An Expanded Sample Exam BUS 302L Statistics Exam

wayne.smith@csun.edu

[updated: Saturday, October 28, 2006]

"You cannot manage what you cannot measure." (paraphrase)
--Peter Drucker (1909-2005)

The purpose of this document is to expand on the existing sample exam for the BUS302L Statistics exam. No questions have been added, changed or deleted. I simply expanded the answer key by providing the "missing logic" (the why?) to help understand the correct answer for each question.

For use before of after the review workshop, I recommend that the sample exam be administered in a strictly timed fashion. This will help you understand how much time you have to answer each question. Subsequent questions on the sample exam should be covered, say, with a piece of blank paper, until each prior question is answered. To simulate the pace in the actual exam (35 minutes divided by 16 questions), each question should be answered in no more than 2 minutes. I recommend that you do not go on to the next question until you "finish" the previous question (i.e., you either know the correct answer or you can eliminate one or two known wrong answers and guess).

Sample Statistics Exam #500

Multiple Choice

Identify the letter of the choice that best completes the statement or answers the question.

1. Which is the following is correct?

a. The probability of a type I error is β.
b. The probability of a type II error is $(1 - \beta)$.
c. The probability of a type II error is α.
d. The probability of a type I error is $(1 - \alpha)$.
e. none of the above

2. A random sample of 10 items is taken from a normal population. The sample had a mean of 82 and a standard deviation is 26. Which is the appropriate 99% confidence interval for the population mean?

a. $82 \pm z_{0.005}(26)$
b. $82 \pm t_{0.005}(26)$
c. $82 \pm t_{0.001}(26/\sqrt{10})$
d. $82 \pm t_{0.005}(26/\sqrt{10})$
e. none of the above

3. A manufacturer of contact lenses is studying the curvature of the lenses it sells. In particular, the last 500 lenses sold had an average curvature of 0.5. The population is

a. the 500 lenses.
b. 0.5.
c. the lenses sold today.
d. all the lenses sold by the manufacturer.
e. none of the above

4. According to the empirical rule, the bell or mound shaped distribution will have approximately 68% of the data within what number of standard deviations of the mean?

a. one standard deviation
b. two standard deviations
c. three standard deviations
d. four standard deviations
e. none of the above

5. A random sample of 5 mosquitos is sampled. The number of mosquitos carrying the West Nile Virus in the sample is an example of which random variable?

a. normal
b. student's t
c. binomial
d. uniform
e. none of the above

6. A political scientist is studying voters in California. It is appropriate for him to use a mean to describe

a. the age of a typical voter.
b. the party affiliation of a typical voter.
c. the sex of a typical voter.
d. the county of residence of a typical voter.
e. none of the above

7. The long-run average of a random variable is

a. the expected value
b. the coefficient of determination
c. the standard deviation
d. the mode
e. none of the above

8. A manufacturer of women's blouses has noticed that 80% of their blouses have no flaws, 15% of their blouses have one flaw, and 5% have two flaws. If you buy a new blouse from this manufacturer, the expected number of flaws will be

a. 0.15
b. 0.20
c. 0.80
d. 1.00
e. none of the above

9. If population A has a larger standard deviation than population B, which of the following is NOT true?

a. Population B has a smaller variance than population A.
b. The mean of a sample of 20 from population A has a larger standard deviation than the mean of a sample of 20 from population B.
c. A typical observation from population A will be farther from the mean of population A than a typical observation from B will be from the mean of population B.
d. The mean of a sample from population A will on average be larger than the mean of a sample from population B.
e. none of the above

10. An inspector needs to learn if customers are getting fewer ounces of a soft drink than the 28 ounces stated on the label. After she collects data from a sample of bottles, she is going to conduct a test of a hypothesis. She should use

a. a two tailed test.
b. a one tailed test with an alternative to the right.
c. a one tailed test with an alternative to the left.
d. either a one or a two tailed test because they are equivalent.
e. none of the above

