# 3: Summary Statistics

Review Questions

1. List three measures of central location.
2. The two distinct measures of spread are:
3. A study shows a mean of 0.98 and median of 0.56. What does this suggest about the shape of the distribution?
4. The line inside a boxplot represents the distribution's _____.
5. The top of the whisker on a box-and-whiskers plot goes up to the [3 words].
6. What visual clues of spread are on a boxplot?
7. Besides being the gravitational center of the distribution, the mean tells you what three things?
8. This is the reason we don't often use the variance for descriptive purposes.
9. The *hinge-spread* is the same as the _____.
10. What five points comprise a 5-point summary?
11. Q1 is the median of the _____ half of the data set.
12. Do the upper and lower fences appear on a boxplot?
13. The denominator of the formula for the population variance is _____.
14. Distinguish between μ and $\bar{x}$. Distinguish between σ and s.
15. Discuss these side-by-side boxplots.
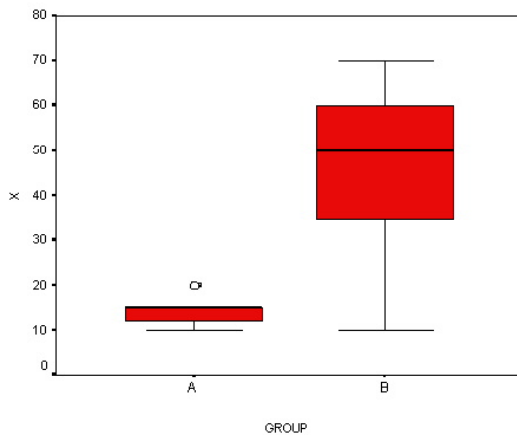
16. When *n* is odd and you are splitting the data into a high group and low group for the purpose of determine Tukey's quartiles, does the median go into the low group, the high group, neither group or both groups?

Exercises begin on the next page…

Exercises

3.1 *Serum polyphenols and red wine consumption.* Drinking moderate amounts of red wine may reduce the risk of heart disease. The proposed mechanism is that polyphenols in red wine reduce serum cholesterol. An experiment is conducted in which 9 healthy volunteers drink half a bottle of red wine daily for two weeks. They have their serum polyphenol levels measured at the beginning and end of this experiment. Percent change in polyphenols levels are shown below (Nigdikar 1998; Moore 2004).

3.5	8.1	7.4	4.0	0.7	4.9	8.4	7.0	5.5

(A) Calculate the mean and variance by hand showing the solution step-by-step. Calculate the standard deviation from the variance. [Use four or five significant digits during calculations and round your final answer to three significant digits.]
(B) Calculate the mean and standard deviation with your calculator.
(C) If answers to (A) and (B) do not match to three significant digits, go back and correct your work!

3.2 Comparing two groups with l*eaves on a common stem*. Take note of the three situations depicted below. In each stemplot, Group 1 has its values to the left of the stem and Group 2 has its values to the right of the stem. For example, in situation (A), Group 1 has these values {10, 30, 50, 70, 90} and Group 2 has these values {30, 40, 50, 60, 70}.

```
(Situation A)              (Situation B)              (Situation C)

Group 1| |Group 2          Group 1| |Group 2          Group 1| |Group 2
--------------             ------------------         ------------------
    0|1|                        |1|                        0|1|
     |2|                        |2|                         |2|
    0|3|0                       |3|0                        0|3|
     |4|0                       |4|0                         |4|
    0|5|0                      0|5|0                        0|5|0
     |6|0                      0|6|0                         |6|0
    0|7|0                      0|7|0                        0|7|0
     |8|                       0|8|                          |8|0
    0|9|                       0|9|                        0|9|0
     ×10                        ×10                         ×10
```

(A) Calculate the mean and standard deviation for group 1 and then for group 2 in Situation A. (Use your calculator, by all means!) Then discuss your results relating them to what you *see* on the plot.
(B) Do the same for Situation B.
(C) Do the same for Situation C.
(D) What does the mean tell you about each distribution? What does the standard deviation tell you?

3.3 *Gravitational center.* This exercise is intended to demonstrate how the mean is the balancing point of a distribution while telling you nothing about its shape or spread.

(A) The values 1 and 5 are marked as "X"s on the number line below. Calculate the mean and mark its location on the number line.

```
X                   X
1----2----3----4----5
```
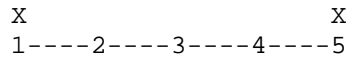
(B) The values {1, 5, 5} are shown on this number line below. Calculate the mean mark its location. Notice how the extra 5 "pulls" the mean to the right.
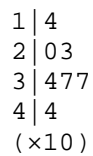
```
                    X
X                   X
1----2----3----4----5
```
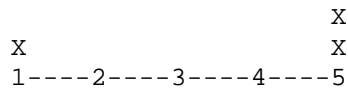
(C) Their values {2.75, 3.00, 3.25} are shown on this number line. Calculate the mean and show its location on the number line.

```
          XXX
1----2----3----4----5
```
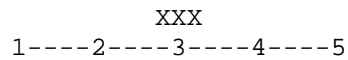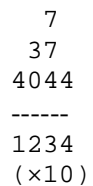
(D) Calculate the mean of these three points and show its location. What have you learned about the behavior of the mean?

```
          X
          X         X
1----2----3----4----5
```
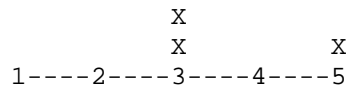
3.4 *Eye-balling the mean.* Consider this stemplot of seven data points:

```
1|4
2|03
3|477
4|4
(×10)
```

```
Here's a horizontally orientated stemplot of same data:

  7
 37
4044
------
1234
(×10)
```

