**SCHULICH SCHOOL OF BUSINESS**
**YORK UNIVERSITY**

**SESSION:** **FALL 2003 Final Examination** **NAME:** **SOLUTION**

**COURSE NO:** **MGTS1000 3.0 C, D, E** **I.D. # :** _____

**COURSE TITLE:** **Statistics for Management Decisions**

**PROFESSORS:** **A. Marshall, D. Nevo**

---

**NUMBER OF PAGES:** **16 pages (NOT including this cover page)**
**LENGTH OF EXAMINATION:** **180 minutes (3 hours)**
**EXAMINATION AIDS ALLOWED:** **Calculator; 1, 8½x11 double-sided formula sheet**

---

**INSTRUCTIONS:**

Please place your I.D. card on your desk.

You are not allowed to leave the examination room until one hour after the start of the exam and you must sign the sign-in sheet before leaving.

Your examination paper must be handed in before you leave. When you are finished please leave the exam room quietly.

Cheating on an examination will result in an "F" grade in the course concerned and possible suspension from the University.

All notes, briefcases, and books must be deposited at the front of the room.

You are not allowed to use your own paper for rough work. You may only use the back of the pages for rough work.

Answer the Multiple Choice and True/False Questions on the bubble sheet supplied. Be sure to put your name and student number on the bubble sheet. In Part IV, place your answers in the answer boxes, where supplied.

Count the pages to be certain that there are no pages missing.

Do not remove the staple from the exam. You may remove the tables on pages 16-23.

Do not begin this examination until you are instructed to do so.

Do not write in the mark summary table below.

| | | | | |
|---|---|---|---|---|
| **Part I – Multiple Choice (Problem Recognition)** | | | **36 marks** | |
| **Part II – True/False** | | | **14 marks** | |
| **Part III – Multiple Choice (Interpretation)** | | | **15 marks** | |
| **Part IV – Problems** | **Questions 46-50** | **6 marks** | | |
| | **Questions 51-53** | **6 marks** | | |
| | **Questions 54-55** | **6 marks** | | |
| | **Question 56** | **7 marks** | | |
| | **Question 57** | **10 marks** | | |
| | **Sub-total** | | **35 marks** | |
| **TOTAL** | | | **100 marks** | |

**Part I - Problem Recognition Multiple Choice (36 marks, 2 marks each)**

Instructions:
1. Indicate your answers on the bubble sheet.
2. Your student number should be **left justified** on the bubble sheet.
3. Only one of the choices is the best answer

**Question 1**

A survey of 1,500 Canadians reveals that 945 believe that there is too much violence on television. In a survey of 1,500 Americans, 810 believe that there is too much television violence. The appropriate test to see if there is a difference in attitudes toward television violence in the U.S. and Canada is:

(a) t-test of $\mu_D$
(b) **z-test of $p_1$-$p_2$**
(c) t-test of $\mu_1-\mu_2$
(d) F-test of $\sigma_1^2/\sigma_2^2$

**The following information relates to Questions 2 and 3:**

A financial advisor claims that over the past 20 years, the Good and Steady Mutual Fund has performed better than the Big Bank Index Fund 75% of the time and has also had less risk than the Big Bank Index Fund.

**Question 2**

In order to test the claim about performance, the appropriate test is:

(a) t-test of $\mu_D$
(b) t-test of $\mu_1-\mu_2$
(c) z-test of $p_1$-$p_2$, $H_O$: $p_1$-$p_2$ = 0
(d) z-test of $p_1$-$p_2$, $H_O$: $p_1$-$p_2$ = D
(e) **z-test of p**

**Question 3**

In order to test the claim about risk, the appropriate test is:

(a) F-test of $\sigma_1^2 - \sigma_2^2$
(b) F-test of $\sigma_1 - \sigma_2$
(c) **F-test of** $\sigma_1^2/\sigma_2^2$
(d) F-test of $\sigma_1/\sigma_2$

**Question 4**

A statistics professor was trying to determine how educational background influences performance in statistics classes. He randomly selected the statistics grades of 200 students from a recent term and classified them into three groups based on the high school math background: (1) one mathematics course counting in their graduating average; (2) two math courses counting in their graduating average; and (3) three more math courses in their graduating average. Also, based on research by a colleague that indicated that female students tend to get lower grades in mathematics courses because of gender bias in the texts and examples used, the gender of the students was also measures and used as a variable. The appropriate test design is:

