Name Jerry

Student Number ______

STA 442/2101 f2013 Quiz 7

1. (4 points) Consider the prediction interval for Y_{n+1}.

   (a) What is the distribution of $Y_{n+1} - \widehat{Y}_{n+1} = Y_{n+1} - \mathbf{x}'_{n+1}\widehat{\beta}$? Your answer includes both the expected value and the variance.

$$E(Y_{n+1} - x'_{n+1}\hat{\beta}) = x'_{n+1}\beta - x'_{n+1}\beta = 0$$

$$V(Y_{n+1} - x'_{n+1}\hat{\beta}) = \sigma^2 + x'_{n+1}\sigma^2(X'X)^{-1}x_{n+1}, \text{ so}$$

$$Y_{n+1} - \hat{Y}_{n+1} \sim N(0, \sigma^2(1 + x'_{n+1}(X'X)^{-1}x_{n+1}))$$

   (b) Now standardize the difference to obtain a standard normal.

$$Z = \frac{Y_{n+1} - \hat{Y}_{n+1}}{\sqrt{\sigma^2(1 + x'_{n+1}(X'X)^{-1}x_{n+1})}}$$

   (c) Divide by the square root of a chi-squared random variable, divided by its degrees of freedom, and simplify. Call it T.

$$T = \frac{Z}{\sqrt{\frac{SSE}{\sigma^2}/(n-p)}} = \frac{Y_{n+1} - \hat{Y}_{n+1}}{\sqrt{\sigma^2(1 + x'_{n+1}(X'X)^{-1}x_{n+1})}\sqrt{\frac{1}{\sigma^2}\frac{SSE}{n-p}}}$$

$$= \frac{Y_{n+1} - \hat{Y}_{n+1}}{\sqrt{MSE(1 + x'_{n+1}(X'X)^{-1}x_{n+1})}}$$

   (d) How do you know that numerator and denominator are independent?

Y_{n+1} is independent of SSE and $\hat{\beta}$ is independent of SSE. Numerator is a function of Y_{n+1} & $\hat{\beta}$, while the denominator is a function of SSE.

2. Based on your regression analysis of the SAT data, please answer the questions below by writing numbers from your printout. In case of doubt, base your answers on the full (unrestricted) model. **Please do NOT use R's scientific notation for the p-values.** So, you would write $p = 0.000712$ rather than $p =$`7.12e-04`.

(a) (1 Point) Controlling for Verbal score, is Math score related to first-year grade point average?

Test Statistic (F or t)	Degrees of Freedom (2 numbers if F)	p-value
t=1.641	197	0.102

(b) (1 Point) Allowing for Math score, is Verbal score related to first-year grade point average?

Test Statistic (F or t)	Degrees of Freedom (2 numbers if F)	p-value
t = 4.178	197	0.0000442

(c) (2 Points) Test $H_0 : \beta_1 = \beta_2$.

Test Statistic (F or t)	Degrees of Freedom (2 numbers if F)	p-value
F = 1.9848 or t = 1.408838	1, 197 or 197	0.1605

(d) (1 Point) Test $H_0 : \beta_1 = \beta_2 = 0$.

Test Statistic (F or t)	Degrees of Freedom (2 numbers if F)	p-value
F = 12.93	2, 197	0.0000052 84

(e) (1 Point) Test $H_0 : \beta_0 = 0$.

Test Statistic (F or t)	Degrees of Freedom (2 numbers if F)	p-value
t=1.374	197	0.171

