**Homework 1**
**Due 4/22/21** (pdf format please)

For all questions, please show your work or include a copy of the output, whichever is relevant.  Please 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).  **Your answers should be in your own words** and most answers should be approximately one paragraph.

For each of the problems below, request a specific analysis for the data set I have supplied, which are available at http://web.pdx.edu/~newsomj/data.htm. However, you also are free to use your own data set for any of the problems. If you use your data set, make sure the variable types match the analysis requested for the problem. For any data set you use, please supply a one paragraph description of the study and the variables you will be using.

1. Data for this question come from a recent poll conducted by CNN asked a series of questions about the American Rescue Plan, the COVID relief bill recently passed (`rescue.sav`). One question posed to respondents was whether they supported the minimum wage provision (later dropped from the plan) requiring raising the federal minimum wage to $15/hr (`minwage`; 0="no", 1="yes").[1] Use SPSS, R, or SAS to conduct a $z$-proportion test to determine whether there were more respondents who favored the plan than opposed it. Calculate the margin of error for this question.  Retest the hypothesis using a chi-square analysis.  Include relevant output from both analyses, but report your results from only the $z$-proportions test.  Be sure to include the relevant descriptive statistics as well as the test value, $p$-value, confidence limits and margin of error.  *(Note: if using your own data, you will need one binary variable for which you can ask whether one category has a higher proportion than the other).*

2. The data for this problem come from a 2014 report from the Portland police related to racial profiling in traffic stops (`race2.sav`).[2]  There are two variables in the data set, whether the offense a pedestrian was suspected of was a minor or major offense (`offense`; 0="minor", 1="major") [3] and the race of the pedestrian (`ped`; 0="White", 1="Black").  Using SPSS, R, or SAS, conduct a 2 × 2 chi-square analysis to check whether there is a significant difference among Black as compared to White pedestrians in whether they were suspected of 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. *(Note: if using your own data, you will need two binary variables for which you can ask whether one group has a higher proportion of an event than another).*

3. 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`). 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).  Use SPSS, R, or SAS to 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

---

[1] See summary of findings at: https://www.cnn.com/2021/03/10/politics/cnn-poll-biden-covid-relief-bill.html?utm_term=image&utm_medium=social&utm_source=twCNNp&utm_content=2021-03-10T11%3A56%3A03. See details of poll at: http://cdn.cnn.com/cnn/2021/images/03/10/biden.economic.relief.bill.poll.pdf. Note that rounding was used in generating these results.
[2] For more information, see the full report: https://www.portlandoregon.gov/police/article/481668.
[3] "Major" offenses include drug offenses, open alcohol container, and violent crimes. "Minor" offenses include jaywalking or other minor infractions.

the likelihood that abuse would occur in households with a boyfriend. *(Note: if using your own data, you will need three binary variables for which you can ask whether the relationship between two of the variables, X and Y, is the same in two levels of the third variable, Z.)*

4. Using the racial profiling report data described above. Use SPSS, R, or SAS to conduct a loglinear analysis of the 2 × 2 contingency table to determine whether there is a difference between Black and White pedestrians in whether they were stopped for a major vs. a minor offense. Report and interpret the appropriate percentages and statistical tests in your results. Briefly compare your results from the loglinear analysis to those obtained with the chi-square analysis. No more than 2-3 sentences are needed, but refer to specific values when making your comparisons. *(Note: if using your won data, please reanalyze the same data as used in the chi-square analysis above).*

5. Use the Early Head Start data set and SPSS, R, or SAS to conduct a loglinear analysis of to examine the three-way contingency table (ABUSE × BOYFRIEND × PROGRAM) to investigate whether the program (PROGRAM) the relationship between abuse and the boyfriend variable differed by program group. Describe whichever counts and percentages are 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. Briefly compare your results from your results from the loglinear analysis to those obtained with the chi-square analysis. No more than 2-3 sentences are needed, but refer to specific values when making your comparisons. *(Note: if using your won data, please reanalyze the same data as used in the chi-square analysis above).*

6. This problem uses data from a drug court survey of individuals arrested for drunk driving in New Mexico and contains two ordinal variables reported by the defendant, whether the defendants believe they are guilty of drinking while intoxicated (dwi, 0="not guilty", 1="somewhat guilty", 2="very guilty") and education level (ed; 1="less than high school degree", 2="high school diploma", 3="some or more college"). Use SPSS, R, or SAS to conduct an ordinal loglinear analysis to see whether there is a relationship between reported guilt level and education level. whether the defendants believe they are guilty of drinking while intoxicated. Obtain one of the measures of association for ordinal variables and report those results too. *(Note: if using your own data, you will need two ordinal variables for which you can ask whether the relationship between them.)*