

# *Categorical Data Analysis: Homework 1*

*Original content by Dr. Jason Newsom. Sidenotes containing mathematical notations, R-Code & Rmarkdown formatting by Rachel Smith*

*Due 24 October 2016*

## *Contents*

|                     |   |
|---------------------|---|
| <i>Instructions</i> | 1 |
| <i>Question-1</i>   | 2 |
| <i>Question-2</i>   | 3 |
| <i>Question-3</i>   | 4 |
| <i>Question-4</i>   | 5 |
| <i>Question-5</i>   | 6 |
| <i>Question-6</i>   | 7 |

---

## *Instructions*

FOR ALL QUESTIONS: Your answers should be in your own words and most answers should be approximately one paragraph, *and* ...

- Show your work or include a copy of the output, whichever is relevant.
- Type your answers in report form, as if you were describing results in a published study.
- Include the relevant descriptive and statistical values in your write-up (e.g., *proportions*, *chi-squared*).

## Question-1

Below is a summary of some data provided in a **2014 report** from the Portland police related to racial profiling in traffic stops. To simplify, *the frequencies below use only data from African American and White drivers or pedestrians.*

- **(A)** The observed proportions ( $P()$ ) for traffic stops are 0.08 for African American drivers (AA) and 0.92 for White drivers (WH), and the sample size is 11,769.
  - Compute the **binomial test (z-test for proportions)**, the **confidence limits**, and the **margin of error by hand** (please show your work).
  - Specify which type of test and confidence limits you used.
  - Use 0.5 as the test proportion.
  - Report and interpret your findings in terms of the research problem.
- **(B)** For the same data, compute a **Pearson  $\chi^2$  test by hand**, also assuming an equal null frequency for the two groups.
  - The corresponding frequencies are 985 African American drivers and 10,782 White drivers.
  - Report and interpret your findings in terms of the research problem.

$$P(AA) = 0.084 \text{ (7.98\%)}$$

$$P(WH) = 0.916 \text{ (87.02\%)}$$

$$\binom{n}{k} \pi^{n-k} = \frac{n!}{k!(n-k)!} \pi^k (1-\pi)^{n-k}$$

*Question-2*

Based results from the same report, I have created a data set and input statements (race.sav, race.sps, race.R, race.sas), available at the [data page](#). Use SPSS, R, or SAS to compute the binomial test to determine whether there was a significantly higher proportion of African American drivers stopped by the police than the benchmark proportion of drivers.

- **(A)** Compute the **95% confidence interval**.
  - Be sure to specify the type of test and confidence limits you used.
- **(B)** The variable of interest is called DRIVER. The report uses the proportion of injury accidents as one possible benchmark to determine the base rate (null hypothesis proportion) of African American vs. White drivers on the road. According to the report, the benchmark proportion to use for African Americans involved in injury accidents is .080389769 (out of only African American and White drivers in injury accidents).
  - Report and interpret your findings in terms of the research problem (you are welcome to cut and paste material from your earlier answer for this question).

## Question-3

A second data set (`race2.sav`), which concerns police stops of pedestrians, is available at the data page) and is derived from the same report. There are five race categories (African American, Asian, Hispanic, Native American, White) and two categories of reasons given for the stop (major offense, minor offense).<sup>1</sup>

<sup>1</sup> "Major" offenses include drug offenses, open alcohol container, and violent crimes. "Minor" offenses include jaywalking or other minor infractions.

- (A) Use *SPSS*, *R*, or *SAS* to compare the five racial groups with *Pearson*  $\chi^2$  to see if there are group differences in the number of pedestrian stops. (You will not need the offense variable for this problem.) What is problematic about this comparison?
- (B) Because they are the same analysis, provide just one write-up for parts B and C.
  - Conduct a  $2 \times 2$   $\chi^2$  *analysis by hand* to check whether there is a significant difference among African American as compared to White pedestrians in whether they are stopped for a major vs. a minor offense.
  - Please show your work.
- (C) Using *SPSS*, *R*, or *SAS*, conduct a  $2 \times 2$   $\chi^2$  *analysis* to check whether there is a significant difference among African American as compared to White pedestrians in whether they were stopped for a major vs. a minor offense.
  - Show how to obtain the phi statistic, the relative risk ratio, and the odds ratio by hand.
  - Interpret all of your results (*percentages*, *Pearson*  $\chi^2$ ,  $\phi$ , *odds ratio*, and *relative risk ratio*) in terms of the research problem.

*Question-4*

The data for this problem come from a randomized experiment investigating the effects of Early Head Start, a preschool program for ages 0 to 3, on child maltreatment (conducted by Professor Beth Green at Portland State). I have created a data set and input statements (`child.sav`, `child.sps`, `child.R`, `child.sas`), <http://web.pdx.edu/~newsomj/data.htm>. In earlier analyses I found that there was a marginally significant difference between whether there was any report of abuse (*ABUSE*) according to whether or not the mother had a boyfriend who was not the biological father (*BOYFRIEND*).

- Investigate whether the program (*PROGRAM*) was able to reduce or eliminate this marginal difference by conducting the most appropriate significance test of a three-way contingency table for these data.
- Present two **two-way tables** (*ABUSE* × *BOYFRIEND*), one for the *control group* and one for the *program group*, that include both the **counts** and **percentages**.
- Describe whichever counts and percentages needed to explain the results in the text of your write-up and give the appropriate statistical values that indicate whether the program made a significant difference in the likelihood that abuse would occur in households with a boyfriend.

USE YOUR OWN DATA SET FOR Problems 5 and 6. In addition to your results paragraph for each, please add a one-paragraph (**no more!**) description of your study design, sample, and pertinent measures. Choose research questions that are of interest to you—your results do not need to be significant.

Please let me know if you need help finding an appropriate data set, variables, or research question.

### *Question-5*

Choose a binary variable to examine the difference in proportions using a binomial test using whatever test proportion is most appropriate in your case.

- Include the **95% confidence limits**, and specify which type of test and confidence limits you used.
- Report and interpret your findings in terms of the research problem.

*Question-6*

Choose two categorical variables and test an interesting hypothesis about their relationship or group differences using a **Pearson  $\chi^2$  for a  $2 \times 2$  or  $I \times J$  table**.

- Include a table of the counts and percentages in addition to your write-up.
- Obtain and interpret the **phi statistic**, the **relative risk ratio**, and the **odds ratio** in addition to the results of the **Pearson  $\chi^2$**  test.