ECO 5350 Prof. Tom Fomby
Introductory Econometrics Fall 2007

TOPICS TO CONSIDER
FOR MID-TERM I

The Mid-term in this class is scheduled for Thursday, September 27, 2007 at the regular class time. Don't spend a whole lot of time memorizing formulas. Following these review points you will see the formula sheet that I will be handing out to you in class. Given that we will not have access to a computer during the test you should review up on how to determine critical values for hypothesis tests directly from statistical tables like the Z, t, and F tables. You should commit to memory the ANOVA table I presented in class.

1. The test will cover the material in Chapters 1 – 5 in your Stock and Watson textbook and Sections I – VIII (parts A and B) in our course outline. Also it will cover the material in Exercises 1 – 3 and all of the previous Quick Quizzes (QQs). See your class website for the Keys to the Exercises and QQs.
2. What are the different types of data sets that economists are called on to investigate? **Cross-Section data**, **Time Series data**, and **Panel data.** What distinguishes these different data sets? Can you give me an example of each? Why is it important to be able to distinguish between them?
3. What is the distinction between a **discrete probability distribution** and a **continuous probability distribution**? Can you give me an example of each? How do you calculate the mean and variance of a discrete distribution? How do you calculate the mean and variance of a continuous distribution?
4. What does the term **homoskedasticity** mean? Draw a population regression function $(E(x \mid y))$ and several conditional density functions at various values of x while at the same time demonstrating the phenomenon of homoskedasticity. Do the same except draw a population regression and several conditional density functions at various values of x that demonstrate the phenomenon of **heteroskedasticity**. In looking at a data scatter of observations of y and x, how would you determine whether the population regression function is subject to homoskedasticity or heteroskedasticity? Is it OK to use OLS when you have heteroskedasticity in the errors of your regression model?
5. What are the **two basic parts** of a SAS computer program? When you run a SAS program, what are the two files that are created? What is a **program file**? A **log file**? An **output (listing) file**? What kind of information do they contain?
6. What are the **basic assumptions** of the simple regression model? Review assumptions 1 – 3 in Section 4.4 of your textbook.
7. The **Ordinary Least Squares** estimators of the coefficients β_1 and β_2 in the simple linear regression model are derived by using the calculus. What function do you minimize in order to get the OLS estimators? The nice thing about the least squares criterion is that it gives rise to **analytic solutions** for the coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$.
8. Define the term **Population Regression Function (PRF)**. What is a **Sample Regression Function (SRF)** and how does it compare to the population regression (conditional mean) function? What is meant by the "**repeated sampling view**" of statistical hypothesis testing? In words, explain what is meant by "the **sampling distribution** of $\hat{\beta}_1$?" What is meant when we say "$\hat{\beta}_1$ is an **unbiased estimator** of β_1?" Think in terms of the **JAVA Applets** that we examined in class and in the

document **Using Applets in Eco 5350.pdf** that can be found under the link "**Applets for Understanding Statistical Concepts**" on the class website.

9. What is the meaning of the term "perfect collinearity?"
10. What are the **properties of the Ordinary Least Squares estimator** $\hat{\beta}_0$? What are the properties of the Ordinary Least Squares estimator $\hat{\beta}_1$? What does it mean for these estimators to be **BLUE**? In other words, what does the **Gauss-Markov theorem** state? Can you give me a drawing that represents the Gauss-Markov Theorem? What is the **Extreme Value estimator**? How do you use it to estimate the intercept β_0 and the slope β_1 of the conditional mean function (E(y|x)? Is the Extreme Value estimator unbiased? How does its sampling variance compare to the corresponding Ordinary Least Squares estimators? What does this have to say about the Gauss- Markov Theorem?
11. Hypothesis Testing. What additional assumption on the error term e allows us to conduct small sample inference in the simple linear regression model?
12. How do you display an estimated regression function when presenting it to someone else for his/her inspection?
13. How do you construct a **95% confidence interval for** β_1 using $\hat{\beta}_1$ and se($\hat{\beta}_1$)? How do you use such a confidence interval to test the null hypothesis that $H_0 : \beta_1 = 0$ versus the alternative hypothesis $H_1 : \beta_1 \neq 0$? How do you construct a t-test of the above hypothesis?
14. What is the difference between testing **a two-sided alternative** and testing **a one-sided alternative**? In a "word problem" would you be able to distinguish between when you would use a two-side test versus using a one-sided test? Can you get the p-value of an observed t-statistic for a two-sided alternative? A one-sided alternative?
14. In the **sum of squares decomposition**, TSS = ESS + SSR, what is TSS, ESS, and SSR? What are their mathematical formulas?
15. What is the **coefficient of determination** (R^2)? How do you interpret $R^2 = 0.90$?
16. What is the test for **Overall Significance** of the Regression equation? How does this relate to the **ANOVA table**?

