Lab 5 - Math 58b: relative risk & odds ratios

your name here

due Feb 25, 2020

Lab Goals

The focus of the lab is on understanding relative risk and odds ratios.

- computing sample (statistic) RR and OR
- finding CI for the population (parameter) RR and OR
- understanding how RR and OR are (mathematically) related
- understanding the disadvantages (and advantages!) of running a case-control study when trying to report how a probability has changed.

n.b. The problems discussed here which are inherent to RR are also inherent to the **difference** in proportions. Only OR is invariant to the choice of explanatory vs. response variable.

Getting started

Load packages

In this lab, generally just use R as a calculator.

The applet

We will spend some time working with the following applet: https://kenkleinman.shinyapps.io/odds-tool/

note1 don't worry about the concept of "bias" which is described on the applet. it is not something we will cover. (Although it is interesting to think about if you are curious!)

note2 the applet calls the "relative risk" the "risk ratio" – they are the same number.

note3 the applet switches which is p_1 and which is p_2 (from what we did in class). In the notation below I have been consistent with the applet (p_2 in the numerator, p_1 in the denominator).

1. Using the tab which says "start with the probabilities," find a set of (p_1 = baseline / denominator, p_2 = exposed / numerator) values where $p_1 \neq p_2$ that give an OR which is at least 5 times bigger than the RR. Report the (p_1,p_2) values, OR, and RR.

2. Using the tab which says "start with the probabilities," find a set of (p_1 = baseline / denominator, p_2 = exposed / numerator) values where $p_1 \neq p_2$ that give an OR which is very close to the RR **AND** for which RR and OR are not close to 1. Report the (p_1,p_2) values, OR, and RR.

3. Using the tab which says "start with the odds ratio" [Note: you can click on "include a bias marker" to see the values along the curve.]

   (a) Can you find a pair of (p_1, p_2) where the OR is smaller than the RR? Why or why not?
   (b) You can see the following in the graph, why? For a fixed OR, as p_1 grows: i. p_2 increases ii. RR decreases

To Turn In

4. Assuming that $p_1 \neq p_2$ and using the formula below & what you found above, describe/explain (in words) the settings for when the RR and OR are very similar (AND they are not equal to 1) and for when RR and OR are very different.

$$RR = p_2/p_1 \qquad OR = \frac{p_2/(1 - p_2)}{p_1/(1 - p_1)}$$

Consider the article (and data therein) on sleepy driving: (Connor et al. "Driver sleepiness and risk of serious injury to car occupants: population based case control study", British Medical Journal, 2002, https://www.bmj.com/content/bmj/324/7346/1125.1.full.pdf)

Although the researchers look at many variables, consider the two following two variables: (1) driver sleepiness score of 1-3 vs 4-7, (2) driver involved in an "injury crash" or not.

5. In the study on sleepy driving, which is the explanatory and which is the response variable?

6. How were the data selected, as a case-control study or a cohort study? What aspect of the population is it impossible to know given how the data were sampled?

7. Find and interpret a 90% CI for the true OR of the variables above. [Note: to complete this question you will need to be able to write out the words which describe the true OR.] Use R as a calculator. Note that there seems to be some missing information with respect to the sleepiness score.

8. Go to *The Journal of the American Medical Association* (https://jamanetwork.com/). Find an article that includes either a relative risk (RR) or an odds ratio (OR).

(a) Provide the article citation.
(b) Copy and paste the sentence that reports the RR or OR.
(c) Can you tell how the observational units were selected? What is case-control, cohort, something else? (Possibly it was an experiment!)
(d) Report/explain one thing from the article that you wouldn't have been able to understand prior to taking this class.