ENTRANCE EXAMINATIONS – 2019

(Ph.D. Admissions - January 2020 Session)

Ph.D. Statistics

Hall Ticket No.	

Time : 2 hours
Max. Marks : 70

PART A: 35 Marks
PART B: 35 Marks

Instructions

1. Write your Hall Ticket Number on the OMR Answer Sheet given to you. Also write the Hall Ticket 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. The question paper consists of 70 objective questions of **one mark** each. There is **Negative** marking of **0.33** for each wrong answer.
8. The appropriate answer(s) should be coloured with either a blue or black ball point or a sketch pen. DO NOT USE A PENCIL.
9. This book contains **13 pages** including this page and excluding pages for the rough work. Please check that your paper has all the pages.
10. 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.

Part-A

1. Sum of three prime numbers is 100. If one of them exceeds another by 36, then one of the numbers is

   (A) 7. (B) 29. (C) 41. (D) 67.

2. What is the harmonic mean of two numbers whose geometric mean and arithmetic mean is 8 and 5 respectively?

   (A) 12.8 (B) 12. (C) 13.5. (D) 14.6

3. Errors may occur in performing numerical computation on a computer due to

   (A) rounding errors. (B) power fluctuation.

   (C) operator fatigue. (D) all of these.

4. If $\phi(x) = \int_0^{x^2} \sqrt(t)dt$, then $d\phi/dx$ is

   (A) $2x^2$. (B) $\sqrt(x)$. (C) 0. (D) 1.

5. The set of all real numbers under the usual multiplication operation is not a group since

   (A) multiplication is not a binary operation.

   (B) multiplication is not associative.

   (C) identity element does not exist.

   (D) zero has no inverse.

6. To perform a Chi-square test of goodness of fit:

   (A) the sampling method is simple random sampling.

   (B) the variable under study is categorical.

   (C) the expected value for any cell should be at least 5.

   (D) all of these.

7. Limit of the following series as x approaches $\pi/2$ is

$$F(x) = x - x^3/3! + x^5/5! - x^7/7! + \dots$$

   (A) $2\pi/3$. (B) $\pi/2$. (C) $\pi/3$. (D) 1.

8. The number of diagonals that can be drawn by joining the vertices of an octagon is

   (A) 28. (B) 48. (C) 20. (D) 56.

V-60

9. The vector $(1, 2, -1)$ is an eigen vector of $A = \begin{bmatrix} -2 & 2 & -3 \\ 2 & 1 & -6 \\ -1 & -2 & 0 \end{bmatrix}$, then one of the eigen values of A is

(A) 1. (B) 2. (C) 5. (D) -1.

10. Let $A = ((a_{ij}))$ be an n-rowed square matrix and I_{12} be the matrix obtained by interchanging the first and second rows of the n-rowed identity matrix. Then the matrix AI_{12} is such that

(A) its first row is the same as its second row.

(B) its first row is the same as the second row of A.

(C) its first column is the same as the second column of A.

(D) its first row is all zero.

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

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

12. Consider the experiment of tossing four coins. Then the complement of the event '4 tails' is given by the event:

(A) all heads. (B) exactly one head.

(C) three tails. (D) at least one head.

13. In the matrix equation $Px = q$. which of the following is a necessary condition for the existence of at least one solution for the unknown vector x?

(A) Augmented matrix $[P|q]$ must have the same rank as matrix P.

(B) Vector q must have only non-zero elements.

(C) Matrix P must be singular.

(D) None of these.

14. The function $f(x) = 3x(x-2)$ has a

(A) minimum at x = 1. (B) maximum at x = 1.

(C) minimum at x = 2. (D) maximum at x = 2.

15. Identify which one of the following is 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 can not simultaneously occur.

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

16. Let $v = (1, 1)$ and $w = (1, -1) \in \mathbb{R}^2$. Then a vector $u = (a, b) \in \mathbb{R}^2$ 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).

V-60

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

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

18. In a hypothesis test, a p-value is

(A) a probability of observing a parameter more extreme than the one observed under the assumption that the null hypothesis is true.

(B) a probability of observing a parameter more extreme than the one observed under the assumption that the null hypothesis is false.

(C) a probability of observing a statistic at least as extreme as the one observed under the assumption that the null hypothesis is true.

(D) a probability of observing a statistic at least as extreme as the one observed under the assumption that the null hypothesis is false.

19. The expected value of the random variable

(A) will also be the most likely value of the random variable.

(B) is another term for the mean value.

(C) is also called the variance.

(D) cannot be greater than 1.

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.

