Entrance Examination – 2017
Ph.D. Statistics

(Ph.D. Admissions - Jan.'18 session)

Hall Ticket No.	

Time : 2 hours
Max. Marks : 80

PART A: 40 MARKS
PART B: 40 MARKS

Instructions

1. Write your Hall Ticket Number on the OMR Answer Sheet given to you. Also write the Hall Number in the space provided above.
2. Answers are to be marked on the OMR sheet.
3. Pleas read the instructions carefully before marking your answers on the OMR answer sheet.
4. Hand over the OMR answer sheet at the end of the examination to the Invigilator.
5. No additional sheets will be provided. Rough work can be done in the question paper itself/space provided at the end of the booklet.
6. Calculators are not allowed.
7. There are a total of 40 questions in **PART A** and **PART B** together.
8. Each correct answer carries 2 marks.
9. The appropriate answer(s) should be coloured with either a blue or black ball point or a sketch pen. DO NOT USE A PENCIL.
10. This paper contains **7 pages** including this page. Please check that your paper has all the pages.
11. Given below are the meanings of some symbols that may have appeared in the question paper:
    $\mathbb{R}$-The set of all real numbers, $E(X)$-Expected value of the random variable X, $V(X)$-Variance of the random variable X, $Cov.(X,Y)$-Covariance of the random variables X and Y, $\rho_{X,Y}$ denotes the correlation coefficient between X and Y, iid-independent and identically distributed, pdf-probability density function, $B(n,p)$, $N(\mu,\sigma^2)$, $U((a,b))$ and $\beta(m,n)$ denote respectively, the Binomial, Normal, Uniform and the Beta distributions with the said parameters. $Rank(A)$ and $det(B)$ mean rank and determinant of the matrices A and B respectively.

P-58

Part - A

1. Which among the following is the odd one out?

$$\pi, \qquad \tfrac{22}{7}, \qquad e, \qquad \sqrt{15}.$$

(A) π. (B) $\frac{22}{7}$. (C) e. (D) $\sqrt{15}$.

2. Which of the options fits with the following collection?

SPSS, SAS, MATLAB, STATA, MINITAB

(A) RATS . (B) MS WORD. (C) LATEX. (D) MS OFFICE.

3. Idetify the odd one out in the following collection

AVEDEV, CHISQ.INV, lm, RSQ

(A) AVEDEV . (B) CHISQ.INV. (C) lm. (D) RSQ.

4. What is the next number in the sequence below?

$$1,\ 4,\ 27,\ 256,\ \ldots$$

(A) 625. (B) 3125. (C) 1250. (D) 6250.

5. The highest power of 10 that is a factor of 25! is
(A) 3. (B) 6. (C) 5. (D) 2.

6. In EXCEL what does the function LARGE(array,k) return?

(A) The largest number in the highlighted array of k numbers.

(B) The specified k^{th} largest number in the highlighted row or column.

(C) Index of the k^{th} largest number in the highlighted row or column.

(D) The smallest to the k^{th} smallest numbers in the highlighted row or column.

7. Which of the following is not a literature survey source for Statistics and Mathematics?

EXCEL, Google Sci-Hub MathSciNet

(A) EXCEL. (B) Google. (C) Sci-Hub. (D) MathSciNet.

8. In a $q-q$ plot for normal distribution, a few of the smaller observations are way below and a few larger values in the data are way above the straight line, this indicates that

(A) The distribution of data values is highly positively skewed and hence reject normality.

(B) The distribution of data values is highly negatively skewed and hence reject normality.

(C) The distribution of data values may be symmetric, but the tails may be heavy, and hence reject normality.

(D) Symmetry and light tails, so accept normality.

P-58

9. The value of $\sum_{m=1}^{n} \binom{2n}{2m}$ is

(A) 2^n. (B) 4^n. (C) 2^{n-1}. (D) $\frac{4^n}{2}$.

10. If every arrangement of the numbers $1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ in a row is equally likely, what is the probability of getting an arrangement in which all the prime numbers are in a row?

(A) 1/2. (B) 2/5. (C) 1/720. (D) 1/30.

