Code No.	L – 4039

Entrance Examination for Admission to the P.G. Courses in the Teaching Departments, 2021

CSS

STATISTICS

&

APPLIED STATISTICS AND DATA ANALYTICS

General Instructions

1. The Question Paper is having two Parts — Part 'A' Objective type (60%) & Part 'B' Descriptive type (40%).
2. Objective type questions which carry 1 mark each are to be (✓) 'tick marked' in the response sheets against the appropriate answers provided.
3. 8 questions are to be answered out of 12 questions carrying 5 marks each in Part 'B'.
4. **Negative marking** :0.25 marks will be deducted for each wrong answer in Part 'A'.

Time : 2 Hours **Max. Marks : 100**

To be filled in by the Candidate									
Register Number	In Figures								
	In words								

PART – A

(Objective Type)

Choose appropriate answer from the options in the questions. **One** mark **each**.

(60 × 1 = 60 marks)

1. Five numbers are in arithmetic progression with common difference equal to 1 and their arithmetic mean equal to 3. The median is

   a) 2 b) 3

   c) 4 d) 5

---

2. For a positively skewed distribution,

   a) Mean = Median = Mode
   b) Mean > Median > Mode
   c) Mode > Median > Mean
   d) Median > Mode > Mean

3. A measure of dispersion which does not depend upon units of measurement of observations

   a) Standard deviation
   b) Quartile deviation
   c) Coefficient of variation
   d) Semi Interquartile range

4. Subjective definition of probability is based on

a) Statistical regularity
b) Empirical theories
c) Axiomatic theories
d) Personal belief of the experimenter

5. The difference between a field and sigma field is that

a) The former is closed under all finite set of operations while the latter is closed under all countable set of operations
b) The former is closed under all countable set of operations while the latter is closed under all finite set of operations
c) The former is closed under countable unions and the latter is closed under finite unions
d) The former is closed under complementations while the latter is not.

6. With the usual notations, $\left(A^C \cap B^C\right)^C$ is equal to

a) $A \cup B$ b) $A \cap B$ c) $(A \cup B)^C$ d) $(A \cap B)^C$

7. A family of ten contains four males and six females. A sample of size four is drawn without replacement. The exact distribution of number of males is

a) Poisson distribution
b) Negative binomial distribution
c) Uniform distribution
d) Hypergeometric distribution

8. For a frequency distribution, with usual notations, we get $\gamma_1 = 1$ and $\beta_2 = 4$. the distribution is

a) Positively skewed and leptokurtic
b) Symmetric and leptokurtic
c) Negatively skewed and plattykurtic
d) Symmetric and plattykurtic

9. Most of the results in statistical inference is based on the assumption that the data obeys the normal law. This can be justified using

a) Weak law of large numbers
b) Chebyshev's inequality
c) Lyaponov's inequality
d) Central limit theorem

10. (X,Y) is a bivariate random vector admitting an absolutely continuous distribution with p.d.f $f(x,y)$, marginals $f_1(x)$ and $f_2(y)$, conditionals $f(y/x)$ and $f(x/y)$ and conditional expectations *E(Y/X)* and *E(X/Y)*. Pick out the correct statement from among those given below.
    a) The marginal distributions $f_1(x)$ and $f_2(x)$ uniquely determines the distribution
    b) The conditional distributions $f(y/x)$ and $f(x/y)$ uniquely determines the distribution
    c) $f_1(x)$ and $f(y/x)$ as well as $f_2(y)$ and $f(x/y)$ uniquely determines the distribution
    d) *E(Y/X)* and *E(X/Y)* uniquely determines the distribution

11. For a bivariate random vector (X,Y), admitting an absolutely continuous distribution, denote by $R(x,y)=P(X>x,Y>y)$. Then the p.d.f $f(x,y)$ is equal to

    a) $-\frac{\partial^2 R(x,y)}{\partial x \partial y}$ b) $\frac{\partial^2 R(x,y)}{\partial x \partial y}$ c) $\int_0^x \int_0^y R(x,y)dx\,dy$ d) $1-R(x,y)$

