# MCDA5520 Practice Questions

# Section I:
## Descriptive Statistics

1. The Highway Loss Data Institute's Injury and Collision Loss Experience report rates car models on the basis of the number of insurance claims filed after accidents. Index ratings near 100 are considered average. Lower ratings are better, indicating a safer car. Shown are ratings for 20 mid-size cars and 20 small cars.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 81 | 91 | 93 | 127 | 68 | 81 | 60 | 51 | 58 | 75 |
| **Midsize** | 100 | 103 | 119 | 82 | 128 | 76 | 68 | 81 | 91 | 82 |
| | 73 | 100 | 127 | 100 | 124 | 103 | 119 | 108 | 109 | 113 |
| **Small** | 108 | 118 | 103 | 120 | 102 | 122 | 96 | 133 | 80 | 140 |

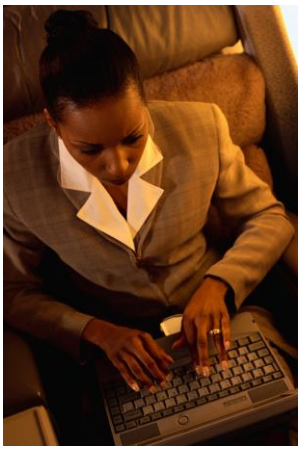
a) Find the mean, median, and mode of the number of accidents filed for small and mid-sized cars.
b) Find the range, inter-quartile range, standard deviation and coefficient of variation of the number of accidents filed for small and mid-sized cars.
c) Using the information in a) and b), offer a viewpoint about the safety of mid-size cars in comparison to small cars. Which type of vehicle would you recommend to a safety conscious buyer?

2. The Nielson Home Technologies Report provided information about home technology and its usage by persons age 12 and older. The following data are hours of personal computer usage during one week for a sample of 50 users.

| **Hours** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 4.1 | 1.5 | 10.4 | 5.7 | 3.0 | 5.9 | 3.4 | 6.1 | 1.6 | 3.7 |
| 3.1 | 4.8 | 2.0 | 4.2 | 11.1 | 14.8 | 5.4 | 4.1 | 3.9 | 3.5 |
| 4.1 | 4.1 | 8.8 | 3.3 | 6.2 | 5.6 | 4.3 | 10.3 | 7.1 | 7.6 |
| 10.8 | 2.8 | 9.5 | 0.7 | 4.4 | 12.9 | 12.1 | 9.2 | 4.0 | 5.7 |
| 7.2 | 6.1 | 5.7 | 3.9 | 6.1 | 5.9 | 4.7 | 3.1 | 3.7 | 3.1 |

a) Draw an ordered stem and leaf plot for the data and describe the shape of the plot.
b) Find the mean, median, mode, standard deviation and coefficient of variation for the data.
c) Assuming that the data was drawn from a bell-shaped curve, within what limits would you expect:
   - 67% of the values to fall?
   - 95% of the values, and,
   - 99.9% of the data to fall?
d) compute the actual proportion of the sample that fall within the limits stipulated by the empirical rule. Is there reason to believe that the data came from a bell-shaped curve?

3. In January 2003, the American worker spent an average of 77 hours logged on to the internet while at work (CNBC, March 15, 2003). Assume times are normally distributed with a standard deviation of 20 hours.



a) What is the probability that a randomly selected worker spent less than 50 hours on the internet?
b) What percentage of workers spent between 60 and 85 hours?
c) A person is classified as a heavy user if he or she is in the upper 20% of usage. How many hours must a worker log on to be considered a heavy user?
d) What is the probability that a random sample of 40 workers will spend an average of 87 hours or more logged on to the internet?
e) Reflecting on your answer in d), is this a highly likely (highly probable) event? Please interpret your answer.
f) An internet analyst projects by the end of 2005, the top 20% of internet users at work will exceed 120 hours. Assuming the standard deviation hasn't changed, what is the new mean time spent on the internet by an American worker?

