No. of Printed Pages : **12**

PHDSTAT

00364

Entrance Test for

Ph.D. (STATISTICS) Programme - 2016

Time : 3 hours *Maximum Marks : **100***

Note :

(i) ***All*** *questions are* ***compulsory****.*

(ii) *Objective type questions given in* ***Part - A*** *carry* ***one*** *mark each.*

(iii) *Descriptive type questions given in* ***Part - B*** *carry 5 marks each.*

(iv) *Calculator (Non-Programmable) is allowed.*

RESEARCH METHODOLOGY

PART - A

1. Arrange the following steps of research in correct sequence :
   (i) Identification of research problem
   (ii) Listing of research objectives
   (iii) Collection of data
   (iv) Methodology
   (v) Data analysis
   (vi) Results and discussion
   (1) (i) - (ii) - (iii) - (iv) - (v) - (vi)
   (2) (i) - (ii) - (iv) - (iii) - (v) - (vi)
   (3) (ii) - (i) - (iii) - (iv) - (v) - (vi)
   (4) (ii) - (i) - (iv) - (iii) - (v) - (vi)

2. Research is done for :
   (1) interest in research
   (2) knowledge of research techniques
   (3) experience in conducting research
   (4) interest in the discipline concerned

3. A good research always begins with :
   (1) an original idea
   (2) preparation of the plan and design for study
   (3) study of relevant research methodology
   (4) study of what research others have done

4. A research problem is feasible only when :
   (1) it as researchable (2) it is new
   (3) it has some utility (4) all of the above

5. The longitudinal approach of research deals with :
   (1) horizontal researches (2) long-term researches
   (3) short-term researches (4) none of the above

6. What is the purpose of research ?
   (1) To describe and explain a new phenomenon
   (2) To verify what has already been established
   (3) To reject what has already been accepted as a fact
   (4) all of the above

7. Which one of the following is not a type of Descriptive Research method ?
(1) Correlational (2) Causal Comparative
(3) Survey (4) Developmental study

8. While conducting experimental research a researcher should control the :
(1) independent variables (2) dependent variables
(3) no variables (4) extraneous variables

9. Which of the following sampling methods is based on probability ?
(1) Convenience sampling (2) Quota sampling
(3) Judgement sampling (4) Stratified sampling

10. Which of the following techniques yields a simple random sample ?
(1) Choosing volunteers from a M.Sc. class to participate.
(2) Listing the individuals by ethnic group and choosing a proportion from within each ethnic group at random.
(3) Numbering all the elements of a sampling frame and then using a random number table to pick cases from the table.
(4) Randomly selecting schools, and then sampling everyone within the school.

11. If we took the 500 people attending a college in New Delhi, divided them by gender, and then took a random sample of the males and a random sample of the females, the variable on which we would divide the population is called the :
(1) independent variable (2) dependent variable
(3) stratification variable (4) sampling variable

12. A number calculated with complete population data and quantifies a characteristic of the population is called :
(1) A datum (2) A statistic (3) A parameter (4) A population

13. The finite population correction factor is :
(1) $\sqrt{\frac{N-1}{N-n}}$ (2) $\sqrt{\frac{N+1}{N+n}}$ (3) $\sqrt{\frac{N-n}{N-1}}$ (4) $\sqrt{\frac{N-n}{n-1}}$

14. A population contains 2 items from which 4 items are selected at random with replacement, then all possible, sample are :
(1) 16 (2) 8 (3) ${}^4C_2$ (4) 4

15. A researcher divides the population of the users of a particular product into three groups based on degree of use. If the researcher then draws a random sample from each user group independently, she has created a :

(1) random sample (2) stratified sample

(3) judgement sample (4) quota sample

16. An unordered sample of size n can occur in :

(1) n ways (2) n! ways (3) one way (4) n^2 ways

17. The discrepancies between sample estimate and population parameter is termed as :

(1) human error (2) formula error

(3) non-sampling error (4) sampling error

18. Which of the following is a qualitative variable ?

(1) Favourite brand of toothpaste

(2) Number of people preferring shopping at City Mall

(3) The marks in the class test in Statistics paper of 100 students

(4) Sales of a departmental store during a particular week

19. The manager of an estate - agency wishes to monitor the performance of her sales staff. She records the number of properties sold by each of the 15 staff members for a randomly chosen period of time. What type of variable is "number of properties sold" ?

