Concordia University
John Molson School of Business

Department of Decision Sciences and Management Information Systems

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 (<u>one mark</u> for each of questions 1 to 20 and <u>two marks</u> for each of questions 21 to 34). Some of the numbers in the provided choices have been rounded.

<u>Indicate your answers on the multiple choice answer sheet provided</u>. 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 than75% 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 $\sigma / \sqrt{n}$ 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?
    a) $P(A)P(A|B)=P(B)P(B|A)$
    b) $P(A \cup B) < P(A) + P(B)$
    c) $P(A \mid B^c) = P(B \mid A^c)$
    d) All of the above are true

13. Which of the following statements is true?
    a) A and B are independent
    b) A and B are mutually exclusive
    c) A and B cannot occur at the same time
    d) 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?
    a) 97% of the sampled total compensation values fell between $2,181,260 and $5,836,180.
    b) 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.
    c) We are 97% confident that the mean of the sampled CEO's falls in the interval $2,181,260 to $5,836,180.
    d) 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?
    a) There is insufficient information to determine the p-value since the sample size is not given
    b) The p-value is equal to 0.0548
    c) The p-value is equal to 0.0274
    d) 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
    a) will have a SSE no bigger than that of the original sample regression equation
    b) will have a smaller coefficient of determination
    c) will have a smaller SSR
    d) 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?
    a) $600
    b) $510
    c) $586
    d) cannot be determined with the given information

Introduction to probability

MULTIPLE CHOICE

1. Each individual outcome of an experiment is called
   a. the sample space
   b. a sample point
   c. an experiment
   d. an individual

   ANS: B PTS: 1 TOP: Probability Concepts

2. A graphical method of representing the sample points of an experiment is
   a. a frequency polygon
   b. a histogram
   c. an ogive
   d. a tree diagram

   ANS: D PTS: 1 TOP: Probability Concepts

3. Any process that generates well-defined outcomes is
   a. an event
   b. an experiment
   c. a sample point
   d. a sample space

   ANS: B PTS: 1 TOP: Probability Concepts

4. In statistical experiments, each time the experiment is repeated
   a. the same outcome must occur
   b. the same outcome can not occur again
   c. a different outcome may occur
   d. a different out come must occur

   ANS: C PTS: 1 TOP: Probability Concepts

5. The counting rule that is used for counting the number of experimental outcomes when n objects are selected from a set of N objects where *order of selection* **is not** important is called
   a. permutation
   b. combination
   c. multiple step experiment
   d. None of these alternatives is correct.

   ANS: B PTS: 1 TOP: Probability Concepts

6. From a group of six people, two individuals are to be selected at random. How many possible selections are there?
   a. 12
   b. 36
   c. 15
   d. 8

   ANS: C PTS: 1 TOP: Probability Concepts

7. A method of assigning probabilities based upon judgment is referred to as the
   a. relative method
   b. probability method
   c. classical method
   d. subjective method

   ANS: D PTS: 1 TOP: Probability Concepts

8. A graphical device used for enumerating sample points in a multiple-step experiment is a
   a. bar chart
   b. pie chart
   c. histogram
   d. None of these alternatives is correct.

   ANS: D PTS: 1 TOP: Probability Concepts

9. The set of all possible outcomes of an experiment is
   a. an experiment
   b. an event
   c. the population
   d. the sample space

   ANS: D PTS: 1 TOP: Probability Concepts

10. If a dime is tossed four times and comes up tails all four times, the probability of heads on the fifth trial is
    a. smaller than the probability of tails
    b. larger than the probability of tails
    c. 1/2
    d. 1/32

    ANS: C PTS: 1 TOP: Probability Concepts

11. Of five letters (A, B, C, D, and E), two letters are to be selected at random. How many possible selections are there?
    a. 20
    b. 7
    c. 5!
    d. 10

    ANS: D PTS: 1 TOP: Probability Concepts

12. Assume your favorite football team has 2 games left to finish the season. The outcome of each game can be win, lose or tie. The number of possible outcomes is
    a. 2
    b. 4
    c. 6
    d. 9

    ANS: D PTS: 1 TOP: Probability Concepts

13. An experiment consists of tossing 4 coins successively. The number of sample points in this experiment is
    a. 16
    b. 8
    c. 4
    d. 2

    ANS: A PTS: 1 TOP: Probability Concepts

14. Since the sun **must** rise tomorrow, then the probability of the sun rising tomorrow is
    a. much larger than one
    b. zero
    c. infinity
    d. None of these alternatives is correct.

    ANS: D PTS: 1 TOP: Probability Concepts

15. If a coin is tossed three times, the likelihood of obtaining three heads in a row is
    a. zero
    b. 0.500
    c. 0.875
    d. 0.125

    ANS: D PTS: 1 TOP: Probability Concepts

16. Of the last 100 customers entering a computer shop, 25 have purchased a computer. If the classical method for computing probability is used, the probability that the next customer will purchase a computer is
    a. 0.25
    b. 0.50
    c. 1.00
    d. 0.75

    ANS: B PTS: 1 TOP: Probability Concepts

17. A six-sided die is tossed 3 times. The probability of observing three ones in a row is
    a. 1/3
    b. 1/6
    c. 1/27
    d. 1/216

    ANS: D PTS: 1 TOP: Probability Concepts

18. A perfectly balanced coin is tossed 6 times and tails appears on all six tosses. Then, on the seventh trial
    a. tails can not appear
    b. heads has a larger chance of appearing than tails
    c. tails has a better chance of appearing than heads
    d. None of these alternatives is correct.

    ANS: D PTS: 1 TOP: Probability Concepts

19. A method of assigning probabilities which assumes that the experimental outcomes are equally likely is referred to as the
    a. objective method
    b. classical method
    c. subjective method
    d. experimental method

    ANS: B PTS: 1 TOP: Probability Concepts

20. The probability assigned to each experimental outcome must be
    a. any value larger than zero
    b. smaller than zero
    c. at least one
    d. between zero and one

    ANS: D PTS: 1 TOP: Probability Concepts

21. Some of the CDs produced by a manufacturer are defective. From the production line, 5 CDs are selected and inspected. How many sample points exist in this experiment?
    a. 10
    b. 25
    c. 30
    d. 32

    ANS: D PTS: 1 TOP: Probability Concepts

22. Assume your favorite football team has 3 games left to finish the season. The outcome of each game can be win, lose, or tie. How many possible outcomes exist?
    a. 7
    b. 27
    c. 36
    d. 64

    ANS: B PTS: 1 TOP: Probability Concepts

23. From nine cards numbered 1 through 9, two cards are drawn. Consider the selection and classification of the cards as odd or even as an experiment. How many sample points are there for this experiment?
    a. 2
    b. 3
    c. 4
    d. 9

    ANS: C PTS: 1 TOP: Probability Concepts

24. If a six sided die is tossed two times, the probability of obtaining two "4s" in a row is
    a. 1/6
    b. 1/36
    c. 1/96
    d. 1/216

    ANS: B PTS: 1 TOP: Probability Concepts

25. The intersection of two mutually exclusive events
    a. can be any value between 0 to 1
    b. must always be equal to 1
    c. must always be equal to 0
    d. can be any positive value

    ANS: C PTS: 1 TOP: Probability Concepts

26. The range of probability is
    a. any value larger than zero
    b. any value between minus infinity to plus infinity
    c. zero to one
    d. any value between -1 to 1

