Stats & Probability
Sept 3rd, 2015

Stats Notes

Stem & Leaf Plots

Measures of Center

- **Stemplots (Stem-and-Leaf Plots)**
  - Another simple graphical display for small data sets is a stemplot. Stemplots give us a quick picture of the distribution while including the actual numerical values.

How to Make a Stem plot

1) Separate each observation into a **stem** (all but the final digit) and a **leaf** (the final digit).

2) Write all possible stems from the smallest to the largest in a vertical column and draw a vertical line to the right of the column.

3) Write each leaf in the row to the right of its stem.

4) Arrange the leaves in increasing order out from the stem.

5) Provide a key that explains in context what the stems and leaves represent.

- **Stemplots (Stem-and-Leaf Plots)**
  - These data represent the responses of 20 female AP Statistics students to the question, “How many pairs of shoes do you have?” Construct a stemplot.

50	26	26	31	57	19	24	22	23	38
13	50	13	34	23	30	49	13	15	51

1	
2	
3	
4	
5	

Stems

1	93335
2	664233
3	1840
4	9
5	0701

Add leaves

1	33359
2	233466
3	0148
4	9
5	0017

Order leaves

Key: 4|9 represents a female student who reported having 49 pairs of shoes.

Add a key

Displaying Quantitative Data

- **Splitting Stems and Back-to-Back Stemplots**
  - When data values are "bunched up", we can get a better picture of the distribution by **splitting stems**.
  - Two distributions of the same quantitative variable can be compared using a **back-to-back stemplot** with common stems.

Females

50	26	26	31	57	19	24	22	23	38
13	50	13	34	23	30	49	13	15	51

Males

14	7	6	5	12	38	8	7	10	10
10	11	4	5	22	7	5	10	35	7

Stem	
0	
0	
1	
1	
2	
2	
3	
3	
4	
4	
5	
5	

"split stems"

Females		Males
	0	4
	0	555677778
333	1	0000124
95	1	
4332	2	2
66	2	
410	3	
8	3	58
	4	
9	4	
100	5	
7	5	

Key: 4|9 represents a student who reported having 49 pairs of shoes.

- **Measuring Center: The Mean**
  - The most common measure of center is the ordinary arithmetic average, or **mean**.

Definition:

To find the **mean** $\bar{x}$ (pronounced "x-bar") of a set of observations, add their values and divide by the number of observations. If the *n* observations are $x_1, x_2, x_3, \ldots, x_n$, their mean is:

$$\bar{x} = \frac{\text{sum of observations}}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

In mathematics, the capital Greek letter Σ (Sigma) is short for "add them all up." Therefore, the formula for the mean can be written in more compact notation:

$$\bar{x} = \frac{\sum \mathrm{x_i}}{n}$$

- **Measuring Center: The Median**
  - Another common measure of center is the **median**. In section 1.2, we learned that the median describes the midpoint of a distribution.

Definition:

The **median M** is the midpoint of a distribution, the number such that half of the observations are smaller and the other half are larger.

To find the median of a distribution:

1) Arrange all observations from smallest to largest.

2) If the number of observations ***n* is odd**, the median *M* is the center observation in the ordered list.

3) If the number of observations ***n* is even**, the median *M* is the *average of the two center observations* in the ordered list.

- **Measuring Center**
  - Use the data below to calculate the mean and median of the commuting times (in minutes) of 20 randomly selected New York workers.

Example

10	30	5	25	40	20	10	15	30	20	15	20	85	15	65	15	60	60	40	45

$$\bar{x} = \frac{10+30+5+25+\ldots+40+45}{20} = 31.25 \text{ minutes}$$

0	5
1	005555
2	00**05**
3	00
4	005
5	
6	005
7	
8	5

Key: 4|5 represents a New York worker who reported a 45-minute travel time to work.

$$M = \frac{20+25}{2} = 22.5 \text{ minutes}$$

Comparing the Mean and the Median

- The mean and median measure center in different ways, and both are useful.
- *Don't confuse the "average" value of a variable (the mean) with its "typical" value, which we might describe by the median.*

Comparing the Mean and the Median

The mean and median of a roughly symmetric distribution are close together.

If the distribution is exactly symmetric, the mean and median are exactly the same.

In a skewed distribution, the mean is usually farther out in the long tail than is the median.

- **Examining the Distribution of a Quantitative Variable**
- A graph is used to help us understand the data. After you make a graph, always ask, “What do I see?”

How to Examine the Distribution of a Quantitative Variable

In any graph, look for the **overall pattern** and for striking **departures** from that pattern.

Describe the overall pattern of a distribution by its:

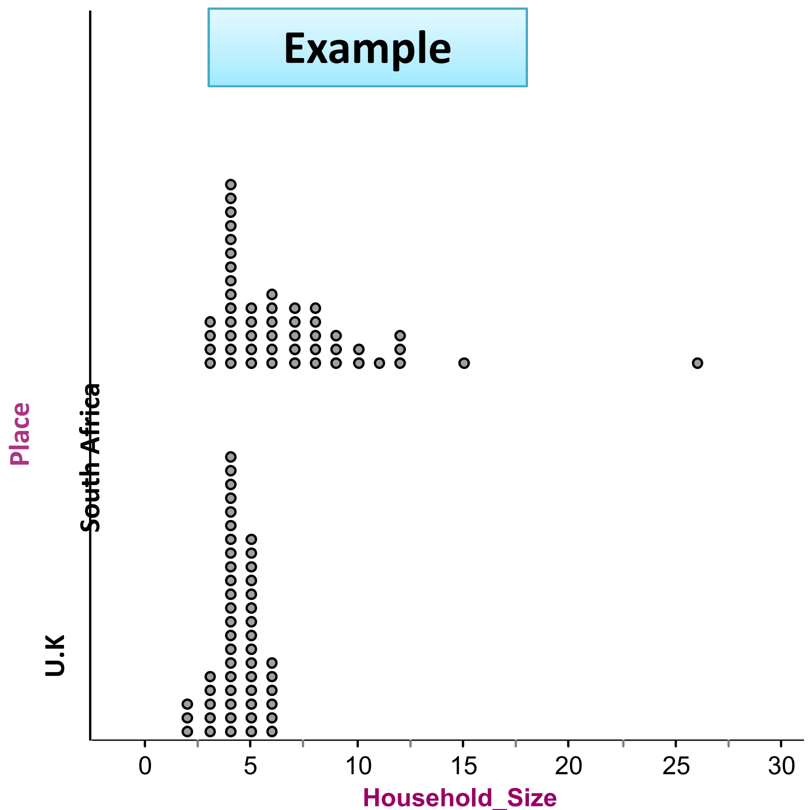
- **Shape**
- **Center**
- **Spread**

Don’t forget your SOCS!

Note individual values that fall outside the overall pattern. These departures are called **outliers**.

- **Comparing Distributions**
  - Some of the most interesting statistics questions involve comparing two or more groups.
  - Always discuss **shape**, **center**, **spread**, and possible **outliers** whenever you compare distributions of a quantitative variable.

Example



Compare the distributions of household size for these two countries. Don't forget your **SOCS**!