

STA 2101/442 Assignment Ten¹

The non-computer questions are just practice for the quiz, and are not to be handed in. Use R for Question 6, and bring your printout to the quiz. **Your printout should show *all* R input and output, and *only* R input and output.** Do not write anything on your printouts except your name and student number.

1. Arsenic is a powerful poison, which is why it has been used on farms for many years to kill insects. Even in very small amounts, arsenic can cause cancer in humans, and recently it has been found that rice and foods made from rice tend to be very high in arsenic. Brown rice is worse, by the way.

In a controlled experiment, pots of rice were prepared by either washing the rice first or not, and by cooking the rice in either a low, a medium or a high amount of water. The response variable is amount of arsenic in the cooked rice.

- (a) Use a regression model with *cell means coding*. That's the model with no intercept, and one indicator dummy variable for each treatment combination. You don't have to say how the dummy variables are defined. That will become clear in the next part. Just give the regression equation.
- (b) Write the expected amounts of arsenic in the table below, in terms of the β parameters of your model.

	Amount of Water		
	Low	Medium	High
Washed			
Unwashed			

- (c) If you wanted to test whether the effect of washing the rice depended on how much water you cook it in, what is the null hypothesis? Give your answer in terms of the β values in your model.
- (d) If you wanted to test whether washing the rice before cooking has any effect if the rice is cooked in a lot of water, what is the null hypothesis? Give your answer in terms of β values.
- (e) Suppose you want to test whether the amount of water used to cook the rice makes any difference if the rice has been washed. What is the null hypothesis? Give your answer in terms of β values.
- (f) Averaging across different amounts of water used to cook the rice, does pre-washing affect the amount of arsenic in the rice. What null hypothesis would you test to answer this question? Give your answer in terms of β values.
- (g) If you wanted to test whether the effect of the amount of water used to cook the rice depends on whether you wash it first, what is the null hypothesis? Give your answer in terms of β values.

¹Copyright information is at the end of the last page.

2. In a study of math education in elementary school, equal numbers of boys and girls were randomly assigned to one of three training programmes designed to improve spatial reasoning. After five school days of training, the students were given a standardized test of spatial reasoning. Score on the spatial reasoning test is the response variable. You will define a regression model for this factorial analysis of variance. Don't write the model yet.

- (a) In the table below, show how your dummy variables are defined. *Use effect coding.* That's the scheme with an intercept and minus ones. Write the name of each dummy variable at the head of its column.

Girls, Programme 1	
Girls, Programme 2	
Girls, Programme 3	
Boys, Programme 1	
Boys, Programme 2	
Boys, Programme 3	

- (b) Give $E[Y_i | \mathbf{X}_i = \mathbf{x}_i]$ for the full model. Include the interaction terms. Notice you are *not* being asked to write expected values in the table. They are too messy.
- (c) Suppose you want to test whether, averaging across training programmes, there is a difference between girls and boys in their average performance on the spatial reasoning test. State the null hypothesis in terms of the β values in your model.
- (d) Suppose you want to test whether, averaging across boys and girls, there is a difference between training programmes in average performance on the spatial reasoning test. State the null hypothesis in terms of the β values in your model.
- (e) Suppose you want to test whether the sex difference in average performance depends on which training programme the children are in. State the null hypothesis in terms of the β values in your model.

3. Consider a two-factor analysis of variance in which each factor has two levels. Use this regression model for the problem:

$$Y_i = \beta_0 + \beta_1 d_{i,1} + \beta_2 d_{i,2} + \beta_3 d_{i,1} d_{i,2} + \epsilon_i,$$

where $d_{i,1}$ and $d_{i,2}$ are dummy variables.

- (a) Make a two-by-two table showing the four treatment means in terms of β values. Use *effect coding*. In terms of the β values, state the null hypothesis you would use to test for
- i. Main effect of the first factor
 - ii. Main effect of the second factor
 - iii. Interaction
- (b) Make a two-by-two table showing the four treatment means in terms of β values. Use *indicator dummy variables* (zeros and ones). In terms of the β values, state the null hypothesis you would use to test for
- i. Main effect of the first factor
 - ii. Main effect of the second factor
 - iii. Interaction
- (c) Which dummy variable scheme do you like more?
4. Effect coding is not the most convenient choice for every purpose. Please consider the two-factor rotten potato example from lecture set 20, the one entitled “Interactions and Factorial ANOVA.” Lecture slide 39 has the expected response for each treatment combination under effect coding. Suppose you wanted to test whether Bacteria Type has an effect just for cool temperatures.
- (a) Write the null hypothesis in scalar form. Simplify as much as possible.
 - (b) Give the \mathbf{L} matrix in $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{h}$.
5. In that last question, your answer may be different from mine, but we both could be right. Let \mathbf{A} be an $r \times r$ matrix with an inverse, and suppose the null hypotheses is written $H_0 : \mathbf{A}\mathbf{L}\boldsymbol{\beta} = \mathbf{A}\mathbf{h}$ instead of $H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{h}$. Show that the F^* statistic for the general linear test is unaffected.

