Name Jerry

Student Number _____

# STA 312f2012 Quiz 6

1. (3 Points) In a study of energy efficiency, data are collected on $n$ houses. The houses have three different chimney types: Rectangular, Round and Square. Consider a regression model in which the only explanatory variable is Chimney Type.

   (a) Write $E[Y|\mathbf{X}]$ for the full model. You do *not* need to say how the dummy variables are defined. You will do that in the table below.

   $E[Y|\mathbf{X}] = \beta_0 + \beta_1 C_1 + \beta_2 C_2$

   (b) In the table below, make columns showing how the indicator dummy variables for Chimney Type are defined. *Make Rectangular the reference category.* At the top of each column, write the name of a dummy variable from your answer to Question 1a. Finally, put $E[Y|\mathbf{X}]$ for each Chimney Type in the last column.

| | $C_1$ | $C_2$ | $E[Y\|\mathbf{X}]$ |
|---|---|---|---|
| Rectangular Chimney | 0 | 0 | $\beta_0$ |
| Round Chimney | 1 | 0 | $\beta_0 + \beta_1$ |
| Square Chimney | 0 | 1 | $\beta_0 + \beta_2$ |

   (c) What null hypothesis would you test to find out whether average energy consumption depends on chimney type? Your answer is a statement involving the $\beta$ symbols from your model.

   $H_0 : \beta_1 = \beta_2 = 0$

2. (1 Point) For the SAT data, consider the model with both the Verbal and Math subtests. Letting $X_1$ denote score on the Verbal sub-test and $X_2$ denote score on the Math sub-test, give the null hypothesis that would be tested in order to answer this question: Controlling for Math score, is Verbal score related to first-year Grade Point Average? Your answer is in symbols, not numbers. You have a lot more room than you need.

   $H_0 : \beta_1 = 0$

3. This question is based on your printout from the SAT data — the model with two explanatory variables, Math test score and Verbal test score. ***Write your answers in the spaces below, and ~~also circle them on your printout. On the printout, label the answers 3a, 3b etc.~~***

(a) (1 point) Give a predicted first-year GPA for a student who got 650 on the Verbal and 700 on the Math SAT. Your answer is a single number. Show a little work and *circle the number.*

$0.6063 + (.001)700 + (.0023)650$

$= 2.8$

(b) (1 point) What proportion of the variation in first-year Grade point Average is explained by Math test score and Verbal test score? The answer is a single number from your printout.

together

$R^2 = 0.1161$

(c) (1 point) Give the test statistic for testing $H_0 : \beta_1 = \beta_2 = 0$. Your answer is a single number, a value of $t$ or $F$, from your printout.

$F = 12.93$

(d) (1 Point) Once you control for score on the Verbal sub-test, is score on the Math sub-test related to Grade Point Average? Give the value of the test statistic. Your answer is a single number, a value of $t$ or $F$, from your printout.

$t = 1.641$

(e) (2 Points) In plain, non-statistical language, what do you conclude from the hypothesis test of Question 3d? Write your answer below.

Once you allow for score on the verbal test, there is no clear evidence that score on the math test is related to first-year GPA.

**Attach your printout for Question 3 (Homework Question 4). Make sure your name is written on the printout.**

# Printout for Quiz Question 3, Assignment 6 Q 4

```
> # 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/appliedf12/data/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
> math550 = data.frame(MATH=550)
> predict(modelA,newdata=math550)
       1
2.460896

> 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

> verbal650math550 = data.frame(VERBAL=650,MATH=550)
> predict(modelB,newdata=verbal650math550)
2.655879
```