12. The justification for using exponential distribution as a potential model in lifetime data analysis is because of
    a) Ease in computation
    b) Reproductive property
    c) Lack of memory property
    d) Existence of UMVUE

13. The probability generating function for the geometric function specified by $p(x)=q^x p; x=1, 2....0<p, q<1, q=1-p$ is
    a) $p(1+sq)^{-1}$
    b) $p(1-sq)^{-1}$
    c) $(q+ps)^{-1}$
    d) $(p+sq)^{-1}$

14. An antigen test declares that a person is COVID positive in 95 % of the cases when the person actually has COVID and in 10 % of the cases when the person does not have the disease. The prior belief is that 20 % of the people in the locality are COVID positive. The probability that a person declared as COVID positive is actually infected is
    a) 0.5 b) 0.95 c) 0.9 d) 0.7

15. A probability distribution for which the moment of generating function does not exist

a) Cauchy distribution
b) Chi-square distribution
c) Beta distribution
d) Hypergeometric distribution

16. Identify the distribution which is symmetric from among the following

a) F distribution
b) Students t distribution
c) Exponential distribution
d) Chi-square distribution

17. A curve which is extensively used to measure inequality

a) Ogive curves
b) Frequency curves
c) Lorenz curves
d) Histogram

18. X and Y are independent random variables following standard normal distribution. Then U = X/Y follows

a) Chi-square distribution
b) Cauchy distribution
c) F distribution
d) Normal distribution

19. X is a random variable following distribution function F(X). Then Y = F(X) follows

a) Normal distribution
b) Cauchy distribution
c) Parsons distribution
d) Uniform distribution

20. If X is a random variable following student t distribution. $Y = X^2$ follows

a) Student *t* distribution
b) Normal distribution
c) Chi-square distribution
d) F distribution

21. If X follows $F_{m,n}$ then $1/X$ follows

a) $F_{m,n}$
b) $F_{n,m}$
c) χ^2_{m+n}
d) t_{mn}

22. A simple random sample of size n is drawn from a population of size N with replacement. The probability that a specified unit in the population is included in the sample is

a) $\frac{1}{N}$ b) $\frac{n}{N}$ c) $1-\left(1-\frac{1}{N}\right)^n$ d) $\frac{1}{NC_n}$

23. In a circular symmetric sampling, if N is population size and n is sample size, the random start is a number between

a) 1 and N b) 1 and n

c) 1 and k where $k=\frac{N}{n}$ d) n and N

24. The arithmetic mean $\overline{X}$ of a sample of size n from the standard Cauchy distribution follows

a) Normal distribution b) Standard Cauchy distribution

c) Chi square distribution with n.d.f d) None of the above

25. X and Y are independent random variables. Pick up the wrong statement from below

a) $F_{Y/X}(y/x)=F_Y(y)$ b) $F(x,y)=F(x)F(y)$

c) $f(x,y)=f_1(x)f_2(y)$ d) $F_x(x)=F_Y(y)$

26. (X,Y) follows the bivariate normal distribution. $\rho_{XY}=0$ implies that X and Y are

a) Uncorrelated but not independent

b) Uncorrelated and independent

c) Independent but not Uncorrelated

d) Correlated and independent

27. In point estimation, the minimum variance unbiased estimator can be determined using

a) Chebyshev's inequality b) Cramer- Rao inequality

c) Rao- Blackwell theorem d) Method of moments

28. If $X_1.X_2.....X_n$ is a random sample from $N(\mu,\sigma^2)$, a sufficient statistic for σ^2 is

a) ΣXi b) ΣXi^2 c) ΣXi^3 d) $\Sigma(Xi-\mu)$

29. If $X_1.X_2.....X_n$ is a random sample from $N(\mu,\sigma^2)$ and $S^2=\frac{1}{n}\sum_1^n(xi-\bar{x})^2$, an unbiased estimator for σ^2 is

a) $\frac{s^2}{n-1}$ b) $\frac{(n-1)s^2}{n}$ c) $\frac{s^2}{n}$ d) $\frac{nS^2}{n-1}$

