KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DEPARTMENT OF MATHEMATICS AND STATISTICS
Term 151
STAT 212: BUSINESS STATISITICS II
Final Major Exam

Tuesday December 22, 2015 7:00 – 9:15 pm

Please circle your instructor name:

R. Anabosi M. Saleh

Name: ___________________________ ID #: ___________________ Serial# _________

Important Note:
Show all your work including formulas, intermediate steps and final answer.
You may assume α = 0.05 for testing if not otherwise stated.

Question No	**Full Marks**	**Marks Obtained**
Q1	6	
Q2	4	
Q3	8	
Q4	20	
Q5	10	
Q6	12	
Q7	10	
Q8	5	
Q9	5	
TOTAL	**80**	

Q1: (6 *points*) A sample of 500 shoppers was selected in Dammam to determine various information concerning customer behavior. Among the questions asked "Do you enjoy shopping clothing?" of 260 males, 140 answered yes, of 240 females, 160 answered yes. Using the p – value approach, is there any evidence of significant difference between males and females in the population that enjoy shopping for clothing at 1% level of significant? (***Clearly write the hypotheses, the decision rule, your decision and the conclusion***)

Q2: An automotive would like to be able to predict automobile mileages. He believes that the three most important characteristics that affect mileage are horsepower, the number of cylinders (4 or 6) of a car and vehicle type (sport or not). He believes that the appropriate model is:

$$\hat{y} = 40 + 15x_1 - 20x_2 + 10x_3$$

Where: Y = mileage

$$x_1 = \text{horsepower},\ x_2 = \begin{cases} 1 & \text{if 4 cylinders} \\ 0 & \text{if 6 cylinders} \end{cases} \text{ and } x_3 = \begin{cases} 1 & \text{if sport car} \\ 0 & \text{if not} \end{cases}$$

a. (3 *points*) Interpret the regression coefficients.

b. (1 *point*) Predict the average for sport car with 100 horsepower and 6 cylinder car.

Q3: A real estate association in a community would like to study the relationship between the size of a single – family house (measured by number of rooms) and the selling price of the house (in thousands of dollars). Two different neighborhoods are included in the study, one on the east side of the community (=0) and the other in the west side (=1). A random sample of size 20 houses was selected.

Consider the Minitab output below, answer the following

a. (3 *points*) Holding constant the effect of variable (Rooms), what is the percentage of the variation in the Price that can be explained by the variation in the variable (Neighborhood)?

b. (5 *points*) Test at the 0.05 level to determine whether the Neighborhood variable significantly improves the model given that Rooms variable is included.

(***Clearly write the hypotheses, the decision rule, your decision and the conclusion***)

Complete Model

Regression Analysis: Price versus Rooms; Neighborhood

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	6635.5	3317.7	55.39	0.000
Residual Error	17	1018.2	59.9		
Total	19	7653.7			

Reduced Model

Regression Analysis: Price versus Rooms

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	5862.9	5862.9	58.93	0.000
Residual Error	18	1790.8	99.5		
Total	19	7653.7			

Q4: An investment company wants to develop a regression model using appropriate independent variables These variables are as follows:

Y: Stock price
X1: Earnings per share annual growth rate as a percentage.
X2: Profits for last four quarters.
X3: Stock price 1 year earlier.
X4: Price earnings (P/E) ratio over last four quarters.

In addition to these variables, there are three mutually exclusive markets where the stock can be traded:

- Over the Counter (OTC)
- New York Stock Exchange (NYSE)
- NASDAQ

a. (2 *points*) Define appropriate variables to indicate the markets where the stock can be traded.

Consider the Minitab output below

Regression Analysis: Y versus X1; X2; X3; X4; X5; X6

Predictor	Coef	SE Coef	T	P
Constant	15.965	4.916	3.25	0.002
X1	0.01943	0.01757	1.11	0.272
X2	0.10782	0.02815	3.83	0.000
X3	0.12385	0.04737	2.61	0.010
X4	0.15853	0.04556	3.48	0.001
X5	-0.615	4.991	-0.12	0.902
X6	-3.591	5.741	-0.63	0.533

