students?

a. Since both mean and standard deviation are equalshould be a symmetric distribution
b. The distribution is likely to be positively skewed
c. The distribution is likely to be negatively skewed
d. Nothing can be concluded

Ans: (d)

Of the given options, nothing can be concluded as the median and mode are not provided, hence comment over shape of curve cannot be given.

3. Which of the following results, provide the best estimates of statistical significance, sample size, and strength of association?

a. A relative risk of 2.5 with a 95 percent confidence interval of 2.0 to 3.1
b. A p value of 0.0004 and and Relative risk of 1
c. A relative risk of 5.0 with a 95 percent confidence interval of 0.1 to 9.8
d. A p value of > 0.05 and a relative risk of 2.5

Ans: (a)

The confidence interval is range of the possible values, within which the true value shall lie.

Remember: The 95% confidence interval corresponds to 1.96 SD (~2SD), which further corresponds to p-value to 0.05. the p-value decreases as the confidence limits increase.

Hence the following information should be helpful:

1. P-value of less than 0,05 is a significant p-value
2. Relative risk (RR) or Odds ratio (OR) equals to one – implies NO association

3. Relative risk (RR) or Odds ratio (OR) more than one – implies positive association
4. Relative risk (RR) or Odds ratio (OR) less than one – implies negative association

Hence, if the p-value is more than 0.05 and / or the RR(OR) = 1, means a non-significant result

a. A relative risk of 2.5 with a 95 percent confidence interval of 2.0 to 3.1	RR > 1	95% CI does not contain value '1'	Significant
b. A p value of 0.0004 and and Relative risk of 1	RR =1 Non-significant	P value < 0.05 (Significant)	Non-significant
c. A relative risk of 5.0 with a 95 percent confidence interval of 0.1 to 9.8	RR > 1 (significant)	95% CI contains the value '1'(Non-significant)	Non-significant
d. A p value of > 0.05 and a relative risk of 2.5	RR >1 (significant)	P value > 0.05 (Non-significant)	Non-significant

4. All the following statements concerning statistical inference are true except

a. A test of statistical significance does not prove causality
b. A statistically significant test assesses the probability of a "chance" occurrence
c. A statistically significant test supports the null hypothesis
d. A statistically significant test result does not relate to association due to bias.

Ans: (c)

A statistically significant test ($p < 0.05$) rejects the *null hypothesis*, which states that no differences of effect was be found. A p-value < 0.05 indicates that a difference was found and there is <5 percent probability that it occurred by chance (corresponding to the alpha error)

5. Normal distribution curve depends on

a. Mean
b. Median and standard deviation
c. Mean and sampling method
d. Mean and standard deviation

Ans: (b)

Note:

The NDC (Normal Distribution Curve) depends on mean and standard deviation.

The Normal distribution means mean = median = mode and the standard deviation is taken as one unit.

Larger the standard deviation, the bigger the curve.

6. p value of a test comparing of two drugs is 0.01; what is the inference?

a. Probability that drug A is better than drug B
b. Probability of finding an insignificant difference is 1%

c. Probability of declaring a significant difference when there is no difference is 1%
d. Probability that drug B is better than drug A

Ans: (b)

- p value indicates the probability of the null hypothesis being true, i.e., p value of 0.1 (i.e., 1%). It means that there is 1% chance of null hypothesis being true and 99% chance of it being false, hence probability of difference between two samples (or two drugs effect) being insignificant is only 1% and there is 99% chance of the difference being significant.
- If p value is <0.05, mean difference is unlikely to be due to by chance and null hypothesis is rejected here.
- Null hypothesis states that there is no significant difference between two samples, i.e., there is no difference between the effects of the different drugs; whatever difference we see is due to chance only and is not significant.

7. When the confidence level of a test is increased which of the following will happen?

a. No effect on significance
b. Previously insignificant value becomes significant
c. Previously significant value becomes insignificant
d. No change in hypothesis

Ans: (c)

On increasing the confidence level, more observations will be under zone of acceptance, and hence previously significant value become insignificant.

8. The Hb level in healthy women has mean 13.5 g/ dl and standard deviation 1.5 g/dl, what is the Z-score for a woman with Hb, level 15.0 g/dl

a. 9.0 b. 10.0 c. 2.0 d. 1.0

Ans: (d)

Z score is:O-E / SD
Where,
O = Observed = 15
E = Expected = mean = 13.5
SD = 1.5
Z = (15-13.5)/1.5 = +1

9. After applying a statistical test, an investigator gets the "p value" as 0.01. It means that

a. The probability of finding a significant difference isl%
b. The probability of declaring a significant difference when there is truly no difference, is 1%
c. The difference is not significant 1% times and significant 99% times
d. The power of the test used is 99%

Ans: (b)

Null hypothesis

Null hypothesis says that there is no difference (or there is no effect) between

groups.

Now, p-value < 0.05 is in the significance zone and p-value > 0.05 is in the non-significant zone.