30. A family of distributions for which a complete sufficient statistic exists

a) Exponential family b) Power series family

c) Pearson's family d) Tucky Lamda family

31. The Pearson's coefficient of correlation is independent of

a) Change of scale only

b) Change of origin only

c) Both change of scale and change of origin

d) Neither change of scale nor change of origin

32. The following provides a method to find the regression lines

a) Method of minimum chi square

b) Method of semi averages

c) Method of minimum variance

d) Method of least squares

33. Let $En=\{x:0<x<\frac{1}{n};$ if n is odd$\}$

$=\{x:\frac{1}{n}\le x<1;$ if n is even$\}$

Then $\{En\}$

a) Converges b) Is monotone

c) Diverges d) Converges and is monotone

34. Let S = (0, 1] and $A_i = \left[\frac{1}{i}, 1\right]$ where S is the sample space and A_i are sets defined over S. Then

a) $\bigcup_{i=1}^{\infty} A_i = \{1\}$ and $\bigcap_{i=1}^{\infty} A_i = (0,1]$

b) $\bigcup_{i=1}^{\infty} A_i = (0,1)$ and $\bigcap_{i=1}^{\infty} A_i = [0,1]$

c) $\bigcup_{i=1}^{\infty} A_i = \{1\}$ and $\bigcap_{i=1}^{\infty} A_i = \{1\}$

d) $\bigcup_{i=1}^{\infty} A_i = (0,1]$ and $\bigcap_{i=1}^{\infty} A_i = \{1\}$

35. Pick up the correct statement from among the following

a) The product of regression coefficients is equal to the correlation coefficient

b) The regression lines intersect an the origin

c) Both regression lines will have same sign

d) The value of both regression coefficients can be individually greater than one

36. Power of the test is the probability of

a) Rejecting the null hypothesis when it is true

b) Rejecting the null hypothesis when it is false

c) Accepting the null hypothesis when it is false

d) Accepting the null hypothesis when it is true

37. Pick up the correct statement from among the following

a) The maximum likelihood estimators (MLE) are always unbiased

b) MLE always exists

c) MLE is always sufficient

d) The MLE is a function of sufficient statistic if it exists

38. The Cramer -Rao inequality provides

a) An upper bound for variance of an unbiased estimator

b) A lower bound for variance of any estimator

c) A lower bound for variance of an unbiased estimator

d) A lower bound for variance of maximum likelihood estimator

39. An estimator which takes into account the prior beliefs of the experimenter about the occurrence of an event

a) Maximum likelihood estimator b) Baye's estimator

c) Moment estimator d) Minimax estimator

40. Neymen -Pearson lemma provides

a) An unbiased test b) A most powerful test

c) An admissible test d) A randomized test

41. Analysis of variance is performed using the

a) F test b) Chi square test

c) t test d) Normal test

42. In analysis of data of a randomized block design with 4 blocks and 3 treatments and having one missing value, the error in degrees of freedom is

a) 12 b) 10 c) 11 d) 9

43. If the interactions AB and BC are confounded with blocks in a 2^n fractional experiment, then the automatically confounded effect is

a) ABC b) AC c) A d) C

44. In the principle of least squares, the constants appearing in the model are obtained by

a) Using the minimax criterion

b) Using the method of minimum variance

c) Using the principle of Lagrangian multipliers

d) Minimizing the error sum of squares

45. Two contrasts of the same treatments are said to be orthogonal if

a) Both of them have the same coefficients of the treatments

b) Both of them have the same coefficients but opposite in sign

c) Sum of squares of coefficients for both treatments is equal to one

d) Sum of cross products of coefficients of same treatment is zero

46. If $X_1, X_2 .....X_n$ be a random sample from (μ, σ^2) distribution and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Xi - \overline{X})^2$. Then

