# KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
# DEPARTMENT OF MATHEMATICAL SCIENCES
# DHAHRAN, SAUDI ARABIA

**STAT 212: BUSINESS STATISTICS II**

Semester 061

FINAL EXAM

Monday, January 29, 2007

7:30 am – 9:30 am

Please **circle** your:

| Instructor's name | & section # (time) |
|---|---|
| Marwan Al-Momani | **1** (9 – 10 am) |
| Raid Anabosi | **2** (10 – 11 am)<br>**3** (11 am – 12 pm)<br>**4** (1 – 2 pm) |

Name: Student ID#: Serial #:

- **Make sure that you have 6 written questions, 1 multiple choice question, 1 True/False question, an answer sheet, and the Durbin-Watson table.**
- **If α is not mentioned in the question, assume it to be 0.05.**

| Question No | Full Marks | Marks Obtained |
|---|---|---|
| 1 | **10** | |
| 2 | **17** | |
| 3 | **11** | |
| 4 | **13** | |
| 5 | **20** | |
| 6 | **11** | |
| 7 | **14** | |
| 8 | **14** | |
| **Total** | **110** | |

**Q1.** (2+7+1 points)

Consider the following MINITAB output on the total sales for the years from 1990 to 2005

$\hat{Y}_t = -709.30 + 553.71\ t$

```
Predictor    Coef  SE Coef      T      P
Constant   -709.3    950.6  -0.75  0.468
Quarter    553.71    98.31   5.63  0.000

S = 1812.68   R-Sq = 69.4%   R-Sq(adj) = 67.2%

Analysis of Variance

Source          DF         SS         MS      F      P
Regression       1  104242884  104242884  31.73  0.000
Residual Error  14   46001509    3285822
Total           15  150244393

Durbin-Watson statistic = 0.397974
```

**a.** What is the forecasted value for the year 2010

| The forecasted Value is: |
|---|
| |

**b.** Test to determine if a positive auto correlation exits between residuals. Use $\alpha = 0.01$

| $H_0$ : | $H_{A:}$ |
|---|---|
| Test Statistic Value: | |
| Critical Values : $d_L$ = | $d_U$= |
| Decision & decision rule: | |

**c.** Based on your conclusion in part (b), which, if any, of the residual assumption is violated?

| Answer: |
|---|
| |

*Q2.* (5+9+1+2 points)

Consider the following set of sales data given in (SR 100,000)

| Year | Quarter | Sales | 4-period (M. Av) | C.M. Av | R. M. Av |
|---|---|---|---|---|---|
| 1997 | 1 | 2191 | | | |
| | 2 | 2423 | 2526.5 | | |
| | 3 | 2677 | 2683.75 | 2605.125 | 1.0276 |
| | 4 | 2815 | | 2705.875 | 1.0403 |
| 1998 | 5 | 2820 | 2777.5 | 2752.75 | 1.0244 |
| | 6 | 2600 | 3051.25 | 2914.375 | |
| | 7 | 2875 | 3801.25 | 3426.25 | 0.8391 |
| | 8 | 3910 | 5118 | 4459.625 | 0.8768 |
| 1999 | 9 | 5820 | | 5839 | 0.9968 |
| | 10 | 7867 | 8747.25 | 7653.625 | 1.0279 |
| | 11 | 8643 | 10684.5 | 9715.875 | 0.8896 |
| | 12 | 12659 | 12637.25 | | 1.0856 |
| 2000 | 13 | 13569 | 14840.5 | 13738.88 | 0.9876 |
| | 14 | 15678 | 17539.75 | 16190.13 | |
| | 15 | 17456 | | | |
| | 16 | 23456 | | | |
| | | | | | |

a. Fill in the blank cells in the above table

b. Find the normalized seasonal index numbers, and interpret the 2nd seasonal index number.

| Normalized Seasonal Index for each quarter | | | |
|---|---|---|---|
| **1st Quarter** | **2nd Quarter** | **3rd Quarter** | **4th Quarter** |
| | | | |

Interpretation:

c. The desesonalized value for the 3rd quarter of the year 2000 =

## *Q*3. (2+7+2 points)

The sales manager of a TV and Stereo company must decide how many 27-inch GB super TVs to order for April. The following sales report shows the sales of the 27-inch GB Super TV for the months of January to March.

| Month | Sales |
|---|---|
| January | 845 |
| February | 825 |
| March | 820 |
| April | |

a. Using $\alpha = 0.3$, the single exponential smoothing forecast for April = ____

b. Using $\alpha = 0.3, \beta = 0.35$, find the double exponential smoothing forecast value for February given that the fitted line equation for the sales is: $\hat{Y} = 832 + 1.30\ t$

| | |
|---|---|
| $C_0 =$ | $T_0 =$ |
| $C_1 =$ | $T_1 =$ |
| $F_2 =$ | |

c. If MAD for the single and double exponential smoothing methods are 14.548, and 15.630 respectively, which method is better? Why?

