PSY653, Unit 11, In class activity – Explore Additive Interactions

Dataframe: asbestos.csv

In this activity you will continue the exploration of the Hilt and colleagues (1986) data. Now, we will consider two risk variables, smoking status (lifetime smoker vs. never smoker) and exposure to asbestos. We want to determine how these two variables come together to influence the likelihood of lung cancer.

**id**: participant’s id

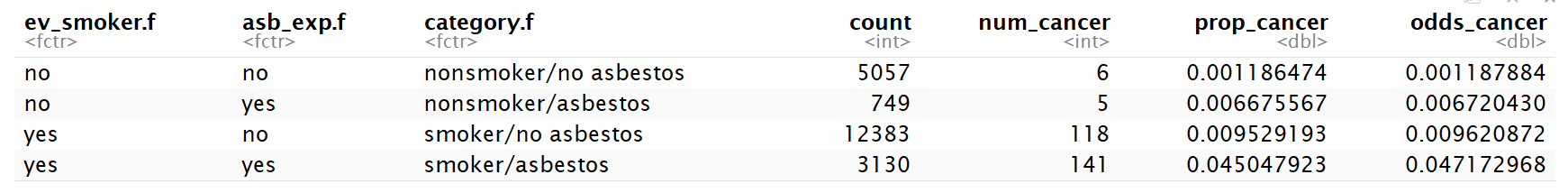
**smoking\_status**: smoking status of participant (1= current smoker, 2 = former smoker, 3 = never smoker)

**asbestos**: binary indicator of exposure to asbestos (0 = no exposure, 1 = exposure)

**lung\_cancer**: binary indicator of lung cancer status (0 = no lung cancer, 1 = lung cancer)

Expand your LungCancerNotebook.

1. Add the code chunks on page 28 of your handout. Execute the code chunks, and the table below will be created:



1. The column labeled prop\_cancer is the predicted probability of cancer for each category (based on the combination of lifetime smoking status and asbestos exposure) . Note how the probability differs across categories. Which group has the highest probability of cancer? What is your intuition about the way that these two risk variables come together to influence lung cancer?
2. The column labeled odds\_cancer is the odds of cancer for each category. Use these values to calculate the odds ratio (OR) to determine if asbestos exposure increases the odds of lung cancer among non-smokers. Now, calculate the OR to determine if asbestos exposure increases the odds of lung cancer among smokers.
3. Use the two ORs from Step 3 to compute a ratio of the ORs – put smokers in the numerator and non-smokers in the denominator.
4. Fit a logistic regression model, regress lung\_cancer on asbestos, ev\_smoker, and the interaction of the two. Make sure that you get the same ratio of the ORs (the OR for the interaction term) as you did in Step 4. Is the ratio of the ORs statistically significant?
5. Evaluate the ORs from Step 3 and the ratio of the ORs you just computed. How does this compare to what you concluded in Step 2?

