Review Questions

Econ 103

Spring 2018

About This Document

Lecture #1 – Introduction

1. Define the following terms and give a simple example: *population*, *sample*, *sample size*.
2. Explain the distinction between a *parameter* and a *statistic*.
3. Briefly compare and contrast *sampling* and *non-sampling* error.
4. Define a *simple random sample*. Does it help us to address sampling error, non-sampling error, both, or neither?
5. A drive-time radio show frequently holds call-in polls during the evening rush hour. Do you expect that results based on such a poll will be biased? Why?
6. Dylan carried polled a random sample of 100 college students. In total 20 of them said that they approved of President Trump. Calculate the margin of error for this poll.
7. Define the term *confounder* and give an example.
8. What is a randomized, double-blind experiment? In what sense is it a "gold standard?"
9. Indicate whether each of the following involves experimental or 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.

10. Based on information from an observational dataset, Amy finds that students who attend an SAT prep class score, on average, 100 points better on the exam than students who do not. In this example, what would be required for a variable to *confound* the relationship between SAT prep classes and exam performance? What are some possible confounders?

11. WRITE SOME EXTENSIONS QUESTIONS ABOUT NON-RESPONSE BIAS, POST-STRATIFICATION, ETC

Lecture #2 – Summary Statistics I

12. 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

13. Explain the difference between a histogram and a barchart.

14. Define *oversmoothing* and *undersmoothing*.

15. What is an *outlier*?

16. Write down the formula for the sample mean. What does it measure? Compare and contrast it with the sample median.

17. Define the *range* and *interquartile range* of a dataset. What do they measure? How do they differ?

18. What is a boxplot? What information does it depict?

19. Write down the formula for variance and standard deviation. What do these measure? How do they differ?

20. Suppose that x_i is measured in inches. What are the units of the following quantities?
    (a) Sample mean of x
    (b) Range of x
    (c) Interquartile Range of x
    (d) Variance of x

(e) Standard deviation of x

Lecture #3 – Summary Statistics II

21. Evaluate the following sums:

    (a) $\displaystyle\sum_{n=1}^{3} n^2$

    (b) $\displaystyle\sum_{n=1}^{3} 2^n$

    (c) $\displaystyle\sum_{n=1}^{3} x^n$

22. Evaluate the following sums:

    (a) $\displaystyle\sum_{k=0}^{2}(2k+1)$

    (b) $\displaystyle\sum_{k=0}^{3}(2k+1)$

    (c) $\displaystyle\sum_{k=0}^{4}(2k+1)$

23. Evaluate the following sums:

    (a) $\displaystyle\sum_{i=1}^{3}(i^2+i)$

    (b) $\displaystyle\sum_{n=-2}^{2}(n^2-4)$

    (c) $\displaystyle\sum_{n=100}^{102} n$

    (d) $\displaystyle\sum_{n=0}^{2}(n+100)$

24. Prove that $\displaystyle\sum_{i=m}^{n}(a_i+b_i) = \sum_{i=m}^{n} a_i + \sum_{i=m}^{n} b_i$.

25. Prove that if c is a constant then $\displaystyle\sum_{i=m}^{n} cx_i = c\sum_{i=m}^{n} x_i$.

26. Prove that if c is a constant then $\displaystyle\sum_{i=1}^{n} c = cn$.

27. Express each of the following using Σ notation:
    (a) $z_1 + z_2 + \cdots + z_{23}$
    (b) $x_1y_1 + x_2y_2 + \cdots + x_8y_8$
    (c) $(x_1 - y_1) + (x_2 - y_2) + \cdots + (x_m - y_m)$
    (d) $x_1^3 f_1 + x_2^3 f_2 + \cdots + x_9^3 f_9$

28. Prove that $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$.

29. Write down the formula for skewness. Why does this formula involve a cubic, and why do we divide by s^2?

30. How do we interpret the sign of skewness, and what is the "rule of thumb" that relates skewness to the mean and median?

31. NEED TO WRITE MORE! Continue from slide 9.

32. Suppose that x_i is measured in centimeters and y_i is measured in feet. What are the units of the following quantities?
    (a) Covariance between x and y
    (b) Correlation between x and y
    (c) Skewness of x

33. THE QUESTIONS FROM HERE DOWN SHOULD BE "EXTENSIONS"

34. 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.

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

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

37. 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.