The better method is _____ because _____

_____.

*Q*4. (1 point for each gap)

A company sells products in several sales territories, each of which is assigned to a single sales representative. A random sample, of 25 sales territories, was drawn and a regression analysis was conducted to determine whether 8 predictor variables could explain sales in each territory. The Minitab output below resulted:

**Regression Analysis: Sales versus Time, Poten, ...**

The regression equation is
Sales = - 1508 + 2.01 Time + 0.0372 Poten + 0.151 AdvExp + 199 Share
+ 291 Change + 5.55 Accounts + 19.8 Work + 8 Rating

| Predictor | Coef | SE Coef | T | P | VIF |
|---|---|---|---|---|---|
| Constant | -1507.8 | 778.6 | -1.94 | 0.071 | |
| Time | 2.010 | 1.931 | 1.04 | 0.313 | 3.3 |
| Poten | 0.037206 | 0.008202 | 4.54 | 0.000 | 2.0 |
| AdvExp | 0.15098 | 0.04711 | 3.21 | 0.006 | 1.9 |
| Share | 199.04 | 67.03 | 2.97 | 0.009 | 3.2 |
| Change | 290.9 | 186.8 | 1.56 | 0.139 | 1.6 |
| Accounts | 5.550 | 4.775 | 1.16 | 0.262 | 5.6 |
| Work | 19.79 | 33.68 | 0.59 | 0.565 | 1.8 |
| Rating | 8.2 | 128.5 | 0.06 | 0.950 | 1.8 |

S = 449.0 R-Sq = 92.2% R-Sq(adj) = 88.3%

Durbin-Watson statistic = 1.87

**Stepwise Regression: Sales versus Time, Poten, ...**

| Step | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Constant | 709.32 | 50.29 | -327.24 | -1441.93 |
| Accounts | 21.7 | 19.0 | 15.6 | 9.2 |
| T-Value | 5.50 | 6.41 | 5.19 | 3.22 |
| P-Value | 0.000 | 0.000 | 0.000 | 0.004 |
| AdvExp | | 0.227 | 0.216 | 0.175 |
| T-Value | | 4.50 | 4.77 | 4.74 |
| P-Value | | 0.000 | 0.000 | 0.000 |
| Poten | | | 0.0219 | 0.0382 |
| T-Value | | | 2.53 | 4.79 |
| P-Value | | | 0.019 | 0.000 |
| Share | | | | 190 |
| T-Value | | | | 3.82 |
| P-Value | | | | 0.001 |
| S | 881 | 650 | 583 | 454 |
| R-Sq | 56.85 | 77.51 | 82.77 | 90.04 |
| R-Sq(adj) | 54.97 | 75.47 | 80.31 | 88.05 |
| C-p | 67.6 | 27.2 | 18.4 | 5.4 |

**Best Subsets Regression: Sales versus Time, Poten, ...**

Response is Sales

| Vars | R-Sq | R-Sq(adj) | C-p | S | Time | Poten | AdvExp | Share | Change | Accounts | Work | Rating |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 56.8 | 55.0 | 67.6 | 881.09 | | | | | | X | | |
| 1 | 38.8 | 36.1 | 104.6 | 1049.3 | X | | | | | | | |
| 2 | 77.5 | 75.5 | 27.2 | 650.39 | | | X | | | X | | |
| 2 | 74.6 | 72.3 | 33.1 | 691.11 | | X | | X | | | | |
| 3 | 84.9 | 82.7 | 14.0 | 545.52 | | X | X | X | | | | |
| 3 | 82.8 | 80.3 | 18.4 | 582.64 | | X | X | | | X | | |
| 4 | 90.0 | 88.1 | 5.4 | 453.84 | | X | X | X | | X | | |
| 4 | 89.6 | 87.5 | 6.4 | 463.93 | X | X | X | X | | | | |
| 5 | 91.5 | 89.3 | 4.4 | 430.21 | X | X | X | X | X | | | |
| 5 | 91.2 | 88.9 | 5.0 | 436.75 | | X | X | X | X | X | | |
| 6 | 92.0 | 89.4 | 5.4 | 427.99 | X | X | X | X | X | X | | |
| 6 | 91.6 | 88.9 | 6.1 | 438.20 | | X | X | X | X | X | X | |
| 7 | 92.2 | 89.0 | 7.0 | 435.66 | X | X | X | X | X | X | X | |
| 7 | 92.0 | 88.8 | 7.3 | 440.29 | X | X | X | X | X | X | | X |
| 8 | 92.2 | 88.3 | 9.0 | 449.02 | X | X | X | X | X | X | X | X |