21. In an experiment to determine if antibiotics increase the final dressed weight of cattle, the following were measured on each animal in the study. Gender, initial weight, weight gain, grade of meat, where grade is recorded as (A, B, or C). The scale of measurement of these variable are:

(A) nominal, ratio, interval, nominal

(B) nominal, ratio, ratio, ordinal

(C) ordinal, ratio, ratio, ordinal

(D) ordinal, ratio, ratio, nominal

22. If most of the measurements in a large data set are of approximately the same magnitude except for a few measurements that are quite a bit larger, how would the mean and median of the data set compare and what shape would a histogram of the data set have?

(A) The mean would be smaller than the median and the histogram would be skewed with a long left tail.

(B) The mean would be larger than the median and the histogram would be skewed with a long right tail.

(C) The mean would be larger than the median and the histogram would be skewed with a long left tail.

(D) The mean would be equal to the median and the histogram would be symmetrical.

V-60

23. 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.

24. The point $(3, 4)$ in the xy-plane is reflected with respect to 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) $(-4, -3)$.

25. In measuring the centre of the data from a skewed distribution, the median would be preferred over the mean for most purposes because:

(A) the median is the most frequent number while the mean is most likely

(B) the mean may be too heavily influenced by the larger observations and this gives too high an indication of the centre

(C) the median is less than the mean and smaller numbers are always appropriate for the centre

(D) the median measures the arithmetic average of the data excluding outliers.

26. Given that we have collected pairs of observations on two variables X and Y , we would consider fitting a straight line with X as an explanatory variable if:

(A) the change in Y is an additive constant.

(B) the change in Y is a constant for each unit change in X

(C) the change in Y is exponential

(D) none of the above

27. The correlation coefficient provides:

(A) a measure of the extent to which changes in one variable cause changes in another variable.

(B) a measure of the strength of the linear association between two categorical variables.

(C) a measure of the strength of the linear association between two quantitative variables.

(D) a measure of the strength of the linear association between a quantitative variable and a categorical variable.

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

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

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

29. The expected value of a random variable is -10 and it's variance is 100, then the value of the second moment is

(A) 10. (B) 100. (C) 200. (D) 1000.

V-60

30. Which of the following is not a measure of dispersion?

(A) Range. (B) Mean deviation about median.

(C) Mode. (D) Mean deviation about mean.

31. Which of the following is equivalent to the statement 'Ashok did not solve all the problems'?

(A) Ashok did not solve any problem.

(B) Ashok did not solve atleast one problem.

(C) Ashok solved at least one problem.

(D) Ashok solved at most one problem.

32. 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.

33. A bag contains 2 red balls and 2 blue balls. If the balls are removed one by one from the bag then what is the probability that the second ball to be drawn will be red?

(A) 2/3. (B) 1/2. (C) 1/3. (D) 1/4.

34. Identify the odd one from the collection: (SPSS, SAS, Minitab, MS-Word)

(A) Minitab. (B) SPSS. (C) MS-Word. (D) SAS.

35. How many 10 digits numbers can be written by using the digits 1 and 2?

(A) ${}^{10}C_1 + {}^{9}C_2$. (B) 2^{10}. (C) ${}^{10}C_2$. (D) 10!.

V-60

Part-B

36. Which one of the following statements 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.

37. 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.

38. 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) $b \leq (r + t - k)$.

39. 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....

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

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

41. 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.

V-60

42. 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.

43. 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.

44. You conduct a hypothesis test based on $n = 25$ and that does not lead to the rejection of the null hypothesis. You note down the values for the sample mean, sample standard deviation and your calculated p-value is 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.

45. 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.

46. 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) $mn-m-n-1$.

V-60

47. Given the ANOVA table

Source	d.f.	SS	MSS	F
Replication	2	0.0971		
Treatment	4	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).

48. 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.

49. In order to be accepted into a program at North Point Institute, 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.

50. 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.

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

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

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

52. 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.

V-60

53. 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) $X_1 X_2 X_3$. (D) $3X_1 + 2X_2 + X_3$.

54. 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) avoid all of the above as none of them are correct.

55. 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 the above.

56. The Central Limit Theorem states that

(A) if n is large, and if the population is normal, then the sampling distribution of the sample mean can be approximated closely by a normal curve.

(B) if n is large, and if the population is normal, then the variance of the sample mean must be small.

(C) if n is large, then the sampling distribution of the sample mean can be approximated closely by a normal curve.

(D) if n is large then the distribution of the sample can be approximated closely by a normal curve.

57. In order for the Poisson to give good approximate values for binomial probabilities we must have the condition(s) that:

(A) the population size is large relative to the sample size.

