

CC-8 : Survey Sampling and Indian Official Statistics

1. What is sampling?
2. What do you mean by sampling error?
3. Which principle governs the theory of sampling?
4. What is non-sampling error?
5. How does complete enumeration differ from sampling?
6. What are the main goals of sampling?
7. Why in most of the cases sampling becomes a necessity?
8. What are the different types of sampling?
9. Give an example of purposive sampling.
10. Give an example of probability sampling.
11. Give an example of mixed sampling.
12. Between SRSWR and SRSWOR which is more efficient and why?
13. Write down the estimator of population mean and its standard error in case of SRSWOR.
14. Write down the estimator of population mean and its standard error in case of SRSWR.
15. Write down the estimator of standard error of the unbiased estimator of population mean in case of SRSWOR.
16. Write down the estimator of standard error of the unbiased estimator of population mean in case of SRSWR.
17. When will stratified random sampling generally be applied?
18. In which situation stratified random sampling becomes really effective?
19. What is proportional allocation formula in case of stratified random sampling?
20. What is Neyman's optimum allocation formula in case of stratified random sampling?
21. What is the estimator of gain in precision in case of stratified random sampling?
22. Write down the estimator of mean in stratified random sampling.
23. Write down the estimator of standard error of the estimator of population mean in stratified random sampling under SRSWOR set up.
24. Write down the variance of the estimator of population mean in stratified random sampling both in SRSWOR and SRSWR.
25. Give an example where systematic sampling is proper to apply.
26. Write down the estimator of population total and standard error of estimator of population mean in linear systematic sampling.
27. Write down the estimator of population total in case of circular systematic sampling.
28. What will be the relative position of efficiencies of systematic, stratified and random sampling when there is a linear trend in the population?

29. Give an example where interpenetrating sub-samples are used in a sampling scheme.
30. Explain the cost function of stratified random sampling scheme.
31. Cite an example where cluster sampling is proper to apply.
32. What is the main disadvantage of cluster sampling in terms of precision?
33. Write any one of the advantages of using cluster sampling.
34. Write down the expression of either estimator T of population mean or variance V of T or estimator of V in case of cluster sampling with equal size cluster.
35. Write down the cost function of two stage sampling.
36. Does two-stage sampling have any advantage over cluster sampling?
37. What is the estimator of population mean in Two-stage sampling?
38. What is the standard error of estimate of population mean in Two-stage sampling?
39. Write down the estimator of variance of the estimator of population mean in Two-stage sampling.
40. Explain Regression Estimator in details.
41. Compare Regression Estimator with mean per unit estimator and ratio estimator.
42. Find the bias of ratio estimator.
43. Explain the cost function in cluster sampling.
44. Write down the efficiency of cluster sampling with respect to SRSWOR and comment on efficiency in detail.
45. What are the major sources of data on labour and employment in India? Cite the different publications related to labour statistics.
46. Discuss the functions of the Central Statistical Organization and National Sample Survey Organization in the Indian statistical system.
47. Write the name of the authority of publication, periodicity and coverage of "Agricultural Situation in India".
48. Write short notes on the following:
 - (a) The annawari method and the crop-cutting experiments for estimating crop yield.
 - (b) Registration of vital events
49. Which organization computes national income in India? Mention one other major activity performed by this organization.
50. Distinguish between the "production approach" and "income approach" of computing national income.
51. Give two reasons why the normal distribution cannot be used to fit the Indian income data.
52. Discuss why and how the Pareto distribution can be used to fit income data.

CC-9 : Statistical Inference-I and Sampling Distributions

1. Discuss the concept of sampling distribution and its role in statistical inference.
2. Derive the m.g.f. of $\chi^2(n)$. Hence derive the same for $Y = \sum_{j=1}^k a_j \chi_j^2(n_j)$ where $\chi_j^2(n_j)$ variables are all independently distributed and a_j 's are positive constants. For what sort of a_j 's would you get Y as a suitable χ^2 variable?
3. If X_1 and X_2 are two independent observations from a rectangular distribution with p.d.f.

$$f(x) = \frac{1}{b-a}, \quad a < x < b$$

find the distribution of $X_1 + X_2$.

4. Express the parameters of F-distribution in terms of its expectation and variance.
5. If $P_x = P[\chi^2(n) \leq x]$, show that $x < \frac{n}{1 - P_x}$.
6. Show that the Student's t-statistic has a p.d.f. symmetric about zero. Obtain its limiting p.d.f. when its parameter approaches infinity.
7. Let X_1 and X_2 be independently and uniformly distributed on the interval (0, 1). Find the probability density function of (i) $|X_1 - X_2|$ and (ii) X_1 / X_2 .
8. Let X have an F-distribution with n_1 and n_2 degrees of freedom. Show that $Y = 1 / \left(1 + \frac{n_2}{n_1} X\right)$ has a beta distribution. Find the parameters of this beta distribution.
9. Obtain the sampling distribution of the sample correlation coefficient in a bivariate normal set up with population correlation coefficient zero. Indicate one use of this distribution.
10. Suppose that X_1 and X_2 are two independent random variables with common p.d.f.

$$f(x) = \frac{\alpha^\nu}{\Gamma(\nu)} e^{-\alpha x} x^{\nu-1}, \quad x > 0, \alpha > 0, \nu > 0.$$

Show that $X_1 + X_2$ and $X_1 / (X_1 + X_2)$ are stochastically independent. Obtain the marginal p.d.f. of X_1 / X_2 .

11. Find the distribution of the sample regression coefficient of Y on X based on n pairs of observations $(x_i, y_i), i = 1, 2, \dots, n$ where X is non-stochastic and conditional distribution of Y given $X = x$ is normal with the expectation linear in x and variance independent of x .
12. Show that the sample mean and sample variance are stochastically independent for a random sample drawn from a normal population.
13. If the p.d.f. of X is an F-distribution with parameters ν_1 and ν_2 , what is the distribution of $\nu_1 X$ when ν_2 approaches infinity?
14. A fair die is thrown. Let X be the face value that turns up. Let X_1, X_2, X_3, X_4 be four independent observations on X. Find the probability mass function of $Y = \text{minimum}\{X_1, X_2, X_3, X_4\}$.

15. Let $X_{(1)} < X_{(2)} < X_{(3)}$ be the order statistics for a random sample of size 3 from an exponential distribution with mean θ^{-1} , $\theta > 0$. Let $Z_1 = X_{(3)} - X_{(2)}$ and $Z_2 = X_{(2)}$. Show that Z_1 and Z_2 are independent.
16. Suppose $X_{(1)} < \dots < X_{(n)}$ represent the order statistics based on a sample of size n from a continuous distribution having c.d.f. $F(x)$. Find the joint distribution of $U_1 = F(X_{(1)})$, $U_2 = F(X_{(2)}) - F(X_{(1)})$, ..., $U_n = F(X_{(n)}) - F(X_{(n-1)})$. Hence find the distribution of $\prod_{i=1}^n U_i$.
17. Explain the distinction between
- standard error and standard deviation
 - null hypothesis and alternative hypothesis
 - simple hypothesis and composite hypothesis
 - confidence interval and acceptance region
 - type I error and type II error
 - level of significance and size of a test.
18. Let X_1 and X_2 follow binomial distributions with parameters (n_1, p_1) and (n_2, p_2) respectively, independently of each other. Describe an exact procedure for testing $H_0: p_1 = p_2$ against $H_1: p_1 < p_2$.
19. Let X and Y be two Poisson random variables with parameters λ_1 and λ_2 respectively. Describe the exact test procedure for testing $H_0: \lambda_1 = \lambda_2$ against $H_1: \lambda_1 > \lambda_2$.
20. Explain the concept of p -value with an example.
21. Describe the Student's t , Fisher's t and Paired t tests in connection with testing of statistical hypotheses problems stating clearly the underlying assumptions. Give one real life example for each.
22. Set up a 95% confidence interval for the ratio of two means of two independent normal populations with common unknown variance.
23. Discuss a procedure for testing the equality of variances of two correlated normal variables.
24. Suggest an exact test of significance for the difference of two normal means. Consider both the correlated and the uncorrelated case.
25. Suggest an exact test of significance for the ratio of two normal variances. Consider both the correlated and the uncorrelated case.
26. Given a pair of random variables X and Y having the bivariate normal distribution $N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, suppose one wants to predict Y for a given value x_0 of X . Give $100(1 - \alpha)\%$ limits of prediction for the value of Y which corresponds to x_0 .

CC-10 : Index Numbers and Time Series Analysis

1. Describe 'formula error', 'sampling error' and 'homogeneity error' in the context of index numbers.
2. What are the different uses of index numbers?
3. How is the base period of an index number chosen?
4. What is a cost of living index number?
5. How can one construct the cost of living index number for middle class families in Kolkata?
6. Show that harmonic mean of price relatives, weighted by the product of base year quantities and current year prices, is the equivalent of an aggregative index weighted with base year quantities.
7. Show that Edgeworth-Marshall index number necessarily lies between Laspeyres' and Paasche's index numbers.
8. The Laspeyres' formula gives 450 and 810 as the price indices for 1993 and 2003, respectively, with 1981 as the base year. Can you find the index for 2003 with 1993 as the base – discuss.
9. Given that the cost of living index for 2004 with 1991 as base year is 250, will Mr. X be satisfied if his income rises to Rs.12,000 in 2004 as compared to Rs.5000 in 1991?
10. What are the different tests for consistency that a price index number should satisfy?
11. Why Fisher's index number is called an ideal index?
12. What are price relatives? Discuss how Laspeyres' and Paasche's index numbers may be looked upon as weighted averages of price relatives.
13. Describe the moving average method of trend fitting. What would be the effect on the moving average series if the original series undergo a base and scale change?
14. Describe a suitable method of fitting a modified exponential trend to a given time-series. Indicate how this method can also be applied for fitting other complicated trend curves.
15. Distinguish between seasonal and cyclical fluctuations in a time series data.
16. Discuss the ratio to trend method to measure the seasonal fluctuations. Does the method depend on the seasonal pattern?
17. When is a time-series said to be stationary?
18. Write down a moving average process of order 2 (MA(2)). Is this process stationary?
19. Define autocorrelation function. Derive the autocorrelation function of a MA(2) process.
20. Describe the exponential smoothing technique for forecasting.
21. Estimate the parameters of an AR(2) process using the Yule-Walker equations.
22. What is an Autoregressive process?
23. Distinguish between Moving average and Autoregressive processes.
24. The number of malaria cases in a particular region is seen to be more or less constant over the years, although there is a variation in the monthly data with a peak either in July or August depending upon the rains. Describe in detail how you will analyse this data.