CC-1 Sem-1 Descriptive Statistics

1. Is the median of the logarithms of a set of positive real numbers equal to the logarithm of the median? What will be your answer for arithmetic mean? Justify your answers.
2. How do ordinal data differ from nominal data? Explain through illustrations.
3. Distinguish between absolute and relative measures of dispersion. What are the percentiles of a set of data?
4. Define Pearson's co-efficient of Skewness. Show that it lies between -1 and $+1$.
5. Describe how one can use "Box plots" to compare two frequency distributions.
6. Discuss the concept of trimmed mean as a measure of central tendency. Why and how is it used?
7. Show that the standard deviation of a set of observations, apart from a numerical constant, may be regarded as the root mean square of all possible pairs of differences of the observations.
8. Define Kendall's tau-coefficient. Show that it is a product moment correlation coefficient.
9. You are given the data $\{(x_i, y_{ij}),\ j = 1(1)n_i,\ i = 1(1)p\}$ on (x, y). Find the best linear regression of y on x and discuss how the square of the correlation coefficient (r^2) can be used as a measure of efficacy of the regression. In this connection define square of the correlation ratio (e^2). Show that $r^2 \le e^2 \le 1$. Discuss the cases of equalities.
10. Define skewness and kurtosis. If β_1 and β_2 are respectively the two measures of skewness and kurtosis of a probability distribution then show that $\beta_2 \ge \beta_1 + 1$.
11. Show that the mean deviation will be least if the deviations are measured about the median.
12. What do you mean by correlation ratio (e_{yx}) of y on x? Show that $0 \le r^2 \le e_{yx}^2 \le 1$, where r is the correlation coefficient. Explain the following situations:

    (i) $e_{yx}^2 = r^2$, (ii) $e_{yx}^2 = 0$.

13. Write the formula of regression line of x on y. Show that the regression coefficient is independent on the change of base but is dependent on the change of scale.
14. Distinguish between:
    (a) Interval scale and Ratio scale of measurement,
    (b) Cross-sectional data and time series data.
15. A manager of a large corporation has recommended that an additional amount of Rs. 1,00,000/- be given to prevent a valued subordinate from moving to another company. What internal and external source of data might be used to decide whether such a salary increase is appropriate?
16. Discuss two merits and two demerits of a boxplot.
17. How can you find the first quartile for an ungrouped data of size n? Explain with examples for n=10 and n=11.
18. Let $\overline{x}, M$ and s be the mean, median and standard deviation respectively, for a set data. Show that $s \geq |\overline{x} - M|$. What is the implication of the above result?
19. How would you design and administer a questionnaire?
20. Discuss the method of obtaining mode for grouped data when the class widths are equal. Indicate two merits and two demerits of mode.
21. Illustrate, with an example, the use of stem and leaf display to summarize data. Discuss its advantages over histogram.
22. Define correlation index and correlation ratio. Show that correlation index is a non-decreasing function of its order.
23. Define intraclass correlation. Suppose a variable is measured for all the members of n families, the i-th family having t_i members, $i = 1, 2, \dots, n$. Deduce the expression for the intraclass correlation coefficient r_I in this case. For $t_1 = t_2 = \cdots = t_n = t$, discuss the cases

    (i) $r_I = -\frac{1}{t-1}$ and (ii) $r_I = 0$.
24. Define interquartile range. What does it represent? Why this is useful? Where do we use it?
25. Show that the standard deviation S of a set of observations $(x_1, x_2, \dots, x_n)$ is given by

$$nS^2 = \sum_{i=2}^{n} \frac{i}{(i-1)} (x_i - \overline{x}_i)^2, \quad \text{where } \overline{x}_i = \frac{1}{i}\sum_{j=1}^{i} x_j, \qquad i = 2(1)n.$$

26. If $u = ax + by,\ v = bx - ay$ and $\text{cov}(u, v) = 0$ then show that

$$\sqrt{V(u).V(v)} = \sqrt{V(x).V(y)}(a^2 + b^2)\sqrt{1 - r_{xy}^2}$$

where the symbols have their usual meaning.