Do **not** attach any printout this time.

```
R version 2.15.1 (2012-06-22) -- "Roasted Marshmallows"
Copyright (C) 2012 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
Platform: i386-apple-darwin9.8.0/i386 (32-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.52 (6188) i386-apple-darwin9.8.0]

[Workspace restored from /Users/brunner/.RData]
[History restored from /Users/brunner/.Rapp.history]

> # Last Question (SAT)  Data have been fixed: No more 100 bottles of beer.
> rm(list=ls())
> dat = read.table("http://www.utstat.toronto.edu/~brunner/appliedf13/code_n_data/hw/sat.data")
> attach(dat)
>
> # a.
> modelA = lm(GPA ~ MATH); summary(modelA)

Call:
lm(formula = GPA ~ MATH)

Residuals:
     Min       1Q   Median       3Q      Max
-2.26114 -0.35543  0.01944  0.36817  1.15075

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.5264336  0.3981176   3.834 0.000169 ***
MATH        0.0016990  0.0006098   2.786 0.005850 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5707 on 198 degrees of freedom
Multiple R-squared: 0.03773,    Adjusted R-squared: 0.03287
F-statistic: 7.764 on 1 and 198 DF,  p-value: 0.00585

> # Prediction interval
> math700 = data.frame(MATH=700)
> predict(modelA,newdata=math700,interval="prediction")
      fit      lwr      upr
1 2.71575 1.585852 3.845648
> summary(modelA)$r.squared #$ Ending syntax colouring
[1] 0.03773038
>
> #b
> modelB = lm(GPA ~ MATH+VERBAL); summary(modelB)
```

```
Call:
lm(formula = GPA ~ MATH + VERBAL)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.24875 -0.35113  0.04659  0.38745  1.03527 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.6062975  0.4414062   1.374    0.171    
MATH        0.0009999  0.0006093   1.641    0.102    
VERBAL      0.0023072  0.0005522   4.178 4.42e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 0.5484 on 197 degrees of freedom
Multiple R-squared: 0.1161,	Adjusted R-squared: 0.1071 
F-statistic: 12.93 on 2 and 197 DF,  p-value: 5.284e-06 

> 
> verbal650math700 = data.frame(VERBAL=650,MATH=700)
> predict(modelB,newdata=verbal650math700,interval="prediction")
       fit      lwr      upr
1 2.805857 1.719307 3.892408
> #           fit      lwr      upr
> # [1,] 2.805857 1.719307 3.892408
> 
> # a = r*F/(n-p + r*F)
> # Can get these quantities by visual inspection without knowing the names.
> NminusP = summary(modelB)$fstatistic[3] #$ Ending syntax colouring
> r=1 # For t-tests
> F_math =  summary(modelB)$coefficients[2,3]^2 #$ Ending syntax colouring
> a_math = r*F_math/(NminusP + r*F_math)
> names(a_math) = NULL; a_math # 0.01348301
[1] 0.01348301
> 
> F_verbal =  summary(modelB)$coefficients[3,3]^2 #$ Ending syntax colouring
> a_verbal = r*F_verbal/(NminusP + r*F_verbal)
> names(a_verbal) = NULL; a_verbal #  0.08139433
[1] 0.08139433
> 
> # Now check a_math=0.01348301 and a_verbal=0.08139433
> 
> verbalmath = anova(lm(GPA~VERBAL+MATH)); verbalmath
Analysis of Variance Table

Response: GPA
           Df Sum Sq Mean Sq F value    Pr(>F)    
VERBAL      1  6.968  6.9682 23.1718 2.945e-06 ***
MATH        1  0.810  0.8097  2.6925    0.1024    
Residuals 197 59.242  0.3007                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 
> mathverbal = anova(modelB); mathverbal
Analysis of Variance Table

Response: GPA
           Df Sum Sq Mean Sq F value    Pr(>F)    
MATH        1  2.529  2.5287  8.4088  0.004158 ** 
VERBAL      1  5.249  5.2492 17.4555 4.418e-05 ***
Residuals 197 59.242  0.3007                      
---
```

```
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> SSTO = sum(verbalmath[,2]); SSTO
[1] 67.02
> var(GPA) * 199 # Just checking
[1] 67.02
>
> a_math; verbalmath[2,2]/(SSTO-verbalmath[1,2])
[1] 0.01348301
[1] 0.01348301
> a_verbal; mathverbal[2,2]/(SSTO-mathverbal[1,2])
[1] 0.08139433
[1] 0.08139433
>
> # Testing equal slopes
> sat = VERBAL+MATH # Total score
> modelC = lm(GPA ~ sat); summary(modelC) # Restricted model under H0: beta1=beta2

Call:
lm(formula = GPA ~ sat)

Residuals:
    Min      1Q  Median      3Q     Max
-2.2309 -0.3609  0.0644  0.3696  1.0874

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.5083622  0.4369807   1.163    0.246
sat         0.0017039  0.0003495   4.875 2.23e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5497 on 198 degrees of freedom
Multiple R-squared: 0.1071, Adjusted R-squared: 0.1026
F-statistic: 23.76 on 1 and 198 DF,  p-value: 2.232e-06

> anova(modelC,modelB)
Analysis of Variance Table

Model 1: GPA ~ sat
Model 2: GPA ~ MATH + VERBAL
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    198 59.839
2    197 59.242  1   0.59688 1.9848 0.1605
> # Can't conclude that expected GPA increases at different rates as a function of Verbal SAT and
Math SAT.
>
>
```

2C