ENTRANCE EXAMINATIONS – 2018

(Ph.D. Admissions - January 2019 Session)

Ph.D. Statistics

Hall Ticket No.	

Time : Two 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. Please 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 and there is **negative marking of 0.66** for each wrong answer.
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 book contains **8 pages** including this page and excluding pages for the rough work. 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, $B(n,p)$ and $N(\mu,\sigma^2)$ denote respectively,the Binomial and the Normal distributions with the said parameters. $Rank(A)$ and $det(B)$ mean rank and determinant of the matrices A and B respectively.

S-54

Part-A

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

$$1, 5, 14, 30, 55, \ldots\ldots$$

(A) 71. (B) 81. (C) 91. (D) 101.

2. The difference between the squares of two consecutive odd numbers is 56, the product of the two numbers is

(A) 105. (B) 143. (C) 195. (D) 255.

3. $A = \{a_1, a_2, a_3, a_4\}$ and $B = \{b_1, b_2, b_3, b_4\}$, how many one-to-one function f exist from A to B?

(A) 20. (B) 60. (C) 100. (D) 120.

4. In an exam 3 marks are given for a correct answer and -1 for a wrong answer and no marks are awarded if a candidate does not attempt a question. A candidate attempted all 50 questions and scored 98 marks. How many questions did this candidate answer incorrectly?

(A) 11. (B) 13. (C) 15. (D) 17.

5. Identify which, if any, of the following are true:

(A) If events A and B are independent then they are also disjoint.

(B) If events A and B are independent then $P(A \cap B) < P(A)P(B)$.

(C) If events A and B are independent then they cant simultaneously occur.

(D) None of the above three statements are true..

6. Manish gets test grades of 71, 76, 80, and 86. He gets a 90 on his final exam. Find the weighted mean if the tests each count for 10% and the final exam counts for 60% of the final grade. Round to one decimal place.

(A) 85.3 . (B) -71.2 . (C) 241.8 . (D) 80.6 .

7. A local country club has a membership of 600 and operates facilities that include an 18-hole championship golf course and 12 tennis courts. Before deciding whether to accept new members, the club president would like to know how many members regularly use each facility. A survey of the membership indicates that 61% regularly use the golf course, 45% regularly use the tennis courts, and 3% use neither of these facilities regularly. What Percentage of the 600 Uses at least one of the golf or tennis facilities?

(A) 97%. (B) 3%. (C) 103%. (D) 9%.

8. The percentage of measurements that are above the 39th percentile is

(A) 71%. (B) 39%. (C) 61%. (D) Cannot be determined.

S-54

9. A statistic is said to be biased if it

   (A) has exactly the same value as the parameter.

   (B) systematically underestimates or overestimates the parameter.

   (C) is determined from a trimmed sample.

   (D) leads to an erroneous conclusion about the sample.

10. Given the following least squares prediction equation, $\hat{y} = -173 + 74x$, we estimate y to ______ by ______ with each 1-unit increase in x

    (A) decrease; 74. (B) decrease; 173.

    (C) increase; 74. (D) increase; 173.

11. Which measure of central tendency may not exist for all numeric data sets?

    (A) Mean. (B) Median. (C) Mode. (D) Variance.

12. The complement of 4 heads in the toss of 4 coins is

    (A) All tails. (B) Exactly one tail.

    (C) Three heads. (D) At least one tail.

13. If x, y are nonzero real numbers, then $x^2 + xy + y^2$ is

    (A) always positive. (B) always negative.

    (C) zero. (D) sometimes positive, sometimes negative.

14. The point $(3, 4)$ in the xy-plane is reflected w.r.t the x-axis and then rotated through 90 degrees in the clockwise direction in the plane about the origin. The final position of the point is

    (A) $(3, -4)$. (B) $(4, -3)$.

    (C) $(-3, -4)$. (D) $(-4, -3)$.

15. If $I = \int_0^1 e^x dx$, then which of the following is true?

    (A) $I < 1$. (B) $1 < I < 2$. (C) $2 < I < e$. (D) $I > e$.

16. Let $v = (1, 1)$ and $w = (1, -1) \in R^2$. Then a vector $u = (a, b) \in R^2$ is in the linear span of v and w

    (A) only when $a = b$. (B) for exactly one value of (a, b).

    (C) always. (D) for at most finitely many values of (a, b).