4. In a recent poll of 1000 teens across North America, 37.3% sent text messages. It is believed that 35% of all teens using cell-phones send text messages.

a) What is the probability that in a random sample of 1000 teens, less than 30% send text messages?
b) What is the probability that the sample proportion of teens sending text messages in a random sample of 1000 would be within $\pm$5% of the population proportion.

c) What is the probability that in a random sample of 1000 teens, 37.3% or more send text messages?
d) Given that the statistic in c) was actually observed, do you believe there is evidence that the true proportion of teens sending text messages has increased?

5. Surf the web and find 3 interesting news items that makes use of statistics. Write one short summary paragraph on each story. You can search websites the focus on health trends, technology, music, automobiles, finance, marketing, etc. Please provide the url reference for your story.

Click this link for an example: http://content.nejm.org/cgi/content/short/336/7/453 . You can't use this story however.

# Section II:
## Sampling Distributions

1. Define the following terms:
   a. Parameter
   b. Statistic
   c. Sampling distribution.
   d. State the First Limit and Central Theorems
   e. Point estimate
   f. Interval estimate
   g. Margin of error

2. Towers Perrin, a New York human resources consulting firm, conducted a survey of 1100 employees at medium-sized and large companies to determine how dissatisfied were employees with their jobs *(The Wall Street Journal,* January 29, 2003). A total of 473 employees indicated they strongly disliked their current work experience.

   a. What is the sample statistic of interest?
   b. What can we say about the nature of the sampling distribution of the sample statistic in (a)?
   c. Compute a *95%* confidence interval for the proportion of the population of employees who strongly dislike their current work experience?
   d. The common view among such firms is that 35% of employees will not like their jobs. Is there any reason to believe that job satisfaction is on the decline? Please explain.
   e. Towers Perrin would like to estimate the true proportion of employees who do not like their jobs with a margin of error of 1%. How many more employees need to surveyed?
   f. Towers Perrin estimates that it costs employers one-third of an hourly employee's annual salary to find a successor and as much as 1.5 times the annual salary to find a successor for a highly compensated employee. What message did this survey send to employers?

3. Audience profile data collected at the ESPN Sports Zone Web site showed that *26%* of the users were women *(USA Toda)* January 21, 1998). Assume that this percentage was based on a sample of 400 users.
   a. At *98%* confidence, what is the margin of error associated with the estimated proportion of users who are women?
   b. What is the *98%* confidence interval for the population proportion of ESPN Sports Zone Web site users who are women?
      Two years ago, ESPN Sports Zone estimated the proportion of women using the website at 15%. Is there sufficient reason to believe the proportion of women using the website has increased? Please explain. How might ESPN Sports Zone use this data?
   c. How large a sample should be taken if the desired margin of error is .03 and the confidence level is 98%?

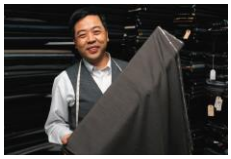
4. A recent study examined the buying preferences of Trinidadian university students for two types of Indian dishes, Dhall and Roti. 225 students were sampled and asked to consider the two types of food.

a. It is believed that students are indifferent to Dhall and Roti. What proportion of students should chose Roti over Dhall?
b. In the above survey, 135 students actually expressed a preference for Roti. Construct a 95% confidence interval for the true proportion of students who would choose Roti.
c. Using your interval in (b), would you conclude that students are indifferent to Dhall and Roti?
d. What is the nature of the sampling distribution for this problem?
e. What theorem if any is being applied here?

5. A clothing manufacturer is interested in knowing the mean height of men in Mexico. They believe that knowing this will help them in their clothes design and manufacturing. They wish to estimate the mean within $\pm$ 0.20 inches with 95 percent confidence. Before actually sampling the population, they randomly selected 15 men and measured their heights (in inches) shown as follows.