(a) **Two-way ANOVA**
(b) Randomized block design ANOVA
(c) One-way ANOVA
(d) Multiple regression with indicator (dummy) variable(s)

**Question 5**

A professor of finance is investigating what influences performance in her finance course. She believes that a student's finance mark can be predicted based on their accounting mark, their quantitative methods mark and the number of hours per week spent studying finance. She has collected data for 100 students. The appropriate test methodology is:

(a) **Multiple regression**
(b) Multiple regression with indicator (dummy) variable(s)
(c) Two-way ANOVA
(d) Randomized block design ANOVA

**Question 6**

Alec Trebex is an ardent fan of the television game show *Jeopardy*. He thinks he has noticed a strange pattern: The more educated the contestant, the less they win. To test his belief, he records the winnings and education level (in years) of the two challenger contestants each day for three weeks. The appropriate test procedure would be:

(a) t-test of $\mu$
(b) **Correlation**
(c) One-way ANOVA
(d) Multiple regression

**Question 7**

A business columnist noticed that there are a large number of women managing mutual funds. Interviewing one of these women, she said laughingly, "I think women make better fund mangers. We don't have testosterone to cloud our thinking." The columnist decided to see if there was any truth in her quip. He chose the ten largest funds managed by men and the ten largest managed by women. Their distributional characteristics looked quite similar, but the funds managed by women had a one-percent better return. The appropriate test procedure is:

(a) t-test of $\mu_D$
(b) **t-test of $\mu_1-\mu_2$, Equal Variances**
(c) t-test of $\mu_1-\mu_2$, Unequal Variances
(d) One-way ANOVA

**Question 8**

The Canadian Securities Course (CSC) is required to obtain the licensing to deal in securities in Canada. An employee of the Canadian Securities Institute (CSI) has been asked to investigate the variation in CSC examination scores. The employee obtained the grades of 100 students who have written the CSC exam in the past six months and conducted a survey to find out (1) the number of years of work experience in an industry related to finance and (2) the total number of hours spent studying for the course. From CSI records, the employee was able to determine if each candidate (1) had a university degree and (2) took a CSI-sponsored seminar. The appropriate test methodology is:

(a) Multiple regression
(b) **Multiple regression with indicator (dummy) variable(s)**
(c) Two-way ANOVA
(d) Randomized block design ANOVA
(e) One-way ANOVA

**Question 9**

A finance researcher was investing the performance of Mutual Fund Portfolio Managers. He looked at the investment performance of 99 managers and found that 1/3 held MBAs, 1/3 had the CFA designation and 1/3 held some other designation or degree. The appropriate test design to see if the type of designation or degree explains their management performance is:

(a) Multiple regression with indicator (dummy) variable(s)
(b) Two-way ANOVA
(c) Randomized block design ANOVA
(d) **One-way ANOVA**

**Question 10**

A finance professor was trying to determine which model of financial calculator students found easiest to learn. He took a group of 30 volunteers and gave them 3, 2-hour seminars on the use of the three models of calculators. A week later, the student volunteers were given 3, 45 minute quizzes, on different days: on Tuesday, the students used model A, on Wednesday, they used model B and on Thursday, they used model C and the test scores were recorded. The appropriate test design is:

(a) Multiple regression with indicator (dummy) variable(s)
(b) Two-way ANOVA
(c) **Randomized block design ANOVA**
(d) One-way ANOVA

**Question 11**

A local business group is concerned about the recent strength of the Canadian dollar. They believe that when the dollar strengthens fewer tourists come to Southern Ontario. To determine the impact that a 1 cent change in the exchange rate has on the number of tourists, the appropriate test procedure is:

(a) Correlation
(b) **Simple regression**
(c) One-way ANOVA
(d) Multiple regression
(e) Multiple regression with indicator (dummy) variable(s)

**Question 12**

A political writer is preparing an article about the latest Gallup poll showing the popular support of the governing party in Canada. The numbers have changed somewhat from the poll released last month, but he wants to test to see if there has been a change. The appropriate test procedure would be:

(a) z-test of p
(b) t-test of $\mu_1 - \mu_2$
(c) **z-test of $p_1$-$p_2$**
(d) z-test of $p_D$

**Question 13**

A statistician wants to test for the equality of means in two independent samples drawn from normal populations. However, he will not perform the equal-variance t-test of the difference between the population means if the condition necessary for its use is not satisfied. The data follow:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Sample 1: | 7 | 9 | 6 | 15 | 7 | 10 | 8 | 12 | |
| Sample 2: | 2 | 25 | 9 | 15 | 10 | 18 | 5 | 22 | 27 |

The appropriate test if the necessary conditions are satisfied is:

(a) F-test of $\sigma_1^2 - \sigma_2^2$
(b) F-test of $\sigma_1 - \sigma_2$
(c) **F-test of** $\sigma_1^2/\sigma_2^2$
(d) F-test of $\sigma_1/\sigma_2$

**Question 14**

Does participation in varsity sports benefit the student's salary? To test this question a researcher obtained the cooperation of the placement office and was able to survey 120 graduates from the 1992 SSB class. Besides the graduates' current salaries, the researcher obtained information on the students' GPA and whether they played on a varsity sports team for two or more years. The appropriate research design is:

(a) Multiple regression
(b) **Multiple regression with indicator (dummy) variable(s)**
(c) Two-way ANOVA
(d) Randomized block design ANOVA
(e) One-way ANOVA

**Question 15**

A Canada Customs and Revenue Agency employee believes that the tendency to file an Income Tax Return late is a function of whether the taxpayer expects a refund or owes money. The policy advisor believes that people who expect a refund have a greater propensity to file on time. People who owe money have a greater tendency to file late. Data from 50,000 returns was gathered and they were categorized by type (refund or payment) and their punctuality. The appropriate research design is:

(a) One-way ANOVA
(b) t-test of $\mu_1 - \mu_2$
(c) **z-test of $p_1$-$p_2$**
(d) z-test of $p_D$

**Question 16**

There are different approaches to fitness training. To judge which one of two approaches is better, 200 twenty-five-year old men are randomly selected to participate in an experiment. For four weeks, 100 men are trained by approach 1 while the other 100 men are trained by approach 2. The percentage improvement in fitness was measured for each man. The percentage figures are known to be normally distributed. The appropriate research design is:

(a) z-test of $p_1$-$p_2$
(b) t-test of $\mu_D$
(c) **t-test of $\mu_1 - \mu_2$**
(d) Two-way ANOVA

**Question 17**

Professor Marshall does educational seminars for the Canadian Securities Institute. He claims that students who take his seminar have a pass rate that is 20% better than those students who do not take the seminar. The appropriate test is:

(a) t-test of $\mu_D$
(b) t-test of $\mu_1 - \mu_2$
(c) z-test of $p_1$-$p_2$, $H_O$: $p_1$-$p_2 = 0$
(d) **z-test of $p_1$-$p_2$, $H_O$: $p_1$-$p_2 = D$ – our text does not cover this test.**
(e) z-test of p

**Question 18**

The Sales Manager of a chain of electronics stores believes that age and family situation (supporting a family or not) are the two best determinants of sales performance of the sale people. With the HR director, they investigate the influence of these variables on sales performance. The appropriate test or research design would be:

(a) Two-way ANOVA
(b) Multiple regression
(c) **Multiple regression with indicator (dummy) variable(s)**
(d) Correlation
(e) Simple regression

**Part II – True/False (14 marks, 1 mark each)**

Instructions:
1. Indicate your answers on the bubble sheet starting at number 19.
2. Use "1" for "true" and "2" for "false"

19. If a small sample size is coupled with an unknown standard deviation, the normal distribution should not be used to establish a confidence interval for the population mean, even if the population is normally distributed. **T**

20. By increasing the confidence level, the probability that the interval actually contains the population parameter decreases, and precision is sacrificed. **F**

21. The variability of the t distribution depends on the size of the sample. **T**

22. A sample of 100 observations reveals that 20% of the voters favored a tax increase bill, and 80% were opposed. The standard error of the point estimator, p, is 0.04. **T**

