Family (Last) Name ______________________

Given (First) Name Jerry

Student Number ______________________

STA 302s13 Quiz 8A

Formula sheet is on the reverse side.

1. (6 points) Suppose you need to test the null hypothesis that a *single* linear combination of regression coefficients is equal to zero. That is, you want to test $H_0 : \mathbf{a}'\boldsymbol{\beta} = 0$. Referring to the formula sheet, verify that $F = T^2$. Show your work.

In the F test, $C = a'$, $t = 0$ and $q = 1$. Then,

$$F = \frac{(a'\hat{\beta})'(a'(X'X)^{-1}a)^{-1}a'\hat{\beta}}{1 \cdot s^2} = \frac{(a'\hat{\beta})^2}{s^2 a'(X'X)^{-1}a}$$

$$= \left(\frac{a'\hat{\beta} - 0}{s\sqrt{a'(X'X)^{-1}a}} \right)^2 = T^2$$

2. (4 points) Using the Census tract data, you fit a model in which dependent variable was crime rate, and the independent variables were `area`, `urban`, `old`, `docs`, `beds`, `hs`, `labor` and `income`. You did a test of `old`, `labor` and `income` controlling for the other independent variables in the model. Attach your R printout to the quiz paper. **Circle the F statistic and the p-value for the test described above.** These are *two numbers* on your printout. *The correct numbers must be circled, and no other numbers must be circled. Otherwise, you get zero marks for this question.*

Family (Last) Name ____________________

Given (First) Name Jerry

Student Number ____________________

STA 302s13 Quiz 8B

Formula sheet is on the reverse side.

1. (6 points) Starting from the formula sheet, show that the F test for comparing full and reduced models may be written
$$F = \left(\frac{a}{1-a}\right)\left(\frac{n-k-1}{q}\right),$$
where $a = \frac{R^2-R_r^2}{1-R_r^2}$ Show your work.

$$\left(\frac{a}{1-a}\right)\left(\frac{n-k-1}{q}\right) \quad \left(\frac{\frac{R^2-R_r^2}{1-R_r^2}}{1-\frac{R^2-R_r^2}{1-R_r^2}}\right)\left(\frac{n-k-1}{q}\right)$$

$$= \left(\frac{\frac{R^2-R_r^2}{\cancel{1-R_r^2}}}{\frac{1-R_r^2-R^2+R_r^2}{\cancel{1-R_r^2}}}\right)\left(\frac{n-k-1}{q}\right) = \frac{\frac{SSR - SSR(\text{reduced})}{SST}}{\frac{SST}{SST} - \frac{SSR}{SST}}\left(\frac{n-k-1}{q}\right)$$

$$= \frac{SSR - SSR(\text{reduced})}{SSE}\cdot\left(\frac{n-k-1}{q}\right) = \frac{SSR - SSR(\text{reduced})}{q \; SSE/(n-k-1)}$$

$$= \frac{SSR - SSR(\text{reduced})}{q\, s^2} = F \quad \text{on formula sheet.}$$

2. (4 points) Using the Census tract data, you fit a model in which dependent variable was crime rate, and the independent variables were `area`, `urban`, `old`, `docs`, `beds`, `hs`, `labor` and `income`. You did a test of `old`, `labor` and `income` controlling for the other independent variables in the model. Attach

```
> census =
read.table("http://www.utstat.toronto.edu/~brunner/302f13/code_n_data/hw/Ce
nsusTract.data")
> attach(census)
> crimerate = crimes/pop
> fullmod = lm(crimerate ~ area + urban + old + docs + beds + hs + labor +
income)
> summary(fullmod)

Call:
lm(formula = crimerate ~ area + urban + old + docs + beds + hs +
    labor + income)

Residuals:
     Min       1Q   Median       3Q      Max
-28.1128  -8.3957  -0.4209   7.1998  31.1864

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 21.0000936 10.2108838   2.057 0.041691 *
area         0.0014182  0.0003977   3.566 0.000506 ***
urban        0.1489428  0.0638183   2.334 0.021114 *
old          0.0858062  0.4465427   0.192 0.847915
docs         0.0042640  0.0019497   2.187 0.030502 *
beds        -0.0015261  0.0006059  -2.519 0.012972 *
hs           0.4475895  0.1415152   3.163 0.001939 **
labor        0.0019947  0.0238075   0.084 0.933354
income       0.0001003  0.0016995   0.059 0.953037
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 12.24 on 132 degrees of freedom
Multiple R-squared: 0.3214,   Adjusted R-squared: 0.2803
F-statistic: 7.815 on 8 and 132 DF,  p-value: 1.472e-08

>
> redmod = lm(crimerate ~ area + urban  + docs + beds + hs )
> anova(redmod,fullmod)
Analysis of Variance Table

Model 1: crimerate ~ area + urban + docs + beds + hs
Model 2: crimerate ~ area + urban + old + docs + beds + hs + labor + income
  Res.Df   RSS Df Sum of Sq      F Pr(>F)
1    135 19817
2    132 19792  3    25.683 0.0571  0.982
>
```

STA 302 Formulas

Columns of $\mathbf{A}$ *linearly dependent* means there is a vector $\mathbf{v} \neq \mathbf{0}$ with $\mathbf{Av} = \mathbf{0}$.

Columns of $\mathbf{A}$ *linearly independent* means that $\mathbf{Av} = \mathbf{0}$ implies $\mathbf{v} = \mathbf{0}$.