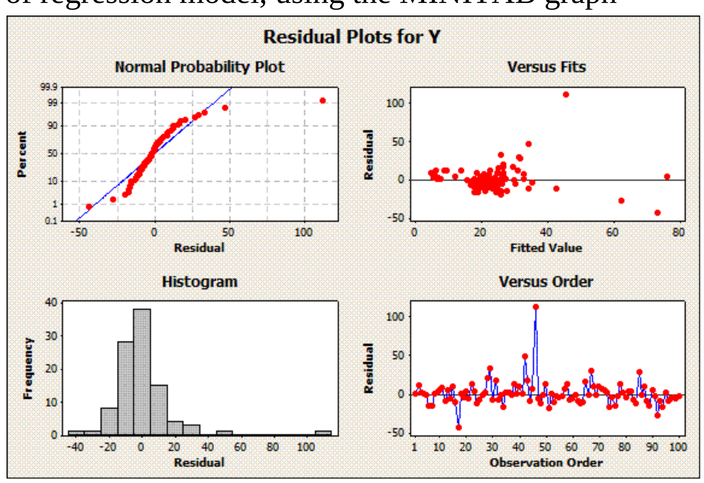
S = 17.2713   R-Sq = 30.3%   R-Sq(adj) = 25.8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	6	12065.4	2010.9	6.74	0.000
Residual Error	93	27741.6	298.3		
Total	99	39807.0			

Answer the questions:

b. (4 *points*) What can you say about the assumptions of regression model, using the MINITAB graph for the residuals? Explain.



c. (2 *points*) Predict the stock price if, X1=99, X2=20, X3=10, X4=14 and the company stock is traded in the NASDAQ.

d. (1 *point*) How much of the total variation in the stock price is explained by the variation in the explanatory variables?

e. (5 *points*) Would you conclude that the model is significant at 10% level of significance? Explain.
(***Clearly write the hypotheses, the decision rule, your decision and the conclusion***)

f. (6 *points*) At 10% level of significance, which of the predictors can be concluded to be significant in explaining the variation in the stock? Explain.
Clearly write the hypotheses, the decision rule, your decision and the conclusion)

Q5: (*10 points*) In reference to question 4, adding the interaction terms to the model, the following Minitab results are given:

Regression Analysis: Y versus X1; X2; ...

Predictor	Coef	SE Coef	T	P
Constant	7.380	6.521	1.13	0.261
X1	0.02143	0.01781	1.20	0.232
X2	0.5262	0.2081	2.53	0.013
X3	-0.0202	0.1644	-0.12	0.903
X4	0.17753	0.04616	3.85	0.000
X5	6.484	6.429	1.01	0.316
X6	8.493	7.312	1.16	0.249
X2*X5	-0.3836	0.2114	-1.81	0.073
X2*X6	-0.4893	0.2120	-2.31	0.023
X3*X5	0.1419	0.1752	0.81	0.420
X3*X6	0.1165	0.1943	0.60	0.550

S = 16.8971   R-Sq = 36.2%   R-Sq(adj) = 29.0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	10	14396.3	1439.6	5.04	0.000
Residual Error	89	25410.7	285.5		
Total	99	39807.0			

At 5% level of significance, determine whether the interactions collectively make a significant contribution to the model.

(***Clearly write the hypotheses, the decision rule, your decision and the conclusion***)

Q6: The following table contains the number of complaints received in a department store from 1999 to 2004

Year	1999	2000	2001	2002	2003	2004
Complaints	432	540	972	1104	1296	1680

a. (*2 points*) Find the first two terms of a three–years moving average.

b. (*2 points*) Using exponential smoothing with smoothing constant of $\frac{1}{3}$, find the first two terms of the exponentially smoothed series.

Model 1; the linear trend forecasting equation is given by

`Complaints = 140 + 247 period time` With MAD = 65.14

Model 2; the exponential regression model is given by

`log(Complaints) = 2.54 + 0.118 period time` With MAD = 85.78

a. (*4 points*)Interpret the coefficients for **<u>both</u>** models.

b. (*2 points*) Calculate the fitted value for 2007 using both Models.

c. (*2 points*) Which model is better? Explain.

Q7: A contractor developed a multiplicative time-series model to forecast the number of contracts in future quarters, using quarterly data on number of contracts during the 3-year period from 1996 to 1998. The following is the resulting regression equation

$$\log \hat{y} = 3.37 + 0.0117X - 0.083\, Q_1 + 1.28\, Q_2 + 0.0617\, Q_3$$

where $\hat{y}$ is the estimated number of contracts in a quarter
X is the coded quarterly value with X = 0 in the first quarter of 1996.
Q_1 is a dummy variable equal to 1 in the first quarter of a year and 0 otherwise.
Q_2 is a dummy variable equal to 1 in the second quarter of a year and 0 otherwise.
Q_3 is a dummy variable equal to 1 in the third quarter of a year and 0 otherwise.

a. (2 *points*) Interpret the constant 3.37 in the regression equation.

b. (2 *points*) interpret the coefficient of X (0.0117) in the regression equation.

c. (2 *points*) interpret the coefficient of Q_1 (- 0.083) In the regression equation.

d. (2 *points*) Interpret the coefficient of Q_3 (0.0617) in the regression equation.

e. (2 *points*) What is the forecast number of contracts in the third quarter of 1999?

