Ecn 102 - Analysis of Economic Data
University of California - Davis
Instructor: John Parman

April 23, 2009

Midterm 1

You have until 10:20am to complete this exam. Please remember to put your name, section and ID number on both your scantron sheet and the exam. Fill in test form A on the scantron sheet. Answer all multiple choice questions on your scantron sheet. Choose the single best answer for each multiple choice question. Answer the long answer questions directly on the exam. Keep your answers complete but concise. For the long answer questions, you must show your work for full credit.

Name: **ID Number:** **Section:**

(POTENTIALLY) USEFUL FORMULAS

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$CV = \frac{s}{\bar{x}}$

$skew = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}(\frac{x_i-\bar{x}}{s})^3$

$kurt = \frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}(\frac{x_i-\bar{x}}{s})^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$

$\mu = E(X)$

$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$

$t = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$

$Pr[T_{n-1} > t_{\alpha,n-1}] = \alpha$

$Pr[|T_{n-1}| > t_{\frac{\alpha}{2},n-1}] = \alpha$

$\sum_{i=1}^{n} a = na$

$\sum_{i=1}^{n}(ax_i) = a\sum_{i=1}^{n} x_i$

$\sum_{i=1}^{n}(x_i + y_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$

$s^2 = \bar{x}(1-\bar{x})$ for proportions data

$t_{\alpha,n-1} = TINV(2\alpha, n-1)$

$Pr(|T_{n-1}| \geq |t^*|) = TDIST(|t^*|, n-1, 2)$

$Pr(T_{n-1} \geq t^*) = TDIST(t^*, n-1, 1)$

(POTENTIALLY) USEFUL EXCEL OUPUT

TINV(.01,499)=2.59	TINV(.01,399)=2.59	TINV(.01,299)=2.59
TINV(.02,499)=2.33	TINV(.02,399)=2.34	TINV(.02,299)=2.34
TINV(.025,499)=2.25	TINV(.025,399)=2.25	TINV(.025,299)=2.26
TINV(.05,499)=1.96	TINV(.05,399)=1.97	TINV(.05,299)=1.97
TINV(.10,499)=1.65	TINV(.10,399)=1.65	TINV(.10,299)=1.65
TINV(.20,499)=1.28	TINV(.20,399)=1.28	TINV(.20,299)=1.28

SECTION I: MULTIPLE CHOICE (55 points)

1. Suppose you have annual observations of obesity rates for the past fifty years and you want a graph of the data that shows the trend in obesity over time. The best graph for this purpose would be:
   (a) A bar chart.
   (b) A pie chart.
   (c) A line chart.
   (d) A histogram.
2. The probability of a Type II error occurring depends on:
   (a) The true population mean.
   (b) The significance level, α.
   (c) The sample size.
   (d) All of the above.
3. Which of the following statements is false?
   (a) The interquartile range is always less than or equal to the range.
   (b) The coefficient of variation will be smaller when the sample mean is larger, all other things held constant.
   (c) The mode is always less than or equal to the median.
   (d) The mean will increase if the largest observation is doubled.
4. The magnitude of the critical value for an upper one-tailed test at a 5% significance level will be:
   (a) Greater than the magnitude of the critical value for a two-tailed test using the same data at a 5% significance level.
   (b) Less than the magnitude of the critical value for a two-tailed test using the same data at a 5% significance level.
   (c) Equal to the magnitude of the critical value for a two-tailed test using the same data at a 5% significance level.
   (d) None of the above.
5. A dataset containing the monthly food expenditures for 10 different families for an entire year (120 observations in total) is an example of:
   (a) Time series data.
   (b) Cross-section data.
   (c) Panel data.
   (d) Categorical data.
6. If X_a and X_b are two different consistent, unbiased estimators for the population mean, we would prefer to use the one with:
   (a) The larger variance.
   (b) The smaller variance.
   (c) Whichever one gives us a larger value for the sample mean.
   (d) Whichever one gives us a smaller value for the sample mean.

