

STA 312f12 Assignment Eight¹

Please bring your R printouts to the quiz. The non-computer questions are practice for the quiz on Friday Nov. 9th, and are not to be handed in. **Bring a calculator to the quiz.**

1. Awards received by students at a particular high school are thought to occur according to a Poisson process. That is, the numbers of awards received by students in one year are independent Poisson random variables, with mean λ that may depend on characteristics of the student. From the [Data Sets](#) link on the course home page, you can find the Awards data. The variables are Student identification code, Number of awards, Program (1=General, 2=Academic, 3=Vocational), and Score on a test of general academic knowledge. If you use `labels = c("General", "Academic", "Vocational")` in your `factor` statement, you will get nicer output.
 - (a) Using `table`, make frequency table of number of awards. Does it look roughly normal?
 - (b) Consider a Poisson regression model, without actually fitting it yet.
 - i. What is the linear predictor? There should be no product terms (yet).
 - ii. What is the link function?
 - iii. Make a table with 3 rows, one for each academic program. Make columns showing how R will define the dummy variables for the variable academic program. If you're not sure, you can check your answer with R.
 - iv. Add another column to your table, showing the expected number of awards given score on the math test, for each academic program.
 - v. The expected number of awards for a student in the Vocational program is _____ times as great as the expected number of awards for a student in the General program with the same score on the general knowledge test.
 - vi. The expected number of awards for a student in the Academic program is _____ times as great as the expected number of awards for a student in the General program with the same score on the general knowledge test.
 - vii. The expected number of awards for a student in the Academic program is _____ times as great as the expected number of awards for a student in the Vocational program with the same score on the general knowledge test.
 - viii. Explain why this model could be called a “proportional means” model.
 - ix. Suppose we wanted to test the proportional means assumption (and it is an assumption).
 - A. Write a linear model for the log of the mean for the full model you would use.
 - B. State the null hypothesis. It is a statement about the β values in the full model.

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

- C. What is the reduced model?
 - D. What are the degrees of freedom of this test?
- (c) Now fit the proportional means Poisson regression model to the awards data. For each question below, state the null hypothesis, give the value of the test statistic (Z or χ^2), the p -value, and be able to state the conclusion in plain language. Give a *directional* conclusion if possible, even though the test is non-directional.
- i. Controlling for academic program, is score on the test of general knowledge related to the expected number of awards?
 - ii. Controlling for score on the test of general knowledge, do students in the Academic program get more awards on average than students in the General program?
 - iii. Controlling for score on the test of general knowledge, do students in the Vocational program get more awards on average than students in the General program?
 - iv. Do any of the explanatory variables matter? You could do this with a calculator from the default output if necessary, but do it with R and get the p -value.
 - v. Controlling for score on the test of general knowledge, do students in the Vocational program get the same number of awards on average as students in the Academic program? I can't get this from the default output.
 - vi. The expected number of awards for a student in the Vocational program is estimated to be _____ times as great as the expected number of awards for a student in the General program with the same score on the general knowledge test.
 - vii. The expected number of awards for a student in the Academic program is estimated to be _____ times as great as the expected number of awards for a student in the General program with the same score on the general knowledge test.
 - viii. The expected number of awards for a student in the Academic program is estimated to be _____ times as great as the expected number of awards for a student in the Vocational program with the same score on the general knowledge test.
- (d) Finally, test the proportional means assumption with a likelihood ratio test. Give the value of G^2 , the degrees of freedom and the p -value. Do the ratios of expected numbers of awards appear to depend on the student's level of general knowledge?

2. In the *Heart attack data* (which you will analyze later), a sample of middle-aged men who had heart attacks were classified into three groups. Either they died of the first heart attack, or they died during the next 10 years, or they were still alive 10 years after the first attack. This is the response variable. Potential explanatory variables include age, blood pressure, and family history of heart disease (Yes-No). Let's just consider these for now. For interpretability, make the probability of being alive 10 years later the denominator in each generalized logit.
- (a) Write the multinomial logit model for these data. How many generalized logits do you have? Of course you must have a regression equation for each one.
 - (b) Solve for the probabilities in terms of the *beta* values in your model. Show your work.
 - (c) Make a table with two rows, one for Family history = Yes, and one for Family history = No. In each row, write *two* probability ratios. Let's call them "relative risks." (The relative risk of dying in a particular way is the probability of dying that way divided by the probability of living.)
 - (d) Controlling for age and blood pressure, the relative risk of dying in the first heart attack is _____ times as great for those with a family history of coronary heart disease.
 - (e) Controlling for age and blood pressure, the relative risk of dying in the next 10 years after the first heart attack is _____ times as great for those with a family history of coronary heart disease.

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/312f12>