1. For each of the following settings:
(i) identify the variable(s) in the study,
(ii) for each variable tell the type of variable (e.g., categorical and ordinal, discrete, etc.),
(iii) identify the observational unit, and
(iv) determine the sample size.

a. A botanist grew 15 pepper plants and measured the stem length in centimeters after 21 days of growth.
Correct:
(i) stem length
(ii) quantitative and continuous
(iii) pepper plant
(iv) 15

b. As a part of a classic experiment on mutations, ten aliquots (a part of the culture) of identical size were taken from the same culture of the bacterium *E. coli*. For each aliquot, the number of bacteria resistant to a certain virus was determined.
Correct:
(i) number of resistant bacteria
(ii) quantitative and discrete
(iii) aliquot
(iv) 10

c. A geneticist observed 234 progeny from self-pollinating pink-flowered snapdragon plants for their color.
Correct:
(i) color
(ii) qualitative and nominal
(iii) snapdragon progeny
(iv) 234

2. Consider the following fictitious data set.
Note, there are no units for this data set.
23, 29, 24, 21, 23

a. Compute the mean.
Correct:
\[
\frac{23 + 29 + 24 + 21 + 23}{5} = 24
\]

b. Compute the sample standard deviation.
Correct:
We compute the mean to be 24.
\[
s^2 = \frac{(23-24)^2 + 29-24^2 + (24-24)^2 + (21-24)^2 + 23-24)^2}{4} = \frac{1 + 25 + 0 + 9 + 1}{4} = 9
\]
\[
s = \sqrt{s^2} = 3
\]

23, 29, 24, 21, 23

a. Compute the mean.
Correct:
\[
\frac{23 + 29 + 24 + 21 + 23}{5} = 24
\]

b. Compute the sample standard deviation.
Correct:
We compute the mean to be 24.
\[
s^2 = \frac{(23-24)^2 + 29-24^2 + (24-24)^2 + (21-24)^2 + 23-24)^2}{4} = \frac{1 + 25 + 0 + 9 + 1}{4} = 9
\]
\[
s = \sqrt{s^2} = 3
\]

c. It's easy to see that the median for this fictitious data set is 23 and you computed the mean in part (a).
Consider replacing the value 24 in this data set with the value 29. Using this new data set, calculate the mean and median again and comment on which of these changed and which didn't.
Correct:
The median stays the same at 23.
The mean becomes \(\frac{23 + 29 + 29 + 21 + 23}{5} = 25\).
The median stayed the same and the mean was affected by this larger value.
3. 2.62. A botanist grew 15 pepper plants on the same greenhouse bench. After 21 days, she measured the total stem length (cm) of each plant, and obtained the following values:
12.4 12.2 13.4 10.9 12.2 12.1 11.8 13.5 12.0 14.1 12.7 13.2 12.6 11.9 13.1
a. Construct an ordered stemplot for these data.

Correct:
Pepper Plant Stem Length After 21 Days
10 | 9
11 | 8 9
12 | 0 1 2 2 4 6 7
13 | 1 2 4 5
14 | 1
KEY: 14 | 1 = 14.1 cm
b. Does this stemplot indicate these data come from a reasonably symmetric bell-shaped distribution?
Correct:
Yes, the stemplot shows a reasonably symmetric bell-shape.
c. Determine the quartiles for these data (including the median).
Correct:
Quartiles are
Q1 = 12.0 cm
Q2 (median) = 12.4 cm
Q3 = 13.2 cm
d. Determine the interquartile range (IQR).
Correct:
IQR = 13.2 - 12.0 = 1.2 cm
e. Determine the range.
Correct:
range = max - min = 14.1 - 10.9 = 3.2 cm

4. Refer to the data from exercise 2.62.
a. How large or small would an observation need to be considered an outlier?
Correct:
The upper fence here is Q3 + 1.5xIQR =
13.2 + 1.5x1.2 =
13.2 + 1.8 = 15.0 cm.
Similarly, the lower fence here is Q1 - 1.5xIQR =
12.0 - 1.5x1.2 =
12.0 - 1.8 = 10.2 cm.
So, to be considered an outlier here, an observation must be larger than 15.0 cm or smaller than 10.2 cm.
b. According to our definition of an outlier, are there any outliers for the pepper plant data?
Correct:
There are no outliers in this data set.

5. Refer to the pepper plant data of exercise 2.62 and note that the mean pepper plant height is 12.54 cm and the standard deviation is 0.813 cm.
a. What percent of the observations are within 1 SD (standard deviation) of the mean?
Correct:
The observations that lie within 1 SD of the mean would lie between 12.54 +or- 0.813 cm
which would lie within the interval (11.727, 13.353).
There are 11 out of 15 observation within this interval, or 73.3%.
b. Compare this to the prediction of the empirical rule.
Correct:
The empirical rule states that for reasonably symmetric, bell-shaped distributions, approximately 68% of the data will lie within 1 SD of the mean. In the pepper plant data, 73% of the data is within 1 SD of the mean.
6. 2.78 (modified) The following boxplots show the mortality rates (deaths within one year per 100 patients) for heart transplant patients at various hospitals. The low-volume hospitals are those that perform between 5 and 9 transplants per year. The high-volume hospitals perform 10 or more transplants per year.

![Boxplot Image]

a. Using the information you can gather from the boxplot, discuss how the shapes of the low and high volume distributions compare to each other.

Correct:
The low-volume hospital data appears to be symmetric where the high-volume hospital data appears less symmetric (a slight skew right?).

Instructor feedback: Remember that when we talk about the "shape" of a distribution in this class, we use words like unimodal/bimodal, symmetric/asymmetric, and skewed right/left. Also make sure to convince yourself that we cannot tell if a distribution is unimodal or bimodal from a boxplot, we can only say something about symmetry or possible skewness.

b. Using the boxplots, discuss how the centers of the two distributions compare.

Correct:
The median for the low-volume hospital appears to be around 21 deaths within a year per 100 patients where the median for the high volume hospitals is lower (somewhere around 17 deaths).

c. Now, discuss the spread of each of the two distributions and how they compare.

Correct:
The low volume distribution appears to have an IQR of approximately 30-12=18 (Range=40-0=40). The high volume distribution appears to have an IQR of approximately 19-11=8 (Range=30-5=25). Based on the range and the IQR, the low volume distribution appears to have more spread than the high volume distribution.

d. There are 32 hospitals in the low volume data set. How many of these low volume hospitals had mortality rates between 30 and 40 percent?

Correct:
8 hospitals had mortality rates between 30 and 40 percent. This is evident by observing that the upper "whisker" (indicating the upper quartile) extends from 30 to 40. And so, 25% of the 32 hospitals is 8 hospitals.