


Name:	 UPES <small>UNIVERSITY WITH A PURPOSE</small>
Enrolment No:	

UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
Online End Semester Examination, Dec 2020

Course: Stats for Data Science	Semester: V
Program: B.Tech (Hons.) Computer Science and Engineering (DevOps)	Time 03 hrs.
Course Code: CSBD3006P	Max. Marks: 100

SECTION A

- 1. Each Question will carry 5 Marks**
- 2. Instruction: Complete the statement / Select the correct answer(s)**

S. No.	Question	CO
Q 1	Which of the following measures of central tendency not get affected by outliers in the data given for statistical computation? a) Mean b) Median c) Mode d) Variance	CO1
Q2	A web portal earns a commission on each affiliate marketing advertisement sold. This person made \$5,000 in commissions during one month and concludes that annual compensation will be \$60,000. This is an example of: a) Descriptive Statistics b) Type-II error c) Secondary Data d) Inferential Statistics	CO2
Q3	A researcher concludes from his analysis that a placebo cures COVID-19. What type of error is he making? a) Type-I error b) Type-II error c) No Error d) Error cannot be determined	CO3
Q4	The p-value in hypothesis testing represents which of the following: Please select the best answer of those provided below. a) The probability of failing to reject the null hypothesis, given the observed results b) The probability that the null hypothesis is true, given the observed results c) The probability that the observed results are statistically significant, given that the null hypothesis is true d) The probability of observing results as extreme or more extreme than currently observed, given that the null hypothesis is true	CO3
Q5	The techniques which provide the decision maker a systematic and powerful means of analysis to explore policies for achieving predetermined goals are called.....	CO4

	a) Correlation Techniques b) Mathematical Techniques c) Quantitative Techniques d) None of them	
Q6	Correlation analysis between one dependent variable with one independent variable by keeping the other independent variables as constant is called a) Partial correlation b) Multiple correlation c) Nonsense correlation d) Simple correlation	CO4

SECTION B

1. Each question will carry 10 marks
2. Instruction: Write short / brief notes

Q 7	<p>Explain the usage of non-parametric tests over parametric tests. Using Wilcoxon Matched Pairs Signed Ranks Test solve the following problem: A firm has decided to select one of two express courier delivery services to provide next-day deliveries to its different offices.</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>District Office</th> <th>DTDC (Hrs.)</th> <th>Overnite Courier (Hrs.)</th> </tr> </thead> <tbody> <tr><td>Delhi</td><td>19</td><td>15</td></tr> <tr><td>Mumbai</td><td>16</td><td>15</td></tr> <tr><td>Bengaluru</td><td>32</td><td>25</td></tr> <tr><td>Chennai</td><td>30</td><td>24</td></tr> <tr><td>Dehradun</td><td>15</td><td>16</td></tr> <tr><td>Jammu</td><td>18</td><td>20</td></tr> <tr><td>Amritsar</td><td>14</td><td>15</td></tr> <tr><td>Hyderabad</td><td>16</td><td>14</td></tr> <tr><td>Ahmedabad</td><td>7</td><td>9</td></tr> <tr><td>Jaipur</td><td>10</td><td>12</td></tr> </tbody> </table>	District Office	DTDC (Hrs.)	Overnite Courier (Hrs.)	Delhi	19	15	Mumbai	16	15	Bengaluru	32	25	Chennai	30	24	Dehradun	15	16	Jammu	18	20	Amritsar	14	15	Hyderabad	16	14	Ahmedabad	7	9	Jaipur	10	12	CO3
District Office	DTDC (Hrs.)	Overnite Courier (Hrs.)																																	
Delhi	19	15																																	
Mumbai	16	15																																	
Bengaluru	32	25																																	
Chennai	30	24																																	
Dehradun	15	16																																	
Jammu	18	20																																	
Amritsar	14	15																																	
Hyderabad	16	14																																	
Ahmedabad	7	9																																	
Jaipur	10	12																																	
Q 8	A sample of the records of motor vehicle bureau shows that 18 drivers in a certain age group received 3, 2, 0, 0, 2, 3, 3, 1, 0, 1, 0, 3, 4, 0, 3, 2, 3, 0 traffic tickets during the last three year. Find the measures of central tendency mean, median and mode?	CO2																																	
Q 9	The company ABC Pvt. Ltd. provides you a chance to test the claim made by a financial accountant that he is able to complete a standard tax return in under an hour. For a random sample of 24 tax returns, the accountant average 63.2 minutes with a standard deviation of 7.7 minutes. Calculate the following: What are the appropriate null and alternate hypothesis? What is the most accurate estimate of the p-value? What is the maximum probability of committing a type 1 error?	CO3																																	
Q 10	Discuss the Structural Equation Modelling (SEM) and its usage in examining the relationships and the limitations of the SEM model. Write all the Steps of SEM analysis.	CO4																																	

Q 11	Explain the concept of decision trees in decision making using data science. How does decision trees are different in approach from cluster analysis to solve a decision-making problem?	CO5
------	--	-----

Section C

- 1. Each Question carries 20 Marks.**
- 2. Instruction: Write long answer.**

Q12	<p>Compute the Karl Pearson Correlation for the following data provided:</p> $N = 10$ $\sum X = 300$ $\sum Y = 450$ $\sum (X-10)^2 = 2500$ $\sum (Y-10)^2 = 3000$ $\sum (X-10)(Y-15) = 2000$ <p>Find “r”</p> <p style="text-align: center;">OR</p> <p>Compute the Spearman Rank Correlation for the following data:</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>X</th> <th>Y</th> </tr> </thead> <tbody> <tr><td>11</td><td>8</td></tr> <tr><td>12</td><td>3</td></tr> <tr><td>10</td><td>2</td></tr> <tr><td>9</td><td>1</td></tr> <tr><td>12</td><td>3</td></tr> <tr><td>11</td><td>8</td></tr> <tr><td>7</td><td>9</td></tr> <tr><td>6</td><td>7</td></tr> <tr><td>8</td><td>9</td></tr> <tr><td>9</td><td>2</td></tr> <tr><td>11</td><td>10</td></tr> <tr><td>12</td><td>11</td></tr> </tbody> </table>	X	Y	11	8	12	3	10	2	9	1	12	3	11	8	7	9	6	7	8	9	9	2	11	10	12	11	CO4
X	Y																											
11	8																											
12	3																											
10	2																											
9	1																											
12	3																											
11	8																											
7	9																											
6	7																											
8	9																											
9	2																											
11	10																											
12	11																											