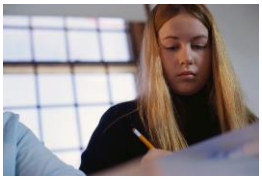
| | | | | |
|---|---|---|---|---|
| 63.5 | 67.4 | 68.3 | 70.5 | 59.3 |
| 59.0 | 66.7 | 72 | 70.1 | 68.3 |
| 66.8 | 70 | 71 | 68.9 | 64.3 |

a. Using the sample data, find a 95% confidence interval for the true mean height of men in Mexico.
b. What assumptions do we have to make for the confidence interval in (a) to be valid?
c. Using the sample data above, how many more men should be sampled to obtain a desired margin of error of $\pm$ 0.20 at a 95% level of confidence?



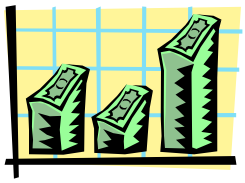
6. An educational organization in Canada is interested in estimating the mean number of minutes per day that children between the age of 6 and 18 spend watching television per day. The organization selected a random sample of n = 200 children between the age of 6 and 18 and recorded the number of minutes of TV that each person watched on a particular day. The mean time was 191.3 minutes with a standard deviation of 21.5 minutes.


a. Find a 90% confidence interval for the true mean number of minutes that a child between the age 6 and 18 watch per day.
b. If the leaders of the organization wish to develop an interval estimate with 90 percent confidence, what will the margin of error be?
**c.** What assumptions are necessary for the interval in (a) be valid?



# Section III:
## Interval Estimation and Hypothesis Testing

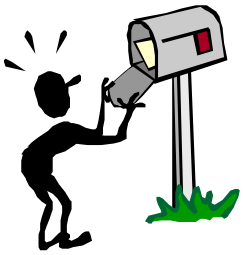
1: Examine the data on the total annual sales (in billions of dollars) for 25 industrial corporations. The data are shown below.



| | | | | |
|---|---|---|---|---|
| 27.0 | 29.4 | 17.6 | 12.0 | 10.2 |
| 96.9 | 24.2 | 15.9 | 12.0 | 10.1 |
| 86.6 | 21.7 | 15.4 | 11.9 | 10.1 |
| 63.4 | 21.7 | 15.2 | 11.7 | 10.0 |
| 60.0 | 20.3 | 15.0 | 11.4 | 9.9 |

a. Find a 95% confidence interval for the mean annual sales for industrial corporations.
b. What assumptions are necessary for the analysis in (a) to be valid?
c. Draw a stem and leaf plot of the data. Is there apparent support for your assumptions in (b)?
d. It is desirable to reduce the error in the estimate by half. What sample size is needed assuming we still want a 95% confidence interval?

2: A mail order company in Dartmouth is attempting a direct marketing strategy on one of its new products. A previous survey of 240 people resulted in 84 willing to purchase the product. The product sells for $14.99.



a. Give a 98% confidence interval for the expected gross revenue if the company will market the product to 10,000 Metro residents.
b. What sample size is needed to estimate the true proportion of people willing to purchase the product with an allowable error of $\pm 5\%$ .

3: A *chewing cycle* is defined as an upward movement followed by a downward movement of the chin. Clinicians have found that the chewing cycles of normal children differ from the chewing cycles of children with eating difficulties. In one study (*The American Journal of occupational Therapy*, May.1984),the number of chewing cycles required for a "normal" preschool child to swallow a bite of graham cracker was found to have a mean of 15.0. In a recent study, a random sample of 20 normal preschool children were found to require a mean of 12 chewing cycles with a standard deviation of 2.5 cycles.

a. Find the probability that a random sample of 20 "normal" preschool children will have a mean of 12 chewing cycles or less for a bite of graham cracker. Interpret your probability.
b. Find a 95% confidence interval for the true mean number of chewing cycles required to chew swallow a bite of a graham cracker.
c. What assumptions are needed for the interval in (b) to be valid?

d. Using your results in (b), is there sufficient evidence that the mean number of chewing cycles necessary to swallow a bite of graham cracker for a "normal" preschool child has decreased?
e. In a separate random sample of 25 preschool children thought to have eating difficulties, a mean of 22 cycles with a standard deviation of 6.4 was observed. Find a 90% confidence interval for the difference between the mean number of chewing cycles required to chew and swallow a bite of graham cracker for the two groups of children.
f. What assumptions are necessary if any for the confidence interval in (e) to be valid?
g. Using the information in (e), is there sufficient evidence that preschool children with eating difficulties require a greater number of cycles to chew and swallow a graham cracker? Assume a 5% significance level.