(B) the sample size is large.

(C) the probability, p, is small and the sample size is large.

(D) the probability, p, is close to 0.5 and the sample size is large.

58. Let $N(t)$ follows a Poisson process and $s < t$. Then the probability $P[N(s) = k|N(t) = n]$ will be

(A) $e^{-s/t}\frac{(s/t)^k}{k!}$.

(B) $\binom{n}{k}\left(\frac{t}{s}\right)^n\left(1-\frac{t}{s}\right)^{n-k}$.

(C) $\binom{n}{k}\left(\frac{s}{t}\right)^k\left(1-\frac{s}{t}\right)^{n-k}$.

(D) $\left(\frac{s}{t}\right)^k\left(1-\frac{s}{t}\right)^{n-k}$.

V-60

59. A researcher is using a simple interval estimator for as ($\bar{Y}$ − 2se, $\bar{Y}$+2se). Which of the following statements is true if the sample size, n, is 'large'?

(A) This interval will contain the true value of approximately 95 times out of 100.

(B) This interval is too narrow to be a useful interval estimator for .

(C) This interval will contain the true value of 997 times out of 1000.

(D) This interval is not an approximate 95% confidence interval for .

60. Suppose we are interested in studying the factors that affect GPA of college students. Fifty colleges are selected at random and we collect the GPA of one male and one female from each college. We also classify each college as public or private. Which of the following procedures is most appropriate to conduct first.

(A) An independent 2 sampled t test on the male and female GPA.

(B) A paired t test on the male and female GPA.

(C) ANOVA on the four sets of GPAs: males at private college, females at private college, males at public college, females at public college.

(D) Two-Way ANOVA using the college status as one factor and gender as the second factor.

61. Which of the following is INCORRECT about the use of a paired experiment?

(A) The analysis of paired data starts by finding the difference between the values of the pair. The order of the difference (as long as it is consistent) is unimportant.

(B) It is crucial to recognize pairing. If pairing is not recognized, the results will not be as accurate and precise as possible.

(C) The degrees of freedom is equal to the number of pairs −1.

(D) Because pairing is beneficial, we can pair all data by matching the smallest value of each sample, the second smallest value of each sample, the third smallest value of each sample, etc.

62. 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%.

63. 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.

64. If X and Y are independent real valued random variables then which of the following statement is not correct.

    (A) $E(XY) = E(X)E(Y)$.

    (B) $V(X+Y) = V(X) + V(Y)$.

    (C) $V(X|Y=y) = E(X)$ for every $y \in \mathbb{R}$.

    (D) $2X$ and $-3Y$ are also independent random variables.

65. Which of the following statements is true. I. The standard error is computed solely from sample attributes. II. The standard deviation is computed solely from sample attributes. III. The standard error is a measure of central tendency.

    (A) I only

    (B) II only

    (C) II and III only

    (D) None of the above

66. Which of the following statements is true. I. When the margin of error is small, the confidence level is high. II. When the margin of error is small, the confidence level is low. III. A confidence interval is a type of point estimate. IV. A population mean is an example of a point estimate.

    (A) I and II only

    (B) III only

    (C) IV only

    (D) None of the above.

67. Which of the following assumptions is not true of the Binomial distribution?

    (A) All trials must be identical.

    (B) Each trial must be classified as a success or a failure.

    (C) The number of successes in the trials is counted.

    (D) The probability of success is equal to 0.5 in all trials.

68. Which of the following is NOT CORRECT?

    (A) The probability of a Type I error is controlled by the selection of the α level.

    (B) The power of a test depends upon the sample size and the distance between the null and the alternate hypothesis.

    (C) The p-value measures the probability that the null hypothesis is true.

    (D) The rejection region is controlled by the α level and the alternate hypothesis.

69. Which of the following is incorrect about constructing histograms?

    (A) The approximate number of classes is $1 + 3.3\log(n)$.

    (B) All class intervals should be of equal width.

    (C) The bars of the histogram are centred over the class mark (midpoint).

    (D) The first and last classes should be open-ended to account for extreme points.

70. A research biologist has carried out an experiment on a random sample of 15 experimental plots in a field. Following the collection of data, a test of significance was conducted under appropriate null and alternative hypotheses and the p-value was determined to be approximately 0.03. This indicates that:

(A) this result is statistically significant at the 0.01 level.

(B) there is some reason to believe that the null hypothesis is incorrect.

(C) if this experiment was repeated, then 3 per cent of the time we would get this same result.

(D) the sample is so small that little confidence can be placed on the result.

V-60