

Problem Set 7

cdurso

06 March, 2019

Tests of Independence, Linear Regression

These questions were rendered in R markdown through RStudio (<https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf>, <http://rmarkdown.rstudio.com>).

Please complete the following tasks, in R where applicable. Please generate a solution document in R markdown and upload the rendered .doc, .docx, or .pdf document. You may add hand computations to a .doc or .docx if you prefer. In the rendered document, please show your code. That is, don't use "echo=FALSE".

In either case, your work should be based on the data's being in the same folder as the R files. Please turn in your work on Canvas. Your solution document should have your answers to the questions and should display the requested plots.

1.a Let X and Y be independent random variables with χ^2 distributions with n and m degrees of freedom respectively. What is the distribution of $X + Y$? (5 points)

1.b. Let X and Y be independent random variables with χ^2 distributions with n and m degrees of freedom respectively. What is the distribution of X/Y ? (5 points)

2. The data in "school_goals.csv" come from <http://lib.stat.cmu.edu/DASL/Stories/Students'Goals.html>. Students in grades 4-6 in selected schools in Michigan were asked the following question: What would you most like to do at school? a) make good grades, b) be good at sports, c) be popular. The results are recorded in "Goals". Demographic information about the students was also collected.

- 2.a. Please display a table of totals for each Goal for each gender (Gender), and a table of proportions for each Goal for gender. The "spread" function may be helpful in generating a two-way table with rows for gender and columns for priorities. (10 points)
- 2.b. Please do a χ^2 test and a Fisher's Exact Test to determine if the goals are independent of Gender, and compare the results (10 points)

3.a. Returning to the precipitation data, use maximum likelihood regression to model total precipitation (total) as a linear function of year. Show the resulting coefficients and their p-values. (10 points)

```
dat<-read.csv("dat_ok.csv")
```

3.b. Is a linear model for the relationship between time and total appropriate? Provide any diagnostic plots you feel are relevant to addressing this question. (5 points)

3.c. Is the assumption of independent, identically distributed Normal errors justifiable? Please show any relevant plots. (5 points)