    ANS: C PTS: 1 TOP: Probability Concepts

27. Two events, A and B, are mutually exclusive and each have a nonzero probability. If event A is known to occur, the probability of the occurrence of event B is
    a. one
    b. any positive value
    c. zero
    d. any value between 0 to 1

    ANS: C PTS: 1 TOP: Probability Concepts

28. The sum of the probabilities of two complementary events is
    a. Zero
    b. 0.5
    c. 0.57
    d. 1.0

    ANS: D PTS: 1 TOP: Probability Concepts

29. One of the basic requirements of probability is
    a. for each experimental outcome E_i, we must have $P(E_i) \geq 1$
    b. $P(A) = P(A^c) - 1$
    c. if there are k experimental outcomes, then $\sum P(E_i) = 1$
    d. $\sum P(E_i) \geq 1$

    ANS: C PTS: 1 TOP: Probability Concepts

30. The symbol $\cup$ shows the
    a. union of events
    b. intersection of two events
    c. sum of the probabilities of events
    d. sample space

    ANS: A PTS: 1 TOP: Probability Concepts

31. The union of events A and B is the event containing
    a. all the sample points belonging to B or A
    b. all the sample points belonging to A or B
    c. all the sample points belonging to A or B or both
    d. all the sample points belonging to A or B, but not both

    ANS: C PTS: 1 TOP: Probability Concepts

32. If A and B are mutually exclusive events with P(A) = 0.3 and P(B) = 0.5, then $P(A \cap B)$ =
    a. 0.30
    b. 0.15
    c. 0.00
    d. 0.20

    ANS: C PTS: 1 TOP: Probability Concepts

33. Events A and B are mutually exclusive with P(A) = 0.3 and P(B) = 0.2. Then, $P(B^c)$ =
    a. 0.00
    b. 0.06
    c. 0.7
    d. 0.8

    ANS: D PTS: 1 TOP: Probability Concepts

34. In an experiment, events A and B are mutually exclusive. If P(A) = 0.6, then the probability of B
    a. cannot be larger than 0.4
    b. can be any value greater than 0.6
    c. can be any value between 0 to 1
    d. cannot be determined with the information given

    ANS: A PTS: 1 TOP: Probability Concepts

35. If P(A) = 0.62, P(B) = 0.47, and $P(A \cup B)$ = 0.88, then $P(A \cap B)$ =
    a. 0.2914
    b. 1.9700
    c. 0.6700
    d. 0.2100

    ANS: D PTS: 1 TOP: Probability Concepts

36. If P(A) = 0.7, P(B) = 0.6, $P(A \cap B)$ = 0, then events A and B are
    a. not mutually exclusive
    b. mutually exclusive
    c. independent events
    d. complements of each other

    ANS: B PTS: 1 TOP: Probability Concepts

37. Two events with nonzero probabilities
    a. can be both mutually exclusive and independent
    b. can not be both mutually exclusive and independent
    c. are always mutually exclusive
    d. are always independent

    ANS: B PTS: 1 TOP: Probability Concepts

38. If A and B are independent events with P(A) = 0.65 and P(A $\cap$ B) = 0.26, then, P(B) =
    a. 0.400
    b. 0.169
    c. 0.390
    d. 0.650

    ANS: A PTS: 1 TOP: Probability Concepts

39. If two events are independent, then
    a. they must be mutually exclusive
    b. the sum of their probabilities must be equal to one
    c. their intersection must be zero
    d. None of these alternatives is correct.

    ANS: D PTS: 1 TOP: Probability Concepts

40. The multiplication law is potentially helpful when we are interested in computing the probability of
    a. mutually exclusive events
    b. the intersection of two events
    c. the union of two events
    d. conditional events

    ANS: B PTS: 1 TOP: Probability Concepts

41. If A and B are independent events with P(A) = 0.4 and P(B) = 0.6, then P(A $\cap$ B) =
    a. 0.76
    b. 1.00
    c. 0.24
    d. 0.20

    ANS: C PTS: 1 TOP: Probability Concepts

42. If A and B are independent events with P(A) = 0.05 and P(B) = 0.65, then P(A $\mid$ B) =
    a. 0.05
    b. 0.0325
    c. 0.65
    d. 0.8

    ANS: A PTS: 1 TOP: Probability Concepts

43. If A and B are independent events with P(A) = 0.4 and P(B) = 0.25, then P(A $\cup$ B) =
    a. 0.65
    b. 0.55
    c. 0.10
    d. 0.75

    ANS: B PTS: 1 TOP: Probability Concepts

44. If A and B are independent events with P(A) = 0.38 and P(B) = 0.55, then P(A $\mid$ B) =
    a. 0.209
    b. 0.000
    c. 0.550
    d. 0.38

    ANS: D PTS: 1 TOP: Probability Concepts

45. If A and B are independent events with P(A) = 0.35 and P(B) = 0.20, then, P(A $\cup$ B) =
    a. 0.07
    b. 0.62
    c. 0.55
    d. 0.48

    ANS: D PTS: 1 TOP: Probability Concepts

COMM 215: Business Statistics
Solution to Practice Problems 2

Sampling Distributions

1 a) $P(\bar{x} < 240)$ or $P(\bar{x} > 260) = P\left(z < \dfrac{240-250}{7.5/\sqrt{20}}\right) + P\left(z > \dfrac{260-250}{7.5/\sqrt{20}}\right)$

$P(z < -5.96) + P(z > 5.96) \approx 0$

b) $z = 1.04, \quad \dfrac{x-250}{7.5} = 1.04, \Rightarrow x = 257.8$

2 $\sigma/\sqrt{n} = 5.5/\sqrt{50} = 0.778$ $\quad P(20 \le x \le 23) =$

$P\left(\dfrac{20-22}{0.778} \le z \le \dfrac{23-22}{0.778}\right) = P(-2.57 \le z \le 1.285) = 0.8964$

3 $P(\bar{p} \ge 0.10) = P\left(z \ge \dfrac{0.10-0.08}{\sqrt{\dfrac{(0.08)(0.92)}{400}}}\right) = 0.0708$ $\quad \sigma_{\bar{p}} = 0.01356$

4 a) $\bar{p} = 15/100 = 0.15$

b) $P(\bar{p} \le .015) = P\left(z \le \dfrac{0.15-0.25}{\sqrt{\dfrac{(.25)(.75)}{100}}}\right)$

$P(z \le -2.31) = .5 - .4896 = .0104$

5 a) $P(\bar{x} \ge 26.2) = P\left(z \ge \dfrac{26.2-25}{3/\sqrt{36}}\right) = P(z \ge 2.4) = .5 - .4918 = .0082$

b) No, the probability is very small and is therefore very unlikely.