4. The following experiment was conducted to compare two coatings designed to improve the durability of the soles of jogging shoes. A 1/8 inch layer of coating 1 was applied to one of a pair of shoes and a layer of equal thickness of coating 2 was applied to the other shoe. Ten joggers were given pairs of shoes treated in this manner and were instructed to record the number of miles covered in each shoe before the 1/8 inch coating was worn through in any one place. The results are listed in the accompanying table.

| Jogger | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Coating 1 | 892 | 904 | 775 | 435 | 946 | 853 | 780 | 695 | 825 | 750 |
| Coating 2 | 985 | 953 | 775 | 510 | 895 | 875 | 895 | 725 | 858 | 812 |

a. Develop a 90% confidence interval for the difference between the mean number of miles covered by the two types of coatings before being worn through.
b. Is there sufficient evidence that coating 2 lasts longer than coating 1?
c. What is the observed significance of the test?
d. Explain why this experimental design is more useful than using two independent groups of joggers, with one group receiving shoes with coating 1 and the other receiving shoes with coating 2.

5: Please do the following questions

a. Outline the phases of a hypothesis test.
b. Define the following terms:
c. Define Null hypothesis and alternate hypothesis
d. Define Type I and Type II Errors
e. Define *p*-value (or observed significance level)

6: The metropolitan airport commission is considering the establishment of limitations on the extent of noise pollution around a local airport. At the present time the noise level per jet takeoff in one neighborhood near the airport is approximately normally distributed with a mean of 100 decibels and a standard deviation of 6 decibels.

a. What is the probability that a randomly selected jet will generate a noise level greater than 108 decibels in this neighborhood?

b. Suppose a regulation is passed that requires jet noise in this neighborhood to be lower than 105 decibels 95% of the time. Assuming the standard deviation of the noise distribution remains the same (6 db), how much will the mean noise level have to be lowered to comply with the regulation?
c. After the regulation was passed, a random sample of 50 jets yielded a mean noise level of 92 decibels with a standard deviation of 4 decibels. At a 5% significance level, is there sufficient evidence that the true mean noise level produced by the jets complies with the regulation implied in (c)?

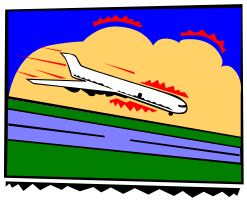
7: The percentage of body fat can be a good indicator of an individual's energy metabolic status and general health. In a recent study conducted by the Brazilian Health Commission of the percentage of body fat of college students, two groups of healthy male students from urban and rural colleges in Eastern Brazil, were randomly and independently selected. The percentage of body fat was measured in each group, and the results summarized in the table below.

| | Urban Students | Rural Students |
|---|---|---|
| Sample Size | 193 | 188 |
| Mean | 12.07 | 11.04 |
| Std. Dev. | 3.04 | 2.63 |


a) If the we believe that there is no difference between the true mean percentage of body fat for urban and rural students, what is the probability that random samples of 193 urban students and 188 rural students will yield an **absolute** difference between the sample means of 1.03% or higher?
b) Would you refer to this sample are a significant sample? Please explain.
c) Find a 98% confidence interval for the difference between the true mean percentage of body fat for urban and rural students. Please interpret the interval.
d) Does the study provide sufficient evidence that there is a difference between the mean percentage of body fat for urban and rural students in Eastern Brazil. Use a significance level $\alpha$=0.02.
e) What is the observed significance of the test (or *p*-value). Give an interpretation of the p-value.

