CX2100 Prob & Stat

Tutorial #1

Descriptive statistics

1. Categorize the following variables as being qualitative or quantitative and specify the level of measurement scale.

   (a) hair colour
   (b) number of television sets in a private home
   (c) possible responses to questionnaire items: strongly agree, agree, etc.
   (d) Intelligence scale of 0 to 100
   (e) the country where you were born in

2. The following data were obtained for a measurement of certain fuel droplet size.

2.1	2.2	2.2	2.3	2.3	2.4	2.5	2.5	2.5	2.8
2.9	2.9	2.9	3.0	3.1	3.1	3.2	3.3	3.3	3.3
3.4	3.5	3.6	3.6	3.6	3.7	3.7	4.0	4.2	4.5
4.9	5.1	5.2	5.3	5.7	6.0	6.1	7.1	7.8	7.9
8.9									

   (a) Group the droplet sizes and obtain a frequency table using class interval of 1 unit. Construct the frequency histogram and comment on the shape of the distribution.
   (b) Construct the Stem-and-leaf diagram.
   (c) Compute the mean, the 25th, 50th and 75th percentile.
   (d) Construct the box plot for the droplet size data.

3. A frequency distribution of the length of telephone calls monitored at the switchboard of an office is given below. Obtain the mean and the mode of the call duration.

Length of Calls (minutes)	Number of Calls
0 and under 2	10
2 and under 4	25
4 and under 6	20
6 and under 8	40
8 and under 10	5
Total	100

4. Find the sample standard deviation for a set of data for which $n = 10$, $\sum x = 50$ and $\sum x^2 = 500$. Explain why it is impossible to have $n = 10$, $\sum x = 50$ and $\sum x^2 = 100$ for a given set of data.

5. Given the following data, determine the Pearson's correlation between variable X and variable Y. Comment on your result.

X	2.5	3.4	5.6	6.7	7.9
Y	10.2	11.8	13.7	19.7	20.6

6. In an attempt to find the mean number of hours his tutorial classmates spent per day preparing for tutorials, John collected data from 10 of his friends in the tutorial group and found that the mean is 2.4 hours with a standard deviation of 0.8 hours. However, a day later he felt that the sample size is too small. So he collected data from another 5 of his friends and found that the mean is 2.0 hours with a standard deviation of 1.2 hours. Find the mean and standard deviation when these 2 sets of data are combined.

Answers

1. (a) qualitative, nominal
   (b) quantitative, ratio
   (c) qualitative, ordinal
   (d) quantitative, interval
   (e) qualitative, nominal

2. (a)

Class	[2,3)	[3,4)	[4,5)	[5,6)	[6,7)	[7,8)	[8,9)
Frequency	13	14	4	4	2	3	1

Histogram and comment on the shape (skewed to the right).

(b) Stem-and-leaf diagram:

(c) mean, 25th, 50th and 75th percentile = 3.97, 2.85, 3.4 and 5.0 respectively.

(d) Box plot:

3. $\bar{x}$ = 5.1 mins, *mode* = 7 mins (middle of the class interval)

4. s = 5.27, Not possible to have variance < 0.

5. ρ = 0.955. X and Y are highly positive correlated.

6. $\bar{x}$ = 2.267 , s = 0.93