Unit 10 Pre-Live

# Introduction

Take a look at Exercise 18.15 for the data (Statistical Sleuth), use the formulas from the video and text and just quickly do the following requests by hand so you can see the process. I know we will be using software in general, but with this section, it’s quite easy to do them by hand.

# Problem

**Perry Preschool Project.** Analyze the data shown in Exercise 4 to determine whether a preschool program like the one used can lead to lower levels of criminal activity, and, if so, by how much.

In a 1962 social experiment, 123 3- and 4-year-old children from poverty-level families in Ypsilanti, Michigan, were randomly assigned either to a treatment group receiving 2 years of preschool instruction or to a control group receiving no preschool. The participants were followed into their adult years. The following table shows how many in each group were arrested for some crime by the time they were 19 years old. (Data reported in *Time*, July 29, 1991).

**Arrested for some crime?**

|  |  |  |
| --- | --- | --- |
|  | Yes | No |
| Preschool | 19 | 42 |
| Control | 32 | 30 |

# Questions

1. What type of study is this? Retrospective, prospective, or completely observational?
   1. **Prospective**
2. Calculate the estimated difference in proportion of arrested between the two groups (preschool vs no preschool)
3. Calculate the estimated odds ratio (remember which one you put on top and bottom results in how you interpret the ratio)
   1. **The odds of being arrested is 0.4241 times lower for students who attended preschool as those who did not.**
4. Calculate 95% confidence intervals for the difference in proportions and the odds ratio. Do either one of them contain 0 or 1 respectively?

   2. **Neither confidence interval contains either 0 or 1.**
5. Optional: Calculate the corresponding z-statistic for each one of the testing procedures (one for difference in proportions and one for the odds ratio) Stop here, if you can figure out how to calculate the p-values using software or a z-table then go ahead, if not no worries, we will get into this with SAS/R later.

# Aside

If you ever wondered why a poll on the news returns a percentage with a MARGIN of error +/-3% when the total number of people in the poll is 1000. This comes directly for the confidence interval formula for a single proportion. If you plug in a proportion of .5 as an estimate, the 95% CI would yield .5+/- 1.96\*sqrt(.5\*.5/1000). The right-hand side part equals .03099 or roughly 3%. The .5 is a conservative estimate that makes the confidence interval as wide as possible. Fun statistical realization 