

COMM 215 Business Statistics

**Section BX
 Summer 2012**

**Final Examination
 August 2012**

Last Name: _____ First Name _____
 (Please Print)

Student No.: _____

INSTRUCTIONS

1. Attempt all questions. Show your work for FULL credit.
2. This is a closed book, closed note examination. You are allowed to use standard, basic calculators during the examination. Sharing of calculators is not allowed. **DO NOT DETACH pages.** Return the exam booklet intact.
3. For **PART I - Multiple Choice Questions:**
 - a. Use **PENCIL** to **fill** the appropriate circles on the blue sheet corresponding to the correct answer choices.
 - b. Use **PENCIL** to **write** your full name and Student ID, and to **fill** the matching circles below your name and ID on the blue sheet.
4. For **PART II – Problem Solving:**
 - a. Use **PEN** to write your answers in the space provided below each question.
 - b. You may use both sides of the paper if necessary. Do not include extra pages.
5. Tables and formulas are appended. **DO NOT DETACH THEM FROM THE BOOKLET.**
6. No questions about the examination are allowed.

		Marks
		obtained
Part I	Sub-Total	/48
Part II		
	Question 1	/14
	Question 2	/12
	Question 3	/12
	Question 4	/14
	Sub-Total	/52

Total

100

PART I: Multiple choice questions (one mark for each of questions 1 to 20 and two marks for each of questions 21 to 34). Some of the numbers in the provided choices have been rounded.

Indicate your answers on the multiple choice answer sheet provided. Use pencil only to make black marks that fill the circle completely. Erase cleanly any answer you wish to change. Make no stray marks on the answer sheet.

1. Which of the following statements is true regarding a simple linear regression model?
 - a) The proportion of variation explained is the correlation coefficient
 - b) SSR must be greater than SSE if the coefficient of determination is larger than 0.5
 - c) The slope coefficient must have the same sign as the coefficient of determination
 - d) None of the suggested answers are correct

2. Which of the following statements is true?
 - a) The central limit theorem states that no more than 75% of the observations lie within 2 standard deviations of the mean
 - b) The central limit theorem states that the sampling distribution of the mean will be approximately normal regardless of the shape of the population distribution as the sample size becomes large.
 - c) The central limit theorem states that the standard error of the sampling distribution of the mean can be approximated by σ as long as the population distribution is symmetric.
 - d) The population must be normal and that σ is not known for the central limit theorem to be valid.

3. In order to use the normal distribution for interval estimation of μ when σ is known and the sample is very small,
 - a) the population must have a t-distribution with at most 30 degrees of freedom
 - b) The population coefficient of variation must be small
 - c) The sample standard deviation must be equal to the population standard deviation
 - d) None of the suggested answers are correct

4. An interval estimate for the mean is a range of values used to estimate
 - a) the sample mean
 - b) the sampling distribution
 - c) sampling error, or the difference between the population mean and the sample mean
 - d) None of the suggested answers are correct

5. Forty shoppers were asked if they preferred the weight of a can of soup to be 6 ounces, 8 ounces, or 10 ounces. Below you are given their responses.

6	6	6	10	8	8	8	10	6	6
10	10	8	8	6	6	6	8	6	6
8	8	8	10	8	8	6	10	8	6
6	8	8	8	10	10	8	10	8	6

Which of the following is an appropriate graphical display of the above data?

- a) Bar chart
- b) Normal distribution curve
- c) Box plot
- d) None of the suggested answers are correct

6. Four hundred people were asked whether gun laws should be more stringent. Three hundred said "yes," and 100 said "no." The point estimate of the proportion in the population who will respond "no" is
 - a) 25
 - b) calculated as 0.25 minus sampling error as determined from the sample
 - c) 0.75
 - d) None of the suggested answers are correct

7. In determining the sample size necessary to estimate a population proportion, which of the following information is not needed?
 - a) the maximum margin of error that can be tolerated
 - b) the confidence level required
 - c) a preliminary estimate of the true population proportion P
 - d) whether or not the population is symmetrically distributed

8. Since the population is always larger than the sample, the value of the sample mean
 - a) is always smaller than the true value of the population mean
 - b) is always larger than the true value of the population mean
 - c) is always equal to the true value of the population mean
 - d) could be larger, equal to, or smaller than the true value of the population mean

9. Sampling distributions are:
 - a) the probability distributions of population parameters.
 - b) the probability distributions of sample statistics.
 - c) referring to the standard errors of sample statistics
 - d) None of the suggested answers are correct

