GU-RET 2016

GAUHATI UNIVERSITY RESEARCH ELIGIBILITY TEST

STATISTICS

Booklet Series : A

Invigilator's Name and Signature	BOOKLET NO.	
	OMR SHEET NO.	
	ROLL NO.	

TIME : 2 HOURS 20 MINUTES **TOTAL MARKS : 80**

Number of Pages in this Booklet : **16**

Instructions for Candidates

1. Write your Roll No. and OMR Sheet No. in the boxes provided above.
2. This paper consists of two sections : **Section B** with 50 (fifty) multiple choice questions (MCQ) and **Section C** with 7 (seven) descriptive questions. Each MCQ has 4 (four) answers, out of which ONLY one is correct. You have to darken the circle (on the OMR Sheet) for the correct answer corresponding to the question given in this booklet.

   Example : (A) (B) (C) (D)

   where (C) is the correct answer. No marks will be given for markings made in this booklet. The descriptive questions in **Section C**, MUST be answered in the space provided in this booklet. **No extra pages will be provided in any case**.
3. Use a BLACK ball point pen in your OMR Sheet.
4. Read the instructions given inside this booklet before attempting to answer any questions.
5. DO NOT write your name, roll no, phone no, or anything, or put any marks anywhere in this booklet, otherwise your candidature will be disqualified.
6. If you are found to resort to any kind of unfair means such as carrying extra material other than pen, pencil, watch, eraser, and scale, or copying from somebody or from external material, your candidature will be disqualified.
7. Use of mobile phones, programmable calculators, log tables or any other tables, wearable smart devices such as smart Android watches or objects of similar nature CAN NOT be used inside the examination hall.
8. At the end of the examination, you have to return this booklet and the OMR Sheet back to the invigilator.
9. There is no negative marks for incorrect answer.

This page is left intentionally blank

Section B (50 Marks)

1. A set of data is found to have a sample standard deviation of 25. Suppose 9 were added to each of the numbers in the data. The standard deviation of the resulting data

   (A) cannot be determined
   (B) would be 28
   (C) would be 34
   (D) would be 25

2. In a particular game, a fair die is tossed. If the number of spots showing is either four or five, you win ₹1. If the number of spots showing is six, you win ₹4. And if the number of spots showing is one, two, or three, you win nothing. You are going to play the game twice. The probability that you win ₹4 both times is

   (A) 1/3
   (B) 1/4
   (C) 1/6
   (D) 1/36

3. The first difference of a stationary AR(1) process can be expressed as

   (A) a white noise process
   (B) an invertible MA(1) process
   (C) a non-stationary AR(2) process
   (D) a stationary ARMA(1,1) process

4. The stages of a malignant disease (cancer) is recorded using the symbols 0, I, II, III, IV. We say that the scale used is

   (A) Alphanumeric
   (B) Numerical
   (C) Ordinal
   (D) Nominal

5. Voltages in a circuit vary between 5 V and 11 V and voltages are spread evenly over the range of possibilities (Uniform Distribution). Then, $P(5.5 \leq X \leq 10)$ is obtained as

   (A) 1/2
   (B) 1/3
   (C) 3/4
   (D) None of the above

6. If the correlation coefficient $(r) = 1.00$, then

   (A) all the data points must fall exactly on a straight line with a negative slope
   (B) all the data points must fall exactly on a straight line with a slope that equals 1.00
   (C) all the data points must fall exactly on a straight line with a positive slope
   (D) all the data points must fall exactly on a horizontal straight line with a zero slope

7. Ignoring twins and other multiple births, assume babies born at a hospital are independent events with the probability that a baby is a boy and a girl, both equal to 0.5. If the first 4 children born are girls, what is the probability the next born child is a boy?

