

Name Jenny

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

If the model has product terms,  
zero marks for ~~this~~ question one: -6

## STA 441s 2018 Quiz 2A

1. (6 points) Hens (female chickens) are fed with one of three different feed types: A, B or C. The response variable is the mean weight of the eggs they lay, based on 100 eggs from each chicken. Hen's age is a covariate. Assume that the relationship between age and expected egg weight can be approximated by a straight line. *over the range of the data*

- (a) Write a regression equation that assumes the lines relating age and expected egg weight have the same slope for each feed type. You need not say how your dummy variables are defined. You will do that in the next part. *age*

$$E(Y|\mathbf{x}) = \beta_0 + \beta_1 f_1 + \beta_2 f_2 + \beta_3 x$$

- (b) Make a table with three rows, one for each feed type. Make columns showing how your dummy variables are defined. Add another, wider column showing expected egg weight for each feed type. The *symbols* for the dummy variables for feed type will not appear in this last column, because they are either zero or one.

| Feed | $f_1$ | $f_2$ | $E(Y \mathbf{x})$               |
|------|-------|-------|---------------------------------|
| A    | 1     | 0     | $\beta_0 + \beta_1 + \beta_3 x$ |
| B    | 0     | 1     | $\beta_0 + \beta_2 + \beta_3 x$ |
| C    | 0     | 0     | $\beta_0 + \beta_3 x$           |

- (c) Controlling for hen's age, is egg weight related to the type of feed? Give the null hypothesis, using symbols from your regression equation.

$$H_0: \beta_1 = \beta_2 = 0$$

- (d) Controlling for hen's age, is expected egg weight different for feed types A and B? Give the null hypothesis, using symbols from your regression equation.

$$H_0: \beta_1 = \beta_2$$

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

2. (4 points) In your analysis of the telephone sales data, you tested whether the lines relating sales last quarter to sales this quarter had the same slope for the three software packages.

1 point

- (a) Give the value of the test statistic ( $t$ , chi-squared or  $F$ .) The answer is a number from your printout. Write the number in the space below. Also, circle the number on your printout and write "Question 2a" beside it.

$$F = 10.30$$

3 points

- (b) You also tested for pairwise differences between slopes. Don't give any numbers. Just state your conclusions (if any) in plain, non-statistical language.

The slope of the line relating sales last quarter was steeper for software package 3 than for package 2. The slope for package 1 was also steeper than for package 2.

or

Sales this quarter increase faster as a function of sales last quarter for software package 1 than for package 2. The rate of increase was also faster for 3 than 2.

No marks if the printout does not have the tests required. Say -2 if no custom test for 1 vs 2

Attach your complete log file and your **COMPLETE** results file to the quiz paper. Make sure your name and student number are written clearly on both printouts.

-1 for accepting  $H_0$

Name Jerry

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

If the model has no product terms, zero marks for question 1: -6

## STA 441s 2018 Quiz 2B

1. (6 points) In data from a study of energy costs, houses have three different types of heating system: forced air, forced water and steam. Consider a regression model in which the response variable is energy consumption, and the explanatory variables are type of heating system and size of house in square metres.

- (a) Write a regression equation that assumes the lines relating size of house to expected energy consumption might have a different slope for each type of heating system. You need not say how your dummy variables are defined. You will do that in the next part.

$$E(Y|x) = \beta_0 + \beta_1 d_1 + \beta_2 d_2 + \beta_3 x + \beta_4 x d_1 + \beta_5 x d_2$$

- (b) Make a table with three rows, one for each heating system type. Make columns showing how your dummy variables are defined. Add another, wider column showing expected energy consumption for each type of heating system. The symbols for the dummy variables for heating system type will not appear in this last column, because they are either zero or one.

|              | $d_1$ | $d_2$ | $E(Y x)$                                    |
|--------------|-------|-------|---------------------------------------------|
| Forced Air   | 1     | 0     | $\beta_0 + \beta_1 + (\beta_3 + \beta_4) x$ |
| Forced Water | 0     | 1     | $\beta_0 + \beta_2 + (\beta_3 + \beta_5) x$ |
| Steam        | 0     | 0     | $\beta_0 + \beta_3 x$                       |

It does not need to be so nicely organized with  $x$  factored out.

- (c) Do differences in energy consumption between heating system types depend on the size of the house? Give the null hypothesis, using symbols from your regression equation.

$$H_0: \beta_4 = \beta_5 = 0$$

- (d) Is the slope of the line relating size of house to energy consumption different for forced air and forced water? Give the null hypothesis, using symbols from your regression equation.

$$H_0: \beta_4 = \beta_5$$



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

2. (4 points) In your analysis of the telephone sales data, you fit a model with software package and sales last quarter as the explanatory variables, and sales this quarter as the response variable. The lines relating sales last quarter to sales this quarter were parallel. You tested whether, controlling for sales last quarter, software package had any effect on sales.

1 point (a) Give the value of the test statistic ( $t$ , chi-squared or  $F$ .) The answer is a number from your printout. Write the number in the space below. Also, circle the number on your printout and write "Question 2a" beside it.

$$F = 0.24$$

3 points (b) You also tested for pairwise differences between software packages, *controlling for sales last quarter*. Don't give any numbers. Just state your conclusions (if any) in plain, non-statistical language.

Allowing for sales last quarter, there is not sufficient evidence to conclude that type of software package has an effect on sales.

or  
Correcting for sales last quarter, there is no evidence that average sales this quarter are different for the 3 software packages.  
*no marks if the printout does not have the tests required. Say -2 if no custom test for 1 vs. 2.*

Attach your complete log file and your **COMPLETE** results file to the quiz paper. Make sure your name and student number are written clearly on both printouts.

*-1 for accepting  $H_0$*