Significant p-value – means: that whatever difference is observed as variation between two groups is a real difference and is 'significantly' due to some external factor influence and not just a 'normal' or 'chance' variation.

Next, the p-value of 0.05 corresponds to alpha value of 5% and p-value 0.01 corresponds to alpha level of 1%.

So, henceforth, Alpha level of 1% (or a p-value of 0.01) means – that the chance of getting a significant result, when in reality the result was non-significant (means no difference) is less than 1%

So, now in the MCQ, the p-value is 0.01 which is lesser than 0.05, hence the observed p-value lies in the significant zone. And the chance of having a non-significant result when in reality the difference was significant (i.e. difference actually present) is less than 1%

10. Standard error is a measure of

a. Sampling error
b. Observer error
c. Conceptual error
d. Instrumental error

Ans: (a)

Standard error is basically a measure of standard deviation for a large sample or the total population. The standard error in principle has tendency to decrease the effect of sampling errors.

SE is standard deviation on larger scale, maybe for multiple populations or multiple studies of the same population

11. In which of the following distribution the mean, median and mode are all at the same point?

a. Binomial data
b. Normal data
c. Poisson data
d. Geometric data

Ans: (b)

Normal distribution has only one peak, corresponding to the mean, mode and median.

Binomial distribution: is a distribution for a binomial data (2 categories) as sick-healthy; dead-alive; pneumonia-no pneumonia.. and so on

Poisson distribution: we sometimes know the value of a variable, however, the other part is not known. Example: we know the number of cases of MI visiting our hospital for emergency care, but at the same we 'do NOT' know the number of cases of MI 'not coming' to our hospital for emergency care.

Geometric distribution: is for a sequence of independent events involving two outcomes.

12. In a normal distribution of data, the maximum values will lie within which of the following confidence limits:

a. $\overline{X} \pm \sigma$

b. $\overline{X} \pm 2\sigma$

c. $\overline{X} \pm 3\sigma$

d. None of the above

Ans: (c)

From the above given options, the maximum values will lie between $\overline{X} \pm 3\sigma$, i.e. mean

± 3 standard deviation

± 3 SD is also known as the universal limits i.e. there is less than 1% chance for a value to lie outside the ± 3 standard deviations

13. The mathematical derivation for a standard score or the Z score is (given that $\overline{X}$ = mean and X = observed value)

a. $\frac{\overline{x} - x}{\sigma}$ b. $\frac{x - \overline{x}}{\sigma}$ c. $\frac{x + \overline{x}}{\sigma}$ d. $\frac{x + \overline{x}}{\sigma}$

Ans: (b)

Z score, also known as standard score is given by the formula (O-E) / SD
The Z score tells us the 'location' of the value in terms of the standard deviations. Hence, the Z score of < 1.96 (~2) SD corresponds to the normal zone and values outside the normal zone are abnormal or different than the actual or the normal values.

14. If all the observations in a set of observations are equal, then the variance of the sample is:

a. Zero b. One c. Infinite d. Cannot conclude

Ans: (a)

The sum of observations taken from the mean is zero, hence the variance is also zero.

The Standard deviation is the mean deviation of values from the central point or the mean.

Hence in the MCQ above, the standard deviation is zero or nil and hence, the variance is also zero.

Variance (V) is SD^2

15. Power of the test is related to:

a. Alpha error b. Beta error
c. P-value d. Precision

Ans: (b)

Power of a test is related to the beta error.

α	The conditional probability of incorrectly rejecting H_0 when it actually is true
β	The conditional probability of failing to reject H_0 when it is false
Power	The complement of β (i.e. 1- β), this is the probability of correctly rejecting H_0 when it is false
H_0	The null hypothesis, usually stated as the population mean being zero, or that there is no difference. However, it does not have to be stated as a zero or no difference hypothesis
H_a	The alternative hypothesis, usually stated as the population mean being non-zero or greater then or less than zero

Power of the test = 1 – beta error

The power of a hypothesis test is affected by three factors.

- Sample size (n). Other things being equal, the greater the sample size, the greater the power of the test.
- Significance level (α). The higher the significance level, the higher the power of the test. If you increase the significance level, you reduce the region of acceptance. As a result, you are more likely to reject the null hypothesis. This means you are less likely to accept the null hypothesis when it is false; i.e., less likely to make a Type II error. Hence, the power of the test is increased.
- The "true" value of the parameter being tested. The greater the difference between the "true" value of a parameter and the value specified in the null hypothesis, the greater the power of the test. That is, the greater the effect size, the greater the power of the test.

16. level of significance relates to the:

a. Type 1 error b. Type II error
c. Random error d. Population mean

Ans: (a)

The significance level corresponds to the α error or the type 1 error

It is taken that the value of 'p' ranges from 0-1. As the p-value is the probability and the probability of an event to happen would be 0% to 100 % chance. Also, it should be noted by the reader that p-values less than 0.05 means that the value for difference is outside the 95% confidence limits zone and hence lies in the significant zone.

