UNIVERSITY of SACRAMENTO

GLOBAL LEADERSHIP MBA

BUSINESS STATISTICS
FINAL EXAM

Name: _____

INSTRUCTIONS

1. Do not open this exam until instructed to do so.
2. Be sure to fill in your name before starting the exam.
3. You have two hours (2:00) to complete the exam.
4. The exam is open book, notes, and any reference materials except your classmates or other people.
5. If you use a laptop, you may *not* use e-mail, instant messaging, or any other type of electronic communication program/facility to communicate with anyone else during the exam
6. This exam has a total 40 questions worth 75 points. The point-value of each question is in parenthesis next to the question number.
7. Read all questions and possible answers *carefully*.
8. For multiple choice questions, mark *only one* letter indicating your answer. Ambiguous selections will be marked wrong!
9. Do all of your work (that you want me to see) on this exam.
10. If you want to check your answers later against the solution set, please make a copy of your answers before turning in your exam.

Good luck!

For questions 1-3, indicate the *type of data* described.

1. **(1 pt)** In a web-based survey, customers are asked to rate your company's product on the following scale: Excellent, good, average, poor.

   (a) Continuous
   (b) Ordinal
   (c) Nominal
   (d) None of the above

2. **(1 pt)** In the same survey, customers are asked to provide their gender: male or female.

   (a) Continuous
   (b) Ordinal
   (c) Nominal
   (d) None of the above

3. **(1 pt)** As part of a quality improvement program, your mail order company records the length of time every customer spends on hold waiting to place their order via the telephone.

   (a) Continuous
   (b) Ordinal
   (c) Nominal
   (d) None of the above

4. **(2 pts)** Calculate the median of the following set of data:

   123, 243, 322, 492, 537, 599, 620, 798, 812, 954.

   (a) 537
   (b) 550
   (c) 568
   (d) 599
   (e) None of the above are correct.

5. **(2 pts)** If you have a data set that consists of the following three values 1, 2, and 3, which of the following statements are true:

   (a) The range of the data is 3.
   (b) The sample standard deviation equals the sample average.
   (c) The sample standard deviation equals the sample variance.
   (d) Both (a) and (b) are true.
   (e) None of the above are true.

6. **(1 pt)** For CEO salaries throughout the entire United States (not just the data set we used in class), which has a distribution that is skewed to the right with some very large outliers, which is greater: the average salary or the median salary?

   (a) Average salary
   (b) Median salary
   (c) Cannot determine from the information given

7. **(1 pt)** The inter-quartile range equals the 75th percentile minus the 25th percentile.

   (a) True
   (b) False

For questions 8-10: You are the senior vice-president in charge of production for a company that manufactures two different types of "widgets." You manage three divisions. You are interested in displaying various types of data about your company. Select the *one* technique from each list that is *most* applicable.

8. **(2 pts)** In order to summarize production for the past month, both broken down by the number of each type of widget produced by each division, as well as the total number of each type of widget produced and the total production for each division, you would use a:

   (a) Scatterplot
   (b) Contingency table
   (c) Confidence interval
   (d) Side-by-side boxplot
   (e) Histogram

9. **(2 pts)** You also want to graphically assess whether there is a positive association between total daily production level (measured in terms of the dollar value of good widgets produced) and daily cost of scrap and re-work (measured in the cost in dollars of lost materials and labor to correct errors) across the divisions for the past 90 days. The best way to evaluate this is to use a:

   (a) Scatterplot
   (b) Contingency table
   (c) Confidence interval
   (d) Side-by-side boxplot
   (e) Histogram

10. **(2 pts)** You now want to graphically compare the *distribution* of salaries for mid-level managers *between* your three manufacturing divisions. Each division has between 37 and 80 mid-level managers. The best way to display the information for this purpose is with a:

    (a) Scatterplot
    (b) Contingency table
    (c) Confidence interval
    (d) Side-by-side boxplot
    (e) Histogram

11. **(1 pt)** A number is drawn at random from a box. There is a 20% chance for it to be less than 10. There is a 10% chance for it to be more than 50. So, the chance of getting a number between 10 and 50 (inclusive) is 70%.

    (a) True
    (b) False

12. **(1 pt)** The Central Limit Theorem says that for large sample sizes the sample mean has an approximately normal distribution.

    (a) True
    (b) False

13. **(1 pt)** From the empirical rule we can deduce that, *for any distribution*, 95% of the observations fall between the mean plus or minus two standard deviations.

    (a) True
    (b) False

14. **(1 pt)** As the number of degrees of freedom increase, the *t* distribution gets closer and closer to the normal distribution.

