

INDIAN STATISTICAL INSTITUTE

Recruitment for posts of Associate Scientist A (*Specialisation: Statistics*)

SECTION II

Syllabus for Online and Skill Tests

Elementary probability theory, Bayes theorem in probability, measures of central tendency, dispersion, correlation and regression, probability distributions, standard distributions: binomial and normal, expectation and variance, joint, conditional and marginal distributions, distribution of function of a random variable, sample survey, unbiased estimator, sufficiency and completeness, consistency, MVUE, maximum likelihood estimator, exponential family of distribution, Bayes estimation, hypothesis testing, Neyman-Pearson lemma, MP and UMP tests, likelihood ratio test, linear models and Gauss-Markov theorem, basics of design of experiments, ANOVA, Monte Carlo simulations, multiple regression model (residuals, diagnostics, multicollinearity, model building), time series analysis (stationarity, trend, seasonality, and ARMA model).

Sample Questions for the Online Test

Note: For each of the questions there are four suggested answers, of which only one is correct. You will score

4 marks for each correctly answered question,

0 mark for each incorrectly answered question, and

1 mark for each unattempted question.

1. Suppose 50 men out of 10,000 and 25 women out of 10,000 are colour blind in a large population, where males and females are in equal proportion. If a colour blind person is chosen at random, the probability of the person being male is

(a) $\frac{1}{3}$ (b) $\frac{2}{3}$ (c) $\frac{3}{4}$ (d) $\frac{4}{5}$

2. Let X be a discrete random variable that takes values 4, 5 and 6 with probabilities 0.2, 0.3 and 0.5, respectively. Then the expected value of $|X - 5|$ is

(a) 0.7 (b) 0.3 (c) -0.3 (d) 0

3. A student tosses an unbiased coin 28 times. Let Y be a random variable denoting the number of heads in first n tosses whereas Z denotes the number of heads in last n tosses. What would be the value of n so that the correlation coefficient between Y and Z is 0.6?

(a) 7 (b) 10 (c) 14 (d) 20

4. A polling agency has decided to increase the size of its random sample of voters from about 2000 people to about 5000 people. The effect of this increase in sample size is to

- (a) reduce the bias of the estimate
- (b) increase the standard error of the estimate
- (c) reduce the variability of the estimate
- (d) increase the width of the confidence interval for the parameter

5. Let X_1, \dots, X_n be a random sample of size n from the Bernoulli distribution with probability of success p ($0 < p < 1$). Then an unbiased estimator for the variance of

$$X = \sum_{i=1}^n X_i \text{ is}$$

(a) $\frac{X(n-X)}{n(n-1)}$ (b) $\frac{X(n-X)}{n}$ (c) $\frac{X(n-X)}{(n-1)}$ (d) $\frac{X(X-1)}{n(n-1)}$

6. Let X_1, \dots, X_n be a random sample of size n from the normal distribution with mean 0 and variance σ^2 . The maximum likelihood estimator of σ^2 for this sample is

(a) $\frac{1}{n} \sum_{i=1}^n X_i^2$ (b) $\frac{1}{n-1} \sum_{i=1}^n X_i^2$ (c) $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ (d) $\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$

7. For a variable having three categories, the observed and expected frequencies are calculated for each cell. Suppose a person wants to apply Pearsonian χ^2 statistic to check the goodness of fit. The value of the test statistic is 3.89 and ν is the degrees of freedom associated with the statistic. The approximate p -value for this test is

(a) $P(\chi_\nu^2 > 3.89 | H_0)$ (b) $P(\chi_\nu^2 > (3.89)^2 | H_0)$
(c) $P(\chi_\nu^2 < 3.89 | H_0)$ (d) none of these

8. In a two-way classified model (without interaction) with 3 blocks and 3 treatments and a single observation per cell, one observation is accidentally lost. Then one can say with certainty that
- (a) difference of effects of at least one pair of treatments will become non-estimable
 - (b) difference of effects of at least one pair of blocks will become non-estimable
 - (c) at least one estimable parameter will become non-estimable
 - (d) no estimable parameter will become non-estimable
9. After fitting a linear regression of Y on X , a statistician plots the residuals against the values of X . It is seen that with increase in values of X , the values of residuals become more scattered. What does this indicate?
- (a) presence of multicollinearity
 - (b) presence of heteroskedasticity
 - (c) autocorrelated errors
 - (d) measurement error of y variable
10. Hourly measurements of temperature are recorded at a weather station over a period of 10 years. Let X_1, \dots, X_{168} be the hourly temperature during the period 1st to 7th February, averaged over 10 years (e.g., X_1 is the average of 10 temperature measurements taken in the first hour of 1st February of the different years). Then the time series X_1, \dots, X_{168} is likely to have
- (a) a dominant trend component
 - (b) a dominant seasonal component
 - (c) a dominant random component
 - (d) no dominant component
-