Homework questions, week 1

Econ 103

1 Daily Homework questions

The questions in bold font are due on **Thursday 26th May**. You do not need to hand in the questions that are not in bold, though these will be useful to complete for your own understanding.

Lecture 1

Textbook questions:
Chapter 1: 3, 5, 7, **9, 13**
Chapter 2: 1

Additional questions:

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.

Lecture 2

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

Additional questions:

1. **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

2. 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.

3. 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.

Lecture 3 - Regression

Textbook questions:
Chapter 11: **1, 3**
Chapter 15: 1(a)

Additional questions:

1. **What value of a minimizes $\sum_{i=1}^{n}(y_i - a)^2$? Prove your answer.**

2. Let
$$z_{x_i} = \frac{x_i - \bar{x}}{s_x}, \text{ and } z_{y_i} = \frac{y_i - \bar{y}}{s_y}.$$
Show that if we carry out a regression with z_{y_i} in place of y and z_{x_i} in place of x, the intercept a will equal zero while the slope b will equal r, the sample correlation.

3. Let $\hat{y}$ denote our prediction of y from a linear regression model: $\hat{y} = a + bx$ and let r be the correlation coefficient between x and y.
   (a) Express b in terms of s_{xy} and s_x.
   (b) Express a in terms of b and the sample means of x and y.
   (c) Express r in terms of the s_{xy}, s_x and s_y.
   (d) Show that
   $$\frac{\hat{y} - \bar{y}}{s_y} = r\left(\frac{x - \bar{x}}{s_x}\right)$$
   (e) (3 points) Using the equation derived in (d), briefly explain "regression to the mean."

2 R Tutorials

You should complete TryR levels 1-6 and R Tutorial #1 by **Monday 30th May**. These will normally be due on Thursdays (and there will be some R material on the quizzes), but I will give you a bit more time for the first week to make sure everyone is set up on R.

TryR can be accessed at `tryr.codeschool.com`.

R tutorials will be posted in Piazza, with solution code.