23. The null hypothesis coupled with its alternative hypothesis must necessarily cover all the possibilities and cannot overlap each other. **T**

24. If the Z-value is used to test a null hypothesis that the population mean equals a given number, large positive and negative values of Z will lead to a rejection of $H_O$. **T**

25. As the area of rejection increases, both the positive and negative values of the test statistic also increase. **F**

26. If a null hypothesis is tested and rejected at the $\alpha = 0.05$ level, the same null hypothesis would also have been rejected had it been tested at the $\alpha = 0.01$ level. **F**

27. In order to conduct a lower-tail test of a population mean, the alternative hypothesis, $H_A$, would be that the population exceeds a specific number. **F**

28. If you fail to reject the null hypothesis when it is actually true, then you have committed neither a Type I nor a Type II error. **T**

29. If the data being analyzed are nominal data, a typical hypothesis would be a statement about the amount of variability in the population. **F**

30. For any hypothesis test, if the value of a parameter being tested falls outside the confidence interval for the same level of $\alpha$, the correct decision is to reject the null hypothesis. **T**

31. Whenever two population means equal each other, the difference between them must be zero. **T**

32. A large-sample test for the difference between two populations can be conducted safely provided either of the two samples exceeds 30. **F**

**Part III – Computer Output Multiple Choice (15 Marks, Individually Weighted)**

Instructions:
1. Indicate your answers on the bubble sheet, starting at Question 33
2. Only one of the choices is the best answer

**The following output relates to Questions 33 though 36:**

| ANOVA | | | | | | |
|---|---|---|---|---|---|---|
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 21.53 | 4 | 5.38 | 0.27 | 0.8989 | 2.42 |
| Within Groups | 3710.43 | 184 | 20.17 | | | |
| | | | | | | |
| Total | 3731.96 | 188 | | | | |

**Question 33 (1 mark)**

The appropriate Null Hypothesis is:

(a) **$H_O$: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$**
(b) $H_O : \mu_1 = \mu_2 = \mu_3 = \mu_4$
(c) $H_O : \sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5$
(d) $H_O : \sigma_1 = \sigma_2 = \sigma_3 = \sigma_4$

**Question 34 (1 mark)**

The appropriate Alternate Hypothesis is:

(a) $H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 \neq \mu_5$
(b) $H_1 : \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$
(c) $H_1 : \sigma_1 \neq \sigma_2 \neq \sigma_3 \neq \sigma_4 \neq \sigma_5$
(d) $H_1 : \sigma_1 \neq \sigma_2 \neq \sigma_3 \neq \sigma_4$
(e) **None of the above (At least one mean is not equal)**

**Question 35 (1 mark)**

How many observations were used in this test?

(a) 188
(b) **189**
(c) 190
(d) 191

**Question 36 (2 marks)**

Which is the appropriate interpretation of the test result:

(a) The test is inconclusive
(b) Reject the Null Hypothesis
(c) Accept the Null Hypothesis
(d) **Do not reject the Null Hypothesis**

**The following output relates to Questions 37 though 41:**

A researcher investigating the impact of classroom style and schedule on student performance in marketing classes obtained the following Excel™ output:

| ANOVA | | | | | | |
|---|---|---|---|---|---|---|
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Room Style | 14.4 | 1 | 14.40 | 0.57 | 0.4548 | 4.15 |
| Time of Day | 581.8 | 3 | 193.93 | 7.71 | 0.0005 | 2.90 |
| Interaction | 548.6 | 3 | 182.87 | 7.27 | 0.0007 | 2.90 |
| Within | 804.8 | 32 | 25.15 | | | |
| | | | | | | |
| Total | 1949.6 | 39 | | | | |

**Question 37 (1 mark)**

Is there evidence that room style influences performance in Marketing class?

(a) Yes
(b) **No**
(c) The test is inconclusive

**Question 38 (1 mark)**

Is there evidence that scheduling (time of day) influences performance in Marketing class?

(a) **Yes**
(b) No
(c) The test is inconclusive

**Question 39 (1 mark)**

Is there evidence that the room style can influence performance in Marketing class, depending on the time of day the class is scheduled?

(a) **Yes**
(b) No
(c) The test is inconclusive

**Question 40 (1 mark)**

What test design was used to obtain this output?