The p-value of 0.05 corresponds to alpha level of 5%

17. A researcher conducted a trial with 5% significance level, to evaluate the use of carnitine containing drug for Alzheimer disease. The regulatory authority granted clearance for the trial on 100 patients as under laboratory circumstances, the drug had shown an effect on the Alzheimer patients. On

further analysis of the research it was found that the carnitine drug showed an improvement in Alzheimer patients with a p-value of 0.82. this is an example of:

a. High Alpha error b. High Beta error
c. High power of the research d. High protective power of the drug

Ans: (b)

NoH (H_o) null hypothesis

What if in Reality,	It Actually means – in reality there was-	So,	And what if WE from our observations,	What have we done..?
The Null hypothesis - H_o was **actually TRUE**	No effect or no difference between two groups	The Alternate hypothesis (H_1) is false	**Accept** the H_o hypothesis	Its good – low ALPHA error
The Null hypothesis - H_o was **actually TRUE**	No effect or no difference between two groups	The Alternate hypothesis (H_1) is false	By chance **Reject** the H_o hypothesis	Its an ERROR! **ALPHA Error**
The Null hypothesis - H_o was **actually FALSE** or wrong	An effect or difference between two groups	The Alternate hypothesis (H_1) is true	**Reject** the H_o hypothesis	Its good – low BETA error
The Null hypothesis - H_o was **actually FALSE** or wrong	An effect or difference between two groups	The Alternate hypothesis (H_1) is true	By chance **Accept** the H_o hypothesis	Its an ERROR! **BETA Error**

18. Which of the following is a more serious error:

a. Type I b. Type II c. Type III d. Type IV

Ans: (a)

Type of errors:

- We reject the null hypothesis and the null hypothesis is true. This is what is known as a Type I error.
- We reject the null hypothesis and the alternative hypothesis is true. In this situation the correct decision has been made.
- We fail to reject the null hypothesis and the null hypothesis is true. In this situation the correct decision has been made.
- We fail to reject the null hypothesis and the alternative hypothesis is true. This is what is known as a Type II error.

Type 1 error is considered to be a more serious error than type II

19. TRUE ABOUT a negatively skewed data:

a. Mode is Less than median. b. Mode is More than median.
c. Mode is equal to median. d. No correlation

Ans: (b)

- Mode>Median>mean

If more outlying values are smaller than the rest, data are said to be skewed to left Data have longer tail among lower values

Skewed **to** Right (Positively skewed)

‘ Mode<Median<mean

- If more outlying values are larger than the rest, data are said to be skewed to right
- Data have longer tail among higher values

20. Which is the best distribution to study the daily admission of head injury patients in a trauma care centre?

a. Poisson distribution b. Nominal distribution
c. Binomial distribution d. Uniform distribution

Ans: (a)

POISSON DISTRIBUTION

In probability theory and statistics, **the Poisson distribution** is a discrete probability distribution that expresses the probability of a number of events occurring in a fixed period of time if these events occur with a known average rate and independently of the time since the last event. The Poisson distribution can also be used for the number of events in other specified intervals such as distance, area or volume.

The distribution was discovered by Simeon-Denis Poisson .The work focused on certain random variables N that count, among other things, a number of discrete occurrences (sometimes called “arrivals”) that take place during a time-interval of given length The Poisson distribution arises in connection with Poisson processes. It applies to various phenomena of discrete nature (that is, those that may happen 0, 1, 2, 3, ... times during a given period of time or in a given area) whenever the probability of the phenomenon happening is constant in time or space. Examples of events that may be modeled as a Poisson distribution include:

- The number of cars that pass through a certain point on a road (sufficiently distant from traffic lights) during a given period of time.
- The number of spelling mistakes one makes while typing a single page.
- The number of phone calls at a call center per minute.
- The number of times a web server is accessed per minute.
- The number of road kill (animals killed) found per unit length of road.
- The number of mutations in a given stretch of DNA afier a certain amount of

radiation.

- The number of unstable nuclei that decayed within a given period of time in a piece of radioactive substance. The radioactivity of the substance will weaken with time, so the total time interval used in the model should be significantly less than the mean lifetime of the substance.
- The number of pine trees per unit area of mixed forest.
- The number of stars in a given volume of space.

21. Not a marker of association

a. P value
b. Odds ratio
c. Correlation coefficient
d. Traditional Alpha

Ans: (d)

P value, OR & correlation Coefficient are all markers of association except the value of alpha which is a test criterion or level for deciding whether to accept or reject a null hypothesis, also known as type Ierror.

According to convention, if P is less than or equal to .05, it is regarded as statistically significant. The smaller the P value, the greater the statistical significance or probability that the association is not due to chance alone. However, the statistical association (p value) does not imply causation. Odds Ratio is a measure of the strength of association between the risk factor and outcome.

A correlation coefficient simply expresses the strength and direction of association or relationship between two quantitative variables, signified by 'r'. Values of r varies from -ito +1; the strength of the relationship is indicated by the size of the coefficient, whereas its direction is indicated by the sign.

The probability level at which it is decided that the null hypothesis is incorrect constitutes a criterion or significance level known as alpha.

