Name Jerry

Student Number

STA 302 f2014 Quiz 6A

1. (5 points) For the usual multiple linear regression model with normal error terms, you already know that $\widehat{\epsilon} \sim N(\mathbf{0}, \sigma^2(\mathbf{I} - \mathbf{H}))$, where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Let $\mathbf{Z} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\widehat{\epsilon}$. Find the distribution of $\mathbf{Z}$. The answer is surprisingly compact, so keep simplifying! Cite facts from the formula sheet when you use them. **Circle your final answer.**

From the formula sheet, if $Y \sim N(\mu, \Sigma)$ then $AY \sim N(A\mu, A\Sigma A')$. This means Z is multivariate normal.

$E(Z) = E\left((X'X)^{-1}X'\hat{\varepsilon}\right) = (X'X)^{-1}X'E(\hat{\varepsilon}) = 0$

$Cov(Z) \overset{\text{Formula sheet}}{=} (X'X)^{-1}X'(I-H)\left((X'X)^{-1}X'\right)'$

$= (X'X)^{-1}X'(I-H)X(X'X)^{-1}$

$= (X'X)^{-1}X'IX(X'X)^{-1} - (X'X)^{-1}X'HX(X'X)^{-1}$

$= \underbrace{(X'X)^{-1}X'X(X'X)^{-1}}_{I} - \underbrace{(X'X)^{-1}X'X(X'X)^{-1}X'X(X'X)^{-1}}_{I \quad\quad I \quad\quad \Sigma}$

$= (X'X)^{-1} - (X'X)^{-1} = \underset{\sim}{0}$, so

$Z \sim N(0, 0)$ Degenerate distribution

They need not say this.

2. (5 points) In your analysis of the Census Tract data, you fit a model in which the dependent variable was crime rate, and the independent variables were `area`, `urban`, `old`, `docs`, `beds`, `hs`, `labor` and `income`.

   Controlling for all the other variables in the model, is Percent of population 25 or older completing 12+ years of school (`hs`) related to crime rate?

   (a) Give the null hypothesis in symbols.

   $H_0 : \beta_6 = 0$

   (b) Write the value of the test statistic in the space below. The answer is a number from your printout. On your printout, circle the test statistic and write "Question 2b" beside it.

   $t = -2.519$

   (c) Write the p-value in the space below. The answer is a number from your printout. On your printout, circle the p-value and write "Question 2c" beside it.

   $P = 0.013$

   (d) Do you reject the null hypothesis at $\alpha = 0.05$? Answer Yes or No.

   Yes

   (e) Allowing for other variables, census regions with higher percentage of High School graduates tend to have lower (higher, lower) crime rates. Be guided by the $\alpha = 0.05$ significance level.

Please attach your printout to the quiz paper. Make sure your name is on the printout.

Name Jerry

Student Number

STA 302 f2014 Quiz 6B

1. (5 points) Suppose data for a regression study are collected at two different locations; n_1 observations are collected at location one, and n_2 observations are collected at location two. The same independent variables are used at each location. We need to know whether the error variance σ^2 is the same at the two locations.

   Recall the definition of the F distribution. If $W_1 \sim \chi^2(\nu_1)$ and $W_2 \sim \chi^2(\nu_2)$ are independent, then $F = \frac{W_1/\nu_1}{W_2/\nu_2} \sim F(\nu_1, \nu_2)$. Suggest a statistic for testing $H_0 : \sigma_1^2 = \sigma_2^2$. Using facts from the formula sheet, show it has an F distribution. Don't forget to state the degrees of freedom. Assume that data coming from the two locations are independent. Circle the test statistic

   From the formula sheet, $W = \frac{SSE}{\sigma^2} \sim \chi^2(n-k-1)$.

   Data from the two locations, so

   $W_1 = \frac{SSE_1}{\sigma_1^2}$ and $W_2 = \frac{SSE_2}{\sigma_2^2}$ are independent

   Then $$F = \frac{\frac{SSE_1}{\sigma_1^2}/(n_1-k-1)}{\frac{SSE_2}{\sigma_2^2}/(n_2-k-1)} \sim F(n_1-k-1, n_2-k-1)$$

   If $H_0: \sigma_1^2 = \sigma_2^2$ is true they cancel, and

   $$F = \frac{MSE_1}{MSE_2} \sim F(n_1-k-1, n_2-k-1)$$

   This is the test statistic $\frac{SSE_1/(n_1-k-1)}{SSE_2/(n_2-k-1)}$ is okay too.

2. (5 points) In your analysis of the Census Tract data, you fit a model in which the dependent variable was crime rate, and the independent variables were `area`, `urban`, `old`, `docs`, `beds`, `hs`, `labor` and `income`.

   Controlling for all the other variables in the model, is Percent of population in cities (`urban`) related to crime rate?

   (a) Give the null hypothesis in symbols.

   $H_0: \beta_2 = 0$

   (b) Write the value of the test statistic in the space below. The answer is a number from your printout. On your printout, circle the test statistic and write "Question 2b" beside it.

   $t = 2.334$

   (c) Write the p-value in the space below. The answer is a number from your printout. On your printout, circle the p-value and write "Question 2c" beside it.

   $p = 0.021$

   (d) Do you reject the null hypothesis at $\alpha = 0.05$? Answer Yes or No.

   Yes

   (e) Allowing for other variables, census regions with higher percentage of population in cities tend to have higher (higher, lower) crime rates. ~~Be guided by the $\alpha = 0.05$ significance level.~~

Please attach your printout to the quiz paper. Make sure your name is on the printout.