a) $\frac{nS^2}{\sigma^2}$ follows the Chi square distribution with $(n-1)df$

b) $\frac{(n-1)S^2}{\sigma^2}$ follows the Chi square distribution with $(n-1)df$

c) $\frac{nS^2}{\sigma^2}$ follows the Chi square distribution with *n df*

d) S^2 follows the Chi square distribution with $(n-1)df$

47. Laspeyre's index number processes

a) No bias

b) Downward bias

c) Upward bias

d) Homogeneity error

48. The method of moving averages for determination of a trend eliminates the effect of

a) Long term variations

b) Short term variations

c) Cyclical variations

d) Residual variations

49. Let $X_i = 1, 2, ..n$ be a random sample from $N(\mu, \sigma)^2$. Denote by $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ then $\frac{(\overline{X} - \mu)}{S/\sqrt{n}}$ follows

a) The student *t* distribution

b) F distribution

c) Normal distribution

d) Chi square distribution

50. Identify the non parametric test from among the following

a) F test
b) t test
c) Normal test
d) Sign test

51. The Pearson's coefficient of correlation is computed based on the assumption that

a) The various forces that contributes to the relationship between the variable are operating independently
b) Associated with increase in one variable, there is a subsequent increase in other also
c) There exists a linear relationship between the variables under consideration
d) There are no extreme items present in the data sets

52. If a simple random sample of size n is drawn from a population of size N, without replacement, the number of possible samples is

a) N^n
b) $N - n$
c) $\frac{N!}{n!}$
d) NC_n

53. X is a continuous non negative random variable with distribution function F(x) and p.d.f $f(x)$. Pick out the wrong expression for E(X) from those given below

a) $\int_0^\infty [1 - F(x)]dx$
b) $\int_0^\infty x\, dF(x)$
c) $\int_0^\infty F(x)dx$
d) $\int_0^\infty xF(x)dx$

54. From the following, pick the correct statement pertaining to Lognormal distribution

a) Lognormal distribution is symmetric
b) If $X_1, X_2, ...X_n$ are lognormal variates, $S_n = \sum_{i=1}^{n} X_i$ follows lognormal
c) The mean, median and mode coincides
d) If $X_1, X_2, ...X_n$ are lognormal variates, $\prod_{i=1}^{n} X_i$ follows lognormal

55. The death rate obtained for a segment of population is known as

a) Specific death rate
b) Crude death rate
c) Standardized death rate
d) Foetal death rate

56. The range of partial correlation coefficient is

a) $-\infty$ to ∞
b) –1 to 1
c) 0 to 1
d) 0 to ∞

57. The inference procedure which takes into account prior belief of the experimenter about the occurrence of an event

a) Classical inference
b) Bayesian inference
c) Non Parametric inference
d) Sequential procedure

58. If the value of correlation coefficient $\gamma = 0$, then the two regression lines

a) Are parallel
b) Are coincident
c) Are perpendicular
d) Do not exist

59. In Neyman allocation, the stratum size n_h is determined such that

a) Variance is minimum subject to fixed cost
b) Cost is minimum subject to fixed Variance
c) Variance is minimum subject to fixed total sample size
d) $n_h / N_h = n / N$ (with usual notations)

60. Deflating the index enables one to

a) Calculate the real wages using an appropriate price index
b) Assign suitable weights to the factors in construction of cost of living index
c) Combine two series of index numbers with different bases into a single series
d) To facilitate accurate comparison when an old series of index numbers is being discontinued and new one is started