11. The manufacturer of Anthony Big's exercise equipment is interested in the relationship between the number of months (X) since the equipment was purchased by a customer and the number of hours (Y) the customer used the equipment last week. The result was the regression equation $Y = 12 - 0.5X$. The number 0.5 in the equation means that the average customer

a. used the equipment for 30 minutes last week.
b. who has owned the equipment an extra month used the equipment 30 minutes less last week than the average customer who has owned it one month less.
c. who just bought the equipment used it 30 minutes last week.
d. bought the equipment one-half month ago.
e. none of the above

12. A researcher is studying students in college in California. She takes a sample of 400 students from 10 colleges. The average age of all college students in California is

a. a statistic.
b. a parameter.
c. the median.
d. a population.
e. none of the above

13. A sample of 150 new cell phones produced by Yeskia found that 12 had cosmetic flaws. A 90% confidence interval for the proportion of all new Yeskia phones with cosmetic flaws is 0.044 to 0.116. Which statement below provides the correct interpretation of this confidence interval?

a. There is a 90% chance that the proportion of new phones that have cosmetic flaws is between 0.044 and 0.116.
b. There is at least a 4.4% chance that a new phone will have a cosmetic flaw.
c. A sample of 150 phones will have no more than 11.6% with cosmetic flaws.
d. If you selected a very large number of samples and constructed a confidence interval for each, 90% of these intervals would include the proportion of all new phones with cosmetic flaws.
e. none of the above

14. The standard deviation of a normal population is 10. You take a sample of 25 items from this population and compute a 95% confidence interval. In order to compute the confidence interval, you will use

a. the t table because the degrees of freedom will be 24.
b. the t table because you have estimated the standard deviation from the sample.
c. the z table because the population standard deviation is known.
d. the z table because the sample size is small.
e. none of the above

15. You are conducting a one-sided test of the null hypothesis that the population mean is 532 versus the alternative that the population mean is less than 532. If the sample mean is 529 and the p-value is 0.01, which of the following statements is true?

a. There is a 0.01 probability that the population mean is smaller than 529.
b. The probability of observing a sample mean smaller than 529 when the population meanis 532 is 0.01.
c. There is a 0.01 probability that the population mean is smaller than 532.
d. If the significance level is 0.05, you will accept the null hypothesis.
e. none of the above

16. Half of the observations in a data set are greater than the

a. mean.
b. median.
c. mode.
d. standard deviation.
e. none of the above

Sample Statistics Exam #500

Answer Section

1. ANSWER: E DIFFICULTY: 2
TOP TEN CONCEPT: 2 (Hypothesis Testing)

Rationale: The probability of obtaining a Type I error is α ("alpha"). The probability of obtaining a Type II error is β ("beta"). Of the two types, the Type I error is the error we emphasize on minimizing. Since the Type I is the "most worst" error, it is identified by the first letter in the Greek alphabet. You will just need to study and memorize these two types errors. Remember: a Type I error is the probability of convicting an innocent, or random, person—a jury would not want to do that. A Type II error is the probability of acquitting a guilty person—a jury would not want that either, but this is "less bad" than convicting an innocent person.

2. ANSWER: D DIFFICULTY: 2
TOP TEN CONCEPT: 3 (Confidence Intervals)

Rationale: A confidence interval is just a point estimate "wrapped" with the standard estimated error of the point estimate. This is sometimes called "constructing a confidence interval." The estimated error is *t* (why? because the population standard deviation is not given in the problem) multiplied by standard error of the mean (which is the sample mean divided by the square root of the sample size). Also, recall that if the *confidence* level is 99%, then *p*=.01, (which is 1 minus .99) is the *significance* level.

But confidence intervals are symmetric, so half of the error is *larger* than sample mean ("to the right") and half of the error is *smaller* than the sample mean ("to the left"). Therefore, *p*=.005 (which is .01 divided by 2).

