

**SUNY-Stony Brook. Economics Department**  
**Economics 323: Fall 2011**  
**Professor: Hugo Benítez-Silva**

Problem Set 4: Due Thursday November 17, 2011.  
Suggested Solutions

**1.Shorter Questions. Remember to justify your answers.**

- a) What are the differences between cross-sectional data and time series data? Give an example of each of them. Use examples different from the ones in the notes.

**This is straight from the notes, cross-sectional data is for different units of observation in a point in time, and time-series follow a unit of observation over a certain period of time.**

- b) Give examples of at least two variables (with some economic meaning) for which using the mode would be better than using the mean or the median to describe the central tendencies of a distribution in a particular population or sample.

**It is not difficult to think of many economic variables nicely described by the mode instead of just the mean or the median. We gave the example of the number of children, also for example the number of cars people have, or the number of computers. The number of cans of Coke people are likely to buy in a trip to the store...etc. Many times the mode is a complement to the mean or other measures not really a substitute.**

2. **The truth will prevail.** (Again, justify your answers).

a) In any given sample, the median is never larger than the mean, unless the distribution is symmetric. **True or False.**

**False. It is perfectly possible that in a given sample the median could be larger than the mean, it means the distribution is not symmetric, in fact it means the distribution is skewed to the left, so some observations have very low values that push the mean towards zero compared to the median.**

b) The sign of the correlation coefficient provides us with information regarding how good a line fits the data points we are studying. **True or False.**

**False. The sign provides information about whether there is a positive or negative relationship between the variables we are studying, in other words, gives information about the slope of that line.**

c) The range is always larger than the inter-quartile range. **True or False**

**In general this is True, assuming that both of them are well defined with  $n \geq 4$ .**

**3. Longer but not more difficult. Show your work: No justification, no credit.**

Suppose you are considering a job offer from a company in London. You are interested in finding out how much employees of the firm spend on housing in order to judge the cost of living in the city. This information will help you decide whether the firm's salary offer is attractive.

There are currently four employees with the firm paying the following rents per month:

<u>Employee</u>	<u>Rent</u> (in British pounds)
Dave	700
Lil	300
Steve	100
Clare	100

Remember, 1 British pound is equal to 1.9 U.S\$

- a) Calculate the population mean rent for employees

**300**

- b) Suppose that you can only talk with 2 employees. Write down the sample means for all possible pairs of employees that you could have asked about rent.

**Fairly mechanical: 500, 400, 400, 200, 200, 100**

- c) Do the same as b) but assuming you could talk with 3 employees.

**Fairly mechanical too: 366.66, 366.66, 300, 166.66**

- d) From the results in b) and c) would you say that the sample mean is a biased or unbiased estimator of the population mean? Why?

**If you look at the sample means individually it looks like we are kind of far from the population mean, but if we take the numbers in b) and c) as data, and compute the mean, or expected value of each of them we get in both cases 300. This means that the sample mean is an unbiased estimator of the population mean. This is an important property of the sample mean.**

- e) Does being able to talk with more employees ensure that you have more accurate information?

**It seems to help if we look at the sample means individually, but it does not help once we compute the expected value of those sample means.**

- f) Calculate the variance of the sample mean for  $n=2$  and  $n=3$  using the results of b) and c) (Note that each sample mean counts as one observation). What do you conclude from these results about being able to talk with more employees?

**The variance is computed following the formulas we mentioned in class, by summing over the squared deviations and dividing by  $(n-1)$ . Notice that here  $n$ , is the number of sample means (pairs or triplets) when  $n=2$  or  $n=3$  respectively.**

**When  $n=2$  the sample variance is 24,000, when  $n=3$  it is 8,888.**

**We can conclude that although having a larger sample does not help to get a good estimate of the mean, it does help to reduce the variance considerably.**

**4. Consider the following five observations on Savings and Income for a given population.**

$\underline{Y}_i$	$\underline{X}_i$
0.0	1.9
0.9	12.4
0.4	6.4
1.2	7.0
0.3	7.0

a) Assuming that  $\underline{Y}_i$  are observations (in thousands of dollars) on savings, and  $\underline{X}_i$  observations on income (also in thousands of dollars), compute by hand and showing all the steps you need to calculate, the regression coefficients  $\beta_0$ , and  $\beta_1$  in the regression

$$\underline{Y}_i = \beta_0 + \beta_1 \underline{X}_i + u_i$$

**Here you just had to follow the formulas in the notes**

$$\beta_0 = -0.0386 \text{ and } \beta_1 = 0.0863$$

b) Write an equation for the fitted regression line considering the regression coefficients you found.

**In fact this was in the notes,  $\hat{Y} = -0.0383 + 0.0863 X$**

c) What is the marginal propensity to save in this model? What happens to saving if income increases by \$3500 dollars?

**Marginal propensity to save is 0.0863 thousands (or \$86.3) per 1000 thousand dollar increase in income.**

**If income increases by \$3500, savings go up by \$302.05**