## *Lab 7:*
## *Regression Topics: Confidence Intervals, Plots and Non-Linear Models*

**Objectives:**

We now assume that the mean of our dependent variable changes as the independent variable changes: $E[\,Y \mid X_t\,] = \beta_0 + \beta_1 X_t$ . The primary objectives for this lab are to: (1) calculate point estimates or predictions for this "population value;" and (2) to create confidence intervals for this "population value" at each level of X. We can create these intervals in Excel using regression results and the formula provided below. This will give us more practice manipulating data in Excel. Minitab actually provides the confidence intervals for us – a good way to check our work. We'll then consider a simple non-linear model that we can estimate using simple regression and learn about interpreting the results of that model.

**Key Terms:**

1. ***Regression.***
2. ***Confidence Intervals.***
3. ***Forecasting and Forecast Intervals.***
4. ***Non-Linear Models.***

**Data**: ***Lab 7 – Regression, Forecasting and Growth Models.xls***

**Exercises:**

- ***Real Interest Rate****.*

1. Open the spreadsheet *du jour*. Start with the **Real Interest Rate** worksheet. Estimate a regression model relating interest rates to inflation. Let's use some different names (X and Y are boring):

$$R_t = \beta_0 + \beta_1 I_t + u_t;$$

where $R_t$ is the T-Bills interest rate (our "*Y* variable") and $I_t$ is the rate of inflation (our "*X* variable"). **You'll need to create % inflation rates** using the CPI. Inflation is the percentage change in the CPI, the following formula can be used to calculate the percentage rate of inflation:

$$I_t = \frac{(CPI_t - CPI_{t-1})}{CPI_{t-1}} \cdot 100 .$$

2. Use the Regression tool (**Data**, **Data Analysis**, **Regression**) to estimate the regression model relating interest rates to the inflation rate. Include a few of the following options available in Excel's regression tool:
   - **Line Fit Plot:** The line fit plot is just a plot of the dependent variable ($R_t$) against the independent variable ($I_t$) with the fitted or predicted values for the dependent variable ( $\widehat{R_t}$ ) included. (Excel gives a funky bar graph – change the chart type to ***Scatter***.) Excel won't put a line in the graph unless you ask for it. Right-click on one of the markers for the ***predicted values*** and then choose **Add Trendline**.
   - **Residual Plot:** The residual plot is a graph of the errors versus the independent variable. Three of the CRM assumptions were related to the disturbances and these errors are our best guesses about the disturbances. Errors should be distributed fairly evenly around the horizontal axis. The amount of variation should not appear to increase as the independent variable increases and there should not be a discernable pattern as the independent variable changes.
   - **Normal Probability Plot:** This plot is used to determine whether the disturbances (and the dependent variable) appear to be normally distributed. (Again, you should change the chart type to Scatter.) In the CRM, we assumed that the disturbances were normally distributed. The

errors, the sample analogues to the disturbances, should reflect the distribution of the disturbances. Expand the size (especially the vertical size) of this graph. The data should appear to be linearly associated. Ideally, you would find that all the data lie on a straight line. These are kind of okay, but they are a bit non-linear looking. We might want to do further tests to see if it's reasonable to assume normally distributed disturbances. What part of our regression analysis is affected if the normality assumption is incorrect?

- ***Prediction and prediction intervals***:

1. **Forecasts**, or **predictions**, are done using the sample regression function. This is an example of ***point estimation or prediction***. **Use the estimated regression model, the sample regression function (SRF), to calculate point estimates or predictions of interest rates for every level of inflation**:

$$\hat{R}_t = \hat{\beta}_0 + \hat{\beta}_1 I_t \ .$$

To calculate predicted values (column E, labeled r-hat), just plug the inflation rate for that year into the ***SRF***. The estimated parameter values are available in the Excel regression output; don't forget to lock those cell addresses using $s when copying.