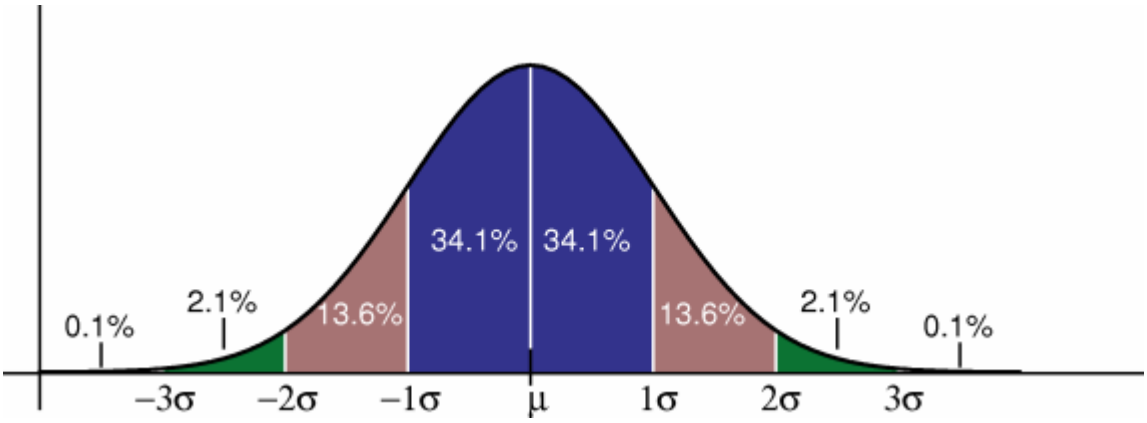
3. ANSWER: D DIFFICULTY: 2
TOP TEN CONCEPT: 9 (Populations and Samples)

Rationale: A. No, because 500 lenses is the sample (the sample is not the population).
B. No, because 0.5 is the mean (the mean is a descriptive statistic)
C. No, because the (number of) lenses sold today (this is irrelevant—the factory is no doubt in production more than just one day).
D. Yes, because all of the lenses of the manufacturer (is the population of interest)

4. ANSWER: A DIFFICULTY: 2

TOP TEN CONCEPT: 1 (Descriptive Statistics)

Rationale: You need to remember your standard normal curve. Sometimes this is called a "bell curve" or a "Gaussian curve" (see below).



The "empirical rule" is the cumulative distribution function (that is, the "area under the curve") in a normal distribution. 34.1% plus 34.1% equals approximately 68% (1 standard deviation on each side of the mean). 13.6% plus 13.6% equals approximately 27%. So, 68% plus 27% equals approximately 95% (2 standard deviations on each side of the mean). And so on. 3 standard deviations on each side of the mean is approximately 99%.

5. ANSWER: C DIFFICULTY: 2
TOP TEN CONCEPT: 6 (Probability Distributions)

Rationale: It does not really matter how many mosquitoes are sampled. Essentially, each mosquito can either *have* the West Nile virus (probably the type of mosquito of most interest to us) or *not have* the West Nile virus (probably the type of mosquito of least interest to us). Or put differently, there are two types of mosquitoes—no other information is given in this problem. Two types of mosquitoes in any sample would be an example of a binominal distribution--"bi" means "two", as in "bicycle".

6. ANSWER: A DIFFICULTY: 2
TOP TEN CONCEPT: 10 (Qualitative and Quantitative Variables)

Rationale: Age is a quantitative variable. It is, for the most part, continuous—that is, you could always determine smaller and smaller parts of it (years, months, days, hours, etc.). Also, you can compute the mean of age. You can not compute a "mean" for variables such as party affiliation (e.g., "Republican", "Democrat", or "Independent"), sex (i.e., "Male" or "Female"), or county of residence (e.g., "Los Angeles" or "San Diego").