22. Normal curve is:

a. Linear
b. Symmetrical
c. Parabolic
d. Curvilinear

Ans: (a)

Normal distribution curve

Also known as Gaussian distribution curve

Features

1. Bilaterally symmetrical
2. Bell shaped curve
3. The ends never touch the baseline
4. The mean = median = mode – i.e. they all coincide
5. The mean is at zero point with standard deviation of 1

6. The Area under curve = 1
7. The variance = 1

23. The mean blood pressure of a group of people is 105. The standard deviation was 10. What is the range within which 95% of the population is expected to lie?

a. 95-115 b. 85-125 c. 75-135 d. 104-106

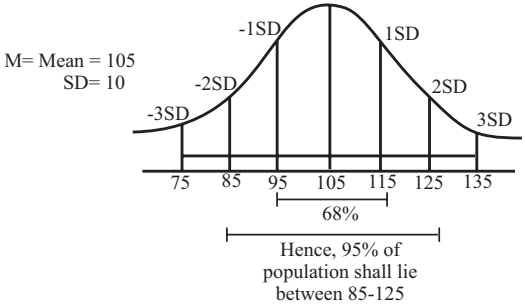
Ans: (b)

95% range includes values within 2 standard deviations of the mean (Z ±2SD). Since the SD is 10, twice of the SD would be 2 * 10 = 20. Hence,

μ - 2SD = 105 – 20 = 85

μ + 2SD= 105 + 20 = 85

So 95% CI would include values ranging from 85 — 125. This is the central 95% of the distribution and the remaining 5% of the distribution would be distributed equally on the two sides of this confidence limit. Hence, there will be 2.5% of the values in this distribution which will be smaller than 2 SD (<85) and another 2.5% of the distribution which will be greater than +2SD (> 125).



24. If the systolic blood pressure in a population has a mean of 130 mmHg and a median of 140 mm Hg. The distribution is said to be:

a. Symmetrical
b. Positively skewed
c. Negatively skewed
d. Either positively or, negatively skewed depending on the standard deviation

Ans: (c)

Negatively skewed distribution – most of the data has higher values.
Skewed to the left (Negative Skewed)
Mode is maximum and mean is lowest - **Mean < median < mode**

Positively skewed distribution - Skewed to the right - (Positive Skew)
Most of the data will have lower values

Mean > median > mode

Mean is highest and mode is of the lowest value in the data central tendencies.

25. Which of the following statements is correct with regard to Confidence limits?

a. Smaller the confidence level, the 'wider the interval.
b. 'The lesser variable our data, the wider is the confidence interval,
c. The width of the confidence interval is independent of sample size.
d. Confidence interval for 95% 'is wider than 70%

Ans: (d)

Option A. Smaller the confidence level, the 'wider the interval' – no, smaller the CI, the narrower would be the Gaussian distribution curve

Option B. 'The lesser is variation in data, the wider is the confidence interval – No, the smaller variations would mean smaller standard deviations and narrow normal distribution curve
Option C. The width of the confidence interval is independent of sample size. No larger the sample size, the larger is the precision and smaller is the error. Thus making the normal curve a narrow curve
Option D. Confidence interval for 95% 'is wider than 70%. Yes correct

26. In a sample of 400 healthy men, the mean serum globulin level is found to be 30 g/L, with a standard deviation of 3 g/L.
The coefficient of variation of serum globulin in this sample of men is:

a. 0.10 b. 0.15 c. 3 d. 15

Ans: (b)

The coefficient of variation is given by the formula = SD / √n. This implies that it directly is related to SD (higher SD means, higher CV),

So, CV = SD / √n * 100 = 3 / √400 = 3/20 = 0.15

27. If it is reasonable to assume that serum globulin levels follow a normal distribution, then approximately 50% of the men will have a value:

a. Between 27 and 33 g/L. b. Between 24 and 36 g/L.
c. Below 27 or above 33 g/L d. Below 30 g/L.

Ans: (d)

Because, the mean = 30 g/L in the above MCQ.. hence in a normal distribution the mean = median = mode.. i.e. = 30 g/L.

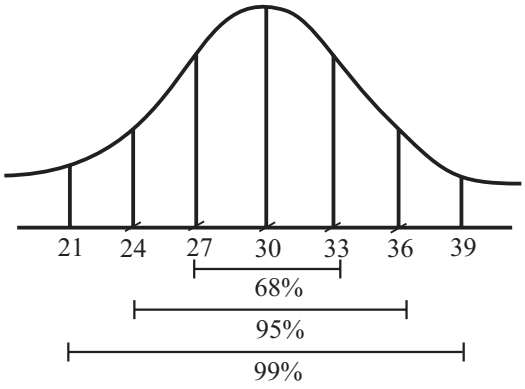
Hence, 50% of the population will lie below 30 g/l as it is also the median (central point)

28. Again, assuming that the distribution of serum globulin levels is normal, a healthy man would have a serum globulin level higher than 36 g/L:

a. 1% of the time b. 2.5% of the time
c. 5% of the time d. 10% of the time

Ans: (b)