7. Which of the following is not a measure of central tendency?
   (a) Interquartile range.
   (b) Sample midrange.
   (c) Median.
   (d) Mean.
8. Which of the following would decrease the likelihood of both Type I and Type II errors?
   (a) Increasing the significance level, α.
   (b) Decreasing the significance level, α.
   (c) Neither (a) nor (b) are correct.
   (d) Both (a) and (b) are correct.
9. Increasing the width of the bins for a histogram will tend to:
   (a) Increase the absolute frequency of observations in each bin.
   (b) Decrease the absolute frequency of observations in each bin.
   (c) Lead to a greater proportion of bins with no observations.
   (d) Shift the center of the distribution to the right.
10. Which of the following is not a random variable?
    (a) The sample mean.
    (b) The sample standard deviation.
    (c) The population mean.
    (d) All of the above are random variables.
11. We will reject the null hypothesis whenever:
    (a) $p < \alpha$.
    (b) $p > \alpha$.
    (c) $t^* > c$.
    (d) $t^* < c$.
12. Suppose that we are interested in the mean level of income, μ, for the entire adult population of the United States. We take a sample of college graduates and calculate the mean income for the sample, $\bar{x}$. Which of the following statements is most likely true about our sample?
    (a) The sample mean would provide a biased estimate of the population mean because $E(\bar{X}) > \mu$ in this case.
    (b) The sample mean would provide a biased estimate of the population mean because $E(\bar{X}) < \mu$ in this case.
    (c) The sample mean would provide an unbiased estimate of the population mean.
    (d) The expected value of the sample mean would be equal to the expected value of the population mean.
13. Which of the following types of data would be the most likely to be represented with a pie chart?
    (a) Categorical, cross-sectional data.
    (b) Continuous, cross-sectional data.
    (c) Continuous, time series data.
    (d) Continuous, panel data.

14. Suppose that we reject the null hypothesis that $\mu = 58$ for a particular random variable X. We used a significance level of 10% and had a positive value for our test statistic t^*. Which of the following statements is definitely true?

    (a) We would reject the null hypothesis that $\mu \leq 58$ at a significance level of 5%.
    (b) We would reject the null hypothesis that $\mu \geq 58$ at a significance level of 5%.
    (c) We would reject the null hypothesis that $\mu = 58$ at a signifiance level of 5%.
    (d) None of the above.

15. If a distribution is symmetric:

    (a) The mean will be equal to the median.
    (b) The 75th percentile will be exactly as far from the mean as the 25th percentile.
    (c) The value for the midrange will be equal to the mean.
    (d) All of the above.

16. On a line graph, the values on the horizontal axis typically correspond to _____ while the values on the horizontal axis of a histogram typically correspond to _____.

    (a) The realized values of the variable (x_i), the realized values of the variable (x_i).
    (b) The realized values of the variable (x_i), the values of the index for observations (i).
    (c) The values of the index for observations (i), the realized values of the variable (x_i).
    (d) The values of the index for observations (i), the values of the index for observations (i).

17. Suppose we are interested in using census data to estimate the mean level of education for the US population. If a 5% sample of the census and a 10% sample of the US census give us the same sample mean and the same sample variance, which will give us a narrower 95% confidence interval for the population mean?

    (a) The 5% sample.
    (b) The 10% sample.
    (c) The confidence intervals will be the same.
    (d) Not enough information.

18. A pollster reports that 60% of the public approves of President Obama's handling of the economy with a margin of error of 4%. What is the probability that the true percentage of the population that approves is greater than 64%?

    (a) 1%.
    (b) 2.5%.
    (c) 5%.
    (d) 10%.

19. Suppose that the absolute frequency of people with a wage equal to \$20 an hour in a sample of 200 people is 40. Which of the following is definitely true?

    (a) 20% of the population earns a wage of \$20 an hour.
    (b) The percentage of people in the sample earning \$40 an hour or less is greater than or equal to 20%.
    (c) The percentage of people in the sample earning \$10 an hour or more is less than or equal to 80%.
    (d) The percentage of people in the population earning \$20 an hour or less is greater than or equal to 20%.

