BUEC 433 Spring 2009
Statistics Review Questions

Some of you have asked me how familiar you should be with the material that we covered in our statistics review. Ultimately, you are responsible for everything that we cover in class, in the assigned readings, and in the questions that we do. Being really familiar with the basic concepts that we covered today will make the rest of the course much easier. So, I'll give you some additional exercises to help you get familiar with the material. The following sample moments are fundamental to applied statistics, so you'll be expected to know how to calculated them.

Sample Moment Formulas:

Sample Mean: $\overline{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$

Sample Variance: $s^2 = \frac{\sum_{t=1}^{T}(y_t-\overline{y})^2}{T-1}$

Sample Standard Deviation: $s = \sqrt{\frac{\sum_{t=1}^{T}(y_t-\overline{y})^2}{T-1}}$

Sample Skewness: $\widehat{S} = \frac{\frac{1}{T}\sum_{t=1}^{T}(y_t-\overline{y})^3}{s^3}$

Sample Kurtosis: $\widehat{K} = \frac{\frac{1}{T}\sum_{t=1}^{T}(y_t-\overline{y})^4}{s^4}$

Sample Covariance: $\widehat{cov} = \frac{\sum_{t=1}^{T}(y_t-\overline{y})(x_t-\overline{x})}{T-1}$

Sample Correlation: $\widehat{corr} = \frac{\widehat{cov}}{s_y s_x} = \frac{\sum_{t=1}^{T}(y_t-\overline{y})(x_t-\overline{x})}{(T-1)s_y s_x}$

Calculate the above sample moments, by hand, for the following small data sets:

$$
\begin{aligned}
x &= \langle 1, 2, 3, 3, 6 \rangle \\
y &= \langle 5, 2, 4, 3, 1 \rangle
\end{aligned}
$$

Now, suppose we fit the following linear regression model:

$$
\begin{aligned}
y_t &= \beta_0 + \beta_1 x_t + \varepsilon_t \\
\varepsilon_t &\sim iid(0, \sigma^2) \text{ for } t = 1, 2, \cdots, 5
\end{aligned}
$$

If our estimates were $\beta_0 = 5$ and $\beta_1 = -0.6$, draw the scatter plot of the (x, y) points as well as the line representing the fitted values. Calculate the residuals of this regression (recall, $e_t = y_t - \widehat{y}_t$). What is the sum of the squared residuals?

Whenever we run a regression, we should examine and understand the regression diagnostic statistics. You don't need to memorize these formulas, but you need to understand what they mean and how to interpretate their values:

Some Regression Diagnostic Formulas

Sum of Squared Residuals: $SSR = \sum_{t=1}^{T} e_t^2$

F-Statistic: $F = \frac{(SSR_{res} - SSR)/(k-1)}{SSR/(T-k)}$

Standard Error of the Regression: $SER = \sqrt{\frac{\sum_{t=1}^{T} e_t^2}{T-k}}$

R-Squared: $R^2 = 1 - \frac{\sum_{t=1}^{T} e_t^2}{\sum_{t=1}^{T} (y_t - \overline{y})^2}$

Adjusted R^2: $\overline{R}^2 = 1 - \frac{\frac{1}{T-k}\sum_{t=1}^{T} e_t^2}{\frac{1}{T-1}\sum_{t=1}^{T} (y_t - \overline{y})^2}$

Akaike Info Criterion: $AIC = e^{(\frac{2k}{T})} \frac{\sum_{t=1}^{T} e_t^2}{T}$

Schwarz Info Criterion: $SIC = T^{(\frac{k}{T})} \frac{\sum_{t=1}^{T} e_t^2}{T}$

Durbin-Watson Stat: $DW = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$

Jarque-Bera Test: $JB = \frac{T}{6}(\widehat{S}^2 + \frac{1}{4}(\widehat{K} - 3)^2)$

Note: in the above formulas, k is the number of right-hand side variables including the intercept. So in a simple regression with an intercept and one independent variable, $k = 2$. Also, for the F-statistic, SSR_{res} is the sum of squared residuals for a restricted regression that only has an intercept on the right-hand side.

If you want some extra practice, calculate the above diagnostic statistics for the simple regression on the previous page. For the F-statistic, let the restricted regression give an estimate of the intercept of 3. So you can calculate the fitted values in the restricted regression from $\widehat{y}_{res} = 3$ for all values of x.