As the normal distribution the curve will be:



Hence higher than 36g/2 will be 2.5% (as between 24 & 36 is 95%, out of remaining 5 % half will be below 24 & another half 2.5% above 36g/2)

29. If the correlation between two measures of mental status is 0.70, we can conclude that:

a. The value of one measure increases by 70% when the other measure increases by 1
b. 49% of the observations fall on the regression line
c. 70% of the variation in one measure is accounted for by the other
d. 49% of the variation in one measure is accounted for by the other

Ans: (d)

Here, in the MCQ, the correlation coefficient is given as 0.7.
There is something known as 'coefficient of determination' (CoD)

CoD = r^2 expressed as percentage.
Hence, the CoD = 49%.

The coefficient determination basically denotes, the change in one variable after a unit change in the other correlated variable. Hence in the given MCQ, the Variable will show 49% change with a unit change in the other variable.

30. A new antibiotic was invented and a trial was planned for evaluation of the drug. 500 children age < 16 years were taken and were randomly allocated to the intervention and control arm. The blood cultures were done before and after drug intervention.
The design used in this study is best described as a:

a. Cohort study
b. Crossover study
c. Clinical trial
d. Randomized clinical trial

Ans: (d)

31. The research showed the difference in proportions as 12% in placebo group and 16.7% in the intervention group with a p-value of 0.87 and the 95%CI of the OR is -1.8 to +21.4

a. There is a statistically significant difference in major infectious morbidity when placebo is compared with antibiotic
b. The proportion of children with major infectious morbidity is the same with placebo and antibiotic.
c. There is no statistically significant difference in the proportions who received

placebo and antibiotic

d. Using a chi-square test to determine significance is preferable to determining a confidence interval for the difference.

Ans: (c)

As seen in the MCQ the intervention group has 16.7% proportion difference while the placebo group has only 12% difference of proportions. But on applying the statistical test, it is seen that the p-value is 0.87, which means it is more than 0.05 (zone of normalcy at 95% CI). Hence the p-value is a non-significant p-value

32. If the relationship between two measures is linear and the coefficient of determination has a value near, 1, a scatterplot of the observations:

a. The scatter plot vector is a horizontal straight line

b. The scatter plot vector is a vertical straight line

c. The scatter plot vector is a straight line that is neither horizontal nor vertical line

d. The scatter plot vector is a random scatter of points about the regression line

Ans: (c)

If the coefficient of determination is 1, the correlation coefficient would be expected to be 1. So, it is presumably a straight line which is not horizontal or vertical (which means 'no' correlation i.e. r = 0)

33. In a placebo-controlled trial of the use of oral amoxicillin to prevent recurrent RTI, 38% of patients receiving the drug did not get more events of RTI, and 39% of patients receiving placebo had repeat infections. In reporting this finding, the researcher reported p > 0.05, which means:

a. Chances of greater than 1 in 20 that a difference would again be found if the study were repeated.

b. The probability is less than 1 in 20 that this difference could occur by chance alone

c. The probability is greater than 1 in 20 that as difference this large could occur by chance alone

d. Treated patients were 1/20 less likely to have re-infections

Ans: (b)

The p-value of 0.05, implies an alpha error of 5% (or 1/20), means:

- There is less than 5% chance that we have observed a difference, which somehow actually does not exist
- There is less than 5% chance that the difference observed is just chance and not the real or actual difference due to the intervention (or risk factor)
- The chances of finding this difference, when in reality there was no difference is less than 5%
- We are more than 5% sure that the difference actually exists

34. A study was undertaken to evaluate any increased risk of breast cancer among women who use oral contraceptive pills. The relative risk was calculated. A type I error in this study consists of concluding:

a. A significant increase in the relative risk when the relative risk is actually 1

b. A significant increase in the relative risk when the relative risk is actually greater than 1
c. No significant increase in the relative risk when the relative risk is actually 1
d. No significant increase in the relative risk when the relative risk is actually greater than 1

Ans: (a)

Type 1 error: is rejecting a hypothesis which was actually true

Type II error: Accepting a hypothesis which was actually false

So, type I error would be finding an effect when in reality there was 'no' effect.

Recall that:

RR = 1 = no association

RR > 1 = Positive association (the factor is a risk effect)

RR < 1 = Negative association (the factor is a protective effect)

35. Birth weights of a population of infants at 42 weeks gestational age are approximately normally distributed, with a mean of 3000 g. Roughly 68% of such infants weigh between 2500 and 3500 g at birth. If a sample of 100 infants were studied, the standard error would be:

a. 50 b. 100 c. 250 d. 500

Ans: (a)

Given in the MCQ that

Mean = 3000 gms

68% of the population lies between 2500 and 3500

sample size (n) 100, calculate the Standard error

SE = SD / $\sqrt{n}$

As, 68% of population corresponds to 1 SD and the range is 2500 and 3500, hence the SD would be 500

Hence, SE = 500 / $\sqrt{100}$ = 500/10 = 50

36. Two research bring the data for the height from two different cities.
City A: 165 ± 8.5 mt
City B: 175 ± 11 mt

Which of following is incorrect:

a. City A has more height variation
b. City B has less height variations compared to city A
c. City B data is more precise
d. City A data is more precise

Ans: (d)

Out of the options given above, the option C does not fit into the scenario provided. From the data we can comment that:

1. City A has lower average height than city B
2. City A data has lesser variations compared to city B
3. Precision is inversely proportional to the Standard error (variations)
4. Hence, Data from City B has greater variations, and lesser precision.

