Problem Set 1 - Variables, Data Sets and Descriptive Statistics

ECO 104 - Statistics for Business and Economics - I
Summer-2025
Faculty: Shaikh Tanvir Hossain
TA: Habiba Afroz
Due Date- 9th June 2025

1. `Tripeasy.com` is one of many online websites that provides hotel reviews throughout the world. Ratings provided by 649 guests at the Lakeview Hotel can be found in the file `question-1.xlsx`. Possible responses were Excellent, Very Good, Average, Poor, and Terrible.

   Now answer the following questions-

   (a) Construct a frequency distribution.

   (b) Construct a percent frequency distribution.

   (c) Construct a bar chart for the percent frequency distribution.

   (d) Comment on how guests rate their stay at the Sheraton Anaheim Hotel.

   (e) Suppose that results for 1679 guests who stayed at the Timber Hotel provided the following frequency distribution.

Rating	Frequency
Excellent	807
Very Good	521
Average	200
Poor	107
Terrible	44

   Compare the ratings for the Timber Hotel with the results obtained for the Lakeview Lodge.

2. Consider the following data.

8.9	10.2	11.5	7.8	10.0	12.2	13.5	14.1	10.0	12.2
6.8	9.5	11.5	11.2	14.9	7.5	10.0	6.0	15.8	11.5

   (a) Construct a frequency distribution.

   (b) Construct a percent frequency distribution.

3. The following observations are for two quantitative variables, x and y.

Observation	**x**	**y**	**Observation**	**x**	**y**
1	28	72	11	13	98
2	17	99	12	84	21
3	52	58	13	59	32
4	79	34	14	17	81
5	37	60	15	70	34
6	71	22	16	47	64
7	37	77	17	35	68
8	27	85	18	62	67
9	64	45	19	30	39
10	53	47	20	43	28

(a) Develop a cross tabulation for the data, with x as the row variable and y as the column variable. For x use classes of 10–29, 30–49, and so on; for y use classes of 40–59, 60–79, and so on.

(b) Compute the row percentages.

(c) Compute the column percentages.

(d) What is the relationship, if any, between x and y?

4. People often wait until middle age to worry about having a healthy heart. However, recent studies have shown that early monitoring of risk factors such as blood pressure can be very beneficial (The Wall Street Journal). Having higher than normal blood pressure, a condition known as hypertension, is a major risk factor for heart disease. Suppose a large sample of male and female individuals of various ages was selected and that each individual's blood pressure was measured to determine if they have hypertension. For the sample data, the following table shows the percentage of individuals with hypertension.

Age	**Male**	**Female**
20–34	11.00%	9.00%
35–44	24.00%	19.00%
45–54	39.00%	37.00%
55–64	57.00%	56.00%
65–74	62.00%	64.00%
75+	73.30%	79.00%

(a) Develop a side-by-side bar chart with age on the horizontal axis, the percentage of individuals with hypertension on the vertical axis, and side-by-side bars based on gender.

(b) What does the display you developed in part (a) indicate about hypertension and age?

(c) Comment on differences by gender.

5. Consider the following data and corresponding weights-

x_i	**Weight** (w_i)
3.2	6/19
2.0	3/19
2.5	2/19
5.0	8/19

(a) Compute the weighted mean.

(b) Compute the sample mean of the four data values without weighting. Note the difference in the results provided by the two computations.

6. Advertising Age annually compiles a list of the 100 companies that spend the most on advertising. Consumer-goods company Procter and Gamble has often topped the list, spending billions of dollars annually. Consider the data found in the file Advertising. It contains annual advertising expenditures for a sample of 20 companies in the automotive sector and 20 companies in the department store sector.

(a) What is the mean advertising spent for each sector?

(b) What is the standard deviation for each sector?

(c) What is the range of advertising spent for each sector?

(d) What is the inter quartile range for each sector?

(e) Based on this sample and your answers to parts (a) to (d), comment on any differences in the advertising spending in the automotive companies versus the department store companies.

7. The New York Times reported that Apple has unveiled a new iPad marketed specifically to school districts for use by students (The New York Times website). The 9.7-inch iPads will have faster processors and a cheaper price point in an effort to take market share away from Google Chromebooks in public school districts. Suppose that the following data represent the percentages of students currently using Apple iPads for a sample of 18 U.S. public school districts.

15 22 12 21 26 18 42 29 64 20 15 22 18 24 27 24 26 19

(a) Compute the mean and median percentage of students currently using Apple iPads.

(b) Compare the first and third quartiles for these data.

(c) Compute the range and interquartile range for these data.

(d) Compute the variance and standard deviation for these data.

(e) Are there any outliers in these data?

(f) Based on your calculated values, what can we say about the percentage of students using iPads in public school districts?

8. Show the five-number summary and the boxplot for the following data: 5, 15, 18, 10, 8, 12, 16, 10, 6.

9. Five observations taken for two variables follow.

x_i	6	11	15	21	27
y_i	6	9	6	17	12

(a) Develop a scatter diagram for these data.

(b) What does the scatter diagram indicate about a relationship between x and y?

(c) Compute and interpret the sample covariance.

(d) Compute and interpret the sample correlation coefficient.