ANSWER SHEET — PART – A

1	A	B	C	D	E	21	A	B	C	D	E	41	A	B	C	D	E
2	A	B	C	D	E	22	A	B	C	D	E	42	A	B	C	D	E
3	A	B	C	D	E	23	A	B	C	D	E	43	A	B	C	D	E
4	A	B	C	D	E	24	A	B	C	D	E	44	A	B	C	D	E
5	A	B	C	D	E	25	A	B	C	D	E	45	A	B	C	D	E
6	A	B	C	D	E	26	A	B	C	D	E	46	A	B	C	D	E
7	A	B	C	D	E	27	A	B	C	D	E	47	A	B	C	D	E
8	A	B	C	D	E	28	A	B	C	D	E	48	A	B	C	D	E
9	A	B	C	D	E	29	A	B	C	D	E	49	A	B	C	D	E
10	A	B	C	D	E	30	A	B	C	D	E	50	A	B	C	D	E
11	A	B	C	D	E	31	A	B	C	D	E	51	A	B	C	D	E
12	A	B	C	D	E	32	A	B	C	D	E	52	A	B	C	D	E
13	A	B	C	D	E	33	A	B	C	D	E	53	A	B	C	D	E
14	A	B	C	D	E	34	A	B	C	D	E	54	A	B	C	D	E
15	A	B	C	D	E	35	A	B	C	D	E	55	A	B	C	D	E
16	A	B	C	D	E	36	A	B	C	D	E	56	A	B	C	D	E
17	A	B	C	D	E	37	A	B	C	D	E	57	A	B	C	D	E
18	A	B	C	D	E	38	A	B	C	D	E	58	A	B	C	D	E
19	A	B	C	D	E	39	A	B	C	D	E	59	A	B	C	D	E
20	A	B	C	D	E	40	A	B	C	D	E	60	A	B	C	D	E

APPLIED STATISTICS AND DATA ANALYTICS

PART – B

(Descriptive Type)

Answer **any eight** questions. Each question carries **5** marks **(8 × 5 = 40 Marks)**

1. (X,Y) follow the trinomial distribution with p.m.f

   $P(X=x,Y=y)=\frac{n!}{x!y!(n-x-y)!}p^x y^x(1-p-q)^{n-x-y}$ where x and y are non negative integers such that $x+y\le n; p,q>0$ with $p+q<1$. Find the marginal and conditional distributions.

2. X follows the geometric distribution with p.m.f. $f(x)=q^x p, x=0,1,2,3,....0<p<1; q=1-p$. Show that for any 2 positive integers m and n, $P(X>m+n \mid X>m)=P(X\ge n)$.

3. Obtain the maximum likelihood estimator for the parameter β while sampling from the exponential distribution specified by $f(x)=\frac{1}{\beta}\exp(-x/\beta), x>0$. Is the MLE unbiased?

4. What are the tests to be satisfied by a good index number? Examine whether Fishers index number satisfies the tests?

5. A truck load of oranges weighs 10 tons. It is intended to measure the extend of sugar content in the entire load. Suggest a suitable sampling design and an appropriate estimator.

6. Write down a model for the analysis of a latus square design. Also give its ANOVA.

7. A fair die is thrown 360 times. Determine a lower bound for probability of getting 50 to 70 ones using Chebyshev's inequality.

8. Describe any one method for estimating trend in a time series.

9. What do you understand by a confidence interval? Let $X_1, X_2 .... X_N$ be a random sample from $N(\mu, \sigma^2)$ and that σ is known. Obtain the $100(1-\alpha)\%$ shortest interval for μ.

10. An man has n keys and he wants to open a door. He tries the keys one by one till the door opens. Assume that only one key suits the door. Find the mean number of trials required to open the door if unsuccessful keys are.

    (a) Eliminated

    (b) Not Eliminated

11. *Let* $X_1, X_2 ... X_m$ *and* $Y_1, Y_2 ... Y_n$ be independent random variables from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ respectively. Explain how you will test the hypothesis $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 \neq \sigma_2^2$.

12. How will you test independence in a contingency table. Illustrate the procedure by using a 2×2 contingency table.
