

## STA 2101/442 Assignment Six<sup>1</sup>

Please bring your R printouts to the quiz. *Your printouts must not contain answers to the non-computer parts of this assignment.* The non-computer questions are just practice for the quiz, and will not be handed in.

1. Men and women are calling a technical support line according to independent Poisson processes with rates  $\lambda_1$  and  $\lambda_2$  per hour. Data for 144 hours are available, but unfortunately the sex of the caller was not recorded. All we have is the number of callers for each hour, which is distributed  $\text{Poisson}(\lambda_1 + \lambda_2)$ . Here are the data, which are also available in the file [poisson.data](#) on the course website:

```
12 9 25 11 22 16 17 8 14 17 14 17 8 14 18 16 13 17 13 11 9 15 18
14 16 17 16 19 13 18 12 12 13 10 8 15 13 11 15 15 6 10 13 13 11 13
9 15 16 9 5 10 8 18 13 17 7 13 13 17 12 17 14 16 6 12 17 10 9
14 11 19 13 17 15 20 14 10 13 14 17 9 13 14 7 16 16 9 25 10 10 9
17 7 15 12 14 21 14 18 14 12 13 15 12 11 16 14 15 16 8 19 13 17 15
11 18 13 12 11 19 14 16 17 13 13 19 19 11 19 10 12 9 18 11 14 9 14
14 14 13 9 13 18
```

- (a) The parameter in this problem is  $\boldsymbol{\theta} = (\lambda_1, \lambda_2)^\top$ . Try to find the MLE analytically. Show your work. Are there any points in the parameter space where both partial derivatives are zero?
- (b) Now try to find the MLE numerically with R's `nlm` function<sup>2</sup>. The Hessian is interesting; ask for it. Try two different starting values. Compare the minus log likelihoods at your two answers. What seems to be happening here?
- (c) Try inverting the Hessian to get the asymptotic covariance matrix. Any comments?
- (d) To understand what happened in the last item, calculate the Fisher information in a single observation from the definition. That is, letting  $\ell = \log f(Y; \boldsymbol{\theta})$ , calculate the elements of the  $2 \times 2$  matrix whose  $(i, j)$  element is

$$-E\left(\frac{\partial^2 \ell}{\partial \theta_i \partial \theta_j}\right)$$

- (e) The Fisher information in the sample is just  $n$  times the Fisher information in a single observation. Using the numerical MLEs from one of your `nlm` runs, estimate this quantity (a  $2 \times 2$  matrix). Compare it to the Hessian. Now do you see what happened when you tried to calculate the asymptotic covariance matrix?

---

<sup>1</sup>Copyright information is at the end of the last page.

<sup>2</sup>When I did this, I got lots of warning messages with some starting values, when the search repeatedly left the parameter space and then bounced back in. For this problem, you don't need to worry about the warnings as long as the exit code is one.

(f) Why did estimation fail for this fairly realistic model?

Bring your printout to the quiz. Your printout must *not* contain answers to the non-computer parts of this question. That is, it must contain only numerical answers.

2. Consider the usual fixed effects multiple regression model in scalar form:

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_{p-1} x_{i,p-1} + \epsilon_i,$$

where the  $x_{i,j}$  quantities are fixed constants, and the  $\epsilon_i$  variables are independent  $N(0, \sigma^2)$ .

- (a) Write the likelihood function.
- (b) The least squares estimates of the regression coefficients are the  $\beta_j$  values that minimize

$$Q(\boldsymbol{\beta}) = \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 x_{i,1} - \cdots - \beta_{p-1} x_{i,p-1})^2$$

Are the least squares and maximum likelihood estimates the same in this case (the normal case), or are they different? You don't need to derive  $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$  in order to see the answer, so don't do it.

