NEW YORK UNIVERSITY
Spring 2003
Econometrics II
G31.2101

Professor: Giorgio Topa
TA: Luca Flabbi

PROBLEM SET 1
due Friday, February 28, 2003

1. Question 5.6 in Goldberger, Ch. 5.

2. Question 6.7 in Goldberger, Ch. 6.

3. For a certain data set with \( n = 100 \) observations, the explanatory variables include:
   
   (a) \( x_1 = 1 \),
   
   (b) \( x_2 \) is a binary variable that is equal to 1 for males and equal to zero for females, and
   
   (c) \( x_3 \) is a binary variable that is equal to 1 for females and equal to zero for males.

   Will the matrix \( X \) have full column rank? Explain.

   What if the regression includes only \( x_2 \) or only \( x_3 \)? Explain.

4. Let \( X \) be an \( n \times k \) matrix whose rank is \( k \), and let

   \[
   Q = X'X, \quad A = Q^{-1}X', \quad H = XA, \quad M = I - H.
   \]

   Recall that

   \[
   b = Ay, \quad \hat{y} = Hy, \quad e = My
   \]

   are the vectors of coefficients, fitted values, and residuals that result when an \( n \times 1 \) vector \( y \) is linearly regressed on \( X \). Show the following as concisely as possible:

   (a) \( AH = A, \quad AM = 0, \quad MH = 0, \quad HM = 0. \)

   (b) \(HX = X, \quad MX = 0. \)
(c) $H\hat{y} = \hat{y}$, $He = 0$.
(d) $M\hat{y} = 0$, $Me = e$.
(e) $X'\hat{y} = X'y$.
(f) $y'\hat{y} = y'Xb = b'X'y = b'Qb = \hat{y}'\hat{y}$.
(g) $e'e = y'My = y'y - \hat{y}'\hat{y}$.

Explain briefly the intuition behind each of the properties (a) – (g).

5. Suppose the Classical Regression Model applies, along with the usual notation. For each of the following statements, indicate whether it is true or false, and justify your answer. No credit will be given to an answer with no explanation.

(a) The random variable $t = b'b$ is an unbiased estimator of the parameter $\beta'\beta$.
(b) Since $\hat{y} = Hy$, it follows that $y = H^{-1}\hat{y}$.
(c) Since $E(\hat{y}) = E(y)$, it follows that the sum of the residuals is zero.
(d) If $b_1$ and $b_2$ are the first two elements of $b$, then for the random variables $t_1$ and $t_2$, defined as

$$t_1 = b_1 + b_2, \quad t_2 = b_1 - b_2,$$

we have

$$Var(t_1) \geq Var(t_2).$$

6. Show that the least squares coefficients $b = Ay$ are uncorrelated with the residuals $e = My$.

7. Show that, in the long/short regression context, the following two procedures to generate $b_1$ are equivalent:

(a) Regress $X_1$ on $X_2$, construct the residuals $X_1^*$, then regress $y$ on $X_1^*$;
(b) Regress $X_1$ on $X_2$ and construct the residuals $X_1^*$, regress $y$ on $X_2$ and construct the residuals $y^*$; then regress $y^*$ on $X_1^*$. 

2