

**Math 214**  
**REVIEW SHEET EXAM #1**  
**Exam: Wednesday March, 2007**

**THOUGHT QUESTIONS:**

1. Suppose you are interested in determining if women are safer drivers than men in New York. Can you go to the Dept. of Motor Vehicles and simply analyze the number of accidents involving men and women drivers to fairly answer this question? Why or why not?
2. BIGOLDCIGARETTES INC. has hired a lobbying firm to help them stop any new anti-smoking laws. These lobbyists hire a statistician who finds out that the average life-span of Americans has almost DOUBLED since the wide-spread introduction of cigarettes a century ago. The lobbyists decide to start an ad campaign pushing the health benefits of smoking. What's wrong with this argument?
3. If the third quartile for scores on a test is 68, what percent of scores are below 68? What is the lowest possible percentile rank of someone with a score of 72?
4. Is it safe for you (height 5 feet) to wade across a puddle with an average depth of 2 feet?
5. Can a test score with a percentile rank of 58 have a negative  $z$ -score? Explain.
6. Freda Friendly, a statistician at Bio-Med Inc, discovers that she has made a mistake in her calculations of fruit-fly weight statistics. In a sample of 10,000 fruit-flies, she forgot to include information about the largest fly. Which of the following statistics is LEAST likely to change when she redoes the analysis:  
a) mean   b) median   c) range   d) standard deviation  
Explain.

**Problem 1 :**

The following comes from an online news source:

TUESDAY, Feb. 27 (HealthDay News) – Aerobic exercise conducted over the long term may reduce a woman's risk of breast cancer, U.S. research suggests.

Reporting in the Feb. 26 issue of the Archives of Internal Medicine, researchers at the University of Southern California, Los Angeles, studied more than 110,000 women, aged 20 to 79, who took part in the California Teachers Study.

At the start of the study, the women were asked about their participation in moderate exercise (such as brisk walking, golf or volleyball) and strenuous

exercise (such as swimming laps, aerobics, and running) from high school up until their current age and within the previous three years.

The study authors also collected information about the women's breast cancer risk factors, including race, family history and use of hormone replacement therapy.

Between the start of the study in 1995-96 and 2002, 2,649 of the women were diagnosed with invasive breast cancer and 593 were diagnosed with in situ (noninvasive) breast cancer.

Women with a long-term history of doing more than five hours per week of strenuous exercise were 20 percent less likely to develop invasive breast cancer and 31 percent less likely to develop in situ breast cancer than those who did less than 30 minutes of strenuous exercise per week, the researchers found.

Based on the reading answer the following:

1. What is an *experimental unit* in the study?
2. What is the sample size?
3. What were the statistics collected?
4. Categorize the statistics as numerical or categorical.
5. What was the target population?
6. Do you think there would be any bias in the sample? Explain.

**Problem 2 :**

A website asks its users for the following information before allowing a download of its software

- The users age
- The users zip code
- The users income
- The users occupation
- The users years of education
- The users time using competitive products

Categorize these 6 variables as either categorical, discrete numeric or continuous numeric.

**Problem 3 :**

The following stem-and-leaf diagram shows the cholesterol levels of for a sample of 62 subjects in the Framingham Heart Study.

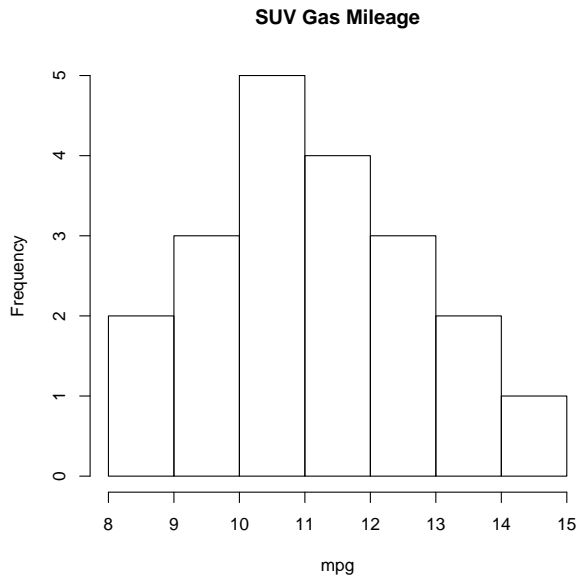
The decimal point is 1 digit(s) to the right of the |

```
16 | 7
18 | 428
20 | 020125678
22 | 05560000122244668
24 | 03678444668
26 | 34777800288
28 | 35
30 | 008
32 | 746
34 | 3
36 |
38 | 3
```

1. Describe the shape of the distribution.
2. Based on the stem and leaf diagram, find the values of the minimum, median, and maximum *Use the correct units.*
3. Based on the stem and leaf diagram, would you expect the mean to be larger or lower than the median? Explain why?
4. Use the stem and leaf diagram to construct a relative frequency polygon. Draw a rough sketch of this. Is the distribution *unimodal*?

