<div align="center">

Math 17: Exercise Set 6 - Posted 10/23
Due Monday 10/29 in class

</div>

15.1: Graphical Summarization of Data

1. The following dataset gives the home-to-school distance for $N = 18$ students.

<div align="center">

5.0, 1.0, 1.5, 1.5, 2.5, 1.5, 6.5, 1.5, 0.5,

3.0, 0.0, 2.0, 0.5, 6.0, 1.0, 2.5, 1.5, 2.0

</div>

   (a) Create a frequency table for the data.

   (b) Create a bar graph for the data.

   (c) Create a histogram with bins of length 1. That is, the bins should be as follows: $1.0 - 1.9$, $2.0 - 2.9$, ..., $6.0 - 7.0$.

   (d) Which of the three methods above do you think summarizes the data best? Briefly justify your answer.

2. Suppose we have the following dataset containing monthly revenue and monthly cost (in thousands of dollars) for a business.

<div align="center">

Month	1	2	3	4	5	6
Revenue	300	250	275	325	375	425
Cost	100	150	150	175	150	275

</div>

   Create a line graph of BOTH the monthly revenue and monthly cost. Without doing any calculations (that is, only using the graph), identify the month that the business profited from most and the month they profited from least.

15.2 & 15.3: Numerical Summarization of Data

3. The following questions reuse the home-to-school distance given in Problem # 1.

   (a) Determine the five number summary for the data. Use this to create a box plot.

   (b) Compute the range and interquartile range.

   (c) Compute the standard deviation.

4. **Outliers** are data points that "lie outside" the rest of the dataset. For example, if twenty students in a class got 60-75 on an exam and Student X got 100, then we would call student X an outlier. Outliers are often due to experimental error or error in recording the data and are usually omitted. A rule of thumb is to consider any points less or equal to Q_1 OR greater or equal to Q_3 to be outliers.

   (a) Determine the outliers.

<div align="center">

1

</div>

(b) Remove the outliers from the data set and again determine the five number summary for the new dataset. Hint: the new dataset should only have 7 data points.

(c) Recompute the standard deviation and compare it to the result from part (c). What do you notice?