10. Last year, 55% of MNM, Inc. employees were female. It is believed that there has been a reduction in the percentage of females in the company. Which of the following gives the correct null and alternative hypotheses in testing the belief?
 - a) The correct hypotheses are: $H_0 : p \leq 0.55$ and $H_a : p > 0.55$
 - b) The correct hypotheses are : $H_0 : p \geq 0.55$ and $H_a : p < 0.55$
 - c) The null and alternative hypotheses cannot be set up because no sample information is given
 - d) None of the suggested answers are correct

Refer to the following in answering questions 11 to 13

In a large corporation, sixty-two percent of the employees are male. Twenty-three percent of the employees earn more than \$30,000 a year. Eighteen percent of the employees are male and earn more than \$30,000 a year. Suppose that an employee is selected at random. Let A be the event that the selected employee is a male and B the event that the employee earns more than \$30,000 a year.

11. Given that the selected employee is a female, what is the probability that the employee will earn \$30,000 or less?
 - a) 38/77
 - b) 0.33
 - c) 33/38
 - d) None of the above is a correct answer

12. Which of the following statements is not true?
- $P(A)P(A|B)=P(B)P(B|A)$
 - $P(A \cup B) < P(A) + P(B)$
 - $P(A | B^c) = P(B | A^c)$
 - All of the above are true
13. Which of the following statements is true?
- A and B are independent
 - A and B are mutually exclusive
 - A and B cannot occur at the same time
 - None of the suggested answers are correct
14. There is a need to estimate the average total compensation of CEO's in the Service industry. Data were randomly collected from 18 CEO's and the 97% confidence interval was calculated to be (\$2,181,260, \$5,836,180). Which of the following interpretations is correct?
- 97% of the sampled total compensation values fell between \$2,181,260 and \$5,836,180.
 - In the population of Service industry CEO's, 97% of them will have total compensations that fall in the interval \$2,181,260 to \$5,836,180.
 - We are 97% confident that the mean of the sampled CEO's falls in the interval \$2,181,260 to \$5,836,180.
 - None of the suggested answers are correct
15. A sample of account balances from a credit company showed an average daily balance of \$1,040. The standard deviation of the population is known to be \$200. We are interested in determining if the mean of all account balances (i.e., population mean) is different from \$1,000. A statistician reported a value of the z statistic of 1.6. Which of the following statements is true?
- There is insufficient information to determine the p-value since the sample size is not given
 - The p-value is equal to 0.0548
 - The p-value is equal to 0.0274
 - None of the suggested answers are correct
16. If a new independent variable is added to an existing regression equation, then the resulting sample regression equation
- will have a SSE no bigger than that of the original sample regression equation
 - will have a smaller coefficient of determination
 - will have a smaller SSR
 - None of the suggested answers are correct
17. Consider a distribution of ten account balances with a mean balance of \$620. If an eleventh account with a balance of \$400 is added to the group, what is the mean balance for the new group consisting of eleven accounts?
- \$600
 - \$510
 - \$586
 - cannot be determined with the given information

18. The least squares criterion is used to find the estimated regression equation. As a result,
- a) the predicted values \hat{y}_i must always be small than the actual y_i .
 - b) The sum of the errors $\sum(\hat{y}_i - y_i)$ must be negative
 - c) The error terms must always be equal to 0
 - d) None of the suggested answers are correct

Refer to the following in answering questions 19 to 20

In the last presidential election, before the candidates started their major campaigns, the percentages of registered voters who favored the various candidates were as follows.

Percentages	
Republicans	34%
Democrats	43%
Independents	23%

After the major campaigns began, a random sample of 400 voters showed that 172 favored the Republican candidate; 164 were in favor of the Democratic candidate; and 64 favored the Independent candidate. We are interested in determining whether the proportions of voters who favored various candidates had changed. The Chi-Square statistic is computed to be 18.42.

19. Which of the following statements is true?
- a) A one-tailed Chi-Square test is not appropriate since the percentage of voters favoring the republicans may be higher or lower after the campaign
 - b) One should conclude that there is evidence at 1% significance level that the percentages are no longer the same as those before the campaign
 - c) The null hypothesis should state that the percentages before the campaign had changed and is accepted at 1% significance level
 - d) None of the suggested answers are correct
20. The expected number of voters in the sample who did not favor the republican after the campaign
- a) is 264 if the percentages had not changed
 - b) is 264 if the percentages had changed
 - c) cannot be determined with the given information
 - d) None of the suggested answers are correct
21. In a multiple regression analysis involving 15 independent variables, SST = 800 and SSE = 240. The coefficient of determination
- a) is equal to 0.3
 - b) is equal to 0.5
 - c) is equal to 0.7
 - d) None of the suggested answers are correct

22. You are given the following stem-and-leaf display of the data on the ages of employees at a company.

```
2 | 67
3 | 2266889
4 | 0122
5 | 235
6 | 1
```

Suppose that a Histogram with a class width of 8 is constructed for the data set above. How many classes are there in the Histogram?