    (a) True
    (b) False

15. **(2 pts)** As the district sales manager for a franchise fast-food company, you prefer to compare sales performance in terms of standardized values. Having just hired a new analyst, you demonstrate for her how to standardize the value $x = 20$ given $\mu = 10$ and $\sigma = 5$. *Show your work.*

    Solution:

    $\boldsymbol{Z = (x - \mu) / \sigma = (20 - 10) / 5 = 2}$

16. **(2 pts)** You are now looking at a computer printout of the daily sales for 100 franchise stores in your sales district prepared by the new analyst. The daily sales values have been standardized by subtracting the average daily sales for all 100 stores and dividing by the standard deviation of the daily sales for the 100 stores. The first 10 entries are:

    -6.2 13.5 12.2 -8.13 14.3 -5.1 -7.2 -11.3 10.8 6.3

    Does the printout look reasonable or is something wrong?

    (a) The numbers look reasonable.
    (b) Something is wrong.
    (c) Need more information about the computer program.

17. **(3 pts)** Assuming the standardized values (call them *X*) have a standard normal distribution, using either the tables in the back of your textbook or Excel, find the following probabilities (to four decimal places):

    (a) $\Pr(X < 0.5)$ = ____**0.6915**_____

    (b) $\Pr(X < 1.5)$ = ____**0.9332**_____

    (c) $\Pr(X < 2.5)$ = ____**0.9938**_____

18. **(2 pts)** Based on your Business Statistics class in the Global MBA program, you know that a confidence interval is wider if:
    (a) A larger sample (*n*) is used.
    (b) A larger *t* or *z* value is used.
    (c) It is changed from a 95% CI to a 90% CI.
    (d) Both (b) and (c).
    (e) All of the above.

19. **(2 pts)** A confidence interval inappropriately using a *z* statistic instead of a *t* statistic will give a _____ interval.
    (a) wider
    (b) narrower

20. **(3 pts)** If a confidence interval has width 1 based on a sample of 50 observations, what is the width if the sample size is increased to 800 (assuming everything else remains constant)?
    (a) 0.25
    (b) 0.5
    (c) 2
    (d) 4
    (e) Not enough information to determine the answer

21. **(1 pt)** In a hypothesis test, assuming the conventional critical value for evaluating *p*-values of 0.05, a *p*-value greater than 0.05 indicates statistical significance.
    (a) True
    (b) False

22. **(1 pt)** In a hypothesis test, the null hypothesis says that the observed difference is just due to chance.
    (a) True
    (b) False

23. **(2 pts)** A paired *t*-test requires:
    (a) One sample with two observations on each unit in the sample.
    (b) Two independent samples.
    (c) Either a one or two samples depending on the hypotheses.
    (d) Either a one-sided or two-sided test depending on the *p*-value.
    (e) Both (c) and (d).

For questions 24-26: You are the sales manager for a large condominium development in Sacramento, California. You are interested in determining how much to price your units for and have collected information on 87 equivalent condominiums that have sold in the past two months in Sacramento.

24. **(2 pts)** To determine a range of plausible sales prices for your condominiums, you would construct a _____ on the data for the 87 equivalents.
    (a) Two-sample t-test
    (b) One-sample t-test
    (c) Contingency table
    (d) Confidence interval
    (e) Mosaic plot

25. **(2 pts)** After having sold 21 of your condominium units (out of 150), you are interested in evaluating whether your units are selling for significantly more than the other 87 Sacramento condos. In order to determine if your condos are, on average, *selling for more* than the other 87 you would use a:

    (a) A one-sided, one-sample t-test
    (b) A two-sided, one-sample t-test
    (c) A one-sided, two-sample t-test
    (d) A two-sided, two-sample t-test
    (e) A paired t-test

26. **(3 pts)** If the average sales price of your 21 sold condos is $350,000 with a standard deviation of $20,000, construct a 95% confidence interval for the average sales price of the population of equivalent condos. *Show your work.*

Solution:

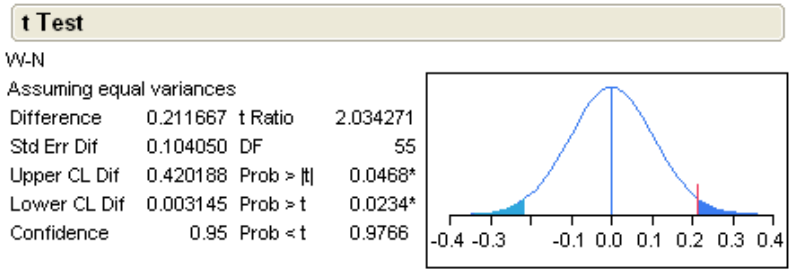
$$x \pm t_{n-1,\alpha/2} \times s/\sqrt{n} = x \pm t_{20,0.025} \times s/\sqrt{n}$$
$$= \$350,000 \pm 2.086 \times 20,000/\sqrt{21}$$
$$= [\$340,896.11, \$359,103.89]$$

For questions 27-29: The Global Leadership MBA program enrolls students directly from undergraduate studies and others after having obtained some work experience. The students with work experience claim that they do better than their peers because of their experience in the business world. However, the students right out of school claim they do better than their peers because they have more recent academic experience.