37. The value of alpha serves as protection against:

a. Type I error b. Selection bias
c. Type II error d. False-negative results

Ans: (a)

Alpha error corresponds to the Type 1 error and the p-value

38. The characteristics of the normal distribution curve includethe following: (Multiple choice correct)

a. The standard deviation is taken as 100
b. The mean and median are found below the peak of the curve
c. 5% of the values lie beyond 2 standard deviationsfrom the median.
d. The curve is perfectly symmetrical
e. All the above

Ans: (b) and (d)

These are the features of the normal distribution curve.

39. In a study conducted in Chennai city, the mean SBP of 500 medical students was 116 ± 4 mmHg. From the above data, 99% of medical students will probably have SBP between what range?

a. 110-130 b. 104-128
c. 112-120 d. 118-122

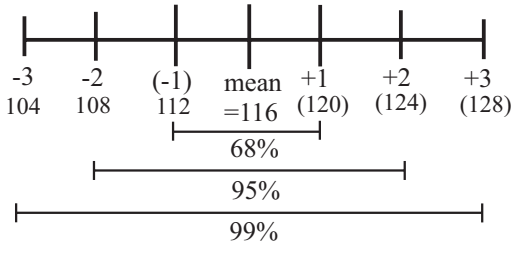
Ans: (b)

Mean = 116

SD = 4

So, the MCQ is asking for 99% population, which we know corresponds to the 3 SD.

Hence, Mean = 116 $\pm$ 4



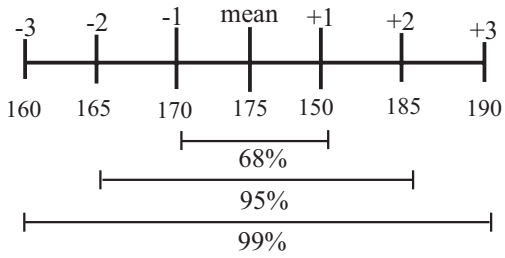
40. In a study involving 150 medical students, the mean height was found to be 175 cms with variance of 25 cms. After analysing the data find out how many students will have height more than 185 cms

a. 34% b. 2% c. 2.5% d. 0.5%

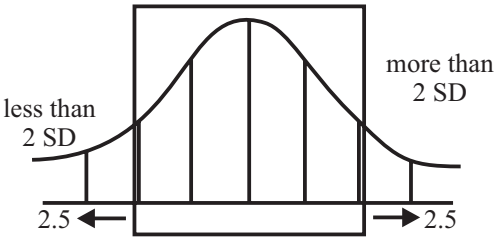
Ans: (c)

We know that variance is = SD^2, hence SD = √variance
So, the SD is 5 (√25)

Hence, Mean = 175, SD = 5



So more than 185 cm is 2.5%



41. Which is the best distribution to study the the admission of breast cancer patients in a cancer care center.

a. Normal distribution b. Binomial distribution
c. Uniform distribution d. Poisson distribution

Ans: (d)

As we know the value of one variable, whereas the value of another variable is not known

42. 'Z score' is for which type of distribution?

a. Normal b. Binomial c. t d. Chi- square

Ans. (c)

Z score or the Z deviate – is a measure of location for normal distributed data.

Remember:

- Z score tells us the location of the value in terms of the standard deviations away from the mean
- Z score 1.96 actually corresponds to 95% confidence limits
- Z score greater than ± 1.96 is taken as in the abnormal zone (or significant difference)

43. In a study conducted in Chennai city, the mean SBP of 500 medical students was 116 ± 4 mm Hg. From the above data, 99% of medical students will probably have SBP between what range?

a. 110-130 b. 104-128 c. 112-120 d. 118-122

Ans: (b)

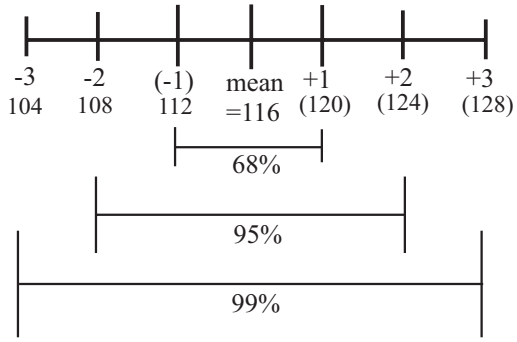
Mean = 116

SD = 4

So, the MCQ is asking for 99% population, which we know corresponds to the 3 SD.

