

Name \_\_\_\_\_

Student Number \_\_\_\_\_

**University of Toronto Mississauga**

**STA442/1008 2012**

**QUIZ 9 (9am)**

1. (4 Points) Suppose you have a data set with a quantitative dependent variable  $Y$ , a quantitative covariate  $X$ , and two factors named  $A$  and  $B$ . Factor  $A$  has 3 levels (categories) and factor  $B$  also has 3 levels. You see this SAS code:

```
proc glm;  
class A B;  
model Y = X A|B;
```

- (a) Write  $E[Y|\mathbf{X}]$  for a regression model equivalent to the `proc glm` model above. Use names like  $a_1$ ,  $a_2$  etc. for the dummy variables, and assume effect coding. You do *not* have to say how the dummy variables are defined. Don't forget the covariate.

$$(E[Y|\mathbf{X}] = \beta_0 + \beta_1 X + \beta_2 a_1 + \beta_3 a_2 + \beta_4 b_1 + \beta_5 b_2 + \beta_6 a_1 b_1 + \beta_7 a_1 b_2 + \beta_8 a_2 b_1 + \beta_9 a_2 b_2)$$

- (b) In terms of the  $\beta$  quantities from your regression model, state the null hypothesis for testing the main effects of factor  $A$  (controlling for the covariate).

$$(H_0 : \beta_2 = \beta_3 = 0)$$

- (c) In terms of the  $\beta$  quantities from your regression model, state the null hypothesis for testing the  $A$  by  $B$  interaction (controlling for the covariate).

$$(H_0 : \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0)$$

- (d) Controlling for the covariate, are there any differences among the 6 treatment means? State the null hypothesis you would test in order to answer this question. Your answer should be in terms of  $\beta$  quantities from your regression model.

$$(H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = 0)$$

2. (3 Points) For the Bunny data, consider your tests for the effect of drug just at day 12.

- (a) On your list file, circle the value of Wilks' Lambda, the  $F$  statistic and the  $p$ -value. Write the question number (2a) beside it. Also, write the words "Statistically significant" or "Non-significant."

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.70621179	6.45	2	31	0.0046

- (b) Now consider the corresponding univariate tests as Bonferroni-protected follow-ups to the multivariate test. On your list file, cross out the two  $p$ -values, and write the Bonferroni-corrected  $p$ -values. Write the question number (2b) beside them.
- (c) Where is the multivariate result coming from? Write the answer to this question in the space below.

(For Force,  $0.0011 * 2 = 0.0022$ ; for Stiffness,  $0.0722 * 2 = 0.1444$ . The effect is coming from Force.)

3. (3 Points) Consider your analysis of the Wine data.

- (a) We want to know if there are any average differences among the wines in how expensive the judges think they should be. On your list file, circle the value of Wilks' Lambda, the  $F$  statistic and the  $p$ -value. Write the question number (3a) beside it. Also, write the words "Statistically significant" or "Non-significant."

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.02314815	42.20	3	3	0.0059

- (b) You followed up that last test with pairwise comparisons. Locate the  $p$ -values on your printout. Beside each  $p$ -value, write an asterisk (\*) if the test is significant with a Bonferroni correction. Write the question number (3b) too.

Variable	Mean	t Value	Pr >  t
d12	-2.0000000	-3.46	0.0180
d13	-6.6666667	-10.00	0.0002 *
d14	-6.0000000	-7.01	0.0009 *
d23	-4.6666667	-14.00	<.0001 *
d24	-4.0000000	-7.75	0.0006 *
d34	0.6666667	1.35	0.2354

- (c) State your findings below in plain, non-statistical language. Base your conclusions on the Bonferroni-protected pairwise comparisons. You have a lot more room than you need.

(Wines 3 and 4 are potentially more expensive than wines 1 and 2.)