

Problem 1: Are the following variables categorical (C), quantitative (Q), or both (B)?

- (a) Q Shoe size
- (b) C Shoe color
- (c) C College majors
- (d) C Gender
- (e) C Gender [Entered into a spreadsheet as Male= 0 and Female= 1.]
- (f) Q SAT score
- (g) C Area code
- (h) B Grade

Problem 2: How do a bar graph and a histogram differ?

A bar graph gives the frequencies for a categorical variable while a histogram is used to give the frequencies for a quantitative variable.

Problem 3: Foziabez gives an exam to her students who receive the following scores:

65, 99, 65, 67, 98, 65, 65, 63, 98, 65.

- (a) What is the mean exam score? What is the median exam score?

$$\bar{x} = \frac{65 + 99 + 65 + \cdots + 65}{10} = \frac{750}{10} = 75$$

In order, the numbers are: 63, 65, 65, 65, 65, 65, 67, 98, 98, 99. There are ten numbers, so the 'middle' number is the average of the $10/2 = 5$ th and 6th number. But $\frac{65+65}{2} = 65$. So the median is 65.

- (b) If you were to take the exam, would the mean or median give you the best "guess" as to what score you would get? Briefly explain your answer.

'Most' of the numbers are 'close' to 65. The mean is larger than the median in this case because the data is right skewed due to the 'large' exam scores of 98, 98, and 99. The median is probably the better estimate of the 'average' exam score.

Problem 4:

- (a) Can the mean be smaller than the median? If not explain why, and if so give an example.
- (b) Can the mean be larger than the median? If not explain why, and if so give an example.
- (c) Can the mean ever be equal to the median? If not explain why, and if so give an example.

Solution.

- (a) Yes. Consider the example 0, 3, 3, whose mean is 2 and whose median is 3.
- (b) Yes. Consider the example 0, 1, 5, whose mean is 2 and whose median is 2.
- (c) Yes. Consider the example 0, 1, 2, whose mean and median is 1.

Problem 5: A sample of employees in a large pharmaceutical company has been obtained. The length of time (in months) they have worked for the company has been obtained. The length of months of employment has been recorded and is given below:

33, 59, 67, 68, 74, 74, 78, 78, 79, 80, 81, 81, 82, 83, 83, 85, 91, 99

- (a) Construct a stem plot of the data.

3	3	3
4		
5	9	
6	78	
7	44889	
8	0112335	
9	19	

- (b) Give the 5-number summary.

Min	Q_1	Median	Q_3	Max
33	74	79.5	83	99

- (c) Are there any outliers? If yes, list them. Justify your answer using the $1.5 \times IQR$ rule.

The IQR is $IQR = Q_3 - Q_1 = 83 - 74 = 9$. Then $1.5 \times IQR = 1.5 \cdot 9 = 13.5$. We have $Q_1 - 1.5 \times IQR = 66$ and $Q_3 + 1.5 \times IQR = 93$. Therefore, the values 33, 59, and 99 are outliers.