6 $P(\bar{p} \ge .75) = P\left(z \ge \dfrac{.75-.70}{\sqrt{\dfrac{(.70)(.30)}{200}}}\right) = P(z \ge 1.543) = 0.0618$

7 a) $\sigma_{\bar{p}} = \sqrt{\dfrac{(.10)(.90)}{400}} = 0.015$

b) $P(.09 \le \bar{p} \le .10) = P\left(\dfrac{.09-.10}{.015} \le z \le \dfrac{.10-.10}{.015}\right) =$

$P(-0.67 \le z \le 0) = 0.2486$

c) $P(\bar{p} < .08) = P\left(z < \dfrac{.08-.10}{.015}\right) = P(z < -1.33) =$

$.5 - .4082 = 0.0918$

Estimation and Hypothesis Testing

1 a) $H_0 : \mu = 105 \quad t_{.01,14d.f.} = -2.264 \quad t = \dfrac{95.53 - 105}{15.39/\sqrt{15}} = -2.38$

$H_1 : \mu < 105$

$-2.38 > -2.626$ do not reject H_0

insufficient evidence to conclude that mean < 105

b) $t_{.05,14d.f.} = -1.761 \quad -2.38 < -1.761$ you would reject H_0

c) .01<p<.025

2 a) $n = \dfrac{(1.645)^2(.6)(.4)}{(.03)^2} = 721.6 = 722$

b) $n = \left(\dfrac{1.645(5)}{(.5)}\right)^2 = 270.6 = 271$

c) $110 \pm 1.96\left(6/\sqrt{70}\right)$, or 110 ± 1.406, or (108.59,111.406)

3 a) $\begin{cases} H_0 : \mu = 10.8 \\ H_1 : \mu < 10.8 \end{cases} \quad t = \dfrac{10.2 - 10.8}{1.25/\sqrt{16}} = -1.92$; p-value: $0.025 < p < 0.05$

REJECT H_0 at α=.05 level of significance.

b) $t_{.025,15d.f.} = 2.131$ Since -1.92>-2.131 DO NOT REJECT H_0

4 type I : Approving a below standard shipment

type II: Refusing a shipment that is up to standard

5 a) $73 \pm 3.182\left(\dfrac{1.414}{\sqrt{4}}\right) \quad 73 \pm 2.25$

b) $H_0 : \mu = 77$

$H_1 : \mu < 77$ reject H_0 if t<-2.353

$t = \dfrac{73 - 77}{1.414/\sqrt{4}} = -5.6577$

Reject H_0 and conclude that there is sufficient evidence to support the mean emission is less than 77 mg per cubic meter of exhaust.

6 a) $H_0: \mu = 5$ Reject H_0 if t >1.1812

$H_1: \mu > 5$

$t = \dfrac{6.09 - 5}{6.41/\sqrt{11}} = 0.564$. DO NOT Reject H_0.

Insufficient evidence to support the airline's claim. p-value >.10

b) $6.09 \pm 2.228\left(\dfrac{6.41}{\sqrt{11}}\right)$ 6.09 ± 4.306 (1.784, 10.396)

7 $H_0: p = 0.04$ Reject H_0 if z<-2.33

H_1 :p<0.04 $z = \dfrac{.10-.04}{\sqrt{\dfrac{(.04)(.96)}{80}}} = 2.74$

Since 2.74 > -2.33, DO NOT reject H_0

No evidence that p< 0.04. p-value= p(z<2.74)=.9969

8 $H_0: p = .51$

$H_1: p \neq .51$

$p = \dfrac{690}{1335} = .5169$ Reject H_0 if $z > |1.645|$

$z = \dfrac{.5169 - .51}{\sqrt{\dfrac{(.51)(.49)}{1335}}} = .5044$. DO NOT REJECT H_0. Insufficient evidence. Reject researcher's claim.

9 Assume standardized test scores follow a normal distribution with μ=100, σ=10.

$\begin{cases} H_0: \mu = 100 \\ H_1: \mu > 100 \end{cases}$ Test Statistic: $z = \dfrac{\bar{X} - 100}{10/\sqrt{n}}$; Rejection region: $Z > Z_{0.025} = 1.96$

with n=25, $\bar{X} = 103$, $z = \dfrac{103-100}{10/\sqrt{25}} = +1.5$. Since $Z=1.5 < Z_{0.025} = 1.96$, do not reject at α=2.5%.

Insufficient evidence to support the claim of above average intelligence. p-value = 0.0668.

10 a) $n = \dfrac{(2.58)^2(.5)(.5)}{(.02)^2} \approx 4160.25$; n=4161

b) $0.48 \pm 2.58\sqrt{\dfrac{(.48)(.52)}{4161}}$, or 0.48 ± 0.01998, or (.46, .50)

11 a) $60 \pm 1.96\left(7.5/\sqrt{100}\right)$

$60 \pm 1.47 \rightarrow 58.53\%$; 61.47% true mean falls between 58.53% and 61.47% 95% of the time

b) $.07 \pm 1.96\sqrt{\dfrac{(.07)(.93)}{100}} \rightarrow .07 \pm .05 \rightarrow (0.02, 0.12) \rightarrow$ 95% confidence

True proportion of failing is between 2% and 12%

c) $n = \dfrac{(2.575)^2(7.5)^2}{(5)^2} = 14.92 = 15$ and $n = \dfrac{(2.575)^2(.07)(.93)}{(.1)^2} = 43.7 = 44$

To satify both conditions n must be at least 44.

12 a) *i*) $H_0 : p = 0.25$ Reject H_0 if z>$|1.96|$

H_1 :p $\neq 0.25$

$\hat{p} = \frac{15}{35} = 0.4286$ $\quad z = \frac{.4286-.25}{\sqrt{\frac{(.25)(.75)}{35}}} = 2.44$

Since 2.44>1.96 reject H_0 and conclude that proportion using visual basic is different from 25%.

ii) p-value=$P(z \geq 2.44) \times 2 = .0073 \times 2 = 0.0146$

b) *i*) average cost > \$40,000 if average day > $\frac{40000-10000}{1200} = 25$

$H_0 : \mu = 25$ Reject H_0 if $Z > Z_{0.05} = 1.645$

$H_1 : \mu > 25$

$$z = \frac{27.2 - 25}{5.5/\sqrt{35}} = 2.37$$

Since $Z=2.37 > Z_{0.05}=1.645$ Reject H_0. Conclude average day >25.

The evidence is sufficient suggesting that mean will exceed \$40,000.

ii) p-value $P(z \geq 2.37)=.5-.4911=0.0089$.

13 a) $\bar{x}=237.829$ s=36.369

$$237.83 \pm 1.645\left(\frac{36.37}{\sqrt{35}}\right) \qquad 237.83 \pm 10.11$$

(227.72, 247.94)

b) $\hat{p}$=more than 240 seconds $= 15/35 = 0.4285$

$$0.43 \pm 2.575\sqrt{\frac{(.43)(.57)}{35}}, \quad \text{or} \quad 0.43 \pm 0.2155, \text{ or } (0.2145, 0.6455)$$

14 $$n = \left(\frac{1.96(175.5)}{20}\right)^2 = 295.8 \rightarrow 296$$

15 $$n = \frac{(1.88)^2(.5)(.5)}{(.055)^2} = 292.099 \rightarrow 293$$

16 i) $\hat{p}=\frac{800-240}{800}=0.70 \quad \sigma_{\hat{p}}=\sqrt{\frac{(.70)(.30)}{800}}=0.0162$

$0.70\pm2.33(.0162)$ or 0.70 ± 0.0377 (.6623, .7377)

ii) $2.45\pm1.96\left(1.3/\sqrt{800}\right)$ $2.45\pm.09$ (2.36, 2.54)

17 $n=\left(\frac{2.575(300)}{45}\right)=294.65=295$

18 a) $18.30\pm t(.025, 8\text{ d.f})\left(6.3/\sqrt{9}\right)$

18.30 ± 4.8426 ($13.46, $23.14)

b) normally distrubuted population

c) $\hat{p}=\frac{49}{80}=0.6125$

$0.6125\pm2.17\sqrt{\frac{.6125(.3875)}{80}}$

$0.6125\pm.1182$

(0.4943, 0.7307)

d) $n=\left(\frac{(1.96)(6.30)}{1.00}\right)^2=152.47$, the sample size should be increased to 153.