**Correlations: Sales, Time, Poten, AdvExp, Share, Change, Accounts, Work, Rating**

| | Sales | Time | Poten | AdvExp | Share | Change | Accounts | Work |
|---|---|---|---|---|---|---|---|---|
| Time | 0.623 | | | | | | | |
| | 0.001 | | | | | | | |
| Poten | 0.598 | 0.454 | | | | | | |
| | 0.002 | 0.023 | | | | | | |
| AdvExp | 0.596 | 0.249 | 0.174 | | | | | |
| | 0.002 | 0.230 | 0.405 | | | | | |
| Share | 0.484 | 0.106 | -0.211 | 0.264 | | | | |
| | 0.014 | 0.614 | 0.312 | 0.201 | | | | |
| Change | 0.489 | 0.251 | 0.268 | 0.377 | 0.085 | | | |
| | 0.013 | 0.225 | 0.195 | 0.064 | 0.685 | | | |
| Accounts | 0.754 | 0.758 | 0.479 | 0.200 | 0.403 | 0.327 | | |
| | 0.000 | 0.000 | 0.016 | 0.338 | 0.046 | 0.110 | | |
| Work | -0.117 | -0.179 | -0.259 | -0.272 | 0.349 | -0.288 | -0.199 | |
| | 0.577 | 0.391 | 0.212 | 0.188 | 0.087 | 0.163 | 0.341 | |
| Rating | 0.402 | 0.101 | 0.359 | 0.411 | -0.024 | 0.549 | 0.229 | -0.277 |
| | 0.046 | 0.630 | 0.078 | 0.041 | 0.911 | 0.004 | 0.272 | 0.180 |

From the previous Minitab output, answer the following:

**a.** The predictor that is ***severely*** related to some of the other predictors is _____.

**b.** From the full regression model, the predictor that is the most significant to the response variable is _____.

**c.** The predictor that is the least significantly related to the response variable is_____.

**d.** The variation in the predictors explains _____ of the variation in the response variable.

**e.** The final model, from the stepwise regression analysis, has ____ predictors with C-p = _____.

**f.** The final model, from the stepwise regression analysis, is:

$\hat{Y}$ = _____.

**g.** The **best** subset regression model has _____ predictors with C-p = _____.

**h.** The **worst** subset regression model has a coefficient of determination = _____ and a standard error = _____.

**i.** The overall regression model is (significant / not significant).

**j.** If a backward elimination regression is to be made, the first predictor to be removed is _____.

**k.** If a forward selection regression is to be made, the first predictor to be selected is _____.

*Q*5. (1 point for each gap)

Given that the variances of the three variables are:

$S_Y^2 = 519.866$, $S_{X_1}^2 = 29.8042$, and $S_{X_2}^2 = 614.123$,

Fill in the blanks in the given Minitab output below:

**Regression Analysis: Y versus X1, X2**

The regression equation is

Y = ____ + ____ X1 + ____ X2

| Predictor | Coef | SE Coef | T | P | VIF |
|---|---|---|---|---|---|
| Constant | 5.877 | 5.545 | ____ | 0.299 | |
| X1 | 2.5356 | 1.1835 | ____ | 0.041 | ____ |
| X2 | 0.4841 | 0.255 | ____ | 0.068 | 4.0 |

S = ____ R-Sq = ____ % R-Sq(adj) = ____ %

Analysis of Variance

| Source | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| Regression | ____ | ____ | ____ | ____ | 0.000 |
| Residual Error | ____ | 6531.9 | ____ | | |
| Total | 29 | ____ | | | |

| Source | DF | Seq SS |
|---|---|---|
| X1 | 1 | 4433.9 |
| X2 | 1 | 4110.4 |

Predicted Values for New Observations

| New obs | Fit | SE Fit | 95% CI | 95% PI |
|---|---|---|---|---|
| 1 | ____ | 8.41 | (____, 65.10) | (11.57 , ____) |

Values of Predictors for New Observations

| New Obs | X1 | X2 |
|---|---|---|
| 1 | 2.33 | 74.5 |

*Q*6. (2+1+3+2+1+1+1 points)

A student is doing a project to compare ice cream flavor preferences at 3 ice cream stores in different cities. He wants to determine if customer flavor preferences vary by location. He selected a sample of customers at each store and compared the proportions of customers who purchase each flavor at each store.