The Global Leadership MBA program is not interested in which group does better, but simply in determining whether the two groups' performance is the same or not. To test this, the program took a random sample of students' grades (from a central university database). Let μ_W denote the population mean GPA for the students with work experience and let μ_N denote the population mean GPA for students with no work experience. Let $\bar{x}_W$ and $\bar{x}_N$ represent the corresponding sample means.

27. **(3 pts)** Write down the null hypothesis and alternative hypothesis that were tested in this study, both in words and using the appropriate notation.

Assume that in answer to the last question, you analyze the data from the random sample of students' grades with a two-sample, two-sided *t*-test (which may or may not be correct), obtaining the following results:



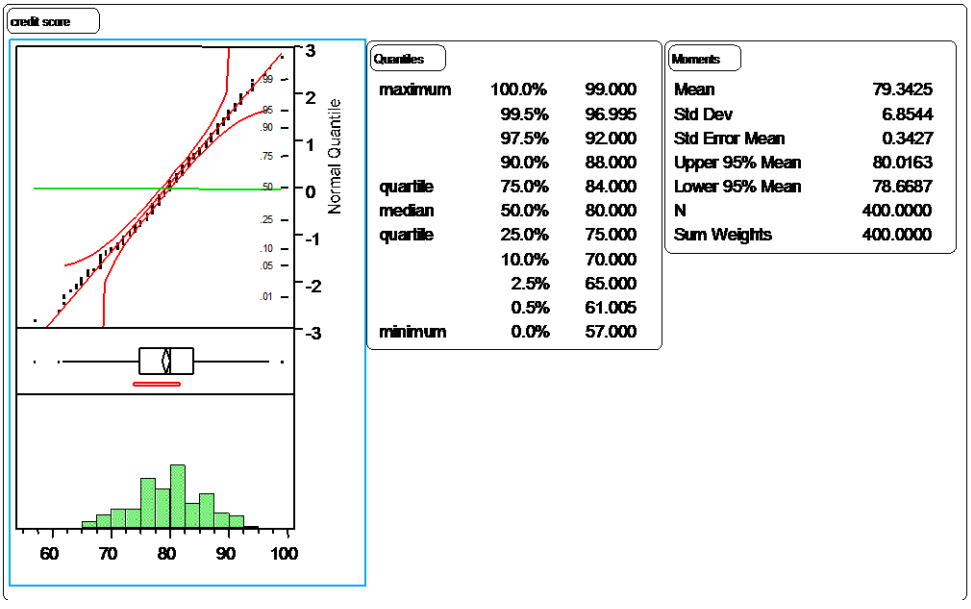
28. **(2 pts)** What conclusion do you reach from the above output?

(a) Reject the null hypothesis and conclude that the mean GPAs are different.
(b) Accept the null hypothesis and conclude that the mean GPAs are different.
(c) Reject the null hypothesis and conclude that the mean GPAs are the same.
(d) Accept the null hypothesis and conclude that the mean GPAs are the same.
(e) None of the above are correct conclusions.

29. **(2 pts)** What is the interpretation of the "Difference" value in the table?

(a) The average GPA of those in the sample without work experience is 0.212 points higher than those with work experience.
(b) The average GPA of those in the sample with work experience is 0.212 points higher than those without work experience.
(c) Can't determine without looking at the original data.

For questions 30-33: Most lenders to individuals use some form of credit scoring system to evaluate applicants for loans. One lender uses such a system with a 0 to 100 scale where higher scores are better. The JMP output below is for a random sample of 400 applicants drawn from a new mailing list.



30. **(2 pts)** The lender regards a mailing list to be desirable if the mean credit score is 82 or higher. What can you conclude from the output (taken at face value, without regard to any criticism of the study) about the mailing list?

    (a) We cannot tell much since the sample size is too small.
    (b) The data are so variable that no conclusion can be reached.
    (c) The mean score of the mailing list is very likely to be less than 82.
    (d) The mean score of the mailing list could reasonably be 82, higher or lower.
    (e) The mean score of the mailing list is very likely to be 82 or higher.

31. **(3 pts)** Using the above output, which *one* of the following steps is *incorrect* (or are all of them are correct) for calculating and interpreting the 95% confidence interval for the population mean credit score of [78.67, 80.02].