Hence,

Mean = 116 ± mmHg



44. Again, assuming that the distribution of serum globulin levels is normal, a healthy man would have a serum globulin level higher than 36 g/L:

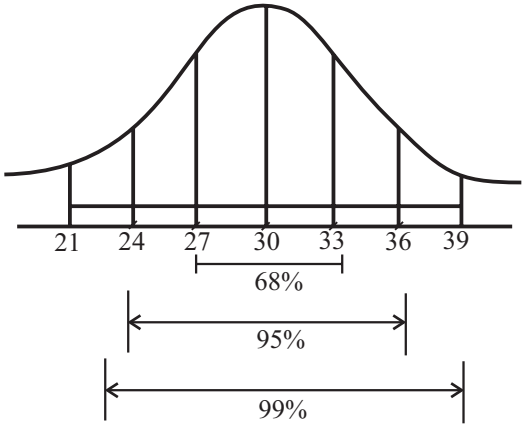
a. 1% of the time b. 2.5% of the time

c. 5% of the time d. 10% of the time

Ans: (b)

As the normal distribution the curve will be:

Given the μ = 30 with σ = 3g/L so the NDC (Normal Distribution Curve) will be



Hence higher than 36 gm/L will be 2.5% (as between 24 and 36 is 95%, out of the

remaining 5%, half will be below value of 24 and another half (i.e. 2.5%) is more than 36 g/L

45. If the correlation between two measures of mental status is .70, we can conclude that:

a. The value of one measure increases by 70% when the other measure increases by 1
b. 49% of the observations fall on the regression line
c. 70% of the variation in one measure is accounted for by the other
d. 49% of the variation in one measure is accounted for by the other

Ans: (d)

Here, in the MCQ, the correlation coefficient is given as 0.7.

We can calculate a mathematical entity known as 'coefficient of determination' (CoD)

CoD = $(r)^2$is expressed as percentage.

Hence, the CoD = 49%.

The coefficient determination basically denotes, the change in one variable after a unit change in the other correlated variable. Hence in the given MCQ, the Variable will show 49% change with a unit change in the other variable.

Multiple Choice Questions for Practice

1. Systolic blood pressure of a group of people between the age of 25-27 years was taken and a mean of 120 mm Hg was found with a Gaussian distribution. What will be the expected number of individuals among a group of 100 people taken for study whose systolic blood pressure will be below 120 mm Hg

a. 25 b. 50 c. 75 d. 100

Ans: (b)

Refer to the fact: mean of the data = 120, which is also the median and the mode (because the data is normally, also known as Gaussian, distributed)

2. The fasting blood levels of glucose for a group of diabetics is found to be normally distributed with a mean of 105 mg per 100 ml of blood and a standard deviation of 10 mg per 100 ml of blood. From this data is can be inferred that approximately 95% of diabetics will have their fasting blood glucose levels within the limits of -

a. 75 and 135 mgs b. 85 and 125 mgs
c. 95 and 115 mgs d. 65 and 145 mgs

Ans: (b)

3. In a group of 100 children, the mean weight of children is 10 kg. The standard error is 1kg. Which one of the following is true?

a. 95% of all children weight between 9.9 and 10.1 kg
b. 95% of all children weight between 7and 13 kg
c. 99% of all children weight between 10 and 12 kg
d. 99% of all children weight between 7and 13 kg

Ans: (d)

4. Which of the following statements is not true about standard normal distribution?

a. Shows a 'bath tub distribution' b. Has mean = 0.0
c. Is bilaterally symmetrical curve d. Has variance = 1

5. For a negatively skewed data mean will be:

a. Less than median b. More than median
c. More than Mode d. One

Ans: (a)

6. A chest physician observed that the distribution of hemoglobin in sample of 300 individuals with a median of 11 gm% with the first and third quartiles being 8 and 14 gm%. Based on this data, how many persons are expected to have Hb between 8 and 14 gm%

a. 7.5 b. 150 c. 225 d. 300

Ans: (b)

7. Normal distribution curve, all are correct except:

a. Mean, median, mode coincide at a given point
b. Bilaterally symmetrical
c. Bell shaped curve
d. SD is zero

Ans. (d)

The salient features of Normal distribution curve:

- Bell-Shaped
- Perfectly symmetrical
- Based on infinitely larger number of observation
- Total area of the curve is 1
- The mean is at Zero point
- The mean = median = mode
- The ends never touch the base line

8. If the distribution of intra-ocular pressure (lOP) seen in 100 glaucoma patients has an average 30 mm with a SD of 1.0, what is the lower limit of the average IOP that can be expected 95% of times?

a. 28 b. 26 c. 23 d. 25

Ans: (a)

9. How much of the sample is included in 1.96 SD?
a. 99% b. 95% c. 68% d. 65%

Ans: (b)

10. If the systolic blood pressure in a population has a mean of 120 mmHg and a median of 140 mm Hg, the distribution is said to be -
a. Symmetrical
b. Positively skewed
c. Negatively skewed
d. Either positively or negatively skewed