   (A) 0.50
   (B) 0.75
   (C) 1/8
   (D) 1/32

8. There are three children in a room, ages three, four, and five. If a four-year old child enters the room the

   (A) mean age will stay the same but the standard deviation will decrease
   (B) mean age will stay the same but the standard deviation will increase
   (C) mean age and standard deviation will increase
   (D) mean age and standard deviation will stay the same

9. Data on which of the following variables are dominated by seasonal component of time series

   (A) Sale of rice
   (B) Sale of edible oil
   (C) Sale of Ice cream
   (D) Sale of milk

10. When water flows across farm land, some of the soil is washed away, resulting in erosion. An experiment was conducted to investigate the effect of the rate of water flow (F) on the amount of soil washed away. Flow is measured in liters per second and the eroded soil (E) is measured in kilograms. The data are given in the following table

F	0.31	0.85	1.26	2.47	3.75
E	0.82	1.95	2.18	3.01	6.07

The association between flow rate and amount of eroded soil is

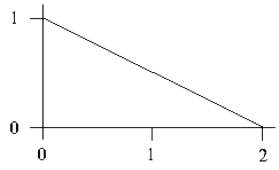
(A) positive

(B) negative

(C) neither positive nor negative

(D) impossible to determine since both variables are categorical

11. Suppose X is a continuous random variable taking values between 0 and 2 and having the probability density function below



then $P(0 \leq X \leq 1)$ has value

(A) 0.25

(B) 0.33

(C) 0.66

(D) 0.75

12. A study was conducted to assess a new surgical procedure designed to reduce the incidence of post-operative complications. The incidence of complications was found to be 40% in 25 patients having the new procedure and 60% in 20 patients having the old procedure. This difference was found to be not statistically significant. It may be concluded that

(A) the new procedure is effective in reducing post-operative complications

(B) the new procedure is as good as the old procedure

(C) the sample is biased

(D) the result is clinically significant

13. In interpreting a box-plot of a data set we note that the median is to the left of the center of the box and the right line is longer than the left line. We can conclude that

(A) the data is symmetric

(B) skewness or symmetry cannot be determined by a box plot

(C) the data is skewed right

(D) the data is skewed left

14. In a chi-square 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

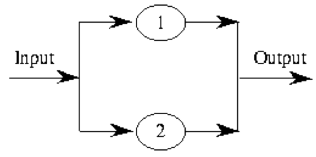
(A) $mn - 1$

(B) $mn - m - n + 1$

(C) $mn - m - n - 1$

(D) None of the above

15. A system has two components that operate in parallel, as shown in the diagram below. Because the components operate in parallel, at least one of the components must function properly if the system is to function properly. The probabilities of *failures* for the components 1 and 2 during one period of operation are 0.20 and 0.03, respectively. Let F denote the event that the component 1 *fails* during one period of operation and G denote the event that component 2 *fails* during one period of operation. The component failures are independent.



The event corresponding to the above system functioning properly during one period of operation is

(A) F and G

(B) F or G

(C) not F or not G

(D) not F and not G

16. Regression analysis was applied to return rates of sparrow hawk colonies. Regression analysis was used to study the relationship between return rate (x : % of birds that return to the colony in a given year) and immigration rate (y : % of new adults that join the colony per year). The following regression equation was obtained

$$\hat{y} = 31.9 - 0.34x$$

Based on the above estimated regression equation, if the return rate were to decrease by 10% the rate of immigration to the colony would

(A) increase by 34%

(B) increase by 3.4%

(C) decrease by 0.34%

(D) decrease by 3.4%

17. Determining the sample interval (represented by k), randomly selecting a number between 1 and k, and including each kth element in your sample are the steps for which form of sampling?

(A) Double Sampling

(B) Stratified Random Sampling

(C) Systematic Sampling

(D) Purposive sampling

18. As a non-parametric method the sign test is based on the following distribution:

(A) Exponential

(B) Poisson

(C) Binomial

(D) Normal

19. Each of the 29 NBA teams has 12 players. A sample of 58 players is to be chosen as follows. Each team will be asked to place 12 cards with its players' names into a hat and randomly draw out two names. The two names from each team will be combined to make up the sample. Which of the following sampling techniques is being used in this situation?

(A) Simple Random Sample

(B) Stratified Sample

(C) Cluster Sample

(D) Two-stage Sample

20. If T_1 is the most efficient estimator of a parameter θ with variance V_1 and T_2 is any other estimator of θ with variance V_2, then the efficiency of T_2 is defined as

(A) $V_1 \times V_2$

(B) V_1/V_2

(C) V_2/V_1

(D) $1/V_2$

21. Which one of the following variables is not categorical?

(A) Age of a person

(B) Gender of a person: male or female

(C) Choice on a test item: true or false

(D) Marital status of a person (single, married, divorced, other)

22. A manufacturing process has a 70% yield, meaning that 70% of the products are acceptable and 30% are defective. If three of the products are randomly selected, find the probability that all of them are acceptable

(A) 0.343

(B) 0.027

(C) 0.429

(D) None of the above

23. The table below describes the smoking habits (Non Smokers, Light Smokers, and Heavy Smokers) of a group of asthma sufferers.