# Section IV:
## Proportions and Analysis of Variance

1: The metropolitan airport commission is considering the establishment of limitations



on the extent of noise pollution around a local airport. At the present time the noise level per jet takeoff in one neighborhood near the airport is approximately normally distributed with a mean of 100 decibels and a standard deviation of 6 decibels. Regulations were passed that requires a mean noise level of 95 decibels. Prior to the regulation, 120 out of a random sample of 300 aircraft met the regulation standard. In the first year after the regulation was imposed, 250 out of a random sample of 450 aircraft met the standard.

a) Find a 95% confidence interval for the difference between the true proportion of aircraft before and after the regulation that meet the new noise standard.
b) At a 5% significance level, is there sufficient evidence that a greater number of aircraft now meet the desired noise level of 95 decibels?
c) What is the observed significance (p-value) of the test?
d) What error do we risk making given our conclusion in (b)?
e) The FAA will consider an aggressive campaign to go after violators if the difference between the true proportion of aircraft meeting the regulation standard before and after the enactment of the standard did not improve by more than 10% in the first year following the new regulation. At a 5% significance level, is there a need for the FAA to get tough with violators?
f) What is the actual risk of committing type I error?

2: A hospital in Nova Scotia would like to determine whether there is a relationship between the level of satisfaction of workers with working conditions and their job category. A random sample of 250 employees were surveyed. The results are summarized in the table below.

| | *Nurses* | *Support Staff* | *Doctors* | *Total* |
|---|---|---|---|---|
| *Satisfactory* | 40 | 80 | 20 | 140 |
| *Indifferent* | 20 | 40 | 20 | 80 |
| *Poor* | 15 | 5 | 10 | 30 |
| *Total* | 75 | 125 | 50 | **250** |

a) Find a 98% confidence interval for the true proportion of all workers that is satisfied with working conditions?

b) Is there sufficient evidence that support staff are more satisfied with working conditions than nursing staff? Use a 90% confidence interval.

c) Using your answer in (b) do you think that job category and worker satisfaction are related? Please explain. Just saying yes or no is not acceptable.

d) At a 5% significance level, is there sufficient evidence of a difference between the level of satisfaction with working conditions for doctors and nurses?

e) Looking at your answer in (d) is there any indication that job category and level of satisfaction are related? Please explain.

3. An Agricultural Lab in Mexico is testing the effect of two types of fertilizers on the growth rate of mango seedlings. A random sample of 20 seedlings were given fertilizer A, and 25 seedlings were given fertilizer B. The increase in the height of the seedlings over a three-week period was measured, and the results summarized as follows:

| | Fertilizer A | Fertilizer B |
|---|---|---|
| Sample size | 20 | 25 |
| Sample mean | 50.5 mm | 57.5 mm |
| Sample standard deviation | 13.2 mm | 8.5 mm |

A statistician wants to perform a t-test to determine whether fertilizer B results in a larger mean growth rate for the seedlings over the three-week period. To do so, she must assume equal population variances. Determine whether the assumption of equal variances is reasonable. Use a 5% significance level

4. High school students planning to attend university were randomly assigned to watch one of four videos about Dalhousie University. Each video differed in the particular aspect of university life of students that was emphasized: athletics, academics, social life, or art and culture. After watching a video, each student was given a questionnaire that measured his or her desire to attend Dalhousie University. Student's answers were analyzed to provide a measure of their desire to attend Dalhousie. (Scores reflect the percentage of questions indicating a strong preference to attend Dalhousie. High scores reflect a greater desire to attend Dalhousie.) A total of 16 students were shown one of the videos.

| Athletics | Social Life | Academics | Art/Cultural |
|---|---|---|---|
| 68 | 89 | 74 | 76 |
| 56 | 78 | 82 | 71 |
| 69 | 81 | 79 | 69 |
| 70 | 77 | 80 | 65 |

   (a) At the 1% level of significance, do these data suggest that the type of activity emphasized in a university video affects the desire of prospective students to attend Dalhousie? (Do all calculations by hand.)
   (b) What is the observed significance of the test?
   (c) Discuss the underlying assumptions of the test used in (a).
   (d) Use Excel to answer part (a). Interpret the results by annotating your printout. Hand in your annotated printout. [Note: Your results for this part of the question must be printed out on a single sheet of computer paper with all excess perforated paper removed. Failure to follow these instructions will result in a mark of zero.]

