MODEL ANSWER

PHYSIOLOGY HONOURS/SEMESTER VI

PAPER DSE-A1 (BIOSTATISTICS)

1. **When would you draw horizontal bar diagram?**
   Horizontal bar diagram is used for graphical representation of frequency in different classes of a sample of nominal variable (mostly). it consists of a set of horizontal bars separated by intervening gaps.
   A horizontal bar diagram is used when the classes have long category names. In a horizontal bar diagram, long labels for the categories are easier to display and read. A horizontal bar diagram can also be extremely helpful for labelling each class in a proportional bar chart of equal width.

2. **Give two physiological examples where mean should be used.**
   The mean is usually the best measure of central tendency when the data distribution is continuous and symmetrical. For example. Body height, Body weight etc.

3. **When is the mode the best measure of central tendency?**
   The mode is the least used of the measures of central tendency. It can be used when dealing with nominal data. For example- Blood Group.

4. **When is the median the best measure of central tendency?**
   The median is usually preferred to other measures of central tendency when the data set is skewed (i.e., forms a skewed distribution) or the data is ordinal in type. This is because the value of the mean can be distorted by the outliers.

5. **For any data set, which measures of central tendency have only one value?**
   The median and mean can only have one value for a given data set. The mode can have more than one value.

6. **What does P_{50} indicate in statistics?**
   Percentile is a score in a frequency distribution below which lie a specific percentage of total number of scores in a sample. Thus, P_{50} (50^{th} percentile) is that score below which lies the lowest 50% scores. P_{50} is identical with median.

7. **State the advantages and disadvantages of mean.**

Advantages	Disadvantages
○ Simple to understand ○ Easy to calculate ○ Rigidly defined ○ Suitable for further algebraic treatment ○ Considers all the values in the series.	○ Highly affected by the presence of a few abnormally high or low scores (extreme scores or outliers) in the distribution especially when the sample size is small. So, mean is not an appropriate measure of central tendency for skewed distribution. ○ Cannot be calculated for nominal or ordinal data. ○ Cannot be determined by simple inspection. ○ Cannot be computed if class intervals have open ends. ○ Cannot be calculated if any of the data is missing.

8. State the advantages and disadvantages of median.

Advantages	Disadvantages
○ Easy to understand. ○ Easy to calculate. In some cases, it is obtained simply by inspection. ○ Not distorted by skewed data (outliers). ○ Can be determined for ordinal, ratio, and interval variables. ○ It can be graphically located by drawing ogive curves of grouped data. ○ It is specifically useful in open-ended distribution.	○ Does not consider the precise value of each observation. ○ Unlike mean, it cannot be used for further algebraic treatment or statistical tests. ○ It is not rigidly defined. In ungrouped data having even number of observations median cannot be exactly found. ○ If ungrouped data contains large number of observations, then the process of arranging them serially and finding median is tedious.

9. State the advantages and disadvantages of mode.

Advantages	Disadvantages
○ Easy to understand and easy to calculate. ○ It is the only measure of central tendency that can be used for nominal variable. ○ Not affected by extreme values in a distribution ○ Can be computed in an open-ended frequency table. ○ It can be located graphically. ○ It can be determined without knowing all the values in a sample.	○ It is not algebraically defined. And therefore, not used in statistical analysis. ○ It cannot be defined if the distribution is bimodal or multimodal. ○ Most affected by fluctuation in sampling. ○ Not based on all values.

10. Why standard deviation score over mean deviation in the statistics?

Standard deviation (SD) score over mean deviation (MD) because of the following two reasons:

a) Mean deviation (MD) measures the absolute deviation of scores from the mean. So, it measures the variations of score in magnitude only, but not in directions. SD measures both.

b) Mean deviation (MD) is unsuitable for further statistical treatment. SD is used for further statistical work.

