Alyssa Semerdjian

Stat 510

19 Feb 2018

Homework 5

1.

a) The odds ratio for mjuser~married is 0.3638 and the odds ratio for mjuser~married+age+educate is

0.5100. They are not the same because the longer model takes age and education into account as well as marital status.

b)

> anova(model1,model2,test="Chisq")

Analysis of Deviance Table

Model 1: mjuser ~ married

Model 2: mjuser ~ married + age + educate

Resid. Df Resid. Dev Df Deviance Pr(>Chi)

1 9095 8197.7

2 9093 7876.6 2 321.1 < 2.2e-16 \*\*\*

The p-value is 2.2e-16, which is less than 0.05, so we should reject the null hypothesis that the short

model is better and conclude that the longer model explains the data better.

c) The odds ratio for the education variable without any other variables taken into account is 1.002. The odds ratio of education with marital status and age taken into account is 1.018. Adjusting for the effects of marriage and education, for each one-year increase in age, the odds of marijuana use are expected to change 0.937 times.

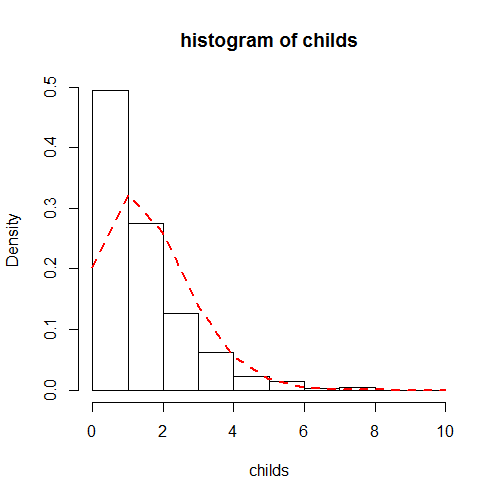
d) The average age for unmarried people is 29.316, the average age for married people is 34.826. The average education level for unmarried people is 12.807 while the average education level for married people is 12.694.

e) The probability of marijuana use for married people of average age and education levels is about 0.096. The probability of marijuana use for unmarried people of average age and education levels is about 0.230. The p-value for ‘married’ is <2e-16. This tells us that there is a significant difference in marijuana use between married and unmarried people when age and education level is take into account. This makes sense given the apparent difference in probabilities.

2.

a) Using complete cases as in data6 only deletes rows that are missing data in the columns you designate. Using na.omit as in data99 deletes any row that’s missing data, including columns you’re not using. I would rather use data6 because more of the data is preserved.

b) The mean for childs is 1.597 while the variance is 2.323. The mean is smaller than the variance, telling us that this is not quite a poisson distribution and that there is overdispersion. The plot shows that there are far more 0’s than would be expected for a poisson distribution. Also there are fewer 3’s 4’s and 5’s.



c) using lm for this model is incorrect because childs is count data, meaning it doesn’t go from negative infinity to infinity. It has a poisson distribution, not a normal distribution.

d) The p-value for the anova comparing the two models is 2.2e-16, which is highly significant. This

means we should reject the null hypothesis that the short model is better and conclude that the longer model explains the data better.

e) > 1-pchisq(3216.9-3102.0,4)

[1] 0

Using the null and residual deviances we find that the model is significant. The p-value is close to 0.

Adjusting for the effects of race, education and income, female respondents are expected to have children about 106% more than male respondents. Adjusting for the effects of sex, race and income, each one unit increase in education is associated with about 6.18% decrease in the number of children.

(For the last part: 1-education coefficient)