7. ANSWER: A DIFFICULTY: 2
TOP TEN CONCEPT: 5 (Expected Value)

Rationale: A. Yes, because this is the definition of the long-run average (or long-run mean)
B. No, because the coefficient of the determination is a measure of well a regression model fits the data used to derive the model
C. No, because the standard deviation is a measure of dispersion (not central tendency)
D. No, because the mode is the most common number; but that is not the same as the average (or mean)

8. ANSWER: E DIFFICULTY: 2
TOP TEN CONCEPT: 5 (Expected Value)

Rationale: Expected value is a weighted, summed measure. The weight is the probability. So, in this problem, the expected value of obtaining a dress with a flaw is (.80 multiplied by 0) plus (.15 multiplied by 1) plus (.05 multiplied by 2). This sums to .25 or 25% (that is, .15 plus .10).

Or put differently, the probability of obtaining a dress with at least one flaw is slightly greater than .20 or 20% (that is, 1 - .80), because some dresses (5% to be exact) have more than one flaw in one dress. So intuitively, this increases the chance of obtaining a (single) dress with (one or more) flaw(s).

9. ANSWER: D DIFFICULTY: 2
TOP TEN CONCEPT: 8 (Variation and Uncertainty)

Rationale: A. No, because the variance is just the square of the standard deviation.
B. No, because over the long run, the sample distribution of the sample means will approach the population distribution (the "central limit theorem").
C. No, because the standard deviation is a measure of dispersion (dispersion is, by definition, how far away, on average, a typical observation is from the mean). Remember that the first part of the formula for variance (and therefore, standard deviation) includes subtracting each observation from the mean)
D. Yes, because the mean of a sample drawn from a population is unrelated to the standard deviation of that same population.

Also, make sure you read the question and each answer carefully. It is easy to get tripped up when the word NOT is in the question.

One thing that might help on a question such as this is make a drawing. For example, you could draw two circles (one for each population) and write a standard deviation (just make up a number) under each circle.

10. ANSWER: C DIFFICULTY: 2
TOP TEN CONCEPT: 2 (Hypothesis Testing)

Rationale: A. No, because the "…inspector needs to learn if customers are getting *fewer* ounces…" (emphasis added). That is, we are only interested in "one tail" or "one side" of the distribution.
B. No, because the "…inspector needs to learn if customers are getting *fewer* ounces…" (emphasis added). The inspector is not interested in results *greater* than 28 ounces.
C. Yes, because the "…inspector needs to learn if customers are getting *fewer* ounces…" (emphasis added). That is, we are (only) interested in "one-tail" or "one side" of the distribution. And the side we are interested in is the "less than" side, which is "to the left" of the center (the mean) of the distribution.
D. No, a "one-tailed" test is not the same as a "two-tailed" test.

The Greek letter μ ("mu") is used to identify the (unknown) population mean. Use a "one-tailed" test (e.g., $\mu<28$) if the hypothesis is "directional" (that is, to "one side and one side only of the distribution"). Use a "two-tailed" test (e.g., $\mu=28$) if the hypothesis is "non-directional" (that is, the result could be on either side of the distribution. If a direction is not given in a problem, use a "two-tailed" test by default.

11. ANSWER: B DIFFICULTY: 2
TOP TEN CONCEPT: 4 (Linear Regression)

Rationale: A. No, because .5 is the relationship (in this case, direct, negative relationship) between the number of months since purchase (X, the independent variable) and the number of hours used (Y, the dependent variable). .5 is the *slope* of the line determined by the equation. A use of 30 minutes (Y) would be the result of a calculation given some value of X (bonus question: what value of X is needed to make Y=30?—yes, you might need a calculator).
B. Yes. Read the answer carefully and slowly. Both "extra month" and "one month less" each mean "an additional month—that is, "moving 1 (X) unit a time". Also, 30 minutes is .5 (one-half) of 1 hour.
C. No, for the same reason as A. Also, the Y-intercept is 12.
D. No, you cannot determine that "the average customer bought the equipment one-half month ago". You can substitute .05 for X into the

equation to calculation the amount usage (again, bonus question—what is Y when X=.5?—no, you don't need a calculator).