5. A manager in a rapidly growing industry wishes to study the effects of different production systems on worker output. Traditionally, all staff has used an assembly line for production. Senior management has been most reluctant to change the existing system. Within his own branch, however, the manager wishes to gather evidence on whether or not a change in the system of production can affect worker output. Over the past 6 months three different production systems – an assembly line in which each worker does only one task, a system in which each worker does many different tasks (and, therefore, requiring more expertise), and a system combining elements of both a single tasking and a multiple tasking system - have been tried with the results presented in the table below. Each of the 9 individuals in the office worked under each of the three production systems (for periods of 1 month each). Each person started on the single task system, then did the combination system, and finished with the multiple-task system. The following data represent output for each worker (in hundreds of units produced) achieved during the last month on the job under each production scheme.

*SYSTEM USED*

| Worker | Combination-Single-Multiple Tasks | Multiple Tasks | Single Task (Assembly Line) |
|---|---|---|---|
| **1** | 256 | 224 | 269 |
| **2** | 239 | 254 | 284 |
| **3** | 222 | 273 | 294 |
| **4** | 207 | 285 | 290 |
| **5** | 228 | 237 | 247 |
| **6** | 241 | 277 | 278 |
| **7** | 212 | 261 | 263 |
| **8** | 216 | 228 | 229 |
| **9** | 236 | 234 | 236 |

(a) Is there evidence that the three production systems differ significantly in their effect on worker output? Use $\alpha = .05$ and find the *p*-value. (Do all calculations by hand.)

(b) Is there evidence that the workers differ significantly in their mean production output? Use $\alpha = .025$ and find the *p*-value. (Do all calculations by hand.)

(c) Can you see any advantages/disadvantages to the use of the type of design used in this study? Discuss in depth. Be clear and concise. Avoid the use of jargon.

(d) Discuss an alternative way this study could have been conducted that would not have had the same major disadvantages.

(e) Use Excel to answer part (a). Interpret the results by annotating your printout. Hand in your annotated printout. [Note: Your results for this part of the question must be printed out on a single sheet of computer paper with all excess perforated paper removed. Failure to follow these instructions will result in a mark of zero.]

# Section V:
## Contingency Tables and Regression Analysis

Please observe the following instructions: 1) Answer each question on a **fresh new page**. 2) Ensure that all questions are clearly labeled. 3) **Clearly label** all Excel printouts where appropriate. 4) Submit only the **relevant portion** of your Excel printout. (5) Submit neat assignments. Illegible assignments will not be marked.

1: A hospital in Nova Scotia would like to determine whether there is a relationship between the level of satisfaction of workers with working conditions and their job category. A random sample of 250 employees were surveyed. The results are summarized in the table below.

| | *Nurses* | *Support Staff* | *Doctors* | *Total* |
|---|---|---|---|---|
| *Satisfactory* | 40 | 80 | 20 | 140 |
| *Indifferent* | 20 | 40 | 20 | 80 |
| *Poor* | 15 | 5 | 10 | 30 |
| *Total* | 75 | 125 | 50 | **250** |

f) Is there clear evidence that employee category and the level of satisfaction with working conditions are related? Use a 1% significance level.

g) Compute and interpret the observed significance of the test.

2. A supermarket chain wanted to know whether there was any relationship between the price (in dollars) set by the chain for its in-house brand of coffee and demand (measured in pounds of coffee). Eight stores in the chain that had nearly equal past histories of demand for this brand of coffee were used in the study. Eight different prices were randomly assigned to the stores (one price per store) and an identical ad campaign was run in each market. The number of pounds of coffee sold during the following week was recorded for each store (see below).

| Store | Demand | Price |
|---|---|---|
| A | 1120 | $ 3.00 |
| B | 999 | $ 3.10 |
| C | 932 | $ 3.20 |
| D | 884 | $ 3.30 |
| E | 807 | $ 3.40 |
| F | 760 | $ 3.50 |
| G | 701 | $ 3.60 |
| H | 688 | $ 3.70 |

(a) What is the coefficient of correlation in this situation? Explain what it means.