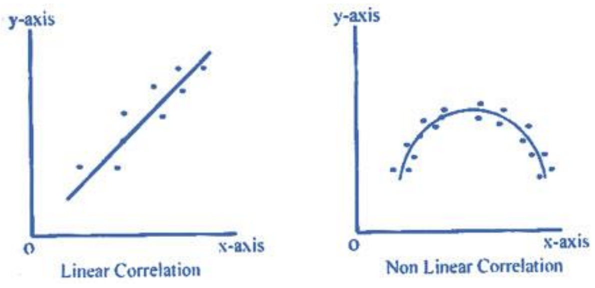
11. How does correlation differ from regression?

Correlation explores the magnitude and direction of relationship between two or more variables. Whereas, regression is a prediction-statistics, which predicts the most likely value of a variable depending upon the value(s) of one or more other variables.

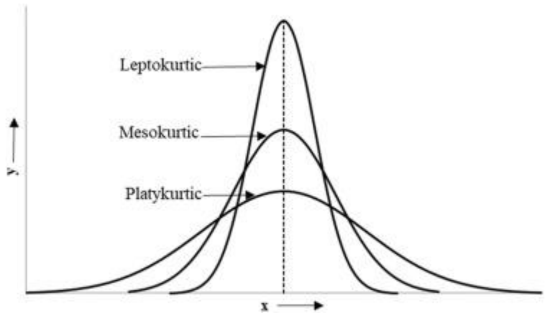
12. How does linear correlation differ from non-linear correlation?

Correlation between two variables can be linear or non-linear. Correlation is said to be linear if the relationship between two variables can be described by a straight line. So, the ratio of change in the two variables stays constant.

Correlation is non-linear if the relation between two variables can be described by a curved line. In this case, ratio of change between two variables is not constant.



13. **Kurtosis:** Kurtosis describes the degree of 'peakedness' of any frequency distribution compared to that of a normal distribution.
    - The normal distribution is said to be **mesokurtic** and its peakedness is taken as standard.
    - A **leptokurtic** distribution has a higher and sharper peak, thicker tails, and narrower body, in comparison to normal distribution. Student's t distribution is generally leptokurtic unless the sample is very large.
    - A **platykurtic** distribution is flatter in centre, broader in the body and thinner at the tails than normal distribution.



The percentile coefficient value of kurtosis (K) is 0.263 in mesokurtic distribution. Higher is the value of K above 0.263, the greater is the platykurtosis. Lower is the value of K below 0.263, greater is the leptokurtosis. (t-distribution has a k value lower than 0.263).

14. **Null hypothesis**: In any experiment, the Null hypothesis (H_0) proposes that the observed result in an experiment has arisen by accidental chances of random sampling and it wouldn't have been occurred if the entire population would have been used in the experiment instead of the sample.

    Null hypothesis is tested by working out the probability(P) of getting the result by chances of random sampling. If this estimated P is lower than a particular chosen level of significance ($p < \alpha$ 0.05), the probability of correctness of Null hypothesis is considered too low. And the H_0 may then be rejected.

    But if the estimated P is exceeds the chosen level of significance ($p > \alpha$ 0.05),, the probability of correctness of null hypothesis is considered too high. In that case Ho is retained.

15. **Degrees of freedom:** The 'degrees of freedom' (df) of a statistic amount to that number of scores of a variable in a sample, which can be changed freely in magnitude and sign without causing any alteration in the values of all such statistics.

    Sometimes in computing a statistic, one or more pre-computed statistics are used. Each such pre-computed statistics must remain unchanged during the computation. This causes the loss of freedom of any one of the scores of the sample.

    If the total no of scores in a sample size is denoted by n, the df is lowered from n by 1 for keeping each pre-computed statistic unaltered. Therefore, if m is the number of pre-computed statistics used in any computation, the degree of freedom (df) will be (n-m). For example, computation of pooled SD of two samples (sample sizes n_1 and n_2) requires the use of two sample means. So, the df will be (n_1+n_2-2).