- a) No more than 4
 - b) 6 since the largest data value is 61
 - c) unknown since the range cannot be determined from the stem-and-leaf display
 - d) None of the suggested answers are correct
23. A group of students had dinner at a local restaurant. The total of all bills for the dinner was \$414.70. Each student will have to pay \$18.85 if the students share equally the total bill. Which of the following statements is true?
- a) \$18.85 is the median value of the bills
 - b) The number of students at the dinner is 22
 - c) At least one of the bills must be equal to \$18.85
 - d) None of the suggested answers are correct
24. In multiple regression analysis, which of the followings is incorrect?
- a) The dependent variable is predicted by the independent variables with an error that is given by the standard error of the estimate s .
 - b) the sample values of the dependent variable is used to calculate SST
 - c) If the F value is significant, some independent variables may still have a non-significant t value.
 - d) All of the above is true
25. In a regression analysis, the error term ε is a random variable. Which of the following statements regarding the mean or expected value of the error term is correct?
- a) the mean is zero since the error of prediction is always equal to 0
 - b) the mean is positive if there is a positive relationship
 - c) the mean can be any value depending on the strength of the relationship
 - d) None of the suggested answers are correct
26. An official of a large national union claims that the fraction of women in the union is not different from one-half. A sample of 100 employees is selected and 47 of them are women. A 5% level of significance is used to test the union's claim. Which of the following statements is true?
- a) The correct hypotheses are: $H_0 : p = 0.5$ and $H_a : p \neq 0.5$
 - b) The null hypotheses should be $H_0 : p \leq 0.5$ since the sample proportion is less than 0.5
 - c) The union's claim is surely wrong because the sample proportion differs from 0.5
 - d) The correct hypotheses are: $H_0 : p = 0.47$ and $H_a : p \neq 0.47$

27. The Bureau of Labor Statistics reported that the average yearly income of dentists in the year 2009 was \$110,000. A sample of 81 dentists taken in 2010 showed an average yearly income of \$120,000. Assume the standard deviation of the population of dentists in 2010 is \$36,000 and the distribution of income is approximately normally distributed. The p-value for testing if the average income in 2010 has increased from 2009 is found to be 0.0062. Which of the following is true?
- The p-value cannot be interpreted since the significance level is not given
 - The p-value can be interpreted as the probability of obtaining a sample with a z value larger than 2.5 if the average income has not changed
 - The probability that the sample mean is greater than \$11,0000 is only 0.0062
 - None of the suggested answers are correct
28. In the past, the average age of employees of a large corporation has been 40 years. In order to determine whether there has been an increase in the average age of all the employees, a sample of 64 employees was selected. The level of significance used is 0.05. Which of the followings is true?
- The p-value is 0.05 since p-value is always equal to the significance level of a test
 - The maximum allowable probability of Type II error is 0.05
 - The probability of rejecting the hypothesis that the average age has not increased when it is true will not exceed 0.05
 - None of the suggested answers are correct
29. The following information regarding a dependent variable Y and an independent variable X is based on 4 pairs of observations: $\sum x = 12$, $\sum y = 23$, $\sum(x - \bar{x})^2 = 12$, $\sum(x - \bar{x})(y - \bar{y}) = -9$. The intercept of the regression function is
- 8
 - 3.5
 - 0.715
 - None of the suggested answers are correct
30. In a local university, 40% of the students live in the dormitories. A random sample of 80 students is selected for a particular study. The probability that the sample proportion (the proportion living in the dormitories) is between 0.30 and 0.50 is
- 0.4664
 - 0.9328
 - 0.0336
 - None of the suggested answers are correct
31. The average life expectancy of a printer is 3.5 years with a standard deviation of 2 years. A random sample of 49 printers was tested. What is the probability that the sample mean will be between 4 and 4.2 years?
- between 0.05 and 0.025
 - 0.0472
 - 0.033
 - None of the suggested answers are correct
32. In a regression analysis situation, the standard error of the slope is:
- a measure of the variation in the estimated regression slope from sample to sample
 - equal to the square root of the standard error of the estimate
 - a measure of the amount of change in y that will occur for a one unit change in x
 - None of the suggested answers are correct

