Program: Computer Engineering
Curriculum Scheme: Rev2019
Examination: MEAI Semester I
Course Code: MEAIC103 and Course Name: Mathematical Foundations of Data Science
Time: 2 hour Max. Marks: 80

==========

=

Qu-1	**Choose the correct option for following questions. All the Questions are compulsory and carry equal marks**
1.	The Central Limit Theorem says that the sampling distribution of the sample mean is approximately normal if
Option A:	all possible samples are selected
Option B:	the sample size is large
Option C:	the standard error of the sampling distribution is small
Option D:	the population size is large
2.	A researcher wishes to find out about the mean of a population. She takes a sample, calculates the sample mean and works out the 95% confidence interval for the population mean. Which of the following results would she prefer?
Option A:	The confidence interval is very wide
Option B:	The confidence interval is very narrow
Option C:	It makes no difference to her what the width of the confidence interval is
Option D:	The mean is large
3.	Parametric test, unlike the non-parametric tests, make certain assumptions about
Option A:	The population size
Option B:	The underlying distribution
Option C:	The sample size
Option D:	The data size
4.	The sampling error is defined as?
Option A:	difference between population and parameter
Option B:	difference between sample and parameter
Option C:	difference between parameter and sample
Option D:	difference between population and sample
5.	Two types of errors associated with hypothesis testing are Type I and Type II. Type II error is committed when
Option A:	We reject the null hypothesis whilst the alternative hypothesis is true
Option B:	We reject a null hypothesis when it is true
Option C:	We accept a null hypothesis when it is not true
Option D:	We accept a null hypothesis when it is true
6.	The area under a standard normal curve is?
Option A:	0

Option B:	∞
Option C:	not defined
Option D:	1
7.	Poisson distribution is applied for
Option A:	Continuous Random Variable
Option B:	Discrete Random Variable
Option C:	Irregular Random Variable
Option D:	Uncertain Random Variable
8.	Consider a random variable with exponential distribution with $\lambda=1$. Compute the probability for P (X>3).
Option A:	e^{-1}
Option B:	e^{-2}
Option C:	e^{-3}
Option D:	e^{-4}
9.	An eigenvalue, corresponding to a real non zero eigenvector, points in a direction in which it is stretched by the transformation and the eigenvector is the factor by which it is stretched.
Option A:	True
Option B:	False
10.	Find the rank of matrix $\begin{bmatrix} -1 & 0 & -1 & 2 \\ 2 & 0 & 2 & 0 \\ 1 & 0 & 1 & -1 \end{bmatrix}$
Option A:	2
Option B:	3
Option C:	0
Option D:	4
11.	What is the principle of factorization?
Option A:	Every square matrix can be expressed as a product of a lower triangular matrix and upper triangular matrix
Option B:	Determinant of an identity matrix is one
Option C:	There exists no inverse for a singular matrix
Option D:	Every matrix can be expressed as a sum of a skew symmetric and a symmetric matrix
12.	Which of the following is an application of the edit distance problem?
Option A:	Approximate string matching
Option B:	Spelling correction
Option C:	Similarity of DNA
Option D:	All of the mentioned

13.	Normal Distribution is symmetric is about
Option A:	Variance
Option B:	Standard Deviation
Option C:	Mean
Option D:	Covariance
14.	Covariance
Option A:	measures the degree to which two variables co-vary
Option B:	measures of the strength of relationship between two variables
Option C:	Dependents on the units of measurement of the variables
Option D:	All of the mentioned
15.	The magnitude of the difference between observed frequencies and expected frequencies is called
Option A:	Chi-square value
Option B:	F value
Option C:	Z value
Option D:	t value
16.	Which of the following statements is false?
Option A:	The *t* distribution is symmetric about zero
Option B:	The *t* distribution is more spread out than the standard normal distribution
Option C:	As the degrees of freedom get smaller, the *t*-distribution's dispersion gets smaller
Option D:	The *t* distribution is mound-shaped
17.	A random sample of size 20 taken from a normally distributed population resulted in a sample variance of 32. The lower limit of a 90% confidence interval for the population variance would be:
Option A:	52.185
Option B:	20.170
Option C:	20.375
Option D:	54.931
18.	A dice is thrown twice. What is the probability of getting two numbers whose product is even?
Option A:	6/4
Option B:	1/2
Option C:	3/4
Option D:	5/4
19.	Let X and Y be two independent random variables. Which one of the relations between expectation (E), variance (Var) and covariance (Cov) given below is FALSE?
Option A:	$E(XY) = E(X)\,E(Y)$
Option B:	$Cov(X,Y) = 0$
Option C:	$Var(X+Y) = Var(X)+Var(Y)$

Option D:	$E(X^2Y^2) = (E(X))^2 (E(Y))^2$
20.	A sample of 15 data is as follows: 17, 18, 17, 17, 13, 18, 5, 5, 6, 7, 8, 9, 20, 17, 3. The mode of the data is
Option A:	4
Option B:	13
Option C:	17
Option D:	20

Qu-2		
A	**Solve any Two**	**5 marks each**
i.	Describe correlation coefficient.	
ii.	Explain global and local optima with graphical representation.	
iii.	Justify the need of statistics from a data science perspective.	
B	**Solve any One**	**10 marks each**
i.	The average number of major storms in your city is 2 per year. What is the probability that exactly 3 storms will hit your city next year?	
ii.	Explain Correlation matrix with example.	

Qu-3		
A	**Solve any Two**	**5 marks each**
i.	Explain Chi-Square test.	
ii.	Describe bootstrap permutation test in detail.	
iii.	Explain Cosine distance with example.	
B	**Solve any One**	**10 marks each**
i.	Calculate Eigenvalue and Eigenvector for the given matrix. $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 4 \\ 0 & 4 & 9 \end{bmatrix}$	
ii.	Illustrate significance of p value and alpha value with example. Describe Type I and Type II error in hypothesis testing.	

University of Mumbai

Examination 2021 under cluster KJSIEIT

Examinations Commencing from 22nd April 2021 to 30th April 2021

Program: Computer Engineering

Curriculum Scheme: Rev2019

Examination: MEAI Semester I

Course Code: MEAIC103 and Course Name: Mathematical Foundations of Data Science

Time: 2 hour Max. Marks: 80

==========

Question Number	Correct Option (Enter either 'A' or 'B' or 'C' or 'D')
Q1.	A
Q2.	B
Q3.	B
Q4	D
Q5	C
Q6	D
Q7	B
Q8.	C
Q9.	B
Q10.	A
Q11.	A
Q12.	D
Q13.	C
Q14.	D
Q15.	A
Q16.	C
Q17.	B
Q18.	C
Q19.	D
Q20.	C