Unit 4: Measures of Center

PREREQUISITES

Students should be able to identify whether distributions are roughly symmetric or skewed given a histogram (Unit 3). The only mathematics prerequisite is knowledge of basic arithmetic operations (ordering, addition, division) needed to calculate the mean and median. Briefly introduce summation notation, $\sum x$, if that notation is new to students.

ADDITIONAL TOPIC COVERAGE

Additional coverage of measures of center can be found in *The Basic Practice of Statistics*, Chapter 2, Describing Distributions with Numbers.

ACTIVITY DESCRIPTION

The purpose of this activity is to help students learn how to assess the relationship between the mean and median based on the shape of the distribution. Students work with the Stemplots interactive from the Interactive Tools menu. The Stemplots interactive generates data and then organizes it into stemplots. Students use information from the graphic display to guess which is larger, the mean or the median. Then they calculate the mean and median. The interactive allows them to check their answers.

MATERIALS

Students will need access to the Stemplots tool from the Interactive Tool's menu online.

THE VIDEO SOLUTIONS

1. The variable is the weekly wages for Americans, separated by gender.

2. The men's distribution is skewed to the right.

3. The medians of the two distributions differ. Median measures the 50-50 point. The median for men's wages was larger than the median for women's wages.

4. A few very large incomes inflate the mean of a group of incomes. Hence, these very large incomes would pull the mean up.

5. A few very large incomes have no effect on the median.

UNIT ACTIVITY SOLUTIONS

1. a. Sample answer (answers differ since the data are randomly generated): For the stemplot below, the mean should be smaller than the median. The few really small test scores should pull the mean down.

1	6
2	
3	56
4	
5	
6	12
7	4
8	235
9	5

b. There were 10 test grades: (10 + 1)/2 = 5.5. So, the median = (62 + 74)/2 = 68. To find the mean, the sum of the test scores is 629. So, the mean = (629)/10 = 62.9. The median is larger than the mean.

2. a. Sample answer: The mean and median should be fairly close. However, the graph is somewhat skewed to the left, so the mean should be lower than the median.

3	26
4	447
5	1258
6	6

b. median = (47 + 51)/2 = 49; mean = 485/10 = 48.5. The median is larger than mean.

3. a. Sample answer:

1	389
2	3
3	
4	2
5	268
6	1
7	59
8	88
9	
10	
11	
12	
13	256
14	3

b. The mean should be larger because there are 4 extremely large test scores in comparison to the other test scores. Those large scores will pull the mean up but not affect the median.

c. median = 61; mean = 71.6. The mean was larger than the median.

4. a. Sample answer:

5	19
6	116
7	256
8	33445
9	57
10	38

b. The mean should be somewhat smaller than the median because the distribution is skewed to the left.

c. median = 83 and mean = 79. The median was larger than the mean.

5. a. The plot has two peaks, one in the 30s and the other in the 60s.

3	2345555555666666666667788999
4	0369
5	08
6	012334445556677799999999999
7	2233444789
8	0

b. median = 61.5; mean ≈ 53.3; there are two modes, one at 36 and the other at 69.

c. The median locates the upper peak, but does nothing to summarize the location of the lower peak. The mean is located where there is little data, and is not close to identifying the location of either of the peaks where there is a lot of data. Using the two modes gives the locations of the two peaks. So, in this case, the modes would be the best choice to describe the location of these data.

6. a. The plot appears roughly symmetric, with a single peak.

6	055
7	0022244445555566668888
8	0055
9	0

b. median = 75; mean = 75.1; mode = 75.

c. In this case, it doesn't really matter which of the three numeric descriptors for the center you choose. They are all about the same.

EXERCISE SOLUTIONS

1. a. Either a histogram or a stemplot would be a good choice of graphic display.
Sample answer based on stemplot below: The distribution is skewed to the right. There is a gap in the 80 thousands, which makes the $90,000 salary appear to be an outlier.

4	89
5	00134569
6	345689
7	678
8	
9	0

b. (59 + 63)/2 = 61; median starting salary is $61,000.

c. (1241/20) = 62.05; mean starting salary is $62,050.

d. The mode is 50 thousand.

e. The mean starting salary is higher than the median. That's largely due to the $90,000 outlier but also due to the shape of the data, which is skewed to the right. The mode does not do a good job in measuring the center or location of these data.

2. The median was $256,900 and the mean is $295,300. The mean is inflated because of a few extremely expensive houses, houses with prices in the millions.