$\boldsymbol{\Sigma} = \mathbf{CDC}'$ $\quad\quad$ $\boldsymbol{\Sigma}^{-1} = \mathbf{CD}^{-1}\mathbf{C}'$

$\boldsymbol{\Sigma}^{1/2} = \mathbf{CD}^{1/2}\mathbf{C}'$ $\quad\quad$ $\boldsymbol{\Sigma}^{-1/2} = \mathbf{CD}^{-1/2}\mathbf{C}'$

$M_Y(t) = E(e^{Yt})$ $\quad\quad$ $M_{aY}(t) = M_Y(at)$

$M_{Y+a}(t) = e^{at}M_Y(t)$ $\quad\quad$ $M_{\sum_{i=1}^n Y_i}(t) = \prod_{i=1}^n M_{Y_i}(t)$

$Y \sim N(\mu, \sigma^2)$ means $M_Y(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$ $\quad\quad$ $Y \sim \chi^2(\nu)$ means $M_Y(t) = (1-2t)^{-\nu/2}$

If $W = W_1 + W_2$ with W_1 and W_2 independent, $W \sim \chi^2(\nu_1 + \nu_2)$, $W_2 \sim \chi^2(\nu_2)$ then $W_1 \sim \chi^2(\nu_1)$

$cov(\mathbf{Y}) = E\left\{(\mathbf{Y} - \boldsymbol{\mu}_y)(\mathbf{Y} - \boldsymbol{\mu}_y)'\right\}$ $\quad\quad$ $C(\mathbf{Y}, \mathbf{T}) = E\left\{(\mathbf{Y} - \boldsymbol{\mu}_y)(\mathbf{T} - \boldsymbol{\mu}_t)'\right\}$

$cov(\mathbf{Y}) = E\{\mathbf{YY}'\} - \boldsymbol{\mu}_y\boldsymbol{\mu}_y'$ $\quad\quad$ $cov(\mathbf{AY}) = \mathbf{A}cov(\mathbf{Y})\mathbf{A}'$

$M_{\mathbf{Y}}(\mathbf{t}) = E(e^{\mathbf{t}'\mathbf{Y}})$ $\quad\quad$ $M_{\mathbf{AY}}(\mathbf{t}) = M_{\mathbf{Y}}(\mathbf{A}'\mathbf{t})$

$M_{\mathbf{Y}+\mathbf{c}}(\mathbf{t}) = e^{\mathbf{t}'\mathbf{c}}M_{\mathbf{Y}}(\mathbf{t})$ $\quad\quad$ $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ means $M_{\mathbf{Y}}(\mathbf{t}) = e^{\mathbf{t}'\boldsymbol{\mu} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}}$

$\mathbf{Y}_1$ and $\mathbf{Y}_2$ are independent if and only if $M_{(\mathbf{Y}_1,\mathbf{Y}_2)'}((\mathbf{t}_1, \mathbf{t}_2)') = M_{\mathbf{Y}_1}(\mathbf{t}_1)M_{\mathbf{Y}_2}(\mathbf{t}_2)$

If $\mathbf{Y} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\mathbf{AY} \sim N_q(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}')$, $\quad\quad$ and $W = (\mathbf{Y} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu}) \sim \chi^2(p)$

$r_{xy} = \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^n (X_i - \overline{X})^2}\sqrt{\sum_{i=1}^n (Y_i - \overline{Y})^2}}$

$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \epsilon_i$ $\quad\quad$ $\epsilon_1, \ldots, \epsilon_n$ independent $N(0, \sigma^2)$

$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ $\quad\quad$ $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2\mathbf{I}_n)$

$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ $\quad\quad$ $\widehat{\mathbf{Y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$

$\sum_{i=1}^n (Y_i - \overline{Y})^2 = \sum_{i=1}^n (Y_i - \widehat{Y}_i)^2 + \sum_{i=1}^n (\widehat{Y}_i - \overline{Y})^2$ $\quad\quad$ $SST = SSE + SSR$ and $R^2 = \frac{SSR}{SST}$

$\widehat{\boldsymbol{\epsilon}} = \mathbf{Y} - \widehat{\mathbf{Y}}$ $\quad\quad$ $\widehat{\boldsymbol{\beta}} \sim N_{k+1}\left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)$

$\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\epsilon}}$ are independent under normality. $\quad\quad$ $SSE/\sigma^2 = \hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}/\sigma^2 \sim \chi^2(n-k-1)$

$T = \frac{Z}{\sqrt{W/\nu}} \sim t(\nu)$ $\quad\quad$ $F = \frac{W_1/\nu_1}{W_2/\nu_2} \sim F(\nu_1, \nu_2)$

$T = \frac{\mathbf{a}'\widehat{\boldsymbol{\beta}} - \mathbf{a}'\boldsymbol{\beta}}{s\sqrt{\mathbf{a}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{a}}} \sim t(n-k-1)$ $\quad\quad$ $T = \frac{Y_0 - \mathbf{x}_0'\widehat{\boldsymbol{\beta}}}{s\sqrt{(1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0)}} \sim t(n-k-1)$

$F = \frac{(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})'(\mathbf{C}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}')^{-1}(\mathbf{C}\widehat{\boldsymbol{\beta}} - \mathbf{t})}{q\,s^2} = \frac{SSR - SSR(reduced)}{q\,s^2} \sim F(q, n-k-1)$, where $s^2 = MSE = \frac{SSE}{n-k-1}$