Q8: (5 *points*) The managing partner of an advertising agency believes that his company's sales (y_t) are related to industry sales (t). He uses MINITAB to analyze the last 15 years of quarterly data. The following regression model has been developed: $\hat{y}_t = 3.862 + 0.040451\, t$ with the standard error = 0.9224 and the Durbin-Watson statistic = 1.59. The partner wants to test if a positive autocorrelation exist. Using 5% level of significance, does a positive autocorrelation exist?

(***Clearly write the hypotheses, the decision rule, your decision and the conclusion***)

Q9: Given below are the prices of a basket of four food items from 1996 to 2000.

Year	Wheat($/Bushel)	Corn($/Bushel)	Soybeans($/Bushel)	Milk($/hundredweight)
1996	4.25	3.71	7.41	15.03
1997	3.43	2.7	7.55	13.63
1998	2.63	2.3	6.05	15.18
1999	2.11	1.97	4.68	14.72
2000	2.16	1.9	4.81	12.32

a. (*1 point*) Construct a simple price indexes for the milk, in 1999 using 1996 as the base year?

b. (*4 points*) What is the Laspeyre's price index for the basket of four food items in 1998 that consisted of 50 bushels of wheat, 30 bushels of corn, 40 bushels of soybeans and 80 hundredweight of milk in 1996 using 1996 as the base year?

The one-sample problem

P-value approach

Hypothesis type	P – value	
Lower tail	$P(Z < z)$	$P(T_v < t)$
Upper tail	$P(Z > z)$	$P(T_v > t)$
2-tailed	$2P(Z > z)$	$2P(T_v > t)$

Test statistics

$$Z_{stat} = \frac{(\bar{x}-\mu_0)\sqrt{n}}{\sigma} \quad or \quad T_{stat} = \frac{(\bar{x}-\mu_0)\sqrt{n}}{s}$$

$$Z_{stat} = \frac{\hat{p} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}} \quad where \quad \hat{p} = \frac{x}{n}$$

For testing hypotheses about *σ*

Test statistic $\chi^2_{stat} = \frac{(n-1)s^2}{\sigma_0^2}$

For testing hypotheses about *σ₁ – σ₂*

Test statistic $F_{stat} = \frac{s_i^2}{s_j^2} \quad where \quad \begin{matrix} fd_1 = n_i - 1 \\ fd_2 = n_j - 1 \end{matrix}$

The two-sample problem

The i[th] **paired difference** $d_i = x_{1i} - x_{2i}$ &

$$T_{stat} = \frac{(\bar{d}-\mu_0)\sqrt{n}}{s_d} \quad \& \quad \bar{d} = \frac{\sum d}{n} \quad \& \quad s_d = \sqrt{\frac{\sum d^2 - n\bar{d}^2}{n-1}}.$$

$$Z_{stat} = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \text{ or } Z_{stat} = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

$$T_{stat} = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \quad where$$

$$s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$$

$$T_{stat} = \frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_1^2}{n_2}}} \quad where$$

$$df = \frac{\left(\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1}+\frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

To test hypotheses about π_1 - π_2

$$Z_{stat} = \frac{(\hat{p}_1 - \hat{p}_2) - \pi_0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

where $\hat{p} = \frac{x_1+x_2}{n_1+n_2}$ & $\hat{p}_1 = \frac{x_1}{n_1}$ $\hat{p}_2 = \frac{x_2}{n_2}$

Test of Independence in *rxc* table

$$\chi^2_{stat} = \sum_{j=1}^{r}\sum_{i=1}^{c}\frac{\left(O_{ij} - e_{ij}\right)^2}{e_{ij}} \quad df = (c-1)(r-1)$$

Marascuillo's Test for Pair-wise Proportions

$$|p_i - p_j| > \sqrt{\chi^2_\alpha}\sqrt{\frac{p_i(1-p_i)}{n_i}+\frac{p_j(1-p_j)}{n_j}}$$