11. The median of a set of 15 distinct numbers is m, now include the average of the seventh smallest and the eighth smallest in this set, what can you say about the median of this set of 16 numbers?

(A) It is equal to m. (B) It is less than m.
(C) It is more than m. (D) Nothing can be said based on the given data.

12. There is a dataset which contains whole numbers, identify which of the following statements regarding Mode is not possible.

(A) Both 36 and 37 which appear with different frequencies are modes.

(B) Both 36 and 57 which appear with different frequencies are modes.

(C) 55 is the mode.

(D) All the observations are modes.

13. The number of bacteria reduce at a rate of 20% per day after treatment begins. Approximately after how many days will there be less than 30% of bacteria as there were just before starting treatment?

(A) 9. (B) 6. (C) 8. (D) 10.

14. A_1, A_2 A_3 are three events with positive probabilities less than 1, $P(A_1|A_2) = P(A_1)$, $P(A_1|A_3) = P(A_1)$, so

(A) A_1 and $A_2 \cap A_3$ are independent.

(B) A_1 and $A_2 \cup A_3$ are independent.

(C) A_2 and A_3 are independent.

(D) A_1 and A_2^c are independent.

15. The ratio of the areas of a circle and a square with the same circumference and perimeter respectively

(A) is equal to 1. (B) equal to 1/2.
(C) more than 1. (D) is more than 1/2, but less than 1.

16. Partition a positive definite real $n \times n$ matrix $\mathbf{A}$ as $\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \vdots & \mathbf{A}_{12} \\ \cdots & & \cdots \\ \mathbf{A}_{21} & \vdots & \mathbf{A}_{22} \end{pmatrix}$, where $\mathbf{A}_1$ is the matrix of the first m rows and columns of $\mathbf{A}$, then, the matrix $\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$

(A) is a singular matrix.

(B) is non negative definite, but may not be positive definite.

(C) does not necessarily exist because inverse of $\mathbf{A}_{11}$ may not exist.

(D) is certainly positive definite.

17. The average marks of students from each of the 4 different schools in a public exam is 65%, so we can say that

(A) there is no variation among the students within the schools.

(B) the variation in the performances of all the students is largely due to the variations among the students and not due to the schools.

(C) there is no variation in the performances of all the students of these 4 schools.

(D) the variation in the performances of all the students is largely due to the schools.

18. There are 10 balls in each of the 10 bags $B_1, \ldots, B_{10}$. The weights of all the balls in 9 of them are $10 gms.$, where as the balls in one of the bags weigh either $9 gms.$ or $11 gms.$ each, to identify the bag which is different an investigator took i balls each from $B_i, \quad i = 1, \ldots, 10$ respectively and weighed all of them together, the total weight of all these 55 balls was $544 gm.$. Identify the correct statement

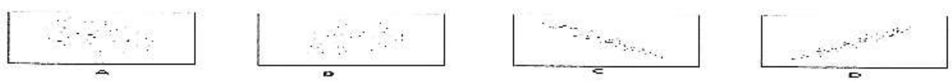
(A) the balls in B_4 weigh $1 gm.$ more.

(B) the balls in B_5 weigh $9 gm.$ each.

(C) the balls in B_6 weigh $9 gm.$ each.

(D) one can't determine which bag contains balls that weigh less or more by this procedure.

19. The correlation coefficient based on the bivariate sample $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ is -0.8, which of the following scatter plots best represents the given data?



20. It is not true that all employees in a company are less than 35 years old, which of the following statements is correct?

(A) At least one employee is more than 35 years old.

(B) All except one person is more than 35 years old.

(C) Everyone in the company is more than 35 years old.

(D) No employee is 35 years old.

Part - B

21. For a random variable X, $P(X \geq k) = (\frac{1}{k+1})$, $k = 1, 2, \ldots$, the expected value of X
(A) is 1. (B) is 2.
(C) is 3. (D) does not exist.

22. $\{X_n\}_1^\infty$ is a sequence of *iid* Poisson random variables with parameter 2,let

$$S_n = X_1 + X_2 + \ldots + X_n = \sum_{j=1}^{n} X_j, \quad n = 1, 2, \ldots$$

(A) $P(\frac{S_n - 2n}{\sqrt{2n}} \leq x) \longrightarrow \Phi(x)$ as $n \to \infty$.