	Non	Light	Heavy	Total
Men	343	42	49	434
Women	352	32	40	424
Total	695	74	89	858

If two different people are randomly selected from the 858 subjects, the probability that they are both heavy smokers is

(A) 0.01076

(B) 0.0001262

(C) 0.003262

(D) 0.01065

24. Which of the following are true statements?
    1. If bias is present in a sampling procedure, it can be overcome by dramatically increasing the sample size
    2. There is no such thing as a 'bad sample'
    3. Sampling techniques that use probability techniques effectively eliminate bias

    (A) 1 only
    (B) 2 only
    (C) 3 only
    (D) None of the statements are true

25. If X is a random variable from Ber(p), its MGF is given by

    (A) $[1 + p(e^t - 1)]^n$
    (B) $[1 + p(e^t - 1)]$
    (C) $pq(e^t - 1)]$
    (D) None of the above

26. A phone-in poll conducted by a newspaper reported that 73% of those who called in liked business tycoon Donald Trump. The number 73% is a

    (A) sample
    (B) parameter
    (C) population
    (D) statistic

27. The series $\sum_{1}^{\infty} \frac{1}{n(n+1)}$

    (A) converges
    (B) diverges
    (C) oscillates
    (D) None of the above

28. The series $\frac{1}{1+x} - \frac{1}{2+x} + \frac{1}{3+x} - \cdots$

    (A) converges
    (B) diverges
    (C) conditionally converges
    (D) none of the above

29. If A is matrix of order 3×3 given by

$$A = \begin{pmatrix} 0.2 & 0.3 & 0.5 \\ 0.5 & 0.1 & 0.4 \\ 0.9 & 0.1 & 0 \end{pmatrix}$$

    then the maximum value of latent root is

    (A) 0.5
    (B) 1
    (C) 3
    (D) 1.3

30. If A is matrix of order 2×2 given by

$$A = \begin{pmatrix} 0.4 & 0.6 \\ 0.2 & 0.8 \end{pmatrix}$$

    then the latent roots are

    (A) $1, 0.2$
    (B) $0.4, 0.8$
    (C) $0.3, 0.9$
    (D) $0, 1.2$

31. The probability generating function variable X is

$$P_X(t) = \frac{p}{1 - tq}, \quad |t| < 1$$

    The corresponding distribution is

    (A) Binomial
    (B) Bernoulli
    (C) Negative binomial
    (D) Geometric

32. For two sets of non-negative real numbers $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ and $p \geq 1$

$$(\Sigma x_i^p)^{\frac{1}{p}} + (\Sigma y_i^p)^{\frac{1}{p}} \geq (\Sigma (x_i + y_i)^p)^{\frac{1}{p}}$$

    is known as

    (A) Cauchy Schwartz inequality
    (B) Holder's inequality
    (C) Minkowski's inequality
    (D) None of the above

33. If X_n is a sequence of independent random variables with mean 0 and variance σ_k^2 such that $\sum (\frac{\sigma_k}{k})^2 < \infty$, the sequence $X_n \to 0$ *almost surely* is a result of

    (A) Kolmogorov SLLN
    (B) Borel's SLLN
    (C) Markoff's WLLN
    (D) Lindeberg Feller theorem

34. If X is a random variable from discrete uniform distribution with parameter K, then its MGF is given by

(A) $\frac{e^t\left(1-e^{Kt}\right)}{K(1-e^t)}$

(B) $\frac{e^t\left(1-e^{Kt}\right)}{(1-e^t)}$

(C) $e^t\left(1-e^{Kt}\right)$

(D) None of the above

35. Which of the following statements is wrong

(A) MLE is the functions of sufficient statistics when they exists

(B) MLE is invariant

(C) MLE is asymptotically efficient

(D) MLE is always UMVUE

36. Moment estimators are

(A) always unique

(B) always consistent

(C) never coincides with MLE under any situation

(D) none of the above

37. Minimum chi-square estimates are

(A) not consistent

(B) unique

(C) MCS and ML estimators are equivalent

(D) not efficient

38. X is a random variable from $U(\mu, \mu+2)$ then the sufficient statistic for μ is

(A) sample mean

(B) $Y(1)$

(C) $Y(n)$

(D) None of the above

39. If $N(t) = N_1(t) - N_2(t)$ where $N_1(t)$ and $N_2(t)$ are two Poisson processes with parameters $\mu_1$and μ_2 then $E(N(t))$ is