This is a difficult question. You need to read carefully. Remember that the LDC exam is about important concepts, not difficult calculations.

12. ANSWER: B DIFFICULTY: 2
TOP TEN CONCEPT: 9 (Populations and Samples)

Rationale: A. No, because a "statistic" is for a sample. The question concerns "the average age of *all* college students in California." (emphasis added).
B. Yes, because a "parameter" is for a population. The question concerns "the average age of *all* college students in California." This is the population, or in the case of counting individuals, sometimes called a "census."
C. No, this question is about the population. Besides, the median is not the "average". We usually use the mean for that.
D. No, the population is the set of all the students. The average age is a single, computed number—it is not a set of numbers.

This is quite nearly a trick question. Read each question carefully. Always answer the correct question.

13. ANSWER: D DIFFICULTY: 2
TOP TEN CONCEPT: 3 (Confidence Intervals)

Rationale: A. No. This is a common, but decidedly wrong, interpretation of the confidence interval. Tip: The term probability (or "chance") goes with samples, not with populations.
B. No. 4.4% is just the "lower bound" of the confidence interval (and remember, a confidence interval is a measure of the error "wrapped" about the mean). 4.4% is not the sample mean, much less the population mean.
C. No, for the same general reason as B.
D. Yes, because this is the definition of a confidence interval. Often, however, we have neither the resources nor the time to "select a very large number of samples", so we just take a single (hopefully, representative) sample and construct a single confidence interval. In general, we want relatively narrow confidence interval, but in practice, this can often be difficult to obtain.

14. ANSWER: C DIFFICULTY: 3
TOP TEN CONCEPT: 6 (Probability Distributions)
Rationale: A. No, because we use the *t* table when we *do not* know the population standard deviation. It is true that we often do not know the population

standard deviation. But in this problem, the population standard deviation is given in the first sentence (and it also says that it is a normal population).
B. No for the same reason as A (above).
C. Yes, because in this problem, the population standard deviation is given in the first sentence (that is, for the same reason as A above) and it is a normal population.
D. No. Yes, we use the *z* table. But the reasons are that in this problem, the population standard deviation is given in the first sentence and it also says that it is a normal population.

Also, as the sample increases, the values of the *t* distribution approach the values of the *z* distribution.

15. ANSWER: B DIFFICULTY: 2
TOP TEN CONCEPT: 7 (*p*-values)

Rationale: A. No, hypothesis tests are for testing the (population mean) parameter of 532, not the (sample mean) statistic of 529.
B. Yes, this is the correct interpretation. It is a little hard to read because the 0.01 is at the end of the sentence.
C. No, because the concept of probability is associated with samples not populations.
D. No, if the significance level is .05, you would *reject* the null hypothesis (why? because p=.01 < α=.05—see slide 25 on the LDC Statistics Review Powerpoint).

Remember to read the question and the answers carefully. The question asks "…which of the following is *true*?" (so try to eliminate false answers).

16. ANSWER: B DIFFICULTY: 1
TOP TEN CONCEPT: 1 (Descriptive Statistics)

Rationale: A. No. While it is possible that the mean splits the data set exactly in "half", this is not the best answer. A mean that splits the data set exactly in "half" would the *exception*, not the *rule*. So therefore, it is not the best answer. The instructions at the beginning of the exam reads "*Identify the letter of the choice that best completes the statement or answers the question.*"
B. Yes, this is definition of the median. If there happen to be an even number of observations in the data set, the median is just the mean of the two middle numbers. Usually, the need to sort or order the data in ascending order to determine the median.
C. No, because the mode is the most common. The word mode is derived from the French word for "fashion". The mode is just the most popular

observation. It is not the descriptive statistic that splits the data set exactly in “half”.
D. No, the standard deviation is a measure dispersion. It is not a measure of central tendency.