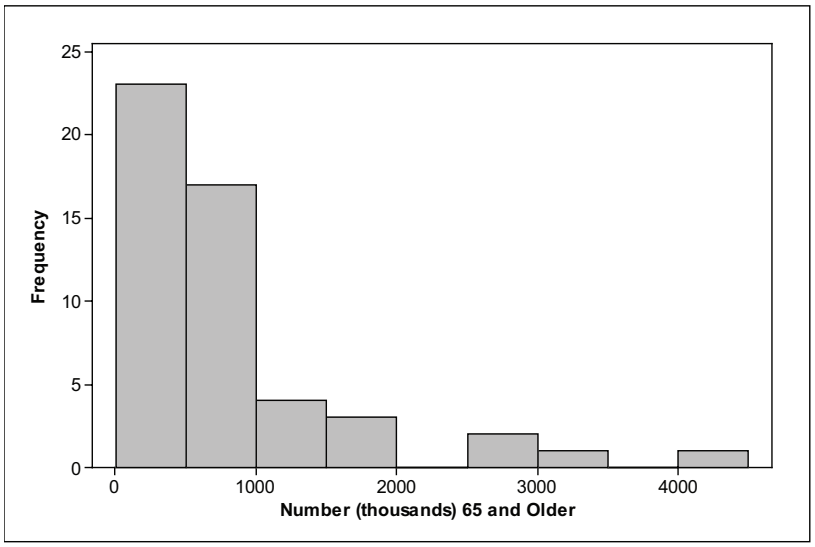
3. a. Mean for all 10 years: (13 + 23 + 26 + 16 + 33 + 61 + 28 + 39 + 14 + 8)/10 = 261/10 = 26.1.
Mean (excluding 61): (13 + 23 + 26 + 16 + 33 + 28 + 39 + 14 + 8)/9 = 200/9 ≈22.22.
Omitting his record year lowers the mean by 3.88.

b. All 10 years: ordered data 8 13 14 16 23 26 28 33 39 61; median = (23 + 26)/2 = 24.5.
Excluding 61: ordered data 8 13 14 16 23 26 28 33 39; median = 23.
Omitting his record year lowers the median by 1.5. Hence, the median is less affected by the record number of home runs than the mean.

c. Sample answer: The mean overstates Maris's usual performance because of the influence of the outlier. But the median doesn't point to the great achievement of his career. Perhaps we

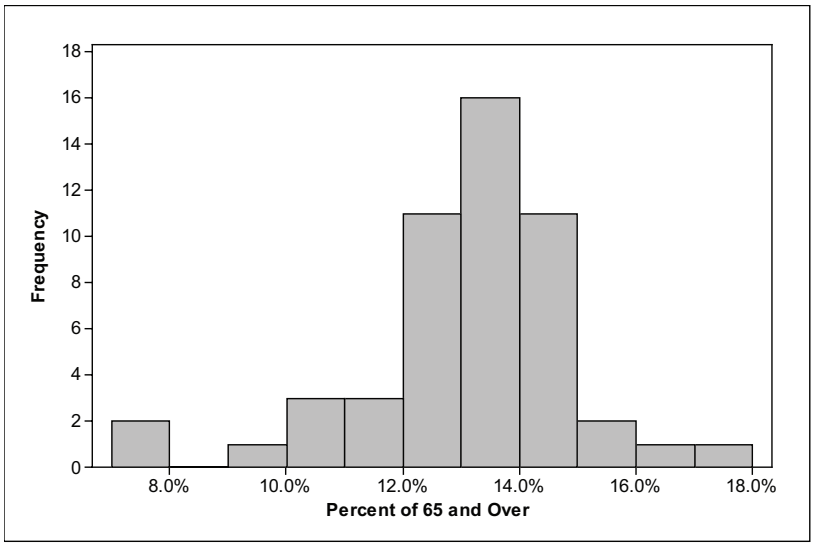
should say “Maris hit 61 home runs in 1961, and averaged about 22 home runs a year in his other 9 years in the American League.”

4. a. The mean is 784,000 and the median is 534,000. Note from the histogram (shown below) that these data are skewed to the right with some possible outliers.



The skewed pattern and outliers inflate the mean. The better choice for describing the location of the 65 and older data is the median.

b. Because the histogram is roughly symmetric, the mean and median should be close in value. The mean is about 13.14% and the median is 13.51%. (The mean is slightly less because of the two percentages below 8%.) In this case, either the mean or median would be appropriate. However, from looking at the histogram below, the median looks slightly more central than the mean.



REVIEW QUESTIONS SOLUTIONS

1. a.

5	66
5	
6	0
6	2
6	4
6	
6	8888
7	000000
7	2222222
7	44444
7	66
7	88

b. Mean ≈ 70.1 beats/min; median = 72 beats/min; mode = 72 beats/min.

c. Sample answer: The median of 72 beats/min best describes a “typical” pulse rate for this man. In addition, the mode is also 72 beats/minute. (The mode is the man’s most frequent pulse rate.) There are a few days when the man’s pulse rate is very low. These low values tend to pull the mean down.

2. a. Approximatly 56 of the fish had mercury levels below 0.30 μg/g.

b. Approximately 27 of the fish from the sample had mercury levels at or above 0.30 μg/g. Hence, around 32.5% of the fish in the sample had levels of mercury concentration above the EPA guidelines.

c. Because the data are skewed to the right, the few high mercury concentration values in the tail will inflate the mean but not affect the median. Hence, the mean mercury concentration will be larger than the median mercury concentration.

3. a. To compute the mean, sum the data and divide by 25: $\bar{x}$ = 1103/25 = 44.12.

To compute the median, order the data from smallest to largest. Select the (25 + 1)/2, or 13th data value from the ordered list:

28	35	37	37	38	38	40	40	42	43	43	44	45
45	46	46	46	47	47	49	49	51	54	55	58	

The median is 45 fries in a bag.

b. A stemplot appears below. We've chosen one that divides the stem into increments of 5.

2	8
3	
3	57788
4	002334
4	556667799
5	14
5	58

c. In this case, the choice between the mean and median is a matter of judgment. The difference between the two is less than one fry per bag. Some students may prefer the median because of the potential outlier of 28, which drags the mean down slightly.