33. If we change a 95% confidence interval estimate to a 99% confidence interval estimate, we can expect
- a) the width of the confidence interval to increase because a higher confidence level is needed
 - b) the width of the confidence interval to decrease because we become more confident that the interval is correct
 - c) the width of the confidence interval to remain the same since the standard deviation has not changed
 - d) None of the suggested answers are correct
34. Which of the following statements is true?
- a) If we select several random samples of the same size from a normally distributed population and we compute the sample means, they must all have the same values
 - b) If we select several random samples of the same size from a normally distributed population, the expected values of the sample means will be different.
 - c) If we select several random samples of the same size from a normally distributed population, the sample means will have different standard deviations.
 - d) None of the suggested answers are correct

e) It is further found out that among these 50 sampled employees, 10 of them did not have a promotion (during their tenure with the company). Give a point estimate for the proportion of employees who did not have a promotion. With a 95% confidence level, the proportion of employees who did not have a promotion falls within what limits?

f) What is the smallest sample size needed to provide a margin of error (with a 0.95 confidence coefficient) of no greater than 0.04 in estimating p , the proportion of employees who did not have a promotion, for all possible values of p ?

Question 2 (12 marks)

- a) A company produces your typical consumer battery. The company claims that their batteries last at least 100 hours, on average. An experiment is conducted to investigate if the company's claim is not true. You believe that the mean life is actually less than the 100 hours the company claims. Data are collected on the battery life (in hours) for a random sample of batteries and the following information regarding the sample is obtained:

Sample mean = 98.5

Estimated standard error of mean = 0.818

Degrees of freedom of t -test for population mean = 19

- i. What assumption is needed for conducting a t -test for the mean in the present case?

- ii. What is the sample size collected?

- iii. At 5% significance level, does the present sample provide sufficient evidence that the company's claim is not true?

- iv. Explain the meaning of type I error in the context of the present application.

- b) The CEO of a software company is committed to expanding the proportion of highly qualified women in the organization's staff of salespersons. He believes that the proportion of women in similar sales positions across the country in 2004 is less than 45%. Hoping to find support for his belief, he directs his assistant to collect a random sample of salespersons employed by his company, which is thought to be representative of sales staffs of competing organizations in the industry. The collected random sample of size 50 showed that 17 were women. Find the p-value of the hypothesis test conducted to investigate if the CEO's claim can be substantiated. At which of the significance levels 0.10, 0.05, 0.01 and 0.001 can one conclude that the claim is true? Justify your answer.

Question 3 (12 marks)

- a) In 2000, computers of Brand A controlled 25% of the market, Brand B 20%, and Brand C 55%. In 2004, sample data was collected from many randomly selected stores throughout the country. Of the 1,200 computers sold, 280 were Brand A, 270 were Brand B, 650 were Brand C. Has the market changed since 2000? Test at the 1% significance level.

b) A sample of an industrial product provided the following data on product quality by production process. Test at 5% significance level the hypothesis that product quality is independent of the production process. State clearly your conclusion and justify your answer.

Process	Defectives	Good
A	36	364
B	17	283
C	27	173

Question 4 (14 marks)

Levesque Realty is frequently asked by his clients about the average monthly cost of heating of a single-family home. The research department at Levesque has been asked to develop a model to predict the average monthly heating cost of single-family homes. Three variables are thought to be relevant:

- 1) TEMP: The mean daily outside temperature,
- 2) INSUL: The thickness (in inches) of insulation in the attic,
- 3) AGE: The age of the furnace.

A random sample of single-family homes was taken and partial information (numbers in some entries are not shown) from a regression analysis with EXCEL is given below.

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Regression	(a)	(c)	57073.3	
Residual	16	(d)	2605.9	
Total	(b)	(e)		

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	427.19	59.6		
TEMP	-4.5827	0.7723		
INSUL	-14.831	4.754		
AGE	6.101	4.021		

a) State the estimated multiple regression equation.

b) Find the values (a), (b), (c), (d), (e) in the ANOVA table above:

(a) =

(b) =

(c) =

(d) =

(e) =

c) Find and interpret the meanings of multiple coefficient of determination R^2 and the standard error of the estimate s in the context of the present application.

d) Test the appropriate hypothesis to determine if there is a significant overall relationship between the monthly cost of heating and the three independent variables. Use a level of significance of .05.

e) Interpret the coefficient of TEMP in the context of the present application.

f) At 5% significance level, is there any evidence that thicker insulation in the attic in general leads to lower cost of heating in single-family homes? Justify your answer with a statistical test.

g) Estimate the heating cost if the mean outside temperature is 30, the thickness of insulation in the attic is 5 inches, and the age of the furnace is 10 years