Extension Problems

Econ 103

Spring 2018

About This Document

Extension problems are designed to give you a deeper understanding of the lecture material and challenge you to apply what you have learned in new settings. Extension problems should only be attempted *after* you have completed the corresponding review problems. As an extra incentive to keep up with the course material, each exam of the semester will contain at least one problem taken *verbatim* from the extension problems. We will circulate solutions to the relevant extension problems the weekend before each exam. You are also welcome to discuss them with the instructor, your RI, and your fellow students at any point.

Lecture #1

1. A long time ago, the graduate school at a famous university admitted 4000 of their 8000 male applicants versus 1500 of their 4500 female applicants.

   (a) Calculate the difference in admission rates between men and women. What does your calculation suggest?

   (b) To get a better sense of the situation, some researchers broke these data down by area of study. Here is what they found:

	Men		Women	
	# Applicants	# Admitted	# Applicants	# Admitted
Arts	2000	400	3600	900
Sciences	6000	3600	900	600
Totals	8000	4000	4500	1500

   Calculate the difference in admissions rates for men and women studying Arts. Do the same for Sciences.

   (c) Compare your results from part (a) to part (b). Explain the discrepancy using what you know about observational studies.

Lecture #2

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. Let m be a constant and $x_1, \ldots, x_n$ be an observed dataset.

   (a) Show that $\displaystyle\sum_{i=1}^{n}(x_i - m)^2 = \sum_{i=1}^{n} x_i^2 - 2m\sum_{i=1}^{n} x_i + nm^2$.

   (b) Using the preceding part, show that $\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$.

Lecture #3

4. 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 #4

5. Define the z-scores
$$w_i = \frac{x_i - \bar{x}}{s_x}, \quad \text{and} \quad z_i = \frac{y_i - \bar{y}}{s_y}.$$
   Show that if we carry out a regression with z_i in place of y_i and w_i in place of x_i, the intercept a^* will be zero while the slope b^* will be r_{xy}, the correlation between x and y.

6. This question concerns a phenomenon called *regression to the mean.* Before attempting this problem, read Chapter 17 of *Thinking Fast and Slow* by Kahneman.

   (a) Lothario, an unscrupulous economics major, runs the following scam. After the first midterm of Econ 103 he seeks out the students who did extremely poorly and offers to sell them "statistics pills." He promises that if they take the pills before the second midterm, their scores will improve. The pills are, in fact, M&Ms and don't actually improve one's performance on statistics exams. The overwhelming majority of Lothario's former customers, however, swear that the pills really work: their scores improved on the second midterm. What's your explanation?

   (b) 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. Show that

   $$\frac{\hat{y} - \bar{y}}{s_y} = r\left(\frac{x - \bar{x}}{s_x}\right)$$

   (c) Using the equation derived in (b), briefly explain"regression to the mean."

No extension problems for Lecture #5

Lecture #6