Ans: (c)

11. For a given set of values, Mean = 28, Median = 24 & Mode = 20. The given distribution is:
a. Symmetric b. Right-skewed
c. Left-skewed d. Can be either symmetric or skewed

Ans: (b)

12. The standard normal distribution:
a. Is skewed to the left b. Has mean = 1
c. Has standard deviation = 0 d. Has variance = 1.0

Ans: (d)

13. In normal curve:
a. Mean = 2 standard deviation
b. Mean = Median
c. Mean = Variance
d. Mean 1 standard deviation

Ans: (b)

14. Regarding the normal curve, which of the following statements is true: (MCC)
a. Both ends of the curve touch the baseline
b. The curve is bilaterally symmetrical
c. There is a skew to the right
d. There is a skew to the left
e. Mean, median and mode concide

Ans: (b) and (e)

15. If the 95% Confidence Interval for prevalence of Cancer in Smokers aged > 70 years is 56% to 76%, the chance that the prevalence could be less than 56% is:
a. Practically NIL b. 44% c. 2.5% d. 5%

Ans: (c)

16. In a standard normal curve, the area between one standard deviation on either side will be:
a. 68% b. 85% c. 99.7% d. None of the above

17. In a standard normal curve, mean +2 standard deviations covers:
a. 60% b. 65%

c. 95% d. 99%

Ans: (c)

18. When the variables follow standard distribution?:

a. Mean = median
b. Median = variance
c. Mean 2 median
d. Standard deviation = 2 variance

Ans: (a)

19. Area covered between two standard deviation in a normal distribution curve:

a. 68% b. 95.4%
c. 99.6% d. 100%

Ans: (b)

20. Normal distribution curve is determined by:

a. Standard deviation and mean b. Standard deviation and mode
c. Mode and median d. Standard deviation and median

Ans: (a)

21. Type I sampling error is taken as:

a. Alpha error b. Beta error
c. Gamma error d. Delta error

Ans: (a)

22. Power of study can be increased by:

a.Increasing α error b. Decreasing β error
c. Decreasing α error d. Increasing β error

Ans: (b)

Remember:

- Beta error is inversely related to power
- Alpha error corresponds to the p-value

23. In assessing the association between maternal smoking status and the birth weight of the newborns, two researchers A and B studied separately and found significant results with p values 0.001 and 0.04 respectively. From this information, what can you infer about the magnitudes of association found by the two investigations?

a. The magnitude of association found by investigator A is more than that found by B, while both are significant
b. The magnitude of association found by investigator B is more than that found by A, while both are non-significant
c. The magnitude of association found by investigator B is more than that found by A, while both are significant
d. The estimates of association obtained by A and B will be equal, since both are significant

Ans. (a)

The p-value is lower for the researcher A. From the data given, it may be commented that a statistical difference was observed by both researchers, however, the study

A yields a more significant (more stronger association) compared to the data from study B

Remember: lower the p-value, the more significant the result is, and stronger would be the association or effect of intervention

66. After applying a statistical test, an investigator gets the 'P value' as 0.01. it means that:

a. The probability of finding a significant difference is 1% when in reality there is no difference.
b. The probability of finding a significant difference is 99% when in reality there is no difference.
c. The probability of finding a non-significant difference is 1% when in reality there is no difference.
d. The difference is not significant 1% times and significant 99% times

Ans: (b)

p-value = 0.01 corresponds to alpha value of 1%, which means that the:

- Probability of finding the effect just by chance, when there was no effect is less than 1% (or we are 99% confident with our results, that it lies in the significant zone)

67. All are true about P-value except

a. Is the probability of committing Type-I error
b. Is equal to 1-β
c. Is the chance that the presence of difference is concluded when actually there is none
d. When P-value is less than α, the result is statistically significant

Ans: (b)

68. Statistical Power of a trial is equal to:

a. I + α
b. 1 — β
c. α + *β*
d. *α/β*

Ans: (b)

69. The risk factor association of smoking with COPD was studied in a case control study. The values are

Group	Odds ratio	95% confidence limits
A	2.9	1.0-3.1
B	5.6	0.1-7.7
C	1.6	1.08 – 2.1

Which of the following is correct:

a. Risk is more associated with Group A
b. Risk is more associated with Group B
c. Risk is more associated with Group C
d. Risk is equally associated with all three groups

Ans: (c)

The RR is always positive with the group C factor

70. All of the following are true about Standard error, except?

a. As the sample size increases, Standard error will also increase
b. Based on Normal distribution
c. It depends on Standard deviation of mean
d. Is used to estimate confidence limit

Ans: (a)

71. In a test of significance, P valuesis 0.023 the observed difference in study can be considered as:

a. Null hypothesis accepted
b. Null hypothesis rejected
c. Alternate hypothesis rejected
d. nothing can be concluded from this data

Ans: (b)

Note: the p-values are significant and the alternate hypothesis is accepted (and null hypothesis is hereby rejected)