Statistics 3540
Assignment #3: Feb. 13, 2008
Due in class: Feb. 25, 2008

If a problem does not specify whether it should be done by hand or using a statistical software package, the choice is up to you.

The data are available at `www.math.mun.ca/~sneddon/st3540`

1. Refer back to the DJIA data used in Assignment #2, and answer the following:

   (a) Fit the regression model
   $$y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + e_t$$
   NOTE: You will need to use **Calc - Calculator** to create a new column in Minitab that contains the t^2 values before fitting this regression model.

   (b) Is this model useful for predicting the DJIA closing values? Base your conclusion on the p-value of the appropriate test.

   (c) Test at $\alpha = 0.10$ if the errors in this model are correlated.

2. In this problem you will analyze a time series of Canadian monthly unemployment rates (numbers and percentages), observed from Jan. ??? to Dec. 2007.

   The data are available on the course website in the Minitab worksheet **unemploy.mtw**, and in the plain text file **unemploy.txt**.

   (a) Plot the unemployment rate versus time and describe any features you see in the plot.

   (b) Fit the model
   $$y_t = \beta_0 + \beta_1 t + \beta_{s1} x_{s1,t} + \beta_{s2} x_{s2,t} + \ldots + \beta_{s11} x_{s11,t} + e_t$$
   to the data, where $x_{s1,t} = 1$ if observation is in January, ..., $x_{s11,t} = 1$ if observation is in November. Note that you have to get Minitab to create your dummy variables, based on the values in the column `Month` in the datafile.
   Report the least squares line.

   (c) Is the model useful in predicting the unemployment rate? Base your conclusion on the p-value of the appropriate test.

   (d) Test at $\alpha = 0.05$ whether we can drop the seasonal terms from the model.

   (e) Find a 99% prediction interval for the Canadian unemploymenet rate in Januaray 2008.

3. For the data in question #2:

   (a) Calculate the least squares regression line for the model

   $$y_t = \beta_0 + \beta_1 t + \beta_2 \sin(2\pi t/L) + \beta_3 \cos(2\pi t/L) + e_t$$

   using the appropriate value for L.

   (b) How does the R^2 value for this model compare to the R^2 value found for the model in Question #2?

   (c) Test at $\alpha = 0.05$ whether or not we can drop the seasonal terms from the model.

   (d) Plot the residuals (or standardized residuals) vs. time, and construct a QQ plot. Interpret these plots.

4. The following are the number of reported cases of a new disease over the last 11 months:

```
Month (t):   1 2 3 4 5 6 7  8  9 10 11
------------------------------------
Cases (yt): 1 1 2 3 4 6 8 13 21 27 45
```

   (a) Plot y_t vs. time. Does the use of a growth curve model for forecasting y_t seem appropriate? Explain.

   (b) Using natural logs, define a transformed growth curve model that will be linear in its parameters.

   (c) Plot the natural logs of the y_t values vs. time. Has the log transformation linearized the data?

   You can find the natural log of a column of values in Minitab by selecting **Calc - Calculator**, and choosing **Natural log** for the expression window. Make sure you specify a new column in which to save these new values.

   (d) Find the least squares regression line for the equation

   $$y_t^* = \beta_0 + \beta_1 t + e_t$$

   where $y_t^* = \ln(y_t)$.

5. In class, we defined the test statistic for the overall F-test in regression as $F_{obs} = MSR/MSE$. Given that we know $R^2 = 1 - (SSE/SS_{yy})$, show that we can write

$$F_{obs} = \frac{R^2/k}{(1 - R^2/(n-k-1)}$$

Hint: You may use the fact that $SSR + SSE = SS_{yy}$.

6. A cubic trend model was fit to the yearly closing stock price of an oil company over a 20 year period, assuming the errors were autocorrelated. The least squares regression line was found to be

$$\hat{y}_t = 160 - 100t + 1.2t^2 + 0.18t^3 + 0.3\hat{e}_{t-1}$$

where $\hat{\sigma}^2 = 49$ and the closing value in year 20 was 70.

Find a 95% prediction interval for the price of the stock at the end of year 21.