2. **Confidence Interval**: The sample regression function is an estimator, so it must have a sampling distribution and a standard error. We can do interval estimation just as we've done for our other estimators. The confidence interval is an interval for the ***population value***, E[ $Y \mid X_t$ ], or in our case E[ $R \mid I_t$ ]. The estimated standard error is, in general:

$$s_{(\hat{Y})} = \sqrt{\hat{\sigma}^2 \left[ \frac{1}{n} + \frac{(X_i - \bar{X})^2}{\sum x_i^2} \right]} \ ,$$

or for our model **using our variable names**,

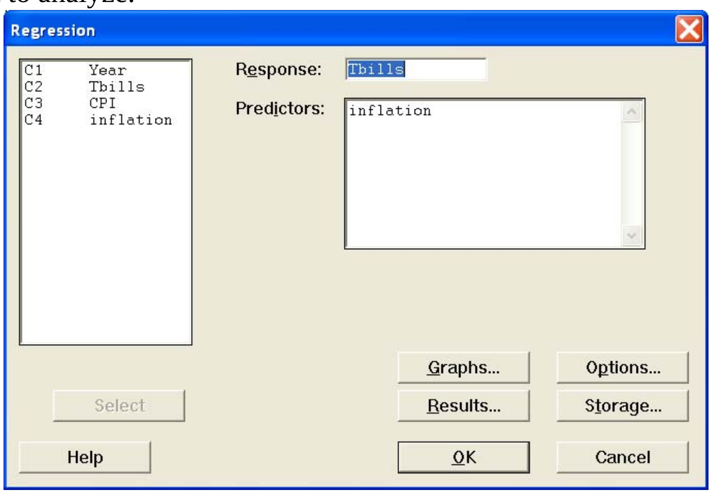
$$s_{(\hat{R})} = \sqrt{\hat{\sigma}^2 \left[ \frac{1}{n} + \frac{(I_t - \bar{I})^2}{\sum i_t^2} \right]}$$

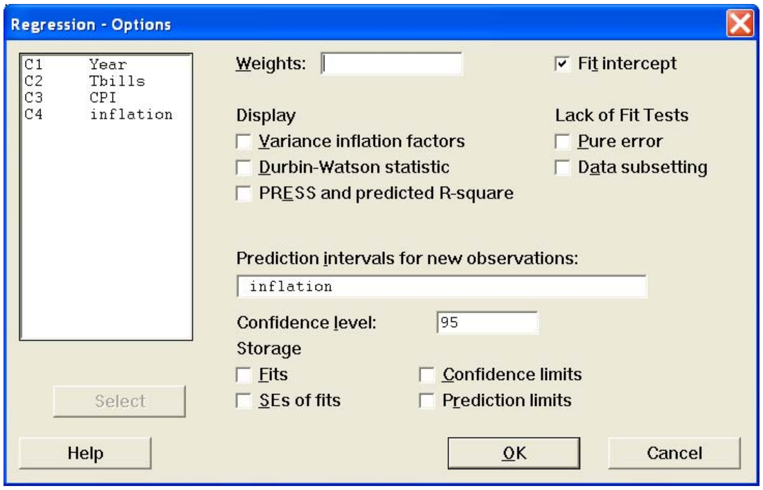
(If you look below the data table, you'll find the numeric value for the denominator of the final term $\sum i_t^2$ .) Use the equation above to **calculate the standard errors for each level of inflation.** Next, create 95% confidence intervals for ***each level of inflation***:

$$\hat{R} \pm t_{(0.025, n-2)} \cdot s_{(\hat{R})} \ .$$

Place the lower and upper limits for the intervals in the columns provided.

3. Check your work. **Open Minitab**. Highlight the data from your Excel spreadsheet and copy. Then, go to Minitab and right-click on the "shaded" cell under C1 and choose ***Paste Cells***. You now have a Minitab Worksheet with a data set to analyze.
4. Minitab's procedures are all in the menus at the top. **Edit**, **Data**, and **Calc** all allow modifications to your data set. When you are ready to analyze your data, you'll find descriptive statistics, correlation, covariance and regression under **Stat**. You'll find graphing tools under, surprise, **Graph**.
5. From the **Stat** menu, choose **Regression**, and then the first option – **Regression**. You'll see the window shown at the right.






You need to select (highlight and click the Select button, or just double-click) Tbills for the Response variable (the dependent variable) and choose ***inflation*** as the Predictor (independent variable). If you click OK, Minitab will use OLS to estimate the regression. Wait!! Select **Options** – you'll see the window shown at the right. About in the middle of that window, there is a box labeled, **Prediction intervals for new observations:** Click in the box and then *Select* the variable ***inflation***. We want to create prediction intervals for each and every value of that variable. Click **OK** once to get back to the Regression window, then click **OK** once again. Minitab will estimate the regression and calculate predicted values, confidence intervals and prediction intervals for each value of the variable inflation. Do the confidence interval limits match your results from Excel? What is the difference between the confidence and prediction intervals?