19 a) $H_0:\mu=2.8$ Reject H_0 if $z<-1.645$

$H_1:\mu<2.8$ or if $t<-1.66$

$z=\frac{2.61-2.8}{0.9/\sqrt{100}}=-2.11$

Since $-2.11<-1.645$ Reject H_0 significant evindence that $\mu<2.8$ and therefore promotion was not profitable.

b) p-value $p(z<-2.11)=0.5-0.4826=0.0174$

or if t $.01<p<.025$

c) Type I since you are rejecting

d) $s/\sqrt{n}=.9/\sqrt{100}=.09$

reduced by half $=.09/2=.045$

$.9/\sqrt{n}=.045$ so $n=400$

Chi-Square Tests

1. $H_0 : p_1 = 0.50,\ p_2=0.25,\ p_3=0.15,\ p_4=0.10$

   H_1 : at least one p ≠ to specified values χ^2

f_i	e_i	χ^2
27	30	0.300
19	15	1.067
11	9	0.444
3	6	1.500

$\chi^2_{0.05,3}$=7.8147. Since 3.311<7.8147, DO NOT REJECT H_0.

Brook's ideas regarding accounts are accurate. (No evidence to reject H_0)

2. H_0: Employment and wage rates are independent

   H_1: Employment and wage rates are NOT independent

	Yes	No	
High	18 (28.74)	40 (29.46)	58
Low	38 (27.26)	17 (27.74)	55
	56	57	113

Re ject if $\chi^2_{0.01,1} = 6.6349$

$\chi^2 = 4.016 + 3.945 + 4.235 + 4.16 = 16.355$

Since 16.355 > 6.6349, reject H_0 and conclude that classification are dependant.

3 a. H_0 : percentage in response categories are independent of type of firm.

H_1 : percentage in response categories are NOT independent.

	Yes	Neutral	No	
US	50 (64.465	57 (46.290)	19 (15.237)	126
Foreign	60 (45.535)	22 (32.702)	7 (10.763)	89
	110	79	26	215

$\chi^2_{.10,2} = 4.605.$ $\chi^2 = 16.06$

Since 16.06 > 4.605, Reject H_0 and conclude H_a. Evidence to indicate the two classification are dependant

b.

Sample proportion: $\bar{p} = 50/126 = .397$

The 90% confidence interval estimate for the percentage of U.S. firms that give hiring preferences to business majors with foreign language skills:

$$0.397 \pm 1.645\sqrt{\frac{(.397)(.603)}{126}}$$

$0.397 \pm .072$ (.325,.469)

4.

H_0 :p_1=p_2=p_3=p_4=p_5=0.20

H_1: at least one p $\neq$ 0.20

e_i for each day =362×.20=72.4 , $\chi^2_{.05,4}$=9.48773

Since χ^2=4.768 <$\chi^2_{.05,4}$= 9.48773, DO NOT REJECT H_0.

Insufficient evidence to say absenteeism is higher on some days.

5.

H_0: preference is independent of experience (no relationship)

$H_{1:}$ preference is NOT independent of experience

$\chi^2_{.05,2}$=5.991. Since χ^2=7.40136 > 5.991, Reject H_0.

Sufficient evidence to conclude that preference and experiences are NOT independent. There is a relationship.

6.

a. H_0 : Scores are independent of gender

H_1 : Scores are NOT independent of gender

Reject H_0 if χ^2>$\chi^2_{.05,4}$=9.48773 [(r-1)×(c-1)=4]

Test statistics: χ^2=1.172 Since 1.172 < 9.487

DO NOT Reject H_0 - Scores are independent of Genders

b. $p(\bar{x} \le 570)=p\left(z \le \frac{570-550}{75/\sqrt{50}}\right)=1.89$

$p(z \le 1.89)=0.470610+0.5=0.9706$

.

7.

a. H_0: Same result obtained by ABC Inc.

H_1: Reason different than obtained by ABC Inc.

Reject H_0 if $\chi^2 > \chi^2_{.05,3}$ = 7.81473

χ^2=1.0526+2.8125+2.3529+6.9231=13.1411

Since χ^2 = 13.1411 > $\chi^2_{.05,3}$ =7.81473,

reject H_0 and conclude that reasons are different.

b.

H_0: Reason for firing is independent to previous warning
H_0: Reason for firing is dependent to previous warning

Reject H_0 if $\chi^2 > \chi^2_{0.10,(4-1)(2-1)} = \chi^2_{0.10,3} = 6.251$.

Since $\chi^2 = \mathbf{5.423} < \chi^2_{0.10,3} = \mathbf{6.251}$, do not reject H_o.

8. a.

$\begin{cases} H_0: \text{Finding a job is independent of working experience} \\ H_1: \text{Finding a job is NOT independent of working experience} \end{cases}$

Reject H_0 if $\chi^2 > \chi^2_{.05,(r-1)(c-1)} = \chi^2_{.05,2} = 5.99$

Test statistics: $\chi^2 = 7.922$

Since $7.922 > 5.99$, Reject H_0.

Finding a job is NOT independent of work experience.

b.

$\begin{cases} H_0: p_1=.70; p_2=.15; p_3=.15 \\ H_1: \text{at least one (or two) p is different} \end{cases}$

Reject H_0 if $\chi^2 > \chi^2_{.10,2} = 4.60517$

Test statistics:

f_i	e_i	$\frac{(f_i - e_i)^2}{e_i}$
52	56	0.28571
16	12	1.33333
12	12	0.00000
80	80	1.61904

Since 1.61904 < 4.605, do not reject H_0; insufficient evidence that proportions differed from the special values.

Simple Linear Regression and Correlation

1 a) $\hat{y}=1.4235+0.53x$

b) $R^2=0821$. That is, 81.2% of the variation in ABC's rate of return is explained by the market of return.

c) $\begin{cases} H_0:\beta_1=0 \\ H_0:\beta_1\neq 0 \end{cases}$ $t=\dfrac{0.53}{0.103}=5.15$. Since $5.15 > 2.447$, reject H_0 and conclude that model is significiant.

d) $\hat{y}=1.4235+0.53(5)=4.07$

$$4.07\pm 3.707(2.8225\sqrt{1+\frac{1}{8}+\frac{(5-2.5)^2}{752}} \quad \Rightarrow \quad 4.07\pm 11.137 \text{ or } (-7.067, 15.207)$$

2 a) $\hat{y}=10.548+0.00578x$

For each additional million dollars, the price per share increases by 0.00578 (in \$) while the initial price (constant) is \$10.55

b) $\begin{cases} H_0:\beta_1=0 \\ H_0:\beta_1\neq 0 \end{cases}$ $t=\dfrac{0.00578}{0.0039}=1.49$

p-value = $p(t \geq 1.49)\times 2 \quad \Rightarrow \quad (.05\times 2) \leq p \leq (.10\times 2)$, or $.10 \leq p \leq .20$

So reject H_0 if $\alpha > .20$, but do not reject H_0 if $\alpha < .10$

c) $R^2=0.219$. That is, 21.9% of the variation in price per share is explained by the size of the offering.

d) $\hat{y}=10.548+0.00578(70)=10.95$. That is, the estimated mean price per share is \$10.95.