(a) One-way ANOVA
(b) **Two-way ANOVA**
(c) Randomized block design ANOVA
(d) Multiple regression

**Question 41 (1 mark)**

How many observations did the researcher use?

(a) 43
(b) **40**
(c) 39
(d) 38

**The following output relates to Questions 42 through 45:**

A number of men took part to see if a two-stage weight-loss and fitness regimen would result in a lowering of blood pressure.

| ANOVA | | | | | | |
|---|---|---|---|---|---|---|
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Rows | 177464.60 | 19 | 9340.24 | 323.16 | 0.0000 | 1.87 |
| Columns | 7131.03 | 2 | 3565.52 | 123.36 | 0.0000 | 3.24 |
| Error | 1098.30 | 38 | 28.90 | | | |
| | | | | | | |
| Total | 185693.93 | 59 | | | | |

**Question 42 (1 mark)**

How many men took part in this test?

(a) 20
(b) **60**
(c) 40
(d) 59

**Question 43 (1 mark)**

Is this the appropriate test design?

(a) **Yes**
(b) No
(c) The test is inconclusive

**Question 44 (1 mark)**

Irrespective of your answer in Question 36, if the results above had indicated that the test design was inappropriate, what test design should be used?

(a) **One-way ANOVA**
(b) Two-way ANOVA
(c) Randomized block design ANOVA

**Question 45 (2 marks)**

Which of the following is the appropriate conclusion from the results:

(a) **There is strong evidence that the exercise and fitness program works**
(b) There is a zero probability of rejecting the null hypothesis
(c) An F-value that high suggests a data problem
(d) More data is needed to obtain meaningful results

**Part IV - Short Answer Questions (35 Marks, Individually Weighted)**

**Answer all questions in the answer boxes, where provided**

**The following output relates to questions 46 to 50:**

A researcher studying the impact of weight on income obtained the following output:

SUMMARY OUTPUT

| *Regression Statistics* | |
|---|---|
| Multiple R | 0.22482262 |
| R Square | 0.05054521 |
| Adjusted R Square | 0.04671676 |
| Standard Error | 8.28206562 |
| Observations | 250 |

ANOVA

| | *df* | *SS* | *MS* | *F* | *Significance F* |
|---|---|---|---|---|---|
| Regression | 1 | 905.5965007 | 905.5965 | 13.20254 | Question 49 |
| Residual | 248 | 17010.9675 | 68.59261 | | |
| Total | 249 | 17916.564 | | | |

| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
|---|---|---|---|---|
| Intercept | 75.321 | 4.362 | 17.269 | 0.00000 |
| Weight | -0.095 | 0.026 | -3.634 | 0.00034 |

**Question 46 (1 mark)**

What is the correlation between weight and income?

Answer

**-0.2248**

**Question 47 (1 mark)**

How much of the variation in income is explained by weight?

Answer

**5.05%**

**Question 48 (2 mark)**

(a) What is the meaning of the value "75.32" listed under "*Coefficients*"? (b) Does it make sense in this model?

Answer

**It is the intercept of the line. It makes no sense in this model since (1) none of the data would have been at the y-axis and (2) it would suggest that a person weighing nothing would earn 75.32 thousand.**

**Question 49 (1 mark)**

What is the significance value of F in the regression output?

Answer

**0.00034**

**Question 50 (1 mark)**

What is the t-statistic for the test: $H_O$: $\rho = 0$?

Answer

**-3.634**

**The following output relates to Questions 51 to 53:**

A consultant for the Forzani Group of stores is trying to develop a model to explain expenditure on sporting goods (in hundreds of dollars) and has obtained the following output:

| SUMMARY OUTPUT | | | | | |
|---|---|---|---|---|---|
| | | | | | |
| *Regression Statistics* | | | | | |
| Multiple R | 0.4455 | | | | |
| R Square | 0.1985 | | | | |
| Adjusted R Square | 0.1903 | | | | |
| Standard Error | 4.19 | | | | |
| Observations | 200 | | | | |
| | | | | | |
| ANOVA | | | | | |
| | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 2 | 856.42 | 428.21 | 24.39 | 0.0000 |
| Residual | 197 | 3458.98 | 17.56 | | |
| Total | 199 | 4315.40 | | | |
| | | | | | |
| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | |
| Intercept | 13.03 | 2.54 | 5.12 | 0.0000 | |
| Age | -0.279 | 0.045 | -6.22 | 0.0000 | |
| Income (x$1000) | 0.094 | 0.031 | 3.00 | 0.0030 | |

**Question 51 (2 marks)**

Interpret the R Square of 0.1985.

