

Statistics – Midterm Review

Directions: This review sheet is intended to help you prepare for the midterm exam.

Part One – Provide a brief definition of the following terms:

- mean, median and mode
- variance and standard deviation
- skewness
- kurtosis
- relative frequency distribution
- probability
- binomial distribution
- expected value
- probability density function
- cumulative distribution function
- the normal distribution and the standard normal distribution
- law of large numbers
- correlation and covariance

Part Two – The cross-tabulation below classifies US states by median household income and by household Gini coefficient (a measure of income inequality). Use the cross-tabulation to answer the questions that follow.

Gini coefficient	median household income		
	under \$35,000	\$35,000 to \$45,000	over \$45,000
0.43 and under	1	9	7
0.44 to 0.46	4	12	6
0.47 and over	5	5	2

- 1 Construct the relative frequency distribution of median household income for US states where the Gini coefficient is:
 - a. 0.43 and under
 - b. 0.44 to 0.46
 - c. 0.47 and over
- 2 Create a table that compares the three distributions. What does your table tell you about the relationship between income inequality and median income?

Part Three – Answer the following questions. When a question asks for an explanation, please provide a detailed explanation. When a question asks you to compute a statistic, please show your work.

- 1 **Explain** why we cannot take the average of nominal and ordinal level measures.
- 2 Suppose that a job seeker has a 10 percent probability of obtaining a job offer for each letter that he/she sends. Suppose further that the job seeker sends 10 letters. What is his/her probability of obtaining at least one job offer? **Show your work.**
- 3 Suppose that you have the following sample: {3,4,4,6,8}.
 - a. Estimate the mean and standard deviation. **Show your work.**
 - b. What is the standard error of your estimate of the mean?

- 4 What is sampling error? How does sampling error relate to the standard error of an estimate? **Why** does the standard error go to zero as the number of observations increases?
- 5 Suppose that you randomly draw a large number of values from a normal distribution with a mean of 5 and a standard deviation of 2. **Explain** why you would expect:
 - a. to draw values between 3 and 7 about 68 percent of the time.
 - b. to draw values between 1 and 9 about 95 percent of the time.
- 6 What is skewness? What is kurtosis? **Explain** why it is not appropriate to assume that your data is distributed normally when the data is skewed or contains “fat tails.”