The 95% confidence interval estimate of the mean price per share for all companies with a size offering of \$70 million:

$$10.95\pm 2.306(.39)\sqrt{\frac{1}{10}+\frac{(70-64.3)^2}{10155.4}} \quad \text{or} \quad 10.95\pm 0.29 \text{ or } (10.66, 11.24)$$

3 a) $\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$, where

$$SS_{xy} = 21115.07 - \frac{(92.93)(2725)}{12} = 12.2158$$

$$SS_{xx} = 720.22 - \frac{(92.93)^2}{12} = 0.554592$$

$$SS_{yy} = 619207 - \frac{(2725)^2}{12} = 404.917$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{12.2158}{0.554592} = 22.0267$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \left(\frac{2725}{12}\right) - 22.0267\left(\frac{92.93}{12}\right) = 56.5048$$

$\hat{y} = 56.5048 + 22.0267x$, For each additional one percent increase in interest rates, the futures index increase by 22.0267 points.

b) $\begin{cases} H_0: \beta_1 = 0 \\ H_0: \beta_1 \neq 0 \end{cases}$ reject H_0 if $t > |t_{.005, 10}| = 3.169$; test stat: $t = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}}$

$$S_{\hat{\beta}_1} = \frac{S}{\sqrt{SS_{xx}}}, \quad S = \sqrt{\frac{(SS_{yy} - \hat{\beta}_1 SS_{xy})}{(n-2)}} = \sqrt{\frac{404.917 - (22.0267)(12.2158)}{10}} = 3.68567$$

$$S_{\hat{\beta}_1} = \frac{3.68567}{\sqrt{0.554592}} = 4.94915, \qquad t = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}} = \frac{22.0267}{4.94915} = 4.45061$$

Since t=4.5061 > 3.169, reject H_0 at 5% level of significance and conclude that interest rate is a significant predict of futures index.

p value: .001 <p-value < .002

c) $r = \frac{SS_{xy}}{\sqrt{SS_{xx} SS_{yy}}} = \frac{12.2158}{\sqrt{0.554592}\sqrt{404.917}} = 0.815180$

The value of r suggests a strong positive correlation between interest rates and future index.

d) The 95% confidence interval estimate for the meanfutures index when interest rate is 7.8% :

$\hat{y} = 56.5048+22.0267(7.8) =228.313;$

$$228.313 \pm 2.228(3.68567)\sqrt{\frac{1}{12}+\frac{(7.8-7.74417)^2}{0.554592}}$$

$228.313\pm2.228(3.68567)(\ 0.298252)$

228.313 ± 2.44915 or (225.864, 230.762)

Multiple Regression

1 a) Coefficient of $x_2 = -0.021$. The age at death will decrease by 0.021 years for every additional unit of cholestrol level.

b) $\begin{cases} H_0: \beta_1=\beta_2=\beta_3=0 \\ H_1: \text{ at least one } \beta_i\neq 0 \end{cases}$ Test statistics: $F=\frac{MSR}{MSE}$

Rejection region: $F > F_{.05,\,3,\,36} = 2.84$

ANOVA table:

df	SS	MS	F
3	939	313	3.49
36	3227	89.64	
39	4166		

Since $F=3.49 > F_{.05,\,3,\,36} = 2.84$

Reject H_0 and conclude that model is significant in predicting lenght of life.

c) $\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$ Test statistics: $t = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}} \sim t_{0.005,36}$

Rejection region: $|t| > t_{0.005,36} = 2.724$

$t = \frac{1.79}{0.44} = 4.068$

Since t= 4.068 > 2.724, reject H_0 at 1% level of significance and conclude that average number of hours of exercise and age at death are linearly related.

d) $R^2 = 0.225$. That is, 22.5% of the variation in age at death is explained by x_1, x_2 and x_3.

e) $\hat{y} = 55.8 + 1.79\,(8) - 0.021\,(0.5) - 0.016\,(10) = 69.95$ years

f) x_1 = average hours of exercise per week because it has largest absolute value of t of 4.068.

2 a) $\hat{y} = 10 + 2.1(1.5) + 13.6(3) = \$53.95$

b) $R^2 = \frac{90.400 - 43.912}{90.400} = 0.5142$. That is, 51.42% of the variation in store price is explained by variations in current dividents and rate of growth.

c) $\begin{cases} H_0: \beta_1 = \beta_2 = 0 \\ H_1: \text{at least one } \beta_i \neq 0 \end{cases}$ Test Statistic: $F = \frac{MSR}{MSE} \sim F_{0.05,2,7}$

Rejection region: $F > F_{0.05,2,7} = 4.47$

$F = \frac{23244}{6273.143} = 3.7$

Since $F = 3.7 < F_{0.05,2,7} = 4.47$, do not reject H_0. The model is not significiant.

d) $2.1(2.25) = \$4.725$ increase

3 a) $15232.5 + 2178.4x_1 + 7.8x_2 + 2675.2x_3 + 1157.8x_4$

b) for each additional room (x_1), the value of the house will increase by $ 2,178.4

c) DF for $\begin{cases} SSR & 4 \\ SSE & 25 \end{cases}$

d) $\begin{cases} H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0 \\ H_1: \text{At least one } \beta_j \neq 0 \end{cases}$ Test statistic: $F = \frac{MSR}{MSE} \square F_{0.05;4,25}$

Rejection region: $F > F_{0.05;4,25} = 2.76$

$$F = \frac{51060.72}{8235.60} = 6.2$$

Since $F = 6.2 > F_{0.05;4,25} = 2.76$, reject H_0 and conclude that the model is significant at 5% level.

e) $\begin{cases} H_0: \beta_1 = 0 \\ H_1: \beta_1 \neq 0 \end{cases}$ Test statistics: $t = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}} \square t_{0.025,25}$

Rejection region: $|t| > t_{0.025,25} = 2.060$

$$t = \frac{2178.4}{778.0} = 2.8$$

Since $t = 2.8 > t_{0.025,25} = 2.060$, reject H_0 at 5% level and conclude that β_1 significant.

The number of rooms (X_1) is an important predictor of value of a house (Y).

f) $R^2 = \frac{204,242.88}{410,132.88} = 0.49799.$

That is, 49.8% of the variation is the value of a house is explained by the 4 independent variables.

g) $\hat{y} = 15,232.5 + 2,178.4(9) + 7.8(7.500) + 2.675.2(2) + 1.157.8$

$= \$99,846.3$

COMM 215: BUSINESS STATISTICS

REVIEW PROBLEMS 2

Sampling Distributions

1. The size of servings of soft ice cream in a fast-food chain is monitored carefully so that the mean serving is 250g, with a standard deviation of only 7.5g. Treat the weights as being normally distributed.
   a. Inspectors select 20 servings at random, and weigh them. What is the probability that the mean weight is either below 240g or above 260g?
   b. At least how much ice cream is there in a serving that is among the heaviest 15% of servings?

2. The length of time required to inspect shipments of transistors at Acme Inc. can be treated as normally distributed with a mean of 22 minutes and a standard deviation of 5.5 minutes. If 50 inspections are selected at random to be timed, what is the standard error of the mean time of inspection? What is the probability that the mean time is between 20 and 23 minutes?

3. The inspections at Acme reveal that 8% of the transistors (in the above problem) are defective. In a sample of 400 transistors, what is the probability that at least 10% of them are defective? What is the standard error of the proportion of defectives?

4. Based on past experience, 25% of the contacts made by a firm's sales representatives result in a sale being made. Jane has contacted 100 potential customers, but has made only 15 sales. Assume that Jane's contacts represent a simple random sample of those who could have been called upon. Given this information
   a. What is the sample proportion p = proportion of contacts that resulted in a sale being made?
   b. For simple random samples of this size, what is the probability of finding a sample this small or smaller?

