Problem Set #2

Econ 103

Part I – Problems from the Textbook

Chapter 2: 1, 3, 5, 7, 13, 15, 17 [in part (b) skip MAD and MSD], 21 23, 27, 35.

Part II – Additional Problems

1. For each variable indicate whether it is nominal, ordinal, or numeric.

    (a) Grade of meat: prime, choice, good.

    (b) Type of house: split-level, ranch, colonial, other.

    (c) Income

2. A drive-time radio show frequently holds call-in polls during the evening rush hour. Explain in no more than two sentences why such polls are likely to be biased.

3. Which of these studies are based on experimental data? Which are based on observational data?

    (a) A biologist examines fish in a river to determine the proportion that show signs of disease due to pollutants poured into the river upstream.

    (b) In a pilot phase of a fund-raising campaign, a university randomly contacts half of a group of alumni by phone and the other half by a personal letter to determine which method results in higher contributions.

    (c) To analyze possible problems from the by-products of gas combustion, people with with respiratory problems are matched by age and sex to people without respiratory problems and then asked whether or not they cook on a gas stove.

    (d) An industrial pump manufacturer monitors warranty claims and surveys customers to assess the failure rate of its pumps.

4. An emergency room institutes a new screening procedure to identify people suffering from life-threatening heart problems so that treatment can be initiated quickly. The procedure is credited with saving lives because in the first year after its initiation, there

is a lower death rate due to heart failure compared to the previous year among patients seen in the emergency room. Do you agree? Explain.

5. Suppose that x_i is measured in centimeters and y_i is measured in feet. What are the units of the following quantities?

   (a) Interquartile Range of x

   (b) Covariance between x and y

   (c) Correlation between x and y

   (d) Skewness of x

   (e) Variance of y

6. The *mean deviation* is a measure of dispersion that we did not cover in class. It is defined as follows:
$$MD = \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{x}|$$

   (a) Explain why this formula averages the absolute value of deviations from the mean rather than the deviations themselves.

   (b) Which would you expect to be more sensitive to outliers: the mean deviation or the variance? Explain.

7. Consider a dataset $x_1, \ldots, x_n$. Suppose I multiply each observation by a constant d and then add another constant c, so that x_i is replaced by $c + dx_i$.

   (a) How does this change the sample mean? Prove your answer.

   (b) How does this change the sample variance? Prove your answer.

   (c) How does this change the sample standard deviation? Prove your answer.

   (d) How does this change the sample z-scores? Prove your answer.