Project #5

STAT 875

Spring 2015

Complete the following problems below. Within each part, include your R program output with code inside of it and any additional information needed to explain your answer. Your R code and output should be formatted in the exact same manner as in the lecture notes.

1. (12 total points) The purpose of this problem is to show how the analysis of the Olestra data (not including the chips variable) can be performed using a loglinear model. Complete the following:
   1. (1 point) Convert the data to the appropriate format needed for the glm() function.
   2. (3 points) Estimate a saturated Poisson regression model using both the row and column variables as explanatory variables in the model.
   3. (2 points) Rewrite the model in a loglinear model format like shown on p. 4.41-4.42 of the notes.
   4. (2 points) Show that the estimated mean counts for the model in b) are equal to the observed counts as should occur with a saturated model.
   5. (4 points) Perform a LRT for independence between the row and column variables using the model in b). Compare your results here to that which can also be obtained using the assocstats() function from the vcd package.
2. (21 total points) On *Star Trek: The Original Series*, the members of the Enterprise crew wore blue, gold, or red shirts as part of their uniform. It is generally thought of by Star Trek fans that crew with red shirts were more likely to be casualties than those wearing the other shirt colors. Below are the casualty data corresponding to the series (source: <http://www.statslife.org.uk/jobs?catid=0&id=458>):

> trek <- data.frame(color = c("blue", "gold", "red"), casualties = c(7, 9, 24),

crew = c(136, 55, 239))

> trek

color causalties crew

1 blue 7 136

2 gold 9 55

3 red 24 239

The crew variable corresponds to the crew complement on the Enterprise with a particular shirt color. Using this data, complete the following using α = 0.05 when needed.

* 1. (3 points) Why can not one simply conclude that the red uniform crew were more likely to be a casualty because they have a larger number of casualties in the data set?
  2. (3 points) Estimate a Poisson rate regression model using causalities as the response variable, color as an explanatory variable, and crew as a baseline measure.
  3. (3 points) Does uniform color truly matter with respect to causality rate? Perform a LRT to answer this question.
  4. (3 points) Find the ratio of the means for the red and gold shirts written in terms of the model (no estimates). Show your work!
  5. (6 points) Compute 95% profile LR intervals for PC comparing:
     1. Red to gold
     2. Red to blue
     3. Gold to blue

Interpret the intervals. Note that I was unable to get mcprofile() to work for this problem.

* 1. (3 points) Use the AIC and BIC to describe why the Poisson rate regression model that takes into account the shirt color is better than the one which does not. Note that Exercise 16c of Chapter 5 discusses a simple function to calculate the BIC.

1. (9 total points) Consider the Poisson regression model . Complete the following:
   1. (3 points) Derive the percent change in the mean response for a c-unit change in x1.
   2. (3 points) Provide a formal interpretation of the expression derived in a).
   3. (3 points) State the Wald interval for the expression derived in a). Derive the necessary variance expression needed for the Wald interval so that the variance is a simple function of variances and/or covariances of individual regression parameter estimators