Sample correlation coefficient

$r = \frac{Sxy}{\sqrt{SxxSyy}}$ *where* $Sxx = \sum(x-\bar{x})^2$

$Syy = \sum(y-\bar{y})^2$ and $Sxy = \sum(x-\bar{x})(y-\bar{y})$

For testing ρ

Test statistic $T_{stat} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ & $df = n-2$

Estimated regression model

$$\hat{y} = b_0 + b_1 x$$

$$where \quad b_1 = \frac{Sxy}{Sxx} \quad \& \quad b_0 = \bar{y} - b_1\bar{x}$$

Total Sum of Squares

$$SST = Syy = \sum(y-\bar{y})^2 = \sum y^2 - n\bar{y}^2$$

$$SSR = b_1 Sxy = \frac{Sxy^2}{Sxx} \quad \& \quad SSE = SST - SSR$$

Coefficient of Determination

$R^2 = \frac{SSR}{SST}$ and $R^2_{adj} = 1 - (1-R^2)\left(\frac{n-1}{n-k-1}\right)$

Standard Error of the model

$$S_\epsilon = S_{y.x} = \sqrt{\frac{SSE}{n-k-1}}$$

Standard Error of the Slope

$$S_{b_1} = \frac{S_\epsilon}{\sqrt{Sxx}}$$

For testing β_1

The test statistic & **C.I. for the slope**

$$T_{stat} = \frac{b_1 - \beta_{10}}{S_{b_1}} \quad where \quad df = n-k-1$$

and $b_1 \pm t_{\frac{\alpha}{2},df} S_{b_1}$

C.I. for the **mean** of y given a particular x_p

$$\hat{y} \pm t_{\frac{\alpha}{2},df} S_\epsilon \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{Sxx}}$$

P.I. estimate for an **Individual value of y** given a particular x_p

$$\hat{y} \pm t_{\frac{\alpha}{2},df} S_\epsilon \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{Sxx}}$$

For testing

H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$ H_A: at least one $\beta_i \neq 0$

Test statistic

$$F_{stat} = \frac{MSR}{MSE} \quad where \quad \begin{matrix} df_1 = k \\ df_2 = n-k-1 \end{matrix}$$

Contribution of a Single Independent Variable X_j

$$SSR(X_j | all\ other\ X's) = SSR_{Full} - SSR_{(X_j)}$$

$$r^2_{Y2.1} = \frac{SSR(X_j | all\ other\ X's)}{SST_{Full} - SSR_{Full} + SSR(X_j | all\ other\ X's)}$$

The Partial F-Test Statistic

$$F_{stat} = \frac{SSR(X_j | all\ other\ X's)}{MSE_{Full}} \quad where \quad \begin{matrix} df_1 = 1 \\ df_2 = n - k_{full} - 1 \end{matrix}$$

For testing H_0: $\beta_{j+1} = \beta_{j+2} = \ldots\ \beta_{j+m} = 0$

against H_A: at least one $\beta_i \neq 0$

Test statistic

$$F_{stat} = \frac{\frac{SSR_{Full} - SSR_{Reduced}}{m}}{MSE_{Full}}$$

$$where \quad \begin{matrix} df_1 = m = k_{Full} - k_{Reduced} \\ df_2 = n_{Full} - k_{full} - 1 \end{matrix}$$

Variance Inflationary Factor $VIF_j = \frac{1}{1-R_j^2}$

$$C_p = \frac{(1-R_k^2)(n-T)}{1-R_T^2} - (n-2k-2)$$

Simple Index number formula & Unweighted aggregate price index formula (respectively)

$$I_t = \frac{y_t}{y_0}(100) \quad \& \quad I_t = \frac{\sum p_t}{\sum p_0}(100)$$

Weighted Aggregate Price Indexes

Paasche $I_t = \frac{\sum q_t p_t}{\sum q_t p_0}(100)$

Laspeyres $I_t = \frac{\sum q_0 p_t}{\sum q_0 p_0}(100)$

Single Exponential Smoothing Model

$$E_{t+1} = w y_{t+1} + (1-w)E_t$$

Exponential Trend Model

$$y_t = \beta_0 \beta_1^{x_t} \varepsilon_t$$

Transformed Exponential Trend Model

$$\log(y_t) = \log(\beta_0) + x_t \log(\beta_1) + \log(\varepsilon_t)$$

Exponential Model for Quarterly data

$$y_t = \beta_0 \beta_1^{x_t} \beta_2^{Q_1} \beta_3^{Q_2} \beta_4^{Q_3} \varepsilon_t$$

***P*th-order Autoregressive Model**

$$y_t = A_0 + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + \varepsilon$$

where A_p : the p^{th} autoregressive parameter