Political Science 686
Spring, 1998
Midterm Examination

1. Imagine you are trying to estimate the following model: $y = b_1 + b_2X_2 + e$. Having forgotten to pay your electric bill, your electricity is turned off, and your computer conks out just before reporting the estimates you want. Spending its last electrons ever so faithfully, the computer prints out:

"Variation in X = 432.19, Covariation of X and Y = - 2212.38"

You find your flashlight and read that printout. Can you compute b_2 from this information? Do so if you can. (At least your pocket calculator is battery powered!) Otherwise, tell me what additional information you would like to have in order to compute it. Can you compute b_1 from this information? Do so if you can. Otherwise, tell me what additional information you would like to have in order to compute it.

2. Imagine you have been given three different samples to estimate one model:

$$Y = b_1 + b_2X_2 + b_3X_3 + e$$

And here are some features of each sample:

Sample #1	Sample #2	Sample #3
N = 80	N = 100	N = 100
$\sigma^2_{x3} = 55.27$	$\sigma^2_{x3} = 107.45$	$\sigma^2_{x3} = 61.33$
$\sigma^2_e = 109.29$	$\sigma^2_e = 87.31$	$\sigma^2_e = 111.87$
$r^2_{x2x3} = 0.35$	$r^2_{x2x3} = 0.09$	$r^2_{x2x3} = 0.33$

Question: Which of these three samples would you expect to give you the *most efficient* parameter estimates for b_1, b_2, and b_3? And which sample would you expect to give you the sloppiest estimates? **Why?**

3. Given:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ . \\ . \\ . \\ y_T \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & X_{12} & X_{13} & \ldots & X_{1k} \\ 1 & X_{22} & X_{23} & \ldots & X_{2k} \\ 1 & X_{32} & X_{33} & \ldots & X_{3k} \\ & . & . & & . \\ & . & . & & . \\ & . & . & & . \\ 1 & X_{T2} & X_{T3} & \ldots & X_{Tk} \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ . \\ . \\ . \\ b_k \end{bmatrix} \quad \mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ . \\ . \\ . \\ e_T \end{bmatrix}$$

Then...

a) Show me the first three elements in the matrix **Xb** .
b) Show me in scalar form the first three lines contained within the equation **y** = **Xb** + **e** .

4) Presidential popularity was tracked on a quarterly basis for 10.75 years (43 quarters). A graduate student, eager to finish her dissertation, estimates a model which holds that presidential popularity is determined by conditions in the national economy--specifically, unemployment and overall living standard. She has obtained quarterly data on the U.S. unemployment rate (measured in percent of the labor force seeking work but unable to find it) and Disposable Income per Household (measured in dollars). Presidential popularity is measured each quarter as the percentage of all respondents in a telephone survey who agreed with the following statement: "I think the President is doing one heck of a swell job!" Her hypotheses claim that unemployment should hurt popularity, and that popularity will rise with the living standard. Adept in the use of canned statistical packages, she obtains the following results. Unfortunately, she did not attend OSU, and thus has no idea how to interpret them. You are asked to help her to the full extent of your ability. Giving her bogus answers is not allowed, even though that might sabotage her chances on the job market, and thus improve yours (it's happened!).

Variable	Coefficient	Std. Error
Constant	8.69473	1.11512
Unemploy	-1.335058	.441278
Inc_House	.002098	.000471

a) Use the t-table on the last page of this exam to tell her whether the results are consistent with her hypotheses or not. What level of significance do you select? Given her hypotheses, should she use a one-tailed test or a two-tailed test? What degrees of freedom are appropriate here?

b) Interpret the *substantive* meaning of the two partial slope coefficients. i.e., In English as plain and precise as you can make it, what do those coefficients say about the relationships being tested?

c) Given an unemployment rate of 6% and Disposable Income per Household of $25,000, what would this model predict the value of Presidential Popularity to be?

5) Here is more output from the regression performed for problem 4:

" Total Sum of Squares = 44211.09 "
" Explained Sum of Squares = 38101.78 "
" Unexplained Sum of Squares = 6109.31 "

a) Compute R^2 .
b) *What*, exactly, is being summed and squared to compute these various "Sums of Squares"?

6) Given two matrices:

$$\mathbf{A} = \begin{bmatrix} 4 & 7 \\ 3 & 2 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 8 \\ 3 \end{bmatrix}$$

a) Find, if it exists, **AB**.
b) Find, if it exists, **BA**.
c) Find, if it exists, **BB'**.
d) Find, if it exists, $\mathbf{A}^{-1}$.

7) We have seen that the general form of the OLS estimator is

$$\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'y}$$

What asssumption(s) did we make while showing that this formula yields parameter estimates that minimize the sum of the squared errors?

8) We have also seen that the OLS estimator for Var $(\mathbf{b}) = \sigma^2 (\mathbf{X'X})^{-1}$.
In deriving this estimator, it helped a great deal that E(**UU'**) reduced to $\sigma^2\mathbf{I}$ *when two assumptions were met*. What were those assumptions?

9) Time for some deep thought, friends. I offer you the following: σ_e = sqrt (σ_e^2). $\sigma_e^2 = (\Sigma e^2)/df$. $\Sigma e^2 = 476.24$. N = 107. The number of parameters estimated for this model (including the intercept) is 5.

a) Calculate the standard error of the *model estimate*.

b) Let's speak in plain english. Standard error means (more or less) "average mistake" or "typical mistake". For example, if I showed you a normally distributed variable with mean = 10 and standard deviation = 5, then, (if we played the guessing game from day 1 of class) we would see that the average mistake we made in our guesses would be 5 units from the *mean*. In plain english, then, tell me the difference between the standard error of the *estimate* and the standard error of a *model parameter* (σ_b). In other words, these two standard errors each tell us about the typical size of our mistake in relation to *what*?

10) Given a column vector of errors, or residuals:

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ . \\ . \\ . \\ e_N \end{bmatrix}$$

a) What is the dimension of **ee'**?

b) Where in **ee'** would I find the error variances?

c) Where in **ee'** would I find the error covariances?

d) What should be the *expected value* of those variances and covariances if OLS assumptions are completely satisfied?