

**Part I:** You have one blue book. That's it! Your answers to these two questions must fit within one blue book. In addition, these questions are worth 40 points, so you should spend no more than 40 minutes on these two questions.

1. We made a number of assumptions in our Classical Regression Model (CRM) that can be succinctly summarized as follows:  $u \sim iid N(0, \sigma^2 I_n)$ .

- (8) a. In 50 words or less, explain what this rather terse “statement” means. What CRM assumptions are included?
- (20) b. **What possible problems might arise** that would lead to our disturbances not being distributed in this manner? (Don't worry about zero mean part.) Craft a concise explanation using the following outline: (1) **what are the problems that might arise and which CRMA would be violated**; (2) **give an explicit example of each problem**; (3) for each problem, **identify and explain the consequences**. You are expected to provide equations to support your answers, but do not do any derivations here – they will count against you in our grading. We want to read your **explanations**.
- (12) 2. I submitted an article for consideration to a journal. One of my reviewers included the following statement: “...In addition, these time-series data are likely fraught with multicollinearity. Thus, you can not trust the estimates.” How would you respond to this chap? Do you think he was on target? Write a concise response that discusses the consequences of multicollinearity and assure him that you indeed do not have a serious problem. (You're not getting a printout for this question; just pretend you have information from your analyses to use in answering the question.)

**Part II:** Applied questions – use the SAS output provided to answer these questions and write all answers in the space provided on these pages.

I had my RA run some regressions on 2004 New England wages for individuals age 18 and over who worked for private firms. She does careful work, so I generally trust that the analyses she's provided me are on target. But, there are some additional questions that I had after she did her work and headed off to Bermuda for the weekend.

Use the first, combined male and female wage model to answer the following questions.

- (5) 1. Interpret the estimated effect of education on New England wages.
- (10) 2. Identify the estimated coefficients that are statistically significant in this first “pooled” model. Comment – how did you arrive at these conclusions?
- (8) 3. Derive the expression for the partial effect of experience on wages. Explain any “peculiarities” of this expression so that we can interpret the partial effects properly.

- (4) 4. The variable “pension” is a binary variable that indicates whether the individual is covered by a pension plan provided by their employer. Please interpret the estimated coefficient for “pension.”
- (4) 5. How well does this model fit the data? Explain.
- (5) 6. Is the regression model statistically significant? Explain.
- (3) 7. She marked the final two regressions as “male” and “female.” I’m worried that she may have marked these incorrectly. Is there any way to tell? Explain.
- (8) 8. I wanted to test whether male and female wages are different. She says it’s no problem with the results I have. Please complete this test for me.
- (i) State the appropriate null and alternative hypotheses.
  - (ii) Calculate the appropriate test statistic.
  - (iii) Complete the test. What do you conclude?
- (8) 9. Okay, now that we’ve done all that, we should decide whether we should have used the results for the pooled male/female wage model. We should test for heteroskedasticity and she said these results could be used to do that. How?
- (i) What test can be applied using these results? What form of heteroskedasticity could be diagnosed by this test? Explain briefly, then state the null and alternative hypotheses for the hypothesis test.
  - (ii) Calculate the appropriate test statistic.
  - (iii) Complete the test. What do you conclude?
- (5) 10. Given the results of your hypothesis tests, which regression results would you use in drawing inferences about New England wages? Why? Explain.