Answer

19.85% of the variation is spending on sporting goods is explained by age and income in this model.

**Question 52 (2 marks)**

What does the ANOVA table tell us?

Answer

It provides us a statistical test to determine whether a significant amount of the variation in the spending on sporting goods is explained by the variables in the model.

**Question 53 (2 marks)**

What is the impact of age on the spending on sporting goods? Is it significant?

Answer

Each additional year of age lowers the spending on sporting goods by $27.90. This is significant as the t-ratio, -6.22, is below the critical value of –1.96, hence we reject the null hypothesis that the slope coefficient equals 0.

**The following relates to questions 54 and 55:**

The sales manager of a textbook publisher was concerned about the high cost of examination copies of textbooks given to professors. He collected the sales revenue data for all of the firm's representatives for the past year, along with the number of examination copies each representative distributed. The output is shown below:

| SUMMARY OUTPUT | | | | | |
|---|---|---|---|---|---|
| | | | | | |
| *Regression Statistics* | | | | | |
| Multiple R | 0.3079 | | | | |
| R Square | 0.0948 | | | | |
| Adjusted R Square | 0.0829 | | | | |
| Standard Error | 28577 | | | | |
| Observations | 78 | | | | |
| | | | | | |
| ANOVA | | | | | |
| | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 1 | 6502546112 | 6502546112 | 7.96 | 0.0061 |
| Residual | 76 | 62067169717 | 816673286 | | |
| Total | 77 | 68569715830 | | | |
| | | | | | |
| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | |
| Intercept | 18018 | 7269 | 2.48 | 0.0154 | |
| Copies | 130.76 | 46.34 | 2.82 | 0.0061 | |

$$\bar{x} = 140.47 \qquad s_x^2 = 4{,}939 \qquad s_\varepsilon = 3{,}278$$

Note: we have deliberately left out the words 'confidence' and 'prediction' in the following questions

**Question 54 (3 marks)**

Construct a 95% interval for the sales revenue generated by a representative that distributed 180 copies.

Answer: This is a prediction interval

$$\hat{y} = b_0 + b_1(x_G) = 18{,}018 + 130.76(180) = 41{,}554.8$$

$$\hat{y} \pm t_{\alpha/2} s_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x_G - \bar{x})^2}{(n-1)s_x^2}} = 41{,}554.8 \pm (2.000)(3{,}278)\sqrt{1 + \frac{1}{78} + \frac{(180 - 140.47)^2}{(77)(4{,}939)}}$$

$$= 41{,}554.8 \pm 6{,}611.26$$

**Question 55 (3 marks)**

Construct a 95% interval for the average sales revenue generated if every representative distributed 180 copies.

Answer: This is a confidence interval

$$\hat{y} = b_0 + b_1(x_G) = 18{,}018 + 130.76(180) = 41{,}554.8$$

$$\hat{y} \pm t_{\alpha/2} s_\varepsilon \sqrt{\frac{1}{n} + \frac{(x_G - \bar{x})^2}{(n-1)s_x^2}} = 41{,}554.8 \pm (2.000)(3{,}278)\sqrt{\frac{1}{78} + \frac{(180 - 140.47)^2}{(77)(4{,}939)}}$$

$$= 41{,}554.8 \pm 853.02$$

**Question 56 (7 marks)**

A university administrator would like to track admissions to the BBA program in order to forecast class size for the next few years and make hiring decisions of new professors. The administrator asks you to collect admissions data from the past ten years and run the regression model to get this forecast. She provides the following information:

> "I suspect that our admissions have slowly increased since 1994 but you will need to verify this. We admit students to the program each term (fall, winter, and summer) but fall admissions are the largest and very few students begin during the summer term. Also, this past fall we had a larger than usual incoming class because of the double cohort. This will likely be the case for the coming fall term as well."

