Name Jerry

Student Number

STA 302 f2014 Quiz 8A

For the general linear model with an intercept (so that $\sum_{i=1}^n Y_i = \sum_{i=1}^n \widehat{Y}_i$), is it true that the sample correlation between the Y_i and $\widehat{\epsilon}_i$ values is equal to zero? Answer Yes or No and prove your answer.

No. Numerator of the sample correlation is $\sum_{i=1}^n (Y_i - \overline{Y})(\hat{\epsilon}_i - 0) = \sum_{i=1}^n Y_i \hat{\epsilon}_i - \overline{Y}\sum_{i=1}^n \hat{\epsilon}_i$

$$= \sum_{i=1}^n Y_i \hat{\epsilon}_i - 0 = Y'\hat{\epsilon} = Y'(Y - X\hat{\beta})$$

$$= Y'Y - Y'X(X'X)^{-1}X'Y$$

$$= Y'IY - Y'HY = Y'(I-H)Y \neq 0$$

in general

Name ______________________________

Student Number ______________________________

STA 302 f2014 Quiz 8B

For the general linear model withan intercept (so that $\sum_{i=1}^n Y_i = \sum_{i=1}^n \widehat{Y}_i$), is it true that the squared sample correlation between the $\widehat{Y}_i$ and $\widehat{\epsilon}_i$ values is equal to R^2? Answer Yes or No and prove your answer.

No. Numerator of the sample correlation is

$$\sum_{i=1}^n (\hat{Y}_i - \bar{Y})(\hat{\varepsilon}_i - 0) = \sum_{i=1}^n \hat{Y}_i \hat{\varepsilon}_i - \bar{Y}\sum_{i=1}^n \hat{\varepsilon}_i$$

$$= \sum_{i=1}^n \hat{Y}_i \hat{\varepsilon}_i - 0 = \hat{Y}'(Y - \hat{Y})$$

$$= (X\hat{\beta})'(Y - X\hat{\beta}) = \hat{\beta}'X'(Y - X\hat{\beta})$$

$$= \hat{\beta}'X'Y - \hat{\beta}'\underbrace{X'X(X'X)^{-1}}_{I}X'Y = 0 \neq R^2$$

in general

Name ______________________________

Student Number ______________________________

STA 302 f2014 Quiz 8C

For the general linear regression model,

1. (2 points) What is $E(\mathbf{YY}')$? Show a little work.

$cov(Y) = E(YY') - \mu_y \mu_y'$

$\Rightarrow E(YY') = \sigma^2 I_n + X\beta\beta'X'$

2. (8 points) Are $\widehat{\mathbf{Y}}$ and $\mathbf{Y}$ independent? Answer Yes or No and show your work.

$C(\hat{Y}, Y) = E(\hat{Y}Y') - E(\hat{Y})E(Y)'$

$= E(X\hat{\beta}Y') - E(X\hat{\beta})(EX\hat{\beta})'$

$= E(X(X'X)^{-1}X'YY') - X\beta\beta'X'$

$= X(X'X)^{-1}X'E(YY') - X\beta\beta'X'$

$= X(X'X)^{-1}X'(\sigma^2 I_n + X\beta\beta'X') - X\beta\beta'X'$

$= \sigma^2 X(X'X)^{-1}X' + X\underbrace{(X'X)^{-1}X'X}_{I}\beta\beta'X' - X\beta\beta'X'$

$= \sigma^2 H + 0 \neq 0$.

Noticing H is not necessary

NO

Name Jerry

Student Number

STA 302 f2014 Quiz 8D

1. (5 points) For the general linear regression model, are the random variables $\widehat{Y}_1, \ldots, \widehat{Y}_n$ independent? Answer Yes or No and prove your answer.

No. $cov(\hat{Y}) = cov(X\hat{\beta})$

Formula Sheet

$\overset{\downarrow}{=} X \sigma^2 (X'X)^{-1} X' = \sigma^2 X (X'X)^{-1} X'$

$= \sigma^2 H \neq 0$

okay

2. (5 points) Before the beginning of the term, students in a first-year Calculus class took a diagnostic test with two parts: Pre-calculus and Calculus. Their High School Calculus marks were also available. The dependent variable was mark in University Calculus. Here is some R output.

```
> fullmodel = lm(UnivCalculus ~ HSCalculus + PrecalcTest + CalcTest)
> summary(fullmodel)

Call:
lm(formula = UnivCalculus ~ HSCalculus + PrecalcTest + CalcTest)

Residuals:
    Min      1Q  Median      3Q     Max
-48.699  -7.954   1.603   9.242  30.260

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.32155    6.01019  -1.052  0.29376
HSCalculus   0.70097    0.08133   8.619  4.4e-16 ***
PrecalcTest  1.87208    0.57572   3.252  0.00128 **
CalcTest     0.79289    0.38927   2.037  0.04257 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 14.16 on 291 degrees of freedom
Multiple R-squared: 0.3583,Adjusted R-squared: 0.3517
F-statistic: 54.17 on 3 and 291 DF,  p-value: < 2.2e-16

> redmodel = lm(UnivCalculus ~ HSCalculus)
> anova(redmodel,fullmodel)
Analysis of Variance Table

Model 1: UnivCalculus ~ HSCalculus
Model 2: UnivCalculus ~ HSCalculus + PrecalcTest + CalcTest
  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1    293 62967
2    291 58375  2    4591.5 11.444 1.643e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Controlling for mark on the Precalculus test and mark on the Calculus test, what proportion of the remaining variation in University Calculus mark is explained by mark in High School Calculus? The answer is a number. Show your work. **circle your answer.**

$$a = \frac{sF}{n-k-1+sF} = \frac{1 \cdot T^2}{291 + 1 \cdot T^2} = \frac{8.619^2}{291 + 8.619^2}$$

$$= 0.20$$