| | | **Location** | | | |
|---|---|---|---|---|---|
| | | Al-Riyadh | Al-Dammam | Al-Khobar | **Total** |
| **Flavor** | Chocolate | **65** | **48** | **4** | 117 |
| | Vanilla | **51** | **30** | **2** | 83 |
| | **Total** | 116 | 78 | 6 | 200 |

Is flavor preference independent of location? Test at $\alpha = 0.05$

**a.** $H_0$ : _____ vs. $H_A$ : _____.

**b.** Assumption(s): _____.

**c.** The test statistic = _____.

**d.** The critical value(s) = _____.

**e.** The decision rule is: _____.

**f.** The decision is: (Reject $H_0$ / Do not reject $H_0$).

**g.** The conclusion is: The two variables are (independent/ not independent).

## *Q7.* (14 points)

Find the correct answer for each of the following and **shade it thoroughly** in the answer sheet:

1. The cost (in US$) of transportation from the airport to the downtown area depends on the method of transportation. One-way costs for taxi and shuttle bus transportation for a sample of 5 major cities follow.

| City | Atlanta | Chicago | Denver | Huston | LA |
|---|---|---|---|---|---|
| **Taxi** | $15 | $22 | $11 | $15 | $26 |
| **Bus** | $7 | $12.5 | $5 | $4.5 | $11 |

For testing the claim that shuttle bus transportation is cheaper than taxi transportation more than $7, the test statistic value =

a. 2.68.
b. 1.86.
c. 6.81.
d. 8.62.
e. None of the above.

2. The coefficient of determination has one of the following properties:

a. Indicates the strength and the direction of the relationship between x and y.
b. It is the ratio of the unexplained variation to the explained variation.
c. Has the same sign as the slope of the regression line.
d. It ranges from zero to one.
e. b and d.

3. The proportion of cars stolen that do not have an alarm installed is 20% in a sample of 40 cars. The proportion of cars stolen that do have an alarm installed is 10%, also in a sample of 50 cars. The p-value for testing that there is a difference between the two population proportions, at 10% level of significance is:

a. 0.1802.
b. 0.2018.
c. 0.1820.
d. 0.2108.
e. None of the above.

4. Using two independent samples, two population means are compared to determine if a difference exists. The number of observations in the first sample is 30 and the number of observations in the second sample is 35. How many degrees of freedom are associated with the critical value?

    a. 65.
    b. 63.
    c. 64.
    d. 67.
    e. None of the above.

5. A large department store wants to test whether the variances of waiting time in two checkout counters is approximately equal. Two independent random samples, of 25 waiting times, for the first checkout counter and, 20 waiting times, for the second checkout counter give $s_1 = 2.5$ minutes and $s_2 = 3.1$ minutes respectively. The conclusion and critical value of the test are:

    a. The two variances are approximately equal with a critical value = 2.040.
    b. The two variances are not equal with a critical value = 2.114.
    c. The two variances are approximately equal with a critical value = 2.345.
    d. The two variances are not equal with a critical value = 2.452.
    e. None of the above.

6. Which of the following forecasting methods allows the decision maker to weigh the past time series differently to make the model more sensitive to more recent data?

    a. Linear trend forecasting model.
    b. Exponential smoothing model.
    c. Moving average model.
    d. Ratio to moving average method.
    e. All of the above.

7. Under which of the following conditions would you suggest that a double exponential smoothing model should be used instead of a single exponential smoothing model?

    a. When the time-series data exhibit a seasonality.
    b. When there are two main time-series variables involved in the forecast application.
    c. When the data exhibit only random variation.
    d. When the time-series data exhibit a trend.
    e. None of the above.

## *Q*8. (14 points)

Mark each statement as True (**T**) or False (**F**) and **shade it thoroughly** in the answer sheet:

1. We cannot, always, test the population mean, using the t-test statistic, if the sample drawn is a small one and the population standard deviation is unknown.

2. Type-I error is defined as; rejecting the null hypothesis while it is really true.

3. In a simple regression model, the coefficient of determination is a measure for the goodness of the regression fit and is not equal to the correlation coefficient.

4. In general, for testing that one population variance is greater than another population variance we use an F test statistic with the first population variance in the numerator.

5. In a multiple regression model, the sample size must be greater than the number of predictors, in the model, by one at least.

6. If the calculated chi-square statistic gets large, this is an evidence to suggest that the fit of the actual data to the hypothesized distribution is not good, and $H_0$ should be rejected.

7. If a time-series plot indicates that the data do not appear to exhibit a trend, then a double exponential smoothing model would likely be the most appropriate to use rather than single exponential smoothing model.