You have data starting fall 1994 and running to fall 2003 (inclusive).

a. Based on what the administrator told you, sketch a time series that would describe the BBA admissions data over the past 10 years. Specify the values on the x-axis (you can leave the y-axis blank). (2 marks)
b. Write the regression model that you will estimate and explicitly define the dependent and independent variables. (2 marks)
c. Assume that you ran the model that you provided as your answer to part b and all the variables are significant. You now wish to predict admissions for the coming fall term: specify the **value** of the variables that you will use in the regression equation in order to get this prediction. (e.g. if x is a variable and x=5 then 5 is the value of the variable x). (3 marks)

**Your answer:**

Not relevant – we did not cover time series this term

**Question 57 (10 marks)**

What are the success factors in a business career? To investigate this question, a staffing consultant has modeled the salaries of 100 people related to the following variables:

Years – number of years since graduation
MBA – 1 if the person has an MBA, 0 otherwise
Gender – 1 if the person is male, 0 otherwise
Cont. Education – the number of continuing education courses attended since graduation
Community Service – 1 if the person is active in at least one community service activity, 0 otherwise

| SUMMARY OUTPUT | | | | | |
|---|---|---|---|---|---|
| | | | | | |
| *Regression Statistics* | | | | | |
| Multiple R | 0.9737 | | | | |
| R Square | 0.9482 | | | | |
| Adjusted R Square | 0.9454 | | | | |
| Standard Error | 3015 | | | | |
| Observations | 100 | | | | |
| | | | | | |
| ANOVA | | | | | |
| | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 5 | 15636303318 | 3127260664 | 344.04 | 0.0000 |
| Residual | 94 | 854451113 | 9089905 | | |
| Total | 99 | 16490754431 | | | |
| | | | | | |
| | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | |
| Intercept | 35916 | 3141 | 11.44 | 0.0000 | |
| Years | 1022 | 48.93 | 20.88 | 0.0000 | |
| MBA | 725.7 | 961.5 | 0.75 | 0.4523 | |
| Gender | 3729 | 619.82 | 6.02 | 0.0000 | |
| Cont. Education | 439.15 | 80.69 | 5.44 | 0.0000 | |
| Community Service | 1089.7 | 632.0 | 1.72 | 0.0879 | |

Analyze this model and interpret its results.

**The model**
Income = 35,916+1,022(Yrs)+725.7(MBA)+3,729(Gender)+439.15(CE)+1,089.7(CS)+$\varepsilon$

**Overall Analysis:**

The $\mathbf{R^2}$ = 94.82% indicates that a large proportion of the total variation in income is explained by the variables in this model. The **F-ratio** of 344.04 > $F_{CRIT}$ = 2.305 (and it's p-value of 0.0000) indicates that a significant amount of the variation in income has been explained in the model. The **Standard Error of Estimate** (SEE) is 3,105, which is a substantial improvement from the Standard Deviation of the income which is $(164,90,754,431/99)^{0.5}$ = 12,906.

**Individual Variable Analysis:**

**Years:** Each year adds $1,022 to the individual's income. This is a plausible amount and the t-ratio of 20.88 (>1.99) indicates that this variable is highly significant.

**MBA:** An MBA adds $725.7 to the individual's income. As the t-ratio is only 0.75 (<1.99) this variable should be dropped from the model and the model re-estimated.

**Gender** (1=M, 0=F): Males are earning an average of $3,729 more than their female counterparts. With a t-ratio of 6.02 (>1.99) this is highly significant, and suggests that there is evidence of gender discrimination.

**Continuing Education:** Each CE course taken increases income by $439. The t-ratio (5.44>1.99) indicates that this is a significant variable.

**Community Service** (1 if involved, 0 otherwise): Employees involved in community service earn $1,089.7 more than those not involved. The t-ratio of this variable (1.66 < t < 1.99) puts this variable in the statistical gray area, we would not reject the null hypothesis that the coefficient equals zero at 5%, but not at 10%. We can also see this from the p-value of 8.79%.

**Recommendation:**
Since we will be re-estimating the model without the MBA variable, we should see is the significance of the Community Service variable increases. If it remains in the statistical gray area, it is the user's judgment call whether to leave it in the model or not.