**Problem 4 :**

A data set on gas mileage for 20 SUVs produces the following histogram



1. What percentage of the SUVs have gas mileage less than 10 mpg (None have exactly 10)
2. Find the median,  $Q_1$  and  $Q_3$ .
3. Construct the corresponding boxplot.

**Problem 5 :**

Ten of the Yankees' salaries (in hundred thousands) are

3 4 7 10 27 30 35 60 160 214

1. Compute the sample mean
2. Compute the 20% trimmed mean
3. Compare these values to the median.
4. Compute the range.
5. Compute the IQR.

**Problem 6 :**

Find the sample standard deviation of these numbers

1, 2, 3, 4, 5, 6, 7

**Problem 7 :**

The 32 salaries of the 2004 Yankees had a mean of  $\$ = 6,105,797$  and a standard deviation of  $\$s = 6,102,199$ .

1. Find the  $z$ -score of Alex Rodriguez's salary of  $\$21,726,881$ .
2. Bubba Crosby had a  $z$ -score of  $-0.95120$ . What was his salary?

**Problem 8 :**

For the Yankees' salaries the following is true

- 81% are between  $-1s$  and  $1s$
  - 93% are between  $-2s$  and  $2s$
  - 100% are between  $-3s$  and  $3s$
1. Are these consistent with a "bell-shaped" data set? Explain.
  2. Are these consistent with Chebyshev's theorem? Explain.

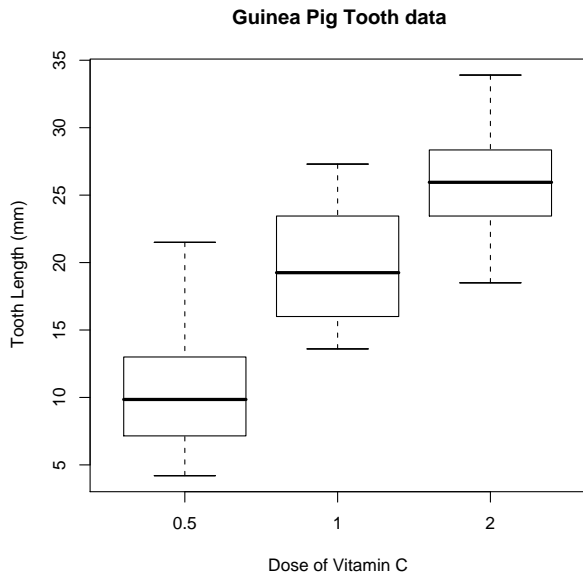
**Problem 9 :**

For each of the following descriptions of the distribution of a univariate data set, draw a rough sketch of a HISTOGRAM and a BOXPLOT:

- symmetric
- skewed left and long-tailed
- symmetric and short-tailed
- bimodal

**Problem 10 :**

The following data comes from a study of the effect injecting vitamin C on the length of a guinea pig's incisors. The data set covers 60 guinea pigs broken up into 3 categories based on vitamin C dosages administered.



Based on the boxplots, answer the following questions:

1. What is the median tooth length of the group with the smallest median tooth length?
2. Estimate the range and the IQR of each of the three groups.
3. True or false: 25% of the pigs given 2 mg of Vitamin C had tooth lengths *smaller* than the 75th percentile of pigs given 1 mg of vitamin C.

**Problem 11 :**

A company collects data lifetime of an electrical component and the temperature of the environment in which it operates.

```
> temp (Celsius Degrees)
  50 100 150 200 250 300
> life (Hours)
  875 884 762 424 365 128
```

1. Draw a scatterplot of the data
2. Compute the Pearson correlation coefficient
3. Compute the Spearman rank correlation coefficient

**Problem 12 :**

A student physics lab has the following data for current versus voltage:

current (mA)	voltage (V)
100	15
200	16
200	16.5
300	16.5

1. Compute the coefficients of the least squares regression line with voltage as the response variable and current as the predictor variable.
2. Find the residual of the data point (100, 15).