27. Distinguish between: (i) Primary data and Secondary data, (ii) Ordinal data and Nominal data. Give examples.
28. What are Ogives? How median is calculated from Ogives?
29. What is coefficient of variation? How does it differ from standard deviation? Mention situations where coefficient of variation is useful.
30. What is scatter diagram? Discuss, in brief, its use in finding a measure of association between two variables.
31. What is regression of one variable on the other in case of two variables? How does it differ from the least square linear regression?
32. Show how the standard deviation can be computed without finding arithmetic mean.
33. Explain the method of least squares. Use this method to find the best fitted line for predicting the value of one variable given the values of the other variables. Suggest a suitable unit free measure of accuracy of such predictor.
34. When is a discrete frequency distribution said to be symmetric? For such a distribution, discuss the relation between mean, median and mode.
35. Define "odds ratio". Show that sample odds ratio does not change when both cell counts within any row are multiplied by a nonzero constant or when both cell counts within any column are multiplied by a nonzero constant. Discuss the implication of the above result using a real life example.

CC-2 Sem-1 Probability and Probability Distributions –I

1. Explain clearly i) Elementary events and sample space or event space ii) Exhaustive and incompatible events.
2. Let A, B, C be three arbitrary events. Find expressions for the probabilities of following events: i) Both A & B occurs but not C (ii) At least two events occur (iii) Exactly two events occur (iv) Only one event occur (v) None of the three events occur (vi) All the three events occur.
3. When are a number of events said to be equally likely? Give example of i) Equally likely cases (ii) Four cases which are not equally likely (iii) five cases in which one case is more likely than the other four.
4. Give example of (i) three mutually exclusive events (iv) three events which are not mutually exclusive.
5. Can i) events be mutually exclusive and exhaustive? (ii) events be mutually exhaustive and independent? (iii) events be mutually exhaustive and independent? (iv) events be mutually exhaustive, exclusive and independent?
6. Give the classical definition of probability and point out its limitations.
7. Give the sample space for each of the following experiments: i) Tossing a coin till the 1st head appears (ii) Tossing a coin till two heads or two tails appear in succession (iii) Noting the error committed in measuring a one-meter rod.
8. Stating the rationale behind the development of axiomatic approach of probability theory give an outline of the axiomatic treatment of probability theory due to Kolmogorov.
9. State and prove the properties of probability function derived from Axiomatic definition of probability.
10. State and prove (i) Boole's Inequality (ii) Bonferroni's inequality (iii) Poincare's theorem (iv) The theorem of probability of occurrence of exactly m out of n events (v) The Theorem of probability of occurrence of at least m out of n events.
11. If three numbers are chosen at random from the 1st thirty natural numbers, hats the probability that they will be in geometric progression?

12. Obtain the probability that in k throws of a die each of the numbers $1,2,\ldots,6$ will appear at least once.
13. Show that the probability of the joint occurrence of the events can be expressed as a series consisting of probabilities of all possible unions of events.
14. Explain and define conditional probability and in this context discuss the notion of 'independence'. Show that conditional probability obeys Axiomatic definition of probability. Also define pairwise and mutual independence with examples.
15. State and prove Bayes' theorem of conditional probability.
16. State and prove the theorem of (i) Compound probability (ii) Total probability.
17. In an examination, the examinee either knows the answer with probability p or he guesses with probability $(1-p)$. Let the probability of answering the question correctly be1for an examinee who knows the answer and $1/m$ for one who guesses (m being the number of multiple choice alternatives). Supposing an examinee answers a question correctly, what is the probability that he really knows the answer?
18. Define a random variable. Give two examples of random variables.
19. If X is a random variable on the probability space $(\Omega, \mathcal{F}, P)$ then show that $Y = X^2$ is also a random variable on the same probability space.
20. Define discrete and continuous random variables. Also discuss their probability distributions.
21. Consider the random experiment of rolling two fair dice simultaneously. Write the corresponding sample space and the σ-field. Define two random variables on the given probability space.
22. Let X be a random variable with cdf

$$F(x) = \begin{cases} 0 & if\ x < 0.5 \\ 0.3 & if\ 0.5 \le x < 1 \\ 0.5 & if\ 1 \le x < 1.5 \\ 0.8 & if\ 1.5 \le x < 2 \\ 1 & if\ x \ge 2 \end{cases}$$

Identify the nature of the random variable and find the pdf/pmf of X. Also find $E(X)$ and $Var(X)$.

23. Find the constant c such that the following is a pdf of a random variable

$$f(x) = c\exp(|x-5|)\ ,x \in \mathbb{R}$$

Also find the median and quartile deviation of the distribution.

24. Show that $MD_{\mu}(X) \leq SD(X)$, for any random variable X. Discuss the equality case.

25. Give an example of a distribution such that mean of the distribution does not exist. Justify your answer.

26. State and prove Markov's inequality.

27. State and prove Chebyshev's inequality.

28. Derive Chebyshev's inequality using Markov's inequality.