- (c) Differentiate  $Q(\boldsymbol{\beta})$  with respect to  $\beta_0$  and set the derivative to zero, obtaining the first *normal equation*.
  - (d) Noting that the quantities  $\widehat{\beta}_0, \dots, \widehat{\beta}_{p-1}$  must satisfy the first normal equation, show that the least squares plane must pass through the point  $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_{p-1}, \bar{Y})$ .
  - (e) Defining "predicted"  $Y_i$  as  $\widehat{Y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i,1} + \cdots + \widehat{\beta}_{p-1} x_{i,p-1}$ , show that  $\sum_{i=1}^n \widehat{Y}_i = \sum_{i=1}^n Y_i$ .
  - (f) The *residual* for observation  $i$  is defined by  $e_i = Y_i - \widehat{Y}_i$ . Show that the sum of residuals equals exactly zero.
3. For the general fixed effects linear regression model in matrix form (see formula sheet), show that the  $n \times p$  matrix of covariances  $C(\mathbf{e}, \widehat{\boldsymbol{\beta}}) = \mathbf{0}$ . Why does this establish that *SSE* and  $\widehat{\boldsymbol{\beta}}$  are independent?

4. The  $t$  distribution is defined as follows. Let  $Z \sim N(0, 1)$  and  $W \sim \chi^2(\nu)$ , with  $Z$  and  $W$  independent. Then  $T = \frac{Z}{\sqrt{W/\nu}}$  is said to have a  $t$  distribution with  $\nu$  degrees of freedom, and we write  $T \sim t(\nu)$ .

For the general fixed effects linear regression model, tests and confidence intervals for linear combinations of regression coefficients are very useful. Derive the appropriate  $t$  distribution and some applications by following these steps. Let  $\mathbf{a}$  be a  $p \times 1$  vector of constants.

- What is the distribution of  $\mathbf{a}^\top \hat{\boldsymbol{\beta}}$ ? Show a little work. Your answer includes both the expected value and the variance.
  - Now standardize the difference (subtract off the mean and divide by the standard deviation) to obtain a standard normal.
  - Divide by the square root of a well-chosen chi-squared random variable, divided by its degrees of freedom, and simplify. Call the result  $T$ .
  - How do you know numerator and denominator are independent?
  - Suppose you wanted to test  $H_0 : \mathbf{a}^\top \boldsymbol{\beta} = c$ . Write down a formula for the test statistic.
  - Suppose you wanted to test  $H_0 : \beta_2 = 0$ . Give the vector  $\mathbf{a}$ .
  - Suppose you wanted to test  $H_0 : \beta_1 = \beta_2$ . Give the vector  $\mathbf{a}$ .
  - Letting  $t_{\alpha/2}$  denote the point cutting off the top  $\alpha/2$  of the  $t$  distribution with  $n - p$  degrees of freedom, give a  $(1 - \alpha) \times 100\%$  confidence interval for  $\mathbf{a}^\top \boldsymbol{\beta}$ .
5. Consider the prediction interval for  $Y_{n+1}$ .
- What is the distribution of  $Y_{n+1} - \hat{Y}_{n+1} = Y_{n+1} - \mathbf{x}_{n+1}^\top \hat{\boldsymbol{\beta}}$ ? Show your work. Your answer includes both the expected value and the variance.
  - Now standardize the difference to obtain a standard normal.
  - Divide by the square root of a chi-squared random variable, divided by its degrees of freedom, and simplify. Call it  $T$ . Compare your answer to a slide from lecture. How do you know that numerator and denominator are independent?
  - Using your result, derive the  $(1 - \alpha) \times 100\%$  prediction interval for  $Y_{n+1}$ .
6. When you fit a full and a reduced regression model, the proportion of remaining variation explained by the additional variables in the full model is  $a = \frac{R_F^2 - R_R^2}{1 - R_R^2}$ . Show

$$F = \frac{(SSR_F - SSR_R)/r}{MSE_F} = \left( \frac{n - p}{r} \right) \left( \frac{a}{1 - a} \right).$$

You are proving a fact that's on the formula sheet.