5. When a production machine is properly calibrated, it requires an average of 25 seconds per unit produced, with a standard deviation of 3 seconds. For a random sample of 36 units, the sample mean was found to be 26.2 seconds per unit.
   a. When the machine is properly calibrated, what is the probability that the mean for a sample of this size will be at least 26.2 seconds?
   b. Based on your answer in part a), does it seem likely that the machine is properly calibrated? Explain.

6. Seventy percent of the registered voters in a large community are members of the Progressive Conservative Party. The mayor will be selecting a random sample of 200 registered voters to serve on an advisory panel that helps formulate policy for the community's parks and recreation facilities. What is the probability that the constituency of the committee will be no less than 75% Progressive Conservative?

7. Historically, 10% of a large shipment of machine parts are defective. If random samples of 400 parts are selected,
   a. compute the standard error of the sampling distribution.
   b. what proportion of the samples will have between 9% and 10% defective?
   c. what proportion of the samples will have less than 8% defective?

Estimation and Hypothesis Testing

1. In 1991, the average number of days that homes in Canada stayed on the market before being sold was 105. A real estate broker in Montreal believes that the average number of days that a house in Montreal stays on the market before being sold is less than 105. Fifteen houses that were sold in Montreal in 1991 were randomly selected and the average number of days that the houses stayed on the market was 95.53 days with a standard deviation of 15.39 days. Assume a normal distribution.
    a. Do the data provide sufficient evidence to conclude that the mean number of days that houses in Montreal stayed on the market before being sold is less than 105? Use $\alpha = 0.01$.
    b. Will the conclusion you made in part a) change if the significance level is changed to 0.05?
    c. What is the p-value of the test?

2. According to a recent report, the federal government is requiring local officials to enforce a 100 kilometer/hour speed limit within their jurisdictions. Accordingly, any province may be in jeopardy of losing millions of dollars in federal road funds if more than 60% of its vehicles are found to be exceeding the speed limit. In one province, 70 radar surveys are conducted each year at a total of 50 different sites in order to estimate the proportion p of the vehicles exceeding the 100 kilometers per hour. Each of the samples involve at least 500 vehicles.
    a. How large of a sample should be selected at site #39 in order to estimate p to within 3% with 90% confidence. Last year approximately 60% of all vehicles exceeded the limit.
    b. The average speed of vehicles on major expressways is also of interest to the government. How large of a sample would be necessary at site #39 in order to estimate the mean to within .5 km. with 90% confidence. Assume the standard deviation of speeds is approximately 5km/hr.
    c. Set up a 95% confidence interval estimate of the true average speed of vehicles on the major highways if the 70 surveys mentioned yield a sample mean equal to 110 kilometers/hour and a standard deviation equal to 6 kilometers/hour.

3. The Human Resources Director of Zubrak Inc. has administered a dexterity test to job applicants for a period of several years. She assumes that the times required to complete the test are normally distributed with a mean of 10.8 minutes. The Director would like to know if applicants who have completed a vocational course tend to complete the test faster than the norm. For a sample of 16 applicants with such a course, she observed the average time of 10.2 minutes and standard deviation of 1.25 minutes to complete the test.
    a. Set up appropriate hypotheses, give the p-value and report your conclusions.
    b. Based on the result in (a), should she reject the null hypothesis at $\alpha = .025$? Explain.

4. In an effort to ensure the quality of incoming goods, a buyer will often do acceptance sampling. Samples of 10 items are chosen for inspection and if 2 or more are defective, the entire shipment may be returned as being below standard; otherwise the shipment is approved as being up to standard. Given the hypotheses:
    H_o: shipment is below standard, H_a: shipment is up to standard,
    what are the consequences of a Type I and Type II error?

5. An engine manufacturer is testing the amount of air pollution (in milligrams per cubic metre of exhaust) emitted by a new rotary engine. The manufacturer randomly selects four rotary engines and subjects them to a pollution test. It is found that these four engines emit 71, 74, 73 and 74 milligrams per cubic metre.
    a. Assuming normality, construct a 95% confidence interval for the mean emission of air pollution by the rotary engine.

b. A competing piston engine has a mean air pollution emission of 77 milligrams per cubic metre of exhaust. Would it be reasonable to claim that the mean emission of air pollution for the rotary engine is less than that of the competing piston? Use $\alpha = .05$.

6. An airline claims that average delay in the arrival time of its local flights is more than 5 minutes. Eleven flights of this airline are randomly selected, showing the following scheduled and actual arrival times. Assume that flight delays are normally distributed.

	ARRIVAL TIME	
FLIGHT	ACTUAL	SCHEDULED
1	16:38	16:30
2	07:35	07:35
3	17:23	17:20
4	18:33	18:10
5	18:45	18:40
6	15:19	15:10
7	12:53	12:50
8	13:21	13:15
9	09:14	09:10
10	23:32	23:25
11	21:09	21:10

a. Test the airline's claim at 5% level of significance. State the p-value of the test.
b. Construct a 95% confidence interval estimate for the mean delay in arrival time.

7. A supplier of inexpensive floppy disks claims that less than 4% of the disks are defective. In a random sample of 80 disks, it is found that 10% are defective, but the supplier claims that this is only a sample fluctuation. At the 1% level of significance, test the supplier's claim. State the p-value.

8. According to data gathered by researchers, the probability of an expectant mother having a boy is 51%. Records from a hospital's maternity ward indicate that 1335 babies were born last year. Of these, 690 were boys. Can you reject the researcher's findings at $\alpha = 0.10$?

9. A student claims that his high school is attended by students of above average intelligence. To test this claim, the standard IQ test is administered to 25 students selected at random. These students produced an average IQ score of 103. If previously published studies have standardized IQ scores at mean = 100 and standard deviation = 10, test this student's claim at a 2.5% level of significance. State the p-value?

10.
a. A new polling company wants to publish confidence intervals that have some advantage over those published by the established companies. The decision has been taken to use a narrower interval, namely one with an error of only 0.02. If the confidence interval for the proportion is to have a confidence level of 99%, what minimum sample size is needed?
b. If a sample is taken, and the proportion of supporters of the Platform Party is found to be 48%, what is a 99% confidence interval estimate for the overall proportion of Platform Party supporters? Use the sample size determined in part (a).

11. Last semester there was a COMM 215 class where 100 students wrote the final exam. The average grade on the final exam for this class was 60 percent with a standard deviation of 7.5%. In this class of 100 students, there were 7 students who failed the course because they did not achieve the required minimum grade of 40% on the final exam. Assuming this class to be representative of all students writing the COMM 215 final exam last semester, determine the following:
    a. A 95% confidence interval for the true average grade on the COMM 215 final exam.
    b. A 95% confidence interval for the true proportion of students failing COMM 215 last semester because they did not achieve a minimum grade of 40% on the final exam.
    c. We want to estimate the true average final exam grade to within 5 percent and the true proportion of students not achieving a minimum final exam grade of 40% to within 0.1. With 99% confidence, what is the smallest sample size that can be used to satisfy both of these criteria?

12. A software developer is in the process of estimating the average cost for its new mid-size projects in the coming year. While it is reasonable to assume that the average time taken to complete a project remains the same for the coming year, the cost (in dollars) incurred for a project should be calculated differently, and is given by **10000+1200*x***, where x is the number of days taken to complete the project. A sample of 35 past mid-size projects is taken. It is found that the mean and standard deviation of project completion time for this sample are, respectively, 27.2 days and 5.5 days. Furthermore, 15 of these projects used Visual Basic as the main tool of the project.

    a. i) At 5% significance level, is there any evidence that the proportion of projects using Visual Basic as the main tool is different from 0.25? Justify your answer.

    ii) Calculate the p-value of the test statistic.

    b. i) At 5% significance level, is there any evidence that the mean cost for the mid size projects will exceed $40000 in the coming year? Justify your answer.

    ii) Calculate the p-value of the test statistic.

