

Homework 7

Public Health 241: Statistical Analysis of Categorical Data

YOUR NAME / YOUR STUDENT ID HERE

TODAY'S DATE

1. The data set `diet.dta` on bCourses contains three variables: `act` measures physical activity (`act=0,1,2,3`), with higher values corresponding to higher activity levels; `diet` is an indicator variable for a low-fat diet (1 = low-fat diet, 0 = other diet); `mort` is an indicator variable for death by the end of the study (1= dead, 0 = alive). We are interested in studying the effect of low-fat diet on all-cause mortality, but are concerned that the relationship might be confounded by physical activity. The table below summarizes the available data. In questions (a)-(i) calculate by hand; in question (j) check your results in R and show your output.

	Mortality		
	Dead	Alive	Total
Activity level 0			
Low-fat diet	17	22	39
Other diet	75	75	150
Activity level 1			
Low-fat diet	28	36	64
Other diet	40	45	85
Activity level 2			
Low-fat diet	10	37	47
Other diet	14	34	48
Activity level 3			
Low-fat diet	4	24	28
Other diet	7	32	39

- (a) Set up a pooled 2×2 table and calculate a point estimate and confidence interval for the crude odds ratio for the risk of mortality comparing low-fat diet to other diets.

```
epitab(c(59, 119, 136, 186))
```

```
## $tab
##           Outcome
## Predictor Disease1      p0 Disease2      p1 oddsratio      lower
##   Exposed1      59 0.3025641      119 0.3901639 1.0000000      