International Institute of Health Management Research Delhi

Term End Exam (Batch- 2020-22)

Data Management and Analysis

Total marks: 70 **Duration: 2 hrs**

SECTION A
(10 marks)

Q1) Which among the following is used to display the frequency distribution of categorical data?

(a) scatter plot (b) steam and leaf plot (c) bar chart (d) none of the above

Q2) An outlier in a dataset is generally a unusual

(a) very small value (b) large value (c) both (d) none of the above

Q3) Range of a dataset is

(a) largest observation – smallest observation (b) smallest observation – median observation

(c) largest observation – mean observation (d) smallest observation – outlier observation

Q4) Which among following can be accepted as variable in SPSS

(a) time out (b) time.out (c) timeout. (d) all of the above

Q5) When all the points in a scatterplot of the data lie exactly on a straight line that slopes upward, the coorelation coefficient value will be

(a) 0.5 (b) -0.5 (c) +1 (d) -1

Q6) In which among the following window, one can enter and view data in SPSS?

(a) data view (b) variable view (c) data editor (d) all of them

Q7) Which option in SPSS allows user to arithmetically combine or alter variables and place the resulting value under a new variable name

(a) transform (b) compute variable (c) recode (d) none of above

Q8) Is it possible to convert a continuous data to categorical data in SPSS? YES/NO

Q9) Cross tabulation can help an individual to look at the relationship between two variables by organizing them in a table YES/NO

Q10) A frequency distribution for categorical data is a table that displays the possible categories along with the associated frequencies and/or relative frequencies. YES/NO

Section B

Short Question (Each carry 6 marks)

Q11) Explain the relationship between data, information and knowledge.

Q12) Explain the importance of "label" option available in variable view window of the SPSS.

Q13) Differentiate between independent and dependent variable by giving an example

Q14) Statistical test shows that two variables are associated with r = 0.6 (co-relation coefficient) and p value of 0.02. What does p-value indicates?

Q15) Briefly mention about DBMS functions.

Section C

Long Questions

Q16) A survey which was conducted among 200 people who had purchased health insurance policy, data was collected on the gender of the insurer, name of companies or health insurers which had provided health insurance, number of health insurance policies the people have, and amount of premium paid for each health insurance policy over the last five years. Answer following:

(10 marks)

a) Which of these variables are categorical?
b) Which of these variables are discrete numerical?
c) Which type of graph would be appropriate to summarize the gender data?
d) What would be the best way to show mean premium paid by the respondent to different provider of health insurance?

Q17) Interpret following outputs:

a) Output 1: **(10 marks)**

Paired Sample Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Extra tutorial class before (marks)	114.84	20	177.82	39.76
	Extra tutorial class after (marks)	371.13	20	195.12	43.63

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Extra tutorial class before (marks) & Extra tutorial class after (marks)	20	.418	.067

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% CI of the difference				
					Lower	Upper			
Pair 1	Extra tutorial class before (marks) & Extra tutorial class after (marks)	256.29	201.75	45.11	161.87	350.71	5.68	19	.000

b) Output 2: **(10 marks)**

Favourite Channel * age crosstabulation

		Elderly	Young	Total
Favourite Channel	Music	12	14	26
	News	23	4	27
	Sports	12	7	19
Total		47	25	72

Chi-Square Tests

	Value	df	Asym. Sig. (2-sided)
Pearson Chi-Square	8.954[a]	2	.011
Likelihood Ratio	9.432	2	.009
N of Valid Cases	72		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6.60.