(A) $\mu_1 + \mu_2$

(B) $\mu_1 - \mu_2$

(C) $(\mu_1/\mu_2)t$

(D) $(\mu_1 - \mu_2)t$

40. If T be a sufficient statistic for Θ, any solution of the likelihood equation will be a

(A) function of Θ

(B) function of t

(C) function of both Θ and t

(D) None of these

41. If $\{Xn\}$ is a Markov chain with state space $S = \{0, 1, 2\}$ and t.p.m.

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 1/4 & 1/4 \\ 0 & 0 & 1 \end{pmatrix}$$

then the chain has

(A) state 0 is a communicating state

(B) state 1 is a communicating state

(C) state 2 is a communicating state

(D) None of the above

42. The critical region W_0 is said to be unbiased if $P_\theta(W_0)$ is

(A) less than $P_{\theta 0}(W_0)$

(B) greater than $P_{\theta 0}(W_0)$

(C) greater than equal to $P_{\theta 0}(W_0)$ for all $\theta = \theta_0$

(D) equal to unity

43. In testing of hypothesis, the lemma due to Neyman and Pearsxon yields

(A) only necessary condition for optimum regions

(B) only sufficient condition for optimum region

(C) necessary and sufficient conditions for optimum regions

(D) necessary condition for any region

44. If we want to confound three effects in a 2^5 factorial experiment then in a replication we would have

(A) three blocks each of size eight

(B) eight blocks each of size four

(C) four blocks each of size eight

(D) eight blocks each of size eight

45. A systematic sampling is a stratified random sampling with

(A) one unit per stratum

(B) infinite units per stratum

(C) 20 units per stratum

(D) no unit per stratum

46. In Bayesian statistical inference, a prior probability distribution, of an uncertain quantity is the probability distribution that would express one's belief about this quantity

(A) after some evidence is taken into account

(B) before some evidence is taken into account

(C) after some testing of the null hypothesis

(D) None of the above

47. To test the equality of the two population mean vectors we use

(A) multivariate normal distribution

(B) principal component analysis

(C) Hotelling T square statistic

(D) factor analysis
break

48. If in a stable population the growth rate is zero then the population becomes

(A) stationary

(B) non-stable

(C) very big

(D) non-existent

49. In the Wilcoxon Signed-Rank test We assume that the parent population is

(A) symmetric

(B) skewed

(C) sometimes symmetric and sometimes skewed

(D) none of these

50. If $\{Xn\}$ is a Markov chain with state space $S = \{0, 1\}$ and t.p.m.

$$P = \begin{pmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \end{pmatrix}$$

then the chain has limiting probabilities

(A) $1, 0$

(B) $0, 1$

(C) $\frac{1}{2}, \frac{1}{2}$

(D) $\frac{2}{3}, \frac{1}{3}$

This page is left intentionally blank

Section C (30 Marks)

Answer any 5 (five) from the following

1. Define Markov Chain. Give examples. State and prove first entrance theorem. (Marks : 6)

2. Describe stationarity in time series analysis. Discuss $AR(P)$ process in time series analysis. (Marks : 6)

3. Formulate the testing of hypothesis testing problem as a decision problem. (Marks : 6)

4. State and prove Jensen's inequality for moments. (Marks : 6)

5. Distinguish between two-state sampling and two-phase sampling. (Marks : 6)

6. Describe the demographic transition theory. (Marks : 6)

7. Describe the Chiang's method of construction of abridge life table. (Marks : 6)

Space for Answers (Section C) : for Questions 1 to 7 (6 pages)

Space for Answers (Section C) : for Questions 1 to 7 (6 pages)

Space for Answers (Section C) : for Questions 1 to 7 (6 pages)

Space for Answers (Section C) : for Questions 1 to 7 (6 pages)

Space for Answers (Section C) : for Questions 1 to 7 (6 pages)

Space for Answers (Section C) : for Questions 1 to 7 (6 pages)