17. Maximum number of positive roots of $x^6 + 9x^5 + 2x^3 - x^2 - 2$ is

    (A) 0. (B) 1. (C) 3. (D) 5.

S-54

18. The value of the following determinant is:

$$\begin{vmatrix} a+b & c+d & e & 1 \\ b+c & d+a & f & 1 \\ c+d & a+b & g & 1 \\ d+a & b+c & h & 1 \end{vmatrix}$$

(A) 0. (B) (a+b)(c+d)+e+f+g+h.

(C) 1. (D) (a+b+c+d)(e+f+g+h).

19. Let X be a set with 30 elements. Let A, B, C be subsets of X with 10 elements each such that $A \cap B \cap C$ has 4 elements. Suppose $A \cap B$ has 5 elements, $B \cap C$ has 6 elements, and $C \cap A$ has 7 elements, how many elements does $A \cup B \cup C$ have?

(A) 16. (B) 14. (C) 15. (D) 30.

20. To determine if there are outliers in a least squares regression models data set, we could construct a Boxplot of the

(A) response variables. (B) predictor variables.

(C) lurking variables. (D) residuals.

S-54

Part-B

21. Which statement is NOT CORRECT?

(A) The sample standard deviation measures variability of our sample values.

(B) A larger sample will give answers that vary less from the true value than smaller samples (assuming both are properly chosen).

(C) The standard error measures how much our estimate (answer) may vary if a new sample of the same size is chosen using the same sampling method.

(D) A large sample size always gives unbiased estimators regardless of how the sample is chosen.

22. In stratified sampling with population size equal to 1000, the population is divided into two strata with sizes $N_1 = 600$ and $N_2 = 400$ respectively. Under Neyman allocation if $S_1 = 24$ and the sample sizes from two strata are in the ration $n_1 : n_2 :: 2 : 1$ then the value of S_2 is

(A) 16. (B) 36. (C) 18. (D) 24.

23. In a BIBD with t treatments in b blocks of k plots each and r replicates, which one of the following is not true?

(A) $rt = bk$. (B) $b \geq t$. (C) $r > k$. (D) $b \leq (r + t - k)$.

24. The following numbers are extracted from a table of random digits:

38683 50279 38224 09844 13578 28251 12708 24684

A scientist will be measuring the total amount of woody debris in a random sample of sites selected without replacement from a population of 45 sites. The sites are labeled $01, 02, \ldots, 45$ and she starts at the beginning of the line of random digits and takes consecutive pairs of digits. Which of the following is correct?

(A) Her sample is 38, 25, 02, 38, 22....

(B) Her sample is 38, 68, 35, 02, 22....

(C) Her sample is 38, 65, 35, 02, 79....

(D) Her sample is 38, 35, 02, 22, 40....

25. For a random variable $X, P(X > k) = \left(\frac{3}{4}\right)^k, k = 1, 2, \ldots,$ the expected value of X

(A) is 1. (B) is 2. (C) is 3. (D) does not exist.

26. The arrivals of customers to a mall are in accordance with a Poisson process at the rate of 100 per minute. An arriving customer to the mall goes to the clothes section with probability 5/8, to the food section with probability 2/8 and to the electronics section with probability 1/8. The expected number of customers that will visit the electronics section in a 12 hour period is

(A) 9000. (B) 12000. (C) 6000. (D) 7200.

S-54

27. Which of the following is NOT true of the confidence level of a confidence interval?

    (A) The confidence level gives us the success rate of the procedure used to construct the confidence interval.

    (B) The confidence level is often expressed as an area $1-\alpha$, where α is the complement of the confidence level.

    (C) The confidence level is also called the degree of confidence.

    (D) There is a $1-\alpha$ chance, where α is the complement of the confidence level, that the true value of the parameter is in the confidence interval produced by the sample.

28. The statement "If there is sufficient evidence to reject a null hypothesis at the 10% significance level, then there is sufficient evidence to reject it at the 5% significance level" is:

    Please select the best answer of those provided below.

    (A) Always True.

    (B) Never True.

    (C) Sometimes True; the p-value for the statistical test needs to be provided for a conclusion.

    (D) Not Enough Information; this would depend on the type of statistical test used.

29. You conduct a hypothesis test and you observe values for the sample mean and sample standard deviation when $n = 25$ that do not lead to the rejection of null hypothesis. You calculate a p-value of 0.0667. What will happen to the p-value if you observe the same sample mean and standard deviation for a sample of size greater than 25?

