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 $\mathbf{10000+1200}x$, 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.