13. The manager of a chain of pizza restaurants monitors operators taking orders by phone. The average time for a phone call, in seconds, is used as one summary of activity for a restaurant. Phone calls that take more than 4 minutes are considered to be inefficient since they tend to tie up phone lines and in the long run result in a significant reduction in revenue. The talk times (in seconds) of a random sample of operators are:

    195 223 230 237 271 239 275 262 226 275 279 214 176 208 189
    245 199 200 302 222 189 223 285 275 235 190 223 300 276 263
    280 199 245 279 195

$$\left(\sum X_i = 8324, \sum X_i^2 = 2024658\right)$$

    a. Estimate with a 90% confidence interval, the average talk time for all orders taken by phone. Interpret your answer.
    b. Estimate with a 99% confidence interval, the proportion of inefficient orders taken by phone. Interpret your answer.

14. Because of delays between the time an invoice is sent and payment is received from foreign customers, fluctuating exchange rates can lead to increased or decreased profits. An accountant decides to evaluate the magnitude of this component of profit during the current year. From the past year's experience, she thinks it is reasonable to assume that the standard deviation is $175.5 per

transaction. How many transactions must the accountant review to have an error margin less than 2000 cents, 19 times out of 20, when estimating the true mean?

15. A researcher is interested in determining the proportion of current users of a product that are likely to try a new related product. How many persons should be included in a sample if the researcher wants to be 94% certain that the error of estimation does not exceed 5.5%?

16. A random sample of 800 graduating university students were interviewed to find out whether or not they were planning to pursue graduate studies. Of the students selected, 240 students were planning to undertake graduate studies. The results also showed that these 800 students had an average GPA of 2.45 with a standard deviation of 1.3.
    i) Obtain a 98% confidence interval estimate for the proportion of graduating university students who were <u>not</u> planning to undertake graduate studies.
    ii) Obtain a 95% confidence interval estimate for the mean GPA of graduating university students.

17. A company is considering a change in group health insurance. The decision rests in part on the average annual family expenditure on medical expenses. The company wants to be 99% confident that the sample estimate will be within $45 of the true average annual family expenditure on medical expenses. A pilot study has been conducted and has yielded a standard deviation of $300. What optimal sample size should the company select in order to make a final decision regarding the group health insurance?

18. Mark Semmers, owner of the Aurora Restaurant, is considering purchasing new furniture. To help him decide on the amount he can afford to invest in tables and chairs, he wishes to determine the average revenue per customer. The checks for 9 randomly sampled customers had an average of $18.30 with a standard deviation of $6.30.
    a. Construct a 95% confidence interval for the size of the average check per customer.
    b. What assumption is needed to be made in part a?
    c. Mr. Semmers conducted a second survey in order to determine the proportion of customers that spend over $20. In a random sample of 80 customers, he found that 49 of them spent over $20. Construct a 97% confidence interval estimate of the proportion of customers that spend over $20.
    d. What additional sample size would be required in part a) if Mr. Semmers would like to be 95% confident that the maximum error in his estimation would not exceed $1.00?

19. A book club advertises an introductory offer, through which new members are entitled to receive an initial package of books at a nominal price with no obligation to purchase additional books. The club estimates that this promotion will be profitable if the new members buy, on average, at least 2.8 books during the following year. A random sample of 100 records of members attracted by this offer showed a mean purchase in the next year of 2.61 books and a standard deviation of 0.9 books.
    a. Can you conclude that the promotion was profitable at a .05 level of significance?
    b. What is the p-value of the above test?
    c. According to your answer in part a), would you be subject to making a Type I or a Type II error? Explain.
    d. If the book club wants to reduce the standard error (used in part a) by half, how many records should be sampled?

Chi Square Tests

1. Brook is in charge of the accounts receivable department of COREX, Inc. An accountant wishes to check on the department and asks Brook for his ideas about outstanding accounts. He replies that he believes that accounts should be about as follows:

Less than 3 months delinquent:	50%
3-6 months delinquent:	25%
6-9 months delinquent:	15%
More than 9 months delinquent:	10%

The accountant, knowing that it is impractical to examine all accounts, chooses a random sample of 60 accounts and finds 27, 19, 11 and 3 accounts in these categories. Can the accountant conclude that Brook's ideas are accurate? Set up appropriate hypotheses and report your conclusion using a 5% level of significance.

2. The theory that changes in employment and wage rates for various occupations are not independent has been proposed. The following data were collected for furniture workers. Do the data support the theory? Test the appropriate hypotheses at a .01 level of significance.

Number of Employees Changing Employment

Wage Rates	Yes	No
High	18	40
Low	38	17

3. As a business major, did you study foreign languages? If so, will your foreign language skills make you more marketable in the business community? To answer these questions, researchers mailed questionnaires to personnel directors of both foreign-based and domestic businesses. The 215 responses to the question of whether a firm would give hiring preference to business majors knowledgeable in foreign languages are summarized in the table:

	YES	NEUTRAL	NO
U.S. FIRMS	50	57	19
FOREIGN FIRMS	60	22	7

a. Conduct an analysis to determine whether the percentages in the response categories for the question depend on the type of firm. Use a 10% level of significance.
b. Construct a 90% confidence interval estimate for the percentage of U.S. firms that give hiring preferences to business majors with foreign language skills.

4. It has been estimated that employee absenteeism costs North American companies more than $100 billion per year. As a first step in addressing the rising cost of absenteeism, the personnel department of a large corporation recorded the weekdays during which individuals in a sample of 362 absentees were away over the past several months. Do these data suggest that absenteeism is higher on some days of the week than on others? Use $\alpha = .05$.

DAYS OF THE WEEK	MON	TUES	WED	THURS	FRI
NUMBER OF ABSENT	87	62	71	68	74

5. An experimental project was undertaken to verify the belief that novice computer users prefer a menu-oriented interface and that experienced users prefer a command-oriented interface. A software product that can be used with both types of interfaces was selected to test this experiment. Novice, trained and experienced computer users were given various tasks to perform and a tally was kept on whether the user selected a menu-oriented or a command-oriented interface. The following table provides the results of the experiment.

EXPERIENCE LEVEL	PREFERENCE	
	COMMAND	MENU
NOVICE	9	25
TRAINED	9	6
EXPERIENCED	8	5

Do the data given provide sufficient evidence to indicate a relationship between experience level and interface preference at the 5% level of significance?

6. The scores on a national achievement test are of considerable interest. The scores from a random sample of 100 students selected from several schools are given in the table below.

Score	**Male**	**Female**
Under 450	**6**	**7**
450 to under 500	**6**	**5**
500 to under 550	**8**	**10**
550 to under 600	**14**	**10**
600 and over	**16**	**18**

a. Do these data provide sufficient evidence to indicate that scores on a national achievement test are related to gender? Use a 5% level of significance.
b. Assume that the population mean of scores is 550 and the standard deviation is 75. From the sample of males given above, what is the probability that the sample mean will not exceed a score of 570?

7. According to an estimate, more than four hundred thousand Canadians are fired from their jobs every year. A study by ABC Inc. identified the following reason for firing (shown in the first column):

	ABC Inc.	Recent Sample
Incompetence	38%	400
Negative attitude	32%	350
Dishonest	17%	150
Other	13%	100

a. Is the current distribution of reasons why employees were fired different from the results obtained by ABC Inc.? Test the hypothesis using a 5% level of significance.
b. Of the 400, 350, 150 and 100 employees interviewed, respectively, 350, 320, 140 and 90 had received a warning from their employer prior to the firing. At a 10% level of significance, can it be concluded that the reason for firing is related to previous warnings given by the employer?