6. I know this is pretty gruesome, but the data are real. An experiment in dentistry seeks to test the effectiveness of a drug (HEBP) that is supposed to help dental implants become more firmly attached to the jaw bone. This is an initial test on animals. False teeth were implanted into the leg bones of rabbits, and the rabbits were randomly assigned to receive either the drug or a saline solution (placebo). Technicians administering the drug were blind to experimental condition.

Rabbits were also randomly assigned to be "sacrificed" after either 3, 6, 9 or 12 days. At that time, the implants were pulled out of the bone by a machine that measures force in newtons and stiffness in newtons/mm. For both of these measurements, higher values indicate more healing. A measure of "pre-load stiffness" in newtons/mm is also available for each animal. This may be another indicator of how firmly the false tooth was implanted into the bone, but it might even be a covariate. Nobody can seem to remember what "preload" means, so we'll ignore this variable for now.

The explanatory variables are Time and Drug. The response variable is Force required to pull out the tooth. There is more than one reasonable way to do this analysis, but just to keep us together please treat Time as a categorical variable.

The data are available in the file [bunnies.data.txt](#). The variables are

- Identification code
 - Time (3,6,9,12 days of healing)
 - Drug (1=HEBP, 0=saline solution)
 - Stiffness in newtons/mm
 - Force in newtons
 - Preload stiffness in newtons/mm
- (a) Use `table` to find out how many rabbits are in each experimental condition.
- (b) Carry out the standard tests of main effects and interactions. Be prepared to answer the following questions about each test.
- i. What is the value of the test statistic? The answer is a number from your printout.
 - ii. What proportion of the remaining variation is explained? You can use R, or just be ready to do it with a calculator.
 - iii. What is the p -value? The answer is a number from your printout.
 - iv. Do you reject the null hypothesis at the 0.05 level? Yes or No.
 - v. What, if anything, do you conclude? This is not the place for statistical jargon. "What do you conclude" means say something about the drug, healing, time – something like that.
- (c) I know this is a bit redundant with the preceding question, but *averaging across time, did the drug help the teeth become more firmly attached to the bone?* If the results justify an answer, then answer Yes or No.
- (d) Make a table with a row for each treatment combination. Make columns showing the dummy variables for effect coding.

- (e) Give $E[Y|\mathbf{X} = \mathbf{x}]$ for a regression model with both main effects and the interaction. Use your variable names from the preceding question.
- (f) In terms of the β values of your regression model, give the null hypothesis you would test in order to answer each of the following questions.
- i. Averaging across time periods, is there a difference between the drug and placebo in mean force required to extract the tooth?
 - ii. Averaging across drug and placebo, does elapsed time affect the mean force required to extract the tooth?
 - iii. Does the effect of the drug depend upon elapsed time?
- (g) Now please return to R. Doing it the easiest way you can, conduct tests to answer the following questions. Just do regular one-at-a-time (custom) tests. Don't bother with any Bonferroni correction this time. Just consider one response variable: Force. As usual, we are guided by the $\alpha = 0.05$ significance level.
- i. Are the marginal means different at 3 and 6 days?
 - ii. Are the marginal means different at 6 and 9 days?
 - iii. Are the marginal means different at 9 and 12 days?
 - iv. Is there a difference between Drug and Placebo just at 3 days?
 - v. Is there a difference between Drug and Placebo just at 6 days?
 - vi. Is there a difference between Drug and Placebo just at 9 days?
 - vii. Is there a difference between Drug and Placebo just at 12 days?
 - viii. Be able to answer questions like these for each test:
 - A. What is the value of the test statistic? The answer is a number from your printout.
 - B. What proportion of the remaining variation is explained? You don't have to do all these calculations in advance; just be ready to do them with a calculator.
 - C. What is the p -value? The answer is a number from your printout.
 - D. Do you reject the null hypothesis at the 0.05 level? Yes or No.
 - E. What, if anything, do you conclude? This is not the place for statistical jargon. "What do you conclude" means say something about the drug, healing, time – something like that.

Please bring your printout for Question 6 to the quiz. **Your printout should show *all* R input and output, and *only* R input and output.** Do not write anything on your printouts except your name and student number.

This assignment was prepared by [Jerry Brunner](#), Department of Statistics, University of Toronto. It is licensed under a [Creative Commons Attribution - ShareAlike 3.0 Unported License](#). Use any part of it as you like and share the result freely. The L^AT_EX source code is available from the course website: <http://www.utstat.toronto.edu/~brunner/oldclass/appliedf17>