(B) $\frac{S_n}{n} \longrightarrow 2$ in probability measure but not with probability one.

(C) $\frac{S_n}{n} \longrightarrow 1$ with probability one.

(D) $P(\frac{S_n - 2n}{n} \leq x) \longrightarrow \Phi(x)$ as $n \to \infty$.

23. X_1 and X_2 are *iid* random variables that are that are normally distributed with means 2 and variances 1/2, the characteristic function of $Y = X_1 - X_2$ is
(A) $\phi_Y(t) = e^{-it}, \ \forall t.$ (B) $\phi_Y(t) = e^{2it - \frac{t^2}{2}}, \ \forall t$
(C) $\phi_Y(t) = e^{-t^2}.$ (D) $\phi_Y(t) = e^{-\frac{t^2}{2}}.$

24. $X_1, X_2, \ldots$ are real valued random variables defined on the σ- probability space $(S, \mathcal{B}, P)$, let A be that subset of S such that for every $s \in A, X_j(s) = 0$ for infinitely many j, if $P(A) = 1$ it means that

(A) $X_n \longrightarrow 0$ almost surely (P).

(B) $X_n \longrightarrow 0$ in probability but not almost surely (P).

(C) $X_n \longrightarrow 0$ almost surely (P) only if $X_1, X_2, \ldots$ are independent.

(D) $X_n \longrightarrow 1$ almost surely (P)

25. Arrivals of customers to a mall which opens at $9.00AM$ are in accordance with a Poisson process, the expected time to the first arrival is $\frac{1}{100}min.$, the expected number of customers that will arrive between $10.30AM$ and $10.35AM$
(A) is 300. (B) can not be determined based on the given information.
(C) is 500. (D) is 600.

26. Consider a homogeneous Markov chain $\{X_n, n = 0, 1, \ldots\}$ with state space $\{1, 2, 3\}$ and the one step Transition probabilities are $p_{12} = p_{23} = p_{31} = 1$ and the others are zero, Read the statements below regarding this Markov chain:

i. It is irreducible and recurrent.

ii. It is aperiodic

iii. It is positive recurrent.

iv. It has a long run distribution..

The correct statements are
(A) (i),(ii) and (iii) only. (B) all (i) to (iv) are correct.
(C) only (i) and (iii) (D) only (ii) and (iv).

27. $3, 7, 9, 2, 4$ is the random sample observed from a Poisson random variable, the maximum likelihood estimate of the parameter in the set $\{1, 2, 3, 4\}$
(A) is 3. (B) is 4. (C) is 2. (D) can not be determined based on the given data.

28. Heads showed up 6 times when a coin was tossed 10 times, if p is the unknown probability of heads showing up, an unbiased estimate of p^2
(A) is 1/3. (B) is 0.6.
(C) is 0.36. (D) can not be determined based on the given data.

29. $X_1, \ldots, X_n$ is a random sample from the $U((0, \theta])$, $\theta > 0$, let $X_{(r)}, \quad r = 1, \ldots, n$ denote the $r^{th.}$ order statistics, identify the correct statement

(A) $X_{(n)}$ in unbiased and sufficient estimator for θ.

(B) $X_{(n)}$ is a sufficient and the maximum likelihood estimator for θ.

(C) $X_{(n)}$ is not sufficient but is an unbiased estimator for θ.

(D) $X_{(1)}$ is a sufficient but not the maximum likelihood estimator for θ.

30. A test based on one sample point X for $H_0 : X \sim U((0, 1))$ vs. $H_1 : X \sim \beta(2, 2)$ is to reject H_0 if the likelihood ratio $(\frac{L_{H_0}}{L_{H_1}})$ is less than $1/1.485$, the size of this test is
(A) 0.0125. (B) 0.025. (C) 0.05. (D) 0.1.

31. $3.1, -1.4, -3.4, 1.7, 1.9$ is a random sample from the $U((-\theta, \theta))$ population, the maximum likelihood estimate of θ based on this sample is
(A) 3.4. (B) 3.1. (C) 3.25. (D) 2.4.

