



WEST BENGAL STATE UNIVERSITY
B.Sc. Honours 5th Semester Examination, 2020, held in 2021

STSADSE01T-STATISTICS (DSE1/2)

Time Allotted: 2 Hours

Full Marks: 40

*The figures in the margin indicate full marks.
Candidates should answer in their own words and adhere to the word limit as practicable.
All symbols are of usual significance.*

Answer any four questions from Question No. 1-6 and any two questions from Question No. 7-9

1. Which scale of measurement is most appropriate for the following variables – 4+1
nominal, or ordinal? Answer with proper reasons.
- (a) Highest educational degree obtained (below secondary, secondary, higher secondary, bachelor's, master's, doctorate).
(b) Patient condition (good, fair, serious, critical).
(c) Hospital location (Kolkata, Delhi, Chennai, Mumbai).
(d) How often feel depressed (never, occasionally, often, always).

Suppose you are going to test the effectiveness of a vaccine on some patients. Can the responses of the patients be nominal? Discuss with reasons.

2. (a) Given n, f_A, f_B and f_{AB} , how would you find the other cell frequencies and marginal frequencies of a 2×2 table? $3\frac{1}{2} + 1\frac{1}{2}$
(b) Show that the number of individuals who have A but not B or have B but not A is

$$f_A + f_B - 2f_{AB}$$

3. (a) Show that $Q = \frac{2Y}{1+Y^2}$, where Q and Y are Yule's coefficient of association $2\frac{1}{2} + 2\frac{1}{2}$
and Yule's coefficient of colligation, respectively.
(b) Using (a) or otherwise show that Q is greater in absolute value than Y , except when both are 0 or ± 1 .

4. State True or False with proper reasons: $2 + 1\frac{1}{2} + 1\frac{1}{2}$
- (a) In 2×2 tables, statistical independence is equivalent to a population log odds ratio value 0.
(b) Interchanging two rows or interchanging two columns in a contingency table has no effect on the value of the Tschuprow's measure. Thus, this statistic treats both the rows and the columns of the contingency table as nominal scale, and if either or both variables are ordinal, the measure ignores that information.
(c) For adults who sailed on the Titanic on its fateful voyage, the odds ratio between gender (female, male) and survival (yes, no) was 11.4. Then we can say that, "The probability of survival for females was 11.4 times that for males."

5. Give one example where Kendall's tau-b (τ_b) takes values 1 and one where it takes value -1 . $2\frac{1}{2}+2\frac{1}{2}$
6. If you are given a dataset with 100 patients who are either suffering from Novel Coronavirus Covid-19 or not along with their age, is it justified fitting a linear regression model in this data? Give reasons. Also discuss if there is any better regression model in this regard. 5
7. (a) Describe log linear model of independence for two way contingency tables. $2\times 5=10$
 (b) Give reasons why the estimates of the parameters are not unique for a 2×2 contingency table.
 (c) When is such a model called "saturated model"?
 (d) Is there a connection with logistic model? If any, discuss briefly.
 (e) State True or False with proper reasons: when both logistic and log linear models are fitted to a contingency table having 30 cells with a binary response, the logistic model treats the cell counts as 15 binomial observations whereas the log linear model treats the cell counts as 30 different count observations.
8. (a) What are the different types of observational studies? Discuss with examples. 8+2
 (b) Give an example which is not an observational study.
9. Suppose the responses (in proportions) of recovery of the patients for two different drug doses are ordinal categorical in nature. The following table gives the details about it. $5+1+1+1+2$

Dose / Response	Worsen	Unaltered	Recovered	Total
Mild	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1
High	p_1	p_2	p_3	1

- (a) Show that, Goodman and Kruskal's γ for this 2×3 table is independent of any one of the parameters p_1 , p_2 and p_3 .
- (b) Give an example of set of values p_1 , p_2 and p_3 for which γ is 0.
- (c) Find values of p_1 , p_2 and p_3 for which γ is 1.
- (d) Find values of p_1 , p_2 and p_3 for which γ is -1 .
- (e) Briefly discuss how do you fit an appropriate regression model in this type of responses when the doses are not categorical but given in some continuous numerical values.

N.B. : Students have to complete submission of their Answer Scripts through E-mail / Whatsapp to their own respective colleges on the same day / date of examination within 1 hour after end of exam. University / College authorities will not be held responsible for wrong submission (at in proper address). Students are strongly advised not to submit multiple copies of the same answer script.

—x—