Ramakrishna Mission Vivekananda University

Belur Math, Howrah, West Bengal

School of Mathematical Sciences, Department of Data Science

M.Sc. in Big Data Analytic 2017, Mid Semester Exam — Date: 20 Sep 2017

Course : **DA102: Basic Statistics** — Time: $1\frac{1}{2}$ hrs

Instructor : Dr. Sudipta Das — *Max marks: 40*

Student signature and Id:

---

1. A researcher has collected the following sample data.

$$3, 5, 12, 3, 2, 9, 3, 5, 7, 16$$

   (a) The variance is

   (b) The coefficient of variation is

   (c) The range is

   (d) The interquartile range is

[4]

2. The following is the frequency distribution for the speeds of a sample of automobiles traveling on an interstate highway.

Spees (mile/hour)	Frequency
50-54	2
55-59	4
60-64	5
65-69	10
70-74	9
75-79	5

   (a) The mean is

   (b) The variance is

   (c) The median is

   (d) The mode is

[4]

3. Choose the correct answers for the following questions

   (a) The measure of dispersion that is influenced most by extreme values is

      i. the mean

      ii. the standard deviation

      iii. the range

      iv. the interquartile range

(b) If a data set has an even number of observations, the median
   i. can not be determined
   ii. is the average value of the two middle items
   iii. must be equal to the mean
   iv. is the average value of the two middle items when all items are arranged in ascending order

(c) The sum of deviations of the individual data elements from their mean is
   i. always greater than zero
   ii. always less than zero
   iii. sometimes greater than and sometimes less than zero, depending on the data elements
   iv. always equal to zero

(d) Since the mode is the most frequently occurring data value, it
   i. can never be larger than the mean
   ii. is always larger than the median
   iii. is always larger than the mean
   iv. must have a value of at least two

(e) Census reports used as a source of data is
   i. Primary source
   ii. secondary source
   iii. Organized data
   iv. None

(f) Which of these represent qualitative data
   i. Height of a student
   ii. Liking or disliking of (500) persons of a product
   iii. Income of a government servant in a city
   iv. Yield from a wheat plot

(g) Patients are asked to express the amount of pain they are feeling on a scale of 1 to 10. The amount of pain, thus reported is
   i. Nominal variable
   ii. Ordinal variable
   iii. Discrete variable
   iv. Continuous variable

[7]

4. In an attempt to implement Swachha Bharat Mission, the Ministry of Drinking Water and Sanitation approached Infomaker Pvt. Ltd. to design different questionnaires for gathering information about quality of sanitation and infrastructure in the locality under each municipal corporation or Panchayat from its residents, cleaning stuffs and administrative officers. This serves to give immediate feedback on the existing issues with sanitation and help the administration to take necessary action. The questionnaires are designed to collect

information which will also help to train the cleaning stuffs and provide adequate equipment. Suppose, you are working as an analyst for Infomaker Pvt. Ltd., and assigned to design suitable questionnaires for the above-mentioned purpose. In order to convince your manager, provide brief justification that your questionnaires divulge important insight for the stipulated policy formulation. [10]

5. Let X, Y and Z be three uncorrelated variables having same standard deviation. What will be the correlation coefficient between $X + Y$ and $Y + Z$? [4]

6. For a set of pairs of observations $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, prove that

$$Total\ Variation = Explained\ Variation + Unexplained\ Variation,$$

where

- Total Variation= $\sum_{i=1}^{n}(y_i - \bar{y})^2$ [and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$]
- Explained Variation= $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ [and $\hat{y}_i$, is obtained by the regression equation of y on x at x_i.]
- Unexplained Variation= $\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$

[8]

7. What is partial correlation? Explain with a short example. [3]

This exam has total 7 questions, for a total of 40 points and 0 bonus points.