FORMULA SHEET FOR
MID-TERM

BASIC STATISTICS:

1. Var(X) = $E(X-\mu_x)^2$

2. Cov(X,Y) = $E(X-\mu_x)(Y-\mu_y)$; $Corr(X,Y) = Cov(X,Y)/(Var(X)\cdot Var(Y))^{1/2}$

3. E(aX + bY) = aE(X) + bE(Y)

4. Var(aX + bY) = $a^2 Var(X) + b^2 Var(Y) + 2Cov(X,Y)$

5. Sample Mean: $\bar{Y} = \sum_1^N Y_i$

6. Sample Variance: $s^2 = \sum_1^N (Y_i - \bar{Y})^2 /(N-1)$

7. t-statistic for testing population mean:

$$t_{N-1} = \frac{\bar{Y} - \mu_{Y,0}}{se(\bar{Y})};\ \text{where}\ se(\bar{Y}) = s/\sqrt{N}$$

8. $(1-\alpha)\%$ confidence interval for μ

$$\Pr(\bar{Y} - t_{N-1,\alpha/2}\cdot se(\bar{Y}) < \mu < \bar{Y} + t_{N-1,\alpha/2}\cdot se(\bar{Y})) = 1-\alpha$$

9. Approximate t-statistic for testing difference in means (variances assumed unequal):

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \to Z = N(0,1)$$

10. Exact t-statistic for testing difference in means (variances assumed equal):

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \to t_{n_1+n_2-2}$$

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

11. F-Test for equal variances across two populations

$$F_{v_1, v_2} = \frac{s_1^2}{s_2^2}$$

where $v_1 = n_1 - 1$ and $v_2 = n_2 - 1$. Also, $F_{1-\alpha/2}(v_1, v_2) = \dfrac{1}{F_{\alpha/2}(v_1, v_2)}$

SOME OLS REGRESSION FORMULAS:

12. $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$; $\quad Var(\hat{\beta}_0) = \dfrac{\sigma^2 \sum_1^N X_i}{N \sum_1^N (X_i - \bar{X})^2}$

13. $\hat{\beta}_1 = \dfrac{\sum_1^N (X_i - \bar{X}) Y_i}{\sum_1^2 (X_i - \bar{X})^2} = \sum_1^N w_i Y_i$; $\quad Var(\hat{\beta}_1) = \dfrac{\sigma^2}{\sum_1^N (X_i - \bar{X})^2}$

14. $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$

15. TSS = ESS + SSR ; $\sum_1^N (Y_i - \bar{Y})^2 = \sum_1^N (\hat{Y}_i - \bar{Y})^2 + \sum_1^N (Y_i - \hat{Y}_i)^2$; $R^2 = \dfrac{ESS}{TSS}$

16. $t = \dfrac{\hat{\beta}_i - \beta_i}{se(\hat{\beta}_i)}$

17. one-tailed p-value: $\Pr(t_0 < t)$ or $\Pr(t < t_0)$

18. two-tailed p-value: $\Pr(|t_0| < t)$

19. $\Pr(\hat{\beta}_i - t_{N-K, \alpha/2} \cdot se(\hat{\beta}_i) < \beta_i < \hat{\beta} + t_{N-K, \alpha/2} \cdot se(\hat{\beta}_i)) = 1 - \alpha$

20. $F_{overall} = \dfrac{R^2 / (K-1)}{(1 - R^2)/(N-K)}$

21. $F = \dfrac{(SSR_R - SSR_U)/J}{SSR_U /(N-K)} = \dfrac{(R_U^2 - R_R^2)/J}{(1 - R_U^2)/(N-K)}$