(b) Is there a significant relationship between demand and price? (Do a complete analysis to justify your conclusions here.) Explain. Use the classical method of testing hypotheses and $\alpha$ = 5%.

(c) Does price predict demand for this coffee? Explain. Assume $\alpha$ = 5% and use the classical method of testing hypotheses.

(d) What is the least squares prediction equation (i.e., the regression equation) here for predicting demand?

(e) Are you justified in using the regression equation to make predictions? Explain.

(f) Is there a straight-line relationship between price and demand for this coffee? Explain.

(g) The chain had been contemplating selling this coffee for $3.99 per pound. Use the regression equation to predict what demand in one of these stores would have been at this price.

(h) In making your prediction for this price of $3.99, did you violate any conditions which must be adhered to when using regression techniques to make predictions? Explain briefly.

(i) What is the coefficient of determination in this situation? Explain what it means.

(j) What assumptions did you have to make to answer part (e)?

(k) What is the percentage of total variation in demand that is not explained by the prediction equation (i.e., what is the error variation here)?

(l) What do the coefficients in the regression equation mean? Explain what each of the 2 coefficients mean.

(m) What is the standard error of estimate in this example?

(n) What is the 95% confidence interval for demand if a store were to price its coffee at $3.00 per pound?

(o) What is the 95% confidence interval for average demand of all stores offering their coffee at $3.00 per pound?

(p) Draw the scatterplot showing the relationship between demand and price. What conclusions can you draw from this diagram?

(q) Reanalyze the data in this question using Excel. Annotate the printout to show r, $r^2$, the regression equation, the t and F tests, sample size, etc.

3. A local express delivery service bases its charge for shipping a package on package weight and distance shipped. The company recently conducted a study to investigate the relationship between the cost of shipment to the delivery company (in dollars) and one of the variables that control the shipping charge to the customer -- package weight (in kilos). Use $\alpha$ = 5%.

| Package | Weight | Cost |
|---|---|---|
| 1 | 5.9 | $ 2.60 |
| 2 | 3.2 | $ 3.90 |
| 3 | 4.4 | $ 8.00 |
| 4 | 6.6 | $ 9.20 |
| 5 | 0.75 | $ 4.40 |
| 6 | 0.7 | $ 1.50 |
| 7 | 6.5 | $ 14.50 |
| 8 | 4.5 | $ 1.90 |
| 9 | 0.6 | $ 1.00 |
| 10 | 7.5 | $ 14.00 |

**Excel printout:**
SUMMARY OUTPUT

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.73449203 |
| R Square | 0.53947854 |
| Adjusted R Square | 0.48191336 |
| Standard Error | 3.64642371 |
| Observations | 10 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 1 | 124.61 | 124.609 | 9.3716 | 0.0156 |
| Residual | 8 | 106.37 | 13.296 | | |
| Total | 9 | 230.98 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| Intercept | 0.3781 | 2.1962 | 0.1721 | 0.8676 |
| Weight | 1.4076 | 0.4598 | 3.0613 | 0.01555 |

**Use Excel to answer the following questions:**

(a) Is the model useful for predicting shipping cost? Explain. Assume $\alpha$ = 5% and use the p-value method of testing hypotheses.

(b) What is the least squares prediction equation (i.e., the regression equation) here for predicting cost?

(c) Are you justified in using the regression equation to make predictions? Explain.

(d) Is there a straight-line relationship between shipping cost and shipping weight? Explain.

(e) Use the regression equation to predict the shipping cost of a package that weighs 5 kilos and has to be shipped 200 kilometers.

(f) In making your prediction in part (e), did you violate any conditions which must be adhered to when using regression techniques to make predictions? Explain briefly.

(g) What is the coefficient of determination in this situation? Explain what it means in this situation.

(h) What is R in this example? Explain what it means.

(i) What is the percentage of total variation in cost that is not explained by the prediction equation (i.e., what is the error variation here)?

(j) What do the coefficients in the regression equation mean? Explain what each of the 2 coefficients mean.