7. In the usual univariate multiple regression model, the  $\mathbf{X}$  is an  $n \times p$  matrix of known constants. But of course in practice, the explanatory variables are random, not fixed. Clearly, if the model holds *conditionally* upon the values of the explanatory variables, then all the usual results hold, again conditionally upon the particular values of the explanatory variables. The probabilities (for example,  $p$ -values) are conditional probabilities, and the  $F$  statistic does not have an  $F$  distribution, but a conditional  $F$  distribution, given  $\mathbf{X} = \mathbf{x}$ .
- Show that the least-squares estimator  $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$  is conditionally unbiased.
  - Show that  $\hat{\boldsymbol{\beta}}$  is also unbiased unconditionally. Use double expectation.
  - A similar calculation applies to the significance level of a hypothesis test. Let  $F$  be the test statistic (say for an  $F$ -test comparing full and reduced models), and  $f_c$  be the critical value. If the null hypothesis is true, then the test is size  $\alpha$ , conditionally upon the explanatory variable values. That is,  $P(F > f_c | \mathbf{X} = \mathbf{x}) = \alpha$ . Find the *unconditional* probability of a Type I error. Assume that the explanatory variables are discrete, so you can write a multiple sum.
8. For this question, you will use the `sat.data` again. There is a link on the course web page in case the one in this document does not work. We seek to predict GPA from the two test scores. Throughout, please use the usual  $\alpha = 0.05$  significance level.
- First, fit a model using just the Math score as a predictor. “Fit” means estimate the model parameters. Does there appear to be a relationship between Math score and grade point average?
    - Answer Yes or No.
    - Fill in the blank. Students who did better on the Math test tended to have \_\_\_\_\_ first-year grade point average.
    - Do you reject  $H_0 : \beta_1 = 0$ ?
    - Are the results statistically significant? Answer Yes or No.
    - What is the  $p$ -value? The answer can be found in *two* places on your printout.
    - What proportion of the variation in first-year grade point average is explained by score on the SAT Math test? The answer is a number from your printout.
    - Give a predicted first-year grade point average and a 95% prediction interval for a student who got 700 on the Math SAT.

- (b) Now fit a model with both the Math and Verbal sub-tests.
- i. Give the test statistic, the degrees of freedom and the  $p$ -value for each of the following null hypotheses. The answers are numbers from your printout.
    - A.  $H_0 : \beta_1 = \beta_2 = 0$
    - B.  $H_0 : \beta_1 = 0$
    - C.  $H_0 : \beta_2 = 0$
    - D.  $H_0 : \beta_0 = 0$
  - ii. Controlling for Math score, is Verbal score related to first-year grade point average?
    - A. Give the null hypothesis in symbols.
    - B. Give the value of the test statistic. The answer is a number from your printout.
    - C. Give the  $p$ -value. The answer is a number from your printout.
    - D. Do you reject the null hypothesis?
    - E. Are the results statistically significant? Answer Yes or No.
    - F. In plain, non-statistical language, what do you conclude? The answer is something about test scores and grade point average.
  - iii. Controlling for Verbal score, is Math score related to first-year grade point average?
    - A. Give the null hypothesis in symbols.
    - B. Give the value of the test statistic. The answer is a number from your printout.
    - C. Give the  $p$ -value. The answer is a number from your printout.
    - D. Do you reject the null hypothesis?
    - E. Are the results statistically significant? Answer Yes or No.
    - F. In plain, non-statistical language, what do you conclude? The answer is something about test scores and grade point average.
  - iv. Math score explains \_\_\_\_ percent of the remaining variation in grade point average once you take Verbal score into account. Using the formula from the slides (see formula sheet), you should be able to calculate this from the output of the `summary` function. You can check your answer using the `anova` function.
  - v. Verbal score explains \_\_\_\_ percent of the remaining variation in grade point average once you take Math score into account. Using the formula from the slides (see formula sheet), you should be able to calculate this from the output of the `summary` function. You can check your answer using the `anova` function.

- vi. Give a predicted first-year grade point average and a 95% prediction interval for a student who got 650 on the Verbal and 700 on the Math SAT. Are you confident that this student's first-year GPA will be above 2.0 (a C average)?
- vii. Let's do one more test. We want to know whether expected GPA increases faster as a function of the Verbal SAT, or the Math SAT. That is, we want to compare the regression coefficients, testing  $H_0 : \beta_1 = \beta_2$ .
  - A. Express the null hypothesis in matrix form as  $\mathbf{L}\boldsymbol{\beta} = \mathbf{h}$ .
  - B. Carry out a two-sided  $t$ -test.
  - C. Carry out an  $F$  test, the easy way. Does  $F = t^2$ ?
  - D. State your conclusion in plain, non-technical language. It's something about first-year grade point average.

Bring your printout to the quiz.

---

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  $\text{\LaTeX}$  source code is available from the course website: <http://www.utstat.toronto.edu/~brunner/oldclass/appliedf14>