



**NAVVRACHANA
UNIVERSITY**

a UGC recognized University

School: School of Engineering and Technology
Program/s: B.Sc Data Science
Year: II **Year: II**
Examination: End Semester Examination
Examination year: December - 2021

Course Code: DS202

Course Name: Regression Analysis and Bayesian Statistics

Date: 03/12/2021

Time: 8:30 am to 10:30 am

Total Marks: 40

Total Pages: 1

Instructions:

- Write each answer on a new page.
- Use of a calculator is permitted/not permitted.
- * COs=Course Outcome mapping, # BTL=Bloom's Taxonomy Level mapping

Q. No.	Details	Marks	COs*	BTL#																						
Q.1	<p>The following table gives the information on incomes (in thousands of dollars) and charitable contributions (in hundreds of dollars) for a random sample of 10 households.</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>Income</td> <td>76</td> <td>57</td> <td>140</td> <td>97</td> <td>75</td> <td>107</td> <td>65</td> <td>77</td> <td>102</td> <td>53</td> </tr> <tr> <td>Contribution</td> <td>15</td> <td>4</td> <td>42</td> <td>33</td> <td>5</td> <td>32</td> <td>10</td> <td>18</td> <td>28</td> <td>4</td> </tr> </table> <p>a) With income as an independent variable and charitable contributions as a dependent variable compute SS_{xx}, SS_{xy}, and SS_{yy}</p> <p>b) Find the regression of charitable contribution on income.</p> <p>c) Briefly explain the values of a and b</p> <p>d) Calculate r and r^2 and briefly explain what they mean</p>	Income	76	57	140	97	75	107	65	77	102	53	Contribution	15	4	42	33	5	32	10	18	28	4	10	CO1 CO2 CO3	BT1, BT2 BT3
Income	76	57	140	97	75	107	65	77	102	53																
Contribution	15	4	42	33	5	32	10	18	28	4																
Q.2	Write a short note of the potential problems which we face while fitting a linear regression model		CO4	BT1, BT3, BT6																						
Q.3	<p>i) Baye's rule is given by $P(H/D) = \frac{P(H) \times P(D/H)}{P(D)}$. Name the terms and give a brief explanation according to Bayesian Statistics.</p> <p>ii) You move into a new house which has a phone installed. You can't remember the phone number, but you suspect it to be some number. To test the hypothesis, you carry out the experiment by picking up the phone and dialing that suspected number. When you do the experiment, you get the busy tone. Now you consider the following four hypothesis and calculate the following:</p> <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>Hypothesis</th> <th>Prior</th> <th>Likelihood</th> <th></th> <th>Posterior</th> </tr> </thead> <tbody> <tr> <td>Phone is working and the number is correct</td> <td>0.4</td> <td>1</td> <td></td> <td></td> </tr> <tr> <td>Phone is working and</td> <td>0.4</td> <td>0.01</td> <td></td> <td></td> </tr> </tbody> </table>	Hypothesis	Prior	Likelihood		Posterior	Phone is working and the number is correct	0.4	1			Phone is working and	0.4	0.01			10	CO5 CO6 CO7 CO8	BT1 BT2 BT3, BT4							
Hypothesis	Prior	Likelihood		Posterior																						
Phone is working and the number is correct	0.4	1																								
Phone is working and	0.4	0.01																								

	<table border="1"> <tr> <td>the number is incorrect</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Phone is not working and the number is correct</td> <td>0.1</td> <td>1</td> <td></td> <td></td> </tr> <tr> <td>Phone is not working and the number is incorrect</td> <td>0.1</td> <td>1</td> <td></td> <td></td> </tr> </table>	the number is incorrect					Phone is not working and the number is correct	0.1	1			Phone is not working and the number is incorrect	0.1	1					
the number is incorrect																			
Phone is not working and the number is correct	0.1	1																	
Phone is not working and the number is incorrect	0.1	1																	
	<p>a) Posterior probabilities of all hypotheses</p> <p>b) Probability of phone working</p> <p>c) Probability of phone working given busy signal is coming</p>																		
Q.4	<p>i) The posterior probability is given by the relation $posterior \propto prior \times \underline{\hspace{2cm}}$.</p> <p>ii) If the values near extremes are more probable, then the beta function which works well as a prior is $\beta(\underline{\hspace{1cm}}, \underline{\hspace{1cm}})$</p> <p>iii) When you have lot of data, the results are more/less sensitive to the choice of prior.</p> <p>iv) Define credible interval</p> <p>v) A term used to describe the case when the independent variables in a multiple regression model are correlated is i) regression ii) correlation iii) multicollinearity iv) none</p> <p>vi) A multiple regression model has the form $y = 2 + 3x_1 + 4x_2$. As x_1 increases by 1 unit (holding x_2 constant) y will i) increase by 3 units ii) decrease by three units iii) increase by 4 units iv) decrease by 4 units.</p> <p>vii) The correct relationship between SSR, SST and SSE is given by i) $SSR=SST+SSE$ ii) $SST=SSR + SSE$ iii) $SSE=SSR-SST$ iv) None of these</p> <p>viii) Which of the following is correct regarding the value of adjusted R^2 in multiple regression model? i) It can be negative ii) It is positive iii) It is larger than the coefficient of determination iv) None of these</p> <p>ix) How many dummy variables should be there in a model for a categorical variable with 3 levels?</p> <p>x) Heteroskedasticity refers to a situation in which the error terms form a regression analysis for equal/unequal variances.</p>	5	<p>C01</p> <p>C02</p> <p>C03,</p> <p>C04</p> <p>C05</p> <p>C06</p> <p>C07</p> <p>C08</p>	<p>BT1</p> <p>BT2,</p> <p>BT3,</p> <p>BT4</p> <p>BT5</p>															

*****End of Question Paper*****