(A) Visually estimate ("eye-ball") the gravitational center of the distribution. This is the approximate location of the mean.
(B) Calculate the mean.
(C) How well did you do in eye-balling the mean?

3.5 *Spread #1.* Each of these batches of numbers has a mean of 100. Which has the largest standard deviation? (Arithmetic *not* required. Calculator not required. Look at data!)

```
Batch A:     0     50    100    150      200
Batch B:    50     75    100    125      150
Batch C:    75   87.5    100  112.5      125
```

3.6 *Spread #2.* Which of the following age distributions has the greatest variability? (Arithmetic *not* required. Calculator not required. Look at data!)

```
Batch A: 0 years, 1 year, 2 years
Batch B: 0 months, 12 months, 24 months
Batch C: 0 days, 365 days, 730 days
```

3.7 *Treatment for arrhythmia.* An individual with an irregular heart rhythm is given a medication that is intended to stabilize their heart rate. The patient's resting heart rate (beats per minute) is monitored before and after treatment, with results as follows:

```
Before:     65     85     90     65     55     60
After :     68     70     69     70     71     72
```

(A) Calculate the mean and standard deviation *before* the patient started taken the medication. [*As always*, use your calculator unless step-by-step calculations are explicitly requested.]
(B) Calculate the mean and standard deviation *after* starting the medication.
(C) The objective of the treatment is to stabilize the heart rate by increasing its consistency. Which statistics will you use to help make this judgment? Did the drug work?

3.8 *Seizures following bacterial meningitis.* Data on the number of months between the onset of bacterial meningitis and seizures in 13 children are shown below.

```
0.10   0.25   0.50   4     12     12     24     24     31     36     42     55     96
```

(A) Calculate the mean and standard deviation for these data.
(B) Determine the median and IQR. Also list the 5-point summary. [Use Tukey's quartiles!]
(C) Compare the mean and median. What does this tell you about the shape of the distribution?

3.9 *Body weight expressed as a percentage of ideal*. Data on body weight expressed as a percentage of ideal are shown on the stemplot below.
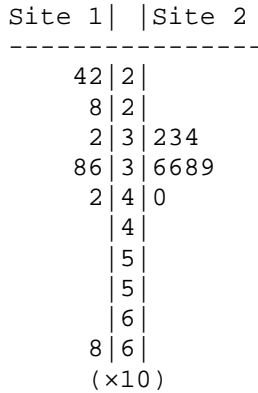
```
08|8
09|59
10|0147
11|444679
12|0145
13|
14|
15|2
×10
```

(A) Report the 5-point summary.
(B) Convert the data to a Tukey boxplot.

3.10 Air pollution at two sites. Suspended particulate matter in air is an indicator of pollution. Particulate air pollution levels (micrograms per cubic meter) from two sites are shown as stemplots on a common stem below.

```
Site 1| |Site 2
----------------
   42|2|
    8|2|
    2|3|234
   86|3|6689
    2|4|0
     |4|
     |5|
     |5|
     |6|
    8|6|
     (×10)
```

(A) Transcribe the values for the two groups and then, using the step-by-step approach demonstrating deviations and the sum of squares, calculate the mean and standard deviation for the values at Site 1.
(B) Use your calculator to determine the mean and standard deviation for the values at Site 2.
(C) Compare the results from the two sites. Do the means differ significantly? How do the groups differ?
(D) Draw a boxplot for the data from site 1. Then, on the same axis, draw the boxplot for site 2. What do you observe?

3.11 Skin fold thickness. Skin-fold thickness is an anthropometric method for assessing fat deposition and its variation with health and disease. It is measured with specialized skin-fold calipers over the triceps muscle in the arm. Measurements (millimeters) at made at the midpoint of the triceps in 5 men with chronic lung disease and 6 normal controls.

```
Chronic lung disease:    9.1    10.9   11.4   15.3   18.4
Normal controls:        10.4   19.6   20.6   23.8   24.7   32.8
```

3.12 FEV. Forced expiratory volume is a measure of respiratory function determined by having a person blow forcibly through a tube. The rate of airflow is measured in "liters per second." FEV values for two independent groups are shown below. Calculate and compare the means and standard deviations of the groups. [By all means, use your calculator!] Comment on your findings.

```
Group 1 (n = 7):  3.94  1.47  2.06  2.36  3.74  3.43  3.78
Group 2 (n = 6):  1.22  3.63  1.95  2.01  2.43  3.02
```