Math 1230, Test #1, September 27, 2012

Part 1: True/False Questions (3 points each):

Instructions: Circle the appropriate answer.

1. True__False__ If the median is closer to the third quartile than to the first quartile, the data set is likely to be skewed to the right. It is likely to be skewed to the right.

2. True__False__ An observational study in which subjects are followed to observe future outcomes is called a retrospective study. It would be a prospective study.

3. True__False__ The sample interquartile range is more easily influenced by outliers than is the sample range. The IQR is outlier-resistant but the range is not.

4. True__False__ The sampling scheme in which every possible sample of the desired size is equally likely to be chosen is called simple random sampling.

5. True__False__ When a data set is skewed to the right, the mean will typically be smaller than the median. For right-skewed data, the mean will exceed the median.

Part 2: Problems:

Instructions: SHOW YOUR WORK to receive full credit.

6. (12 points) The values listed below are the game by game attendance totals for the Villanova field hockey team in 2011. Make an appropriate stem-and-leaf plot. Using statistical terminology, describe the shape of the distribution. Be sure to touch on all the areas that we discussed in class.

| 72 | 20 | 80 | 92 | 40 | 98 | 137 | 75 | 207 | 171 |
| 64 | 197 | 247 | 132 | 213 | 206 | 193 | 123 | 1307 |

```
Stem | Leaves

0    | 72 75 69
1    | 25 37 71 97 72 93 23
2    | 01 07 97 13 06
3    | 50
4    | 89 08
5    | 6
6    | 7
7    | 8
8    | 9
9    | 10
10   | 11
12   | 13
13   | 07
```

Unimodal

Skewed to the right

Extreme outlier at 1307

Gap from 489 to 1307
7. (23 points) The box plots given below are for runs scored and runs batted in for all full-time players in major league baseball as of early September, 2012.

(a) Estimate the median and the IQR for each data set. Label your answers clearly.

**Runs Scored**

- **Median:** \( \text{median} \approx 65 \)
- **IQR:** \( 78 - 58 = 20 \)

**RBI**

- **Median:** \( \text{median} \approx 65 \)
- **IQR:** \( 79 - 52 = 27 \)

(b) Compare the two data sets in terms of where they are centered.

The two data sets are very similar in terms of center. For example, the medians are about the same.

(c) Compare the two data sets in terms of spread.

The runs batted in values are more spread out. For example, the IQR for runs batted in is bigger.

(d) Using statistical terminology, describe the shape of each data set. Give a separate answer for each data set.

**Runs Scored:** Roughly symmetric, with an outlier at about 117.

**RBI:** Roughly symmetric, with an outlier at about 120.

There is a hint of right skew since the right tail is longer than the left.
8. (14 points) Scores on a certain IQ test are normal with mean 100 and standard deviation 15.

(a) What percentage of scores exceed 118?

\[ z = \frac{118 - 100}{15} = 1.2 \]

\[ 1 - 0.8849 = 0.1151 \]

\[ 11.51\% \]

(b) Find the middle 80% of test scores.

\[ z = -1.28; \quad x = 100 + 15(-1.28) = 80.8 \]

\[ z = 1.28; \quad x = 100 + 15(1.28) = 119.2 \]

The middle 80% is between 80.8 and 119.2.

9. (14 points) There are 30 different major league baseball teams, each of which has an active roster of 25 players. Given each sampling method, briefly explain how you might use the method to draw a sample of 150 major league baseball players.

(a) Cluster sampling

Use teams as clusters. Select 6 teams at random and put each player from a selected team into the sample. \( (6 \times 25 = 150) \)

(b) Stratified random sampling

Use teams as strata. Select 5 players at random from each team. \( (5 \times 30 = 150) \).
10. (8 points) An incoming freshman took placement exams both in French and in German. She scored 82 in French and 86 in German. The mean score on the French test was 72, and the standard deviation was 8. The mean score on the German test was 68, and the standard deviation was 12. Which test score is more impressive? Briefly explain your answer.

\[
\text{French: } t = \frac{82 - 72}{8} = 1.25 \\
\text{German: } t = \frac{86 - 68}{12} = 1.50
\]

Since the t score for her score on the German test is higher, her test score in German is more impressive.

11. (14 points) Do students learn statistics better in a distance learning course or in a regular course? One hundred students have volunteered to participate in an experiment to decide.

(a) Briefly describe an appropriate experiment that does not use blocking.

Randomly assign 50 students to the distance learning course and 50 to the regular course. Compare scores on a final exam that is common to both groups.

(b) Briefly describe an appropriate experiment that does use blocking. What are the blocks?

Divide the volunteers into a high-GPA group and a low-GPA group, each of size 50. Within each group, randomly assign 25 students to the distance learning course and 25 to the regular course.

The two blocks are (1) the high-GPA group and (2) the low-GPA group.