6. Lastly, use an X,Y scatter diagram to ***plot the predicted values (*** $\widehat{R_t}$ ***) and the lower and upper limits of the confidence intervals*** for each level of inflation. You should see that the interval becomes wider at lower and higher levels of inflation. Why? Where is the confidence interval the narrowest? Is this expected? Look at the formula for $s_{(\hat{Y})}$ to answer this question.

- ***Non-Linear Regressions***

1. The simple regression model can be used to estimate non-linear relationships ***if*** the non-linear relationship can be made to look linear in parameters to the software. To illustrate one simple non-linear relationship, we'll use the data in the worksheets: **Consumer Price Index** and **Per Capita Income**.
2. First, work with the CPI series. Create an XY Scatter Diagram of the ***CPI*** (*the Y variable*) versus the ***year*** (the *X variable*) and **add a trendline**. Use the ***exponential*** trendline to fit a non-linear relationship to these data. Non-linear equations describe these relationships, for example:

$$CPI_t = \beta_0 \exp^{\{\beta_1 t + u_t\}},$$

is one form of non-linear model that gives an exponential or growth curve. A convenient feature of this equation is that $\beta_1$ represents ***the rate of growth in prices*** for the period 1967 – 2006. If we can find a way to estimate this model, we'll be able to estimate the rate at which prices have increased during the past 40 years. Taking ***natural logs*** of both sides of this equation, we get a model that looks linear and can be estimated:

$$\ln CPI_t = \ln \beta_0 + \beta_1 t + u_t.$$

This equation looks just like our simple linear regression model, with a few changes.

3. ***Data manipulation***: First, we need to create two new variables: ln*CPI* and *trend*. The variable, ln*CPI* is created using the ***natural log*** function in Excel and that becomes our dependent variable (the Y-values). Go to C2 and type " **= ln(b2)**." Hit return and Excel calculates the natural log for the CPI value in 1967. Copy this formula to the remaining cells. Now, this new variable is not particularly useful on its own, but it allows us to estimate the growth model. The

independent variable in this model is the trend variable (*t*), which simply identifies the year in the time-series of data. Create a trend variable; we want $t = 1$ in 1967. Just type the first few values, i.e., 1, 2, 3. Then highlight these three cells, grab the lower right corner and copy down. Excel will know you want to create a trend variable.

4. ***Estimation***: Apply OLS using the two new variables. Use **Data**, **Data Analysis** and **Regression** to estimate the model. Ln*CPI* is the dependent variable (the ***Y-values***) and *t* is the independent variable (the ***X-values***).
5. ***Results***: You should find that the estimate of $\beta_1$ is $\hat{\beta}_1 = 0.0475$. The interpretation is that prices (in general) have increased at the ***annual growth rate of 0.0475***, or about 4.75% annually, since 1967.

♦ ***Comparing Price Inflation to Growth in Income***

1. Let's use this simple non-linear model to determine whether prices or income have grown faster over time. I've gathered per capita income data from 1967 – 2006 from the Statistical Abstract of the US web site. To compare growth rates, we estimate the following model:

$$\ln I_t = \ln \beta_0 + \beta_1 t + u_t ,$$

where $I_t$ is per capita income and $\ln I_t$ is the natural log of income, calculated in Excel using the natural log function.

2. You should find that the estimate of $\beta_1$ is $\hat{\beta}_1 = 0.0604$. It appears that the growth in per capita income has been greater than the growth in prices. However, we're comparing two different periods. Let's go back and estimate a CPI growth model for the period 1967 – 2006.
3. Results: So, what's the verdict? It appears that the growth in prices has been less than the growth in income during the period 1967 – 2006. Why? What are you comparing? In general, are consumers in 2006 better off or worse off than they were in 1967? Is there a statistically important difference? How might you turn this into a hypothesis test?

These simple regression models are useful for comparing growth rates for any number and types of variables: sales, wages, income, prices, etc.