

Homework #4

Student name: Alex Towell (*atowell@siue.edu*)

Course: STAT 579 - Discrete Multivariate Analysis – Professor: Dr. Andrew Neath

Due date: Feb 20, 2022

Question 1

Consider data from a retrospective study on the relationship between daily alcohol consumption and the onset of esophagus cancer.

| | | cancer | no cancer |
|---------------------------|-------|--------|-----------|
| daily alcohol consumption | > 80g | 71 | 82 |
| | < 80g | 60 | 441 |
| | total | 131 | 523 |

- Compute an estimate of the odds ratio.
- Interpret the direction of the association, stated in the context of the problem.
- What assumption is necessary for the sample odds ratio to serve as an estimate of the relative risk?

Answer.

- The odds ratio is defined as

$$\theta := \frac{\Omega_1}{\Omega_2}$$

where Ω_j is the odds of outcome of column 1 given row i , which is defined as

$$\Omega_j := \frac{\pi_{1|i}}{1 - \pi_{1|i}}.$$

Observe that the odds ratio has the interesting property of being invariant to the type of study, whether cross-sectional, prospective, or retrospective. An estimator of θ is given by

$$\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}} = \frac{71 \times 441}{82 \times 60} \approx 6.36.$$

- Since $\hat{\theta} > 1$, we estimate that those consuming > 80g of alcohol per day are more likely to have the onset of esophagus cancer than those who consume less than 80g per day.
- For the odds ratio θ to serve as a reasonable estimator of the relative risk, $\pi_{1|2}$ and $\pi_{1|1}$ must be close to zero since

$$\theta = \text{RR} \times \left(\frac{1 - \pi_{1|2}}{1 - \pi_{1|1}} \right).$$

1. Question 2

The following table summarizes the responses of $n = 91$ couples to the questionnaire item “Sex is fun for me and my partner.”

- What type of sampling was used in collecting the above data? What characteristic do the above variables have that allows us the potential to describe the association with a single parameter? In general, how many parameters would be needed to describe the association?
- Compute $\hat{\gamma}$. Provide an interpretation of your result, stated in the context of the problem.

| husband's rating | wife's rating | | | |
|--------------------|--------------------|--------------|------------|---------------|
| | never/occasionally | fairly often | very often | almost always |
| never/occasionally | 7 | 7 | 2 | 3 |
| fairly often | 2 | 8 | 3 | 7 |
| very often | 1 | 5 | 4 | 9 |
| almost always | 2 | 8 | 9 | 14 |

Answer.

- (a) The data is a cross-sectional. The variables are *ordinal* and the expected relationship is *monotonic*. Ordinal variables have categories that may be placed in a natural *order* and therefore support relational predicates like less-than in addition to equality.

In general, $(4 - 1) \times (4 - 1) = 9$ parameters are necessary to describe all the possible associations.

- (b) The characteristic of interest, γ , is defined as

$$\gamma := \frac{\pi_C - \pi_D}{\pi_C + \pi_D},$$

which is a measure of correlation for ordinal variates analagous to Pearson's correlation coefficient.

Modifying the R code by populating the *counts* matrix with the data in the given table, when I run the R program I get the following output:

```

             wife's rating
husband's rating  never/occasionally  fairly often  very often  almost always
never/occasionally      7             7           2           3
fairly often            2             8           3           7
very often              1             5           4           9
almost always           2             8           9          14
[1] 1508 709
[1] 0.368254 0.173138
[1] 0.36

```

The number of concordances is $C = 1508$ and the number of discordances is $D = 709$. Estimators for π_C and π_D are respectively given by

$$\hat{\pi}_C = \frac{C}{\binom{n}{2}} = \frac{1508}{\binom{91}{2}} \approx .368$$

and

$$\hat{\pi}_D = \frac{D}{\binom{n}{2}} = \frac{709}{\binom{91}{2}} \approx .173.$$

An estimator for γ is given by

$$\hat{\gamma} = \frac{\hat{\pi}_C - \hat{\pi}_D}{\hat{\pi}_C + \hat{\pi}_D} = \frac{0.368 - 0.173}{0.368 + 0.173} \approx 0.36.$$

We say that this value of gamma represents a *medium effect*. Thus, we estimate that there is a medium size, positive association between the husband's and wife's rating of sex.