Name _____

Student Number _____

University of Toronto Mississauga
STA442/1008 2012
QUIZ 5 (9 a.m.)

$$F = \left(\frac{n-p}{s}\right)\left(\frac{a}{1-a}\right) \qquad a = \frac{sF}{n-p+sF}$$

1. (2 Points) In the Furnace data, houses have three different chimney types: Rectangular, Round and Square. Consider a regression model in which the only independent 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.

   $$E[Y|\mathbf{X}] = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

   (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.

	x_1	x_2	$E[Y\|\mathbf{X}]$
Rectangular Chimney	0	0	β_0
Round Chimney	1	0	$\beta_0 + \beta_1$
Square Chimney	0	1	$\beta_0 + \beta_2$

2. (2 Points) For the SAT data, consider the model with both the Verbal and Math sub-tests. Letting X_1 denote score on the Verbal sub-test and X_2 denote score on the Math sub-test, give $E[Y|\mathbf{X}]$ for the *reduced* model that would be used 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.

   $$E[Y|\mathbf{X}] = \beta_0 + \beta_2 X_2$$

3. Now refer to your computer output for the SAT data. The full model includes *both* independent variables. **Except for part 3e, you will write your answers on the printout, and NOT on the quiz paper.**

   (a) (1 Point) What proportion of the variation in first-year Grade Point Average is explained by Math score and Verbal score together? The answer is a single number. *Circle the number on your printout, and write the question number (3a) beside it.* If you have to calculate the number, show a little work. ($R^2 = 0.1161$; a percentage is okay too.)

   (b) (1 Point) Give a predicted first-year grade point average for a student who got 600 on the Verbal SAT and 750 on the Math SAT. The answer is a single number between 0.00 and 4.00. Show the calculation and *write the answer on your printout* beside the table of estimated regression coefficients. *Circle the number, and write the question number (3b) beside it.* ($0.60630 + 0.00231 * 600 + 0.00099985 * 750 = 2.742188$)

   (c) (1 Point) Once you control for score on the Verbal sub-test, what proportion of the *remaining* variation in Grade Point Average is explained by score on the Math sub-test? The answer is a single number between 0.00 and 1.00. *Circle the number on your printout, and write the question number (3c) beside it.* If you have to calculate the number, show a little work. ($F = 1.64^2; a = F/(197 + F) = 0.0135$; `percentage is okay too.`)

   (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? On your printout, *circle the numerical value of the test statistic, and write the question number (3d) beside it.* $t = 1.64; F = 2.69$ (`or so is okay too, because some people will do a custom contrast even theough they don't need to.`)

   (e) (2 Points) In plain, non-statistical language, what do you conclude from the significance test of Question 3d? Write your answer below.
   (`Once we allow for score on the Verbal sub-test, there is not enough evidence to conclude that score on the Math sub-test is related to Grade Point Average.`)

Please turn in your log file and your list file with the quiz. Make sure your name and student number are written on them.