    (a) The normal quantile plot shows that it is reasonable to assume that the credit scores have a normal distribution.
    (b) To be most precise, because the population standard deviation is unknown, the *t* distribution with 399 degrees of freedom should be used in the confidence interval calculation.
    (c) Given the correct *t*-value ("*t*" below), you calculate the confidence interval as

$$CI = \left(79.3425 - t \times 0.3427, 79.3425 + t \times 0.3427\right)$$

    (d) You can interpret the confidence interval to say that 95 percent of the credit scores in the population will fall within the confidence interval.
    (e) All of the above are true.

32. **(2 pts)** Your boss, not having taken this class, isn't sure what to make of a confidence interval. She says that what she really wants to know is whether she can be relatively sure the average credit score is 82 or higher. Based on what you know from the confidence interval, what *can* you tell your boss?

    (a) Well, 82 is outside of the confidence interval, so you think there's a good chance that the average population score is above 82, but you can't be sure without doing a hypothesis test.
    (b) Since 82 is above the upper bound of the 95% confidence interval, then a hypothesis test would reject the null hypothesis that the average population score is equal to 82.
    (c) Since 82 is above the upper bound of the 95% confidence interval, then a hypothesis test would fail to reject the null hypothesis that the average population score is equal to 82.
    (d) You can't conclude anything about a hypothesis test just from the confidence interval.

33. **(2 pts)** The manager who selected the sample later said that he had discarded the obvious low and high score and replaced them with scores nearer the average. What is the consequence of this action, as compared with truly random sampling?

    (a) The mean will definitely be too high.
    (b) The mean will definitely be too low.
    (c) The confidence interval will be too wide.
    (d) The confidence interval will be too narrow.
    (e) The additional information is irrelevant.

For questions 34-40: As the regional manager of a pizza franchise business, you are interested in understanding how income in a region affects pizza sales. Below is a regression output for pizza sales (in thousands of dollars) regressed on the average household income of an area (also in thousands of dollars).

Linear Fit

Pizza Sales ($000) = 14.577381 + 2.9047619*Income ($000)

Summary of Fit

RSquare	0.96832
RSquare Adj	0.96304
Root Mean Square Error	3.108329
Mean of Response	43.625
Observations (or Sum Wgts)	8

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1771.9048	1771.90	183.3946
Error	6	57.9702	9.66	Prob > F
C. Total	7	1829.8750		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	14.577381	2.410088	6.05	0.0009*
Income ($000)	2.9047619	0.214495	13.54	<.0001*

34. **(2 pts)** What is the average pizza sales across all eight regions?

(a) $43.63
(b) $2,904.76
(c) $14,577.38
(d) $43,625
(e) Cannot determine from the JMP output above.

35. **(2 pts)** What does the *p*-value for the income variable ("Income ($000)") mean?

(a) The slope of the regression line is significantly different from zero.
(b) The intercept of the regression line passes through the origin
(c) The income of a region is *not* significant in explaining pizza sales.
(d) Both a and c are correct.
(e) None of the above.

36. **(2 pts)** What is the interpretation of the slope?

(a) For each $1,000 increase in average household income, pizza sales increase by $2.90.
(b) For each $1,000 increase in average household income, pizza sales increase by $2,905.
(c) For each $2,905 increase in average household income, pizza sales increase by $1,000.
(d) For each dollar increase in average household income, pizza sales increase by $14.58.
(e) None of the above.

37. **(3 pts)** What does the model predict for pizza sales in a region with an average household income of $40,000?

(a) $116
(b) $131
(c) $116,205
(d) $130,768
(e) None of the above.

38. **(2 pts)** What can you conclude from/about the estimated intercept?

(a) For the model's fitted line, when x=0 then the model predicts that y=14.577381.
(b) The model predicts that regions with an average household income of $0 will still have pizza sales of $14,577.
(c) The intercept must be an extrapolation from the data, since we could not have observed any regions with average household incomes of $0 (or less).
(d) All of the above are appropriate conclusions.
(e) None of the above are appropriate conclusions.

39. **(2 pts)** What is the interpretation of the R^2?

(a) 0.968% of the variation in pizza sales is explained by income.
(b) 0.968% of the variation in income is explained by pizza sales.
(c) 96.8% of the variation in pizza sales is explained by income.
(d) 96.8% of the variation in income is explained by pizza sales.
(e) None of the above.

40. **(2 pts)** What is the Root Mean Square Error?

(a) It is how far off the intercept is from the origin on average.
(b) It is the estimated standard deviation of the error term in the regression model.
(c) It is calculated as the square of the mean of the independent variable.
(d) Both a and c.
(e) None of the above.

End of Exam