    (A) Increase. (B) Decrease. (C) Remain same. (D) Cannot be said exactly.

30. Let $x_1, x_2, \ldots, x_n$ be a random sample from a $N(0, \theta)$ distribution where the variance θ is unknown. The UMP test for $H_0 : \theta = \theta_0 (> 0)$ against $H_1 : \theta > \theta_0$ is of the form:

    (A) $\sum_1^n x_i^2 = C$ where C is some constant.

    (B) $\sum_1^n x_i \geq C$ where C is some constant.

    (C) $\sum_1^n x_i \leq C$ where C is some constant.

    (D) $\sum_1^n x_i^2 \geq C$ where C is some constant.

31. In a χ^2 test of independence, with m rows and n columns in the contingency table, the number of degrees of freedom associated with the test statistic is

    (A) mn-1. (B) mn+1. (C) mn-m-n+1. (D) mn-m-n-1.

32. Given the ANOVA table

Source	d.f.	SS	MSS	F
Replication	2	0.0971		
Treatment	2	0.7324	G	4.19
Error	8	H		
Total	14	1.1790		

The values of (G, H) are:

(A) 0.1831, 0.3495. (B) 0.1731, 0.3395.

(C) 0.1732, 0.3385. (D) 0.1631, 0.3285.

33. For a linear regression model, how does a confidence interval differ from a prediction interval?

(A) Confidence intervals are used to measure the accuracy of the mean response of all the individuals in the population, while a prediction interval is used to measure the accuracy of a single individuals predicted value.

(B) Confidence intervals are used to measure the accuracy of a single individuals predicted value, while a prediction interval is used to measure the accuracy of the mean response of all the individuals in the population.

(C) Confidence intervals are constructed about the predicted values of y while prediction intervals are constructed about a particular value of x.

(D) Confidence intervals are constructed about the predicted values of x while prediction intervals are constructed about a particular value of y.

34. In order to be accepted into a program at North Point, a student must score in the top 2% of a standardized test on general knowledge. Historically, the scores for this test are found to be normally distributed with mean score of 70 and variance of 9. What would be the minimum score (rounded to the nearest integer) on this test in order for a student to be accepted into the program?

(A) 83 (B) 89 (C) 81 (D) 76.

35. The 95% asymptotic confidence interval for θ of the Poisson distribution is given by

(A) $\bar{x} \pm 2.58\sqrt{\frac{\bar{x}}{n}}$. (B) $\bar{x} \pm 1.96\sqrt{\frac{\bar{x}}{n}}$.

(C) $\bar{x} \pm 1.96\sqrt{\frac{n}{\bar{x}}}$. (D) None of these.

36. $X_1, \ldots, X_n$ are i.i.d. random variables with absolutely continuous distribution function $F(x;\theta)$, then $-\sum_{i=1}^{n} logF(X_i;\theta)$ has

(A) Normal distribution. (B) Beta distribution.

(C) Gamma distribution. (D) Weibull distribution.

37. In a Markov chain with state space $\{0, 1, 2\}$ and one step transition matrix given by $P = \begin{bmatrix} 1/4 & 3/4 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/4 & 3/4 \end{bmatrix}$, the value of $p_{01}^{(2)}$ will be: .

(A) 3/4. (B)9/16. (C) 3/16. (D) 7/16.

S-54

38. Let X_1, X_2 and X_3 be independent random variables with $X_k, k = 1, 2, 3$ having probability density function $f_k(x) = k\theta e^{-k\theta x}, 0 < x < \infty, \theta > 0$. Then a sufficient statistic for θ is

(A) $X_1 + X_2 + X_3$. (B) $X_1 + 2X_2 + 3X_3$.

(C) $X_1X_2X_3$. (D) $3X_1 + 2X_2 + X_3$.

39. To establish causation, a statistician must

(A) take surveys in multiple locations. (B) use a sample size of at least n=30.

(C) use regression analysis. (D) none of the above are correct.

40. An experiment was conducted in which $n = 15$ subjects were timed in the 400 meter run on Monday, fed a carbohydrate heavy diet for the week, and timed in the 400 meter run on Friday. To determine if there is a difference in average performance the appropriate hypothesis testing procedure would be

A) a two sample F-test. B) matched pairs t-test.

C) 2 sample z test. D) none of these is appropriate.

S-54