32. $\Phi(-1.96) = 0.025, \Phi(-1.64) = 0.05$, based on a random sample of size 16 from the $N(\mu, 25)$ population, the length of the 95% confidence interval for μ is about
(A) 3. (B) 5 . (C) 9. (D) 12.

33. X is a random variable whose second moment exists and, $P(-4 < X < 4) = 0.8$,
(A) $E(X^2) \leq 2$. (B) $E(X^2)$ could be 3.
(C) $E(X^2)$ is 2.6.. (D) Nothing can be said about $V(X)$.

34. $(X_1, X_2)^T$ is a bivariate normal random vector with mean vector $(12, 14)^T$ and dispersion matrix $\begin{pmatrix} 16 & 12 \\ 12 & 18 \end{pmatrix}$, $V(2X_2 | X_1 = 13)$ is
(A) 18. (B) 6 . (C) 36. (D) 12.

35. Consider the linear model

$$y_1 = -\beta_1 + 2\beta_2 + \epsilon_1$$
$$y_2 = \beta_1 + 3\beta_3 + \epsilon_2$$
$$y_3 = 2\beta_1 - 2\beta_2 + 3\beta_3 + \epsilon_3$$

$\beta_1, \beta_2, \beta_3$ are the parameters, $\epsilon_i, \quad i = 1, 2, 3$ which are the unobservable, uncorrelated error components with mean 0 and variance σ^2

(A) The $BLUE$s of $\beta_1, \beta_2, \beta_3$ exist.

(B) The $BLUE$ of β_1 exists but the $BLUE$ of β_2 does not exist.

(C) The $BLUE$ of $\beta_1 + \beta_2$ exists.

(D) None of the above is correct.

P-58

36. In a large farm, there are N rows of mango trees, each row has M trees, a simple random sample of $n(< N)$ rows is selected, the yields of each tree in the selected rows are observed, this sampling scheme is

(A) a stratified sampling scheme and we can estimate the average yield per tree in the entire farm.

(B) a two stage cluster sampling scheme with distinct cluster sizes and we can estimate the average yield per tree in the entire farm.

(C) a two stage clustering sampling scheme with equal cluster sizes and we can not estimate the average yield per tree in the entire farm.

(D) a single stage cluster sampling scheme and we can estimate the average yield per tree in the entire farm.

37. 4 medicines for diabetes were used on patients belonging to 4 different age groups and 4 different weight groups in such a way that in every age group one and only patient received each treatment and in every weight group also one and only patient received each treatment. This is a

(A) Randomized Block Design with 4 treatments and 4 blocks with each treatment appearing in a block 4 times.

(B) Latin Square Design requiring 16 patients.

(C) BIBD in which the number of treatments and number of blocks is 4.

(D) 4^4 factorial design.

38. In a 2^4 factorial design, the 2 factor interaction effect between A and B is

(A) $a - (1) + ab - b + ac - c + abc - bc + ad - d + abd - bd + acd - cd + abcd - bcd$.

(B) $b - (1) + ab - a + bc - c + abc - ac + bd - d + abd - ad + bcd - cd + abcd - acd$.

(C) $abcd - bcd - acd + cd + abc - bc - ac + c + abd - bd - ad + d + ab - b - a + (1)$.

(D) none of the above.

39. The empirical distribution function is

(A) not unbiased.

(B) unbiased, but is a not consistent estimator for the distribution function.

(C) unbiased and a consistent estimator for the distribution function.

(D) not an efficient estimator for the distribution function.

40. Consider the following two-variable problem:

$$\begin{aligned} \textit{Maximize}: \quad & z = 5x_1 + 4x_2 \\ \textit{Subject to}: \quad & 6x_1 + 4x_2 \leq 24 \\ & x_1 + 2x_2 \leq 6 \\ & -x_1 + x_2 \leq 1 \\ & x_2 \leq 2 \\ & x_1, \ x_2 \geq 0 \end{aligned}$$

The optimal solution (x_1, x_2) is

(A) $(4, 0)$.

(B) $(3, 1.5)$.

(C) $(2, 2)$.

(D) None of the above is an optimal solution for the given problem.