8. A job-training program offers training in advance computer programming. It is suspected that the probability that an individual will be able to find a programming job within six months (after the

completion of the program) depends on previous related working experience.A random sample of 80 individuals trained in the program yields the following frequency table:

Number of Years of Experience	Number of Individuals
Zero	52
One year	16
More than one year	12

Of the 52, 16 and 12 individuals in the three experience categories listed in the table, respectively, 20, 8 and 10 were able to find a programming job within six months.

a. At a 5% level of significance, is there any evidence to support the hypothesis that the probability of finding a job (within six months) depends on prior related working experience?
b. At a 10% level of significance, is there any evidence against the hypothesis that the percentage of individuals in the program with "zero", "one year" and "more than one year" of experience are respectively 70%, 15%, and 15%?

Simple Linear regression and Correlation

1. The desirability of any financial investment depends heavily upon its risk level. Many financial analysts, consultants and investors use a model known as the beta model to identify the risk level of securities. The dependent variable in this model is the rate of return for ABC, while the average rate of return on a portfolio of all securities traded in the market (also known as the market rate of return) is used as the independent variable. The table below shows the rate of return of the common stock of ABC Inc. against the market rate of return:

Rate of return for ABC Inc.(%)	Market rate of return (%)
1	2
-5	-10
12	18
5	9
6	12
7	-1
-6	-12
2	2

Calculations for the above data are provided:

$\Sigma x = 20$ $\Sigma y = 22$ $\Sigma xy = 454$ $\Sigma x^2 = 802$ $\Sigma y^2 = 320$

a. Assuming that the relationship between the two variables is linear, estimate the regression equation.
b. Compute the coefficient of determination and interpret it in the context of this problem.
c. At the 5% level of significance, can you conclude that the simple linear regression model is significant?
d. Construct the 99% prediction interval for ABC Inc.'s rate of return when the market rate of return is 5%.

2. A study of companies going public for the first time is interested in the relationship between the size of the offering and the price per share. A sample of 10 companies that recently went public revealed the following information:

Company	Size in $ millions (X)	Price per Share (Y)
1	9.0	10.8
2	94.4	11.3
3	27.3	11.2
4	71.9	11.1
5	97.9	11.2
6	70.0	10.7
7	96.5	10.6
8	23.5	10.1
9	58.7	10.7
10	93.8	11.5

$\Sigma x = 643$; $\Sigma x^2 = 51{,}500.3$; $\Sigma y = 109.2$; $\Sigma y^2 = 1{,}194.02$; $\Sigma xy = 7{,}080.29$.

a. Determine the estimated regression equation for the price per share on size. Interpret the meaning of the slope and intercept in the context of the problem.
b. Test for the significance of the relationship between the price per share and the size of the offering. Should the size of an offering be used as a predictor of the price per share? Use the p-value to report your conclusion.
c. Compute and interpret the coefficient of determination.
d. Construct a 95% confidence interval estimate of the mean price per share for all companies with a size offering of $70 million.

3. A study was conducted in order to determine the relationship between the interest rate of federal funds and the commodities futures index.

Day	Interest Rate (x, in %)	Futures Index (y)
1	7.43	221
2	7.48	222
3	8.00	226
4	7.75	225
5	7.60	224
6	7.63	223
7	7.68	223
8	7.67	226
9	7.59	226
10	8.07	235
11	8.03	233
12	8.00	241

In addition, the following sums of squares are provided:
$\Sigma x = 92.93$; $\Sigma x^2 = 720.22$; $\Sigma xy = 21115.07$; $\Sigma y = 2725$; $\Sigma y^2 = 619207$

a. Determine the estimated regression equation for predicting the futures index based on the interest rate. Interpret the meaning of the slope in the context of the problem.
b. At the 1% level of significance, can you conclude that the interest rate is a significant predictor of the futures index? Provide the approximate p-value.
c. Compute the correlation coefficient between the interest rate and the futures index. Comment on the meaning of this statistic.
d. Construct the 95% confidence interval estimate for the mean futures index when the interest rate is 7.8%.

Multiple Regression

1. An actuary wanted to develop a model to predict how long individuals will live. After consulting a number of physicians, she collected the age at death (y), the average number of hours of exercise per week (x_1), the cholesterol level (x_2), and the number of points that the individual's blood pressure exceeded the recommended value (x_3). A random sample of 40 individuals was selected. The following table gives the values of the constant and the multiple regression coefficients, as well as their respective standard deviations:

	Coef	*StDev*
Constant	**55.8**	**11.8**
x_1	**1.79**	**0.44**
x_2	**-0.021**	**0.011**
x_3	**-0.016**	**0.014**

Furthermore, the following sums of squares are given:
Sum of Squares due to Regression (SSR) = 939 Total Variation (SST) = 4166

a. Interpret the meaning of the regression coefficient of x_2 in the context of the problem.
b. Is there enough evidence at the 5% significance level to infer that the model is useful in predicting length of life?
c. Is there enough evidence at the 1% significance level to infer that the average number of hours of exercise per week and the age at death are linearly related?
d. What is the coefficient of determination? Explain its meaning in the context of the problem.
e. Predict the age at death of a person who on average spends 8 hours on exercise per week, has a cholesterol level of 0.5, and whose blood pressure exceeds the recommended value by 10 points.
f. Which variable is the best predictor of age at death? Explain.

2. A researcher has developed the following regression equation relating the current price of a company's stock (Yin dollars), its current dividend per share (X_1 in dollars), and its rate of growth of dividends based on previous 5 years (X_2 in percent). The equation is based on a sample of 10 companies.

$$\hat{Y} = 10 + 2.1\ X_1 + 13.6\ X_2$$

For these data, Total Variation = 90,400 and SSE = 43,912

a. Estimate the mean current price of a common stock when the current dividend is $1.50 per share and the rate of growth of dividends is 3 percent.

b. Estimate and interpret the sample coefficient of multiple determination.
c. Is the model significant at 5% level of confidence?
d. If the current dividend per share increases by $2.25 while the rate of growth remains constant, what change would you predict in the price of the stock?

3. A sample of 30 houses that were sold in the last year was taken. The value of the house (Y) was estimated. The independent variables included in the analysis were the number of rooms (X1), the size of the lot (X2), the number of bathrooms (X3), and a dummy variable (X4), which equals 1 if the house has a garage and equals 0 if the house does not have a garage. The following results were obtained:

	Coefficients	Standard Error
Intercept	15,232.5	8,462.5
X1	2,178.4	778.0
X2	7.8	2.2
X3	2,675.2	2,229.3
X4	1,157.8	463.1

ANOVA

	df	*SS*	*MS*
Regression		204,242.88	51,060.72
Error (Residuals)		205,890.00	8,235.60

a. Write out the estimated equation.
b. Interpret the coefficient on the number of rooms (X1).
c. What are the degrees of freedom for the sum of squares explained by the regression (SSR) and the sum of squares due to error (SSE)?
d. Test whether or not there is a significant relationship between the value of a house and the independent variables. Use a .05 level of significance.
e. Test the significance of X1 at the 5% level.
f. Compute the coefficient of determination and interpret its meaning.
g. Estimate the value of a house that has 9 rooms; a lot with an area of 7,500; 2 bathrooms, and a garage.