(1) Quantitative and discrete (2) Quantitative and continuous

(3) Qualitative and discrete (4) Qualitative and continuous

20. Data taken from the publication, 'Agricultural Situation in India' will be considered as :

(1) primary data (2) secondary data

(3) primary and secondary data (4) neither primary nor secondary data

21. If the actual value of a unit is 415 and its estimated value is 400, the absolute error is :

(1) -15 (2) 15 (3) 0.0375 (4) -0.0361

22. The series

Place	No. of accidents per day
Delhi	10
Kolkata	15
Mumbai	18
Chennai	17
Indore	7

is of the type :

(1) Spatial (2) Geographical (3) Industrial (4) Time series

23. Consider the following data :

14, 16, 16, 22, 25, 38, 38, 38, 38, 2000

Which of the measures of central tendency would be the most useful ?

(1) mean (2) mode (3) median (4) HM

24. If ρ is the simple correlation, the quantity $(1-\rho^2)$ is called :

(1) coefficient of determination (2) coefficient of non-determination

(3) coefficient of alienation (4) none of the above

25. If the standard deviation of a random variable X is σ, then what is the standard deviation of $(\alpha+\beta X)/\gamma$ where α, β and γ are constants ?

(1) $6/\gamma$ (2) $(\beta^2\sigma)/\gamma$ (3) $(\beta\sigma)/\gamma$ (4) $(|\beta|\sigma)/|\gamma|$

26. Consider the following assertions about p-value :

(i) The smaller the p - value, the stronger is the evidence against H_0.

(ii) For given α, we reject H_0 if p-value $\leq \alpha$.

Which of the above statements is/are correct ?

(1) (i) only (2) (ii) only (3) both (i) and (ii) (4) neither (i) nor (ii)

27. The average and variance of rainfall at four stations A, B, C and D based on one month data are given below :

Station	A	B	C	D
Mean rainfall in mm	10	14	8	16
Variance	2.5	3.6	4.9	4.4

In which station was there consistent rainfall ?

(1) Station A (2) Station B (3) Station C (4) Station D

28. For a frequency distribution, a two parameter normal distribution was fitted. The observed and expected frequencies in the various classes are given below :

Class	1	2	3	4	5	6
Observed Frequency	6	14	20	19	16	5
Expected Frequency	4	15	21	21	15	4

To test the goodness of fit using the χ^2 - statistic, what is/are the degree(s) of freedom of the χ^2 - statistic ?

(1) 5 (2) 3 (3) 2 (4) 1

29. Which of the following statements about hypothesis testing is true ?

(1) If the p-value is greater than the significance level, we fail to reject H_0.

(2) A Type - II error is rejecting the null when it is actually true.

(3) If the alternative hypothesis is that the population mean is greater than a specified value, then the test is a two-tailed test.

(4) The significance level equals one minus the probability of a Type-I error.

30. Weighted mean gives a higher value than unweighted mean if :

(1) all the items have equal weights

(2) larger items have higher weights and smaller items have lower weights.

(3) larger items have lower weights and smaller items have higher weights.

(4) none of the above

31. If each value of a series is multiplied by 10, the coefficient of variation will be increased by :

(1) 5 percent (2) 10 percent (3) 15 percent (4) 0 percent

32. Whether a test is one-sided or two-sided depends on :

(1) alternative hypothesis (2) composite hypothesis

(3) null hypothesis (4) simple hypothesis

33. A population is distributed as $N(\mu, 10.24)$. A sample of 576 items has a mean 4.7. The value of the statistic to test $H_0 : \mu = 5.2$ is :

(1) 3.75 (2) 28.125 (3) -3.75 (4) none of the above

34. Assume that the daily sales of petrol follows exponential distribution. The hypothesis that the sales of petrol is 1000 litres per day is tested against the hypothesis that it is 1500 litres per day. If the sales on a day is 1200 litres or more. H_0 is rejected, the size of type I error is :

(1) $1-e^{-1.2}$ (2) $e^{1.2}$ (3) $e^{-1.2}$ (4) none of the above

35. If β_{YX} and β_{XY} are two regression coefficients, they have :

(1) same sign (2) opposite sign

(3) either same or opposite signs (4) nothing can be said

PART - B

1. Identify a research problem in the field of sampling and briefly explain the steps involved in defining a research problem. Explain the necessity of identifying a research problem in sampling. Also, discuss the characteristics of a good research design.

2. A leading automobile company plans to introduce a new car. Design a suitable questionnaire to identify its demand. Also, state any four statistical analysis methods that are to carried out for effective decision making.

3. Fifty students appeared in a PH.D. entrance examination of a university, the regression equation of the marks in Research Methodology (X) on the marks in Sampling (Y) is $3Y-5X+180=0$. The mean marks in Research Methodology is 44 and the variance of the marks in Sampling is $9/16^{th}$ of the variance of the marks in Research Methodology. Find the mean marks in sampling and the coefficient of correlation between marks in two papers.

STATISTICS

PART - A

1. In a popular shopping centre, the waiting time for an ATM machine is found to be uniformly distributed between 1 and 5 minutes. What is the probability of waiting between 2 and 3 minutes to use the ATM ?

(1) 0.25 (2) 0.50 (3) 0.75 (4) 0.20

2. A new car salesperson knows that he sells cars to one in every twenty customers who enter the showroom. What is the probability that he will sell a new car to exactly two of the next three customers ?

(1) 0.007 (2) 0.021 (3) 0.003 (4) 0.010

3. If a random sample $x_1, x_2, \ldots, x_n$ of size n is taken from Exp (λ) then $Y=\min(x_1, x_2, \ldots, x_n)$ follows :

(1) Log-normal distribution (2) Gamma distribution

(3) Chi-square distribution (4) Exponential distribution

4. If X and Y follows BVN $\left(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho\right)$ distribution, the condition distribution $(X/Y=y)$ follows :

(1) $N\left[\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y), \sigma_X^2\left(1-\rho^2\right)\right]$

(2) $N\left[\mu_X + \rho\frac{\sigma_Y}{\sigma_X}(x-\mu_X), \sigma_Y^2\left(1-\rho^2\right)\right]$

(3) $N\left[\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(y-\mu_Y), \sigma_Y^2\left(1-\rho^2\right)\right]$

(4) $N\left[\mu_Y + \rho\frac{\sigma_X}{\sigma_Y}(x-\mu_X), \sigma_X^2\left(1-\rho^2\right)\right]$

5. For a random variable X, if E(X) and E(|X|) exist, then :

(1) $|E(X)| \geq E(|X|)$ (2) $|E(X)| \leq E(|X|)$ (3) $|E(X)| < E(|X|)$ (4) $|E(X)| > E(|X|)$

6. If a random variable X is symmetric about 0, then :

(1) $\phi_X(t)$ is real and odd function of t
(2) $\phi_X(t)$ is complex and odd function of t
(3) $\phi_X(t)$ is real and even function of t
(4) $\phi_X(t)$ is complex and even function of t

7. A researcher wants to investigate the amount of lead per litre of waste water produced by her company. She plans to use statistical methods to estimate the population mean of lead content per litre of water. Based on previous recordings she assumes that the lead content is normally distributed with a standard deviation of 20 mg per litre. How large a sample should she take to estimate the mean lead content per litre of water to within 1 mg with 95% confidence ?

(1) 2041 (2) 1537 (3) 385 (4) 865

8. A 95% confidence interval for the population mean is calculated to be 75.29 to 81.45. If the confidence level is increased to 98%, the confidence interval will :

(1) become narrower (2) remain the same
(3) become wider (4) double in size

9. If X and Y are random variables such that $E(Y)=\mu$, Var (Y) > 0, and $E(Y/X)=\phi(X)$, then :

(1) Var $[\phi(X)] \geq$ Var(Y) (2) Var $[\phi(X)] \leq$ Var(Y)
(3) Var $[\phi(X)] = 0$ (4) Var $[\phi(X)]$ does not exist

10. If X is Binomial (1, θ), then unbiased estimator of θ^2 based on X is :

(1) $\overline{X}^2$ (2) $1/\overline{X}^2$ (3) $\overline{X}$ (4) Does not exist

11. If $x_1, x_2, \ldots, x_n$ is a random sample taken from U(0, θ), then consistent estimator of θ is :

(i) $2\bar{x}$
(ii) max $(x_1, x_2, \ldots, x_n)$
(iii) min $(x_1, x_2, \ldots, x_n)$

Which of the above statements is/are correct ?

(1) (i) only (2) (ii) only (3) (i) and (ii) only (4) (i) and (iii) only

12. A critical region is called unbiased if :

(1) $\alpha+\beta \geq 1$ (2) $\alpha+\beta \leq 1$ (3) $\alpha \geq \beta$ (4) $\alpha \leq \beta$

13. In non-parametric theory, the most frequently used measure of location is :

(1) arithmetic mean (2) median
(3) mode (4) geometric mean

14. If random sample $(x_1, x_2, \ldots, x_n)$ of size n is taken from :

$$f(x,\theta)=\begin{cases}\dfrac{1}{\theta} & ; \quad x = 1,2\ldots, \theta\\ 0 & ; \quad \text{elsewhere}\end{cases}$$

then max $(x_1, x_2, \ldots, x_n)$ is :

(1) sufficient statistic (2) complete statistic
(3) both (1) and (2) (4) neither (1) nor (2)

15. If T is an unbiased estimator of the function of a parameter θ, say γ(θ) then :

(1) $\text{Var (T)} \geq \dfrac{[\gamma'(\theta)]^2}{E\left[\dfrac{\partial}{\partial\theta}(\log L)\right]^2}$ (2) $\text{Var (T)} \leq \dfrac{[\gamma'(\theta)]^2}{E\left[\dfrac{\partial}{\partial\theta}(\log L)\right]^2}$

(3) $\text{Var (T)} > \dfrac{[\gamma'(\theta)]^2}{E\left[\dfrac{\partial}{\partial\theta}(\log L)\right]^2}$ (4) $\text{Var (T)} < \dfrac{[\gamma'(\theta)]^2}{E\left[\dfrac{\partial}{\partial\theta}(\log L)\right]^2}$

16. Friendman's F is distributed as :

(1) snedecor's F (2) student's t (3) chi-square (4) none of the above

17. The most important factor in determining the size of a sample is :
(1) the availability of resources (2) purpose of the survey
(3) heterogeneity of population (4) none of the above

18. If a sample x_1, x_2, ..., x_n from a dichotomous population has n_1 items of type C_1 with proportion p and n_2 items of type C_2 with proportion q. Also.
$x_i = 1$ if $x_1 \in C_1$
$x_i = 0$ if $x_1 \in C_2$
Then which of given four relations does not hold good ?

(1) $\bar{x} = p$ (2) $q = 1 - p$ (3) $q = \frac{n_2}{n}$ (4) $p = \frac{n}{n_1}$

19. A systematic sample does not yield good results if :
(1) variation in units is periodic
(2) units at regular intervals are correlated
(3) both (1) and (2)
(4) neither (1) nor (2)

20. Which of the following statements distinguishes cluster sampling from stratified sampling ?
(1) Clusters are preferably heterogeneous whereas strata are taken as homogeneous as possible.
(2) A sample is always drawn from each stratum whereas no sample of elementary units is drawn from clusters.
(3) Small size clusters are better whereas there is no such restriction for stratum size
(4) All the above

21. The two stage sampling is better than single stage sampling :
(1) when the elements in the same stage are positively correlated
(2) when the elements in the same stage are negatively correlated
(3) when elements in the same stage are uncorrelated
(4) none of the above

22. The selected items of a sample resulted into same values pertaining to a character. The variance of the sample is :
(1) 1 (2) 0 (3) ∞ (4) not determinable

23. In case of inverse sampling, the proportion 'p' of m units of interest contained in a sample of n units is :
(1) m/n (2) $(m-1)/n$ (3) $(m-1)/(n+1)$ (4) $(m-1)/(n-1)$

24. If an investigator selects districts from a state, Panchayat samities from districts and farmers from Panchayat samities, then such a sampling procedure is known as :

(1) two stage sampling (2) three stage sampling

(3) cluster sampling (4) stratified sampling

25. A random sample of a reasonable large size possessing almost all properties of the population confirms to the principle of :

(1) inertia of large numbers (2) statistical regularity

(3) optimisation (4) Newton's first law of inertia

26. Let the standard error of an estimator T under SRSWOR is more than the standard error of T under stratified randomly sampling. Then T under stratified sampling as compared to T under SRSWOR is :

(1) more reliable (2) less reliable (3) equally reliable (4) not comparable

27. If σ_1^2 is the error variance of design - I and σ_2^2 of design - II utilising the same experiment variable, the efficiency of design - I over II is :

(1) $\frac{1}{\sigma_1^2} \Big/ \frac{1}{\sigma_2^2}$ (2) $\frac{1}{\sigma_2^2} \Big/ \frac{1}{\sigma_1^2}$ (3) σ_1^2/σ_2^2 (4) none of the above

28. In case of a random effect model, the hypothesis which is to be tested with regard to the treatment is :

(1) $\sigma_\tau^2 = 0$ (2) $\tau_i = 0$ (3) $\Sigma\tau_i^2 = 0$ (4) $\Sigma\tau_i = 0$

29. In a randomised block design with 4 blocks and 5 treatments having one missing value, the error degrees of freedom will be :

(1) 12 (2) 11 (3) 10 (4) 9

30. While analysing the data of a k $\times$ k Latin square, the error df is equal to :

(1) $(k-1)(k-2)$ (2) $k(k-1)(k-2)$ (3) k^2-2 (4) k^2-k-2

31. If the responses for treatments in a factorial experiment with factors A and B each at two levels from three replications are $a_0b_0 = 18$, $a_1b_0 = 17$, $a_0b_1 = 25$ and $a_1b_1 = 30$, the sum of square for the interaction AB is equal to :

(1) 90 (2) 20 (3) 324 (4) 675

32. If a multiple regression analysis is based on 10 independent variables collected from a sample of 125 observations, what will be the value of the denominator in the calculation of the multiple standard error of estimate ?

(1) 125 (2) 10 (3) 114 (4) 115

33. What can we conclude if the global test of regression rejects the null hypothesis ?

(1) Strong correlations exist among the variables
(2) No relationship exists between the dependent variable and any of the independent variables
(3) At least one of the net regression coefficients is not equal to zero.
(4) Good predictions are not possible

34. If there are 2 equations having 3 variables in an LPP, then the maximum number of possible basic solutions is :

(1) 8 (2) 6 (3) 3 (4) 9

35. If the arrival rate is 6 per hour and service rate is 2 per hour, then what is the probability of no customer in queue ?

(1) 0.7 (2) 0.5 (3) 0.3 (4) 0.6

PART - B

1. Define a regression estimator of Y using the information on variable X. Show that it is an unbiased estimator of population mean. Derive its variance.

2. Explain the Sampling technique of drawing the stratified random sample. Define the estimator of Population mean under this scheme and show that it is unbiased. Also obtain its sampling variance.

3. Find the Likelihood Ratio Test of $H_0 : \mu = \mu_0$ against $H_0 : \mu \neq \mu_0$ based on the sample size n from $N(\mu, \sigma^2)$. When σ is unknown.