20. If most apartment rents in Davis are between \$600 and \$800 but there is a long tail of more expensive apartments in the rent distribution, which of the following is most likely to be true?
    (a) The distribution of apartment rents is left skewed.
    (b) The mean apartment rent is less than the median apartment rent.
    (c) The value for the skewness of the rent distribution is positive.
    (d) The 50th percentile of the rent distribution is greater than the mean apartment rent.
21. Which of the following would be a way of getting a random sample of independent observations of Davis students?
    (a) Sample students on the basis of the last digit of their student ID number.
    (b) Choose a class at random and use the students in the class as the sample.
    (c) Choose all students born in March of 1989 as the sample.
    (d) All of the above.
22. The distribution of the sample mean depends on:
    (a) The mean of the population.
    (b) The variance of the population.
    (c) The sample size.
    (d) All of the above.
23. A 90% confidence interval will:
    (a) Always be centered at the population mean.
    (b) Contain the population mean with a probability of 10%.
    (c) Be wider if the sample variance is larger.
    (d) Be wider than a 95% confidence interval.
24. When testing the null hypothesis that $\mu = 40$ we get a p-value equal to .06. We can say that:
    (a) $\mu = 40$ at a 10% significance level.
    (b) $\mu > 40$ at a 5% significance level.
    (c) We reject the null hypothesis at the 5% significance level.
    (d) We fail to reject the null hypothesis at the 1% significance level.
25. Suppose that we have a sample of heights measured in centimeters. If we switch to measuring heights in meters:
    (a) The new sample mean will be 100 times the old one, the new sample standard deviation will be 100 times the old one.
    (b) The new sample mean will be 100 times the old one, the new sample standard deviation will be 100^2 times the old one.
    (c) The new sample mean will be $\frac{1}{100}$ times the old one, the new sample standard deviation will be $(\frac{1}{100})^2$ times the old one.
    (d) The new sample mean will be $\frac{1}{100}$ times the old one, the new sample standard deviation will be $\frac{1}{100}$ times the old one.

SECTION II: SHORT ANSWER (45 points)

1. (15 points total) Suppose that you are interested in the percentage of Davis students who would favor switching to a semester system. You take a random sample of 500 students. 300 students in the sample say that they would favor switching to the semester system. 100 students in the sample say that they want to stay with the quarter system. The remaining 100 students say that they don't care.
   (a) Suppose that the administration decides they will switch to semesters only if the percentage of students who favor switching is greater than 55%. The administration would prefer to error on the side of not switching. That is to say, if it is hard to tell whether the percentage of students favoring semesters is greater than 55%, the administration would not switch. Write down the null hypothesis and alternative hypothesis the administration should use to test whether the switch should be made.
   (b) Calculate the t statistic you would use to test the hypothesis you wrote down in part (a).
   (c) Based on your answer in part (b), use the critical value approach to test the hypothesis from part (a) at a 10% significance level. Be clear about exactly what critical value you are using. Clearly state your conclusion and what it implies for the administration's decision.
   (d) Calculate a 95% confidence interval for the percentage of students who favor the staying with the quarter system.

2. (6 points total) You take a sample of the weights of 10 males in our ECN 102 class. The observed weights (in pounds) are:

   165, 170, 185, 180, 165, 180, 170, 175, 170, 190

   (a) Calculate two different measures of the central tendency of the data.
   (b) Calculate two different measures of the dispersion of the data. One of your measures should be in the same units as weight. The other measure should be unitless.

3. (9 points total) For this question, consider a situation in which you are doing an upper one-tailed hypothesis test.
   (a) Describe a scenario in which a researcher would be very concerned about Type I error when doing an upper one-tailed hypothesis test.
   (b) On a graph showing the distribution of t^*, show how the choice of a small value versus a large value of the significance level α influences the likelihood of a Type I error when doing an upper one-tailed hypothesis test.
   (c) Based on the situation you described in part (a), if the researcher switched to a lower one-tailed hypothesis test rather than an upper one-tailed hypothesis test, would the researcher be likely to change the choice of α? Explain your answer.

4. (15 points total) You are given a dataset with annual observations of average real wage in the United States for the years 1950 to 1999. Write down the steps you would take in Excel to test whether the mean annual growth rate of real wages is equal to 5% using a 10% significance level. Be as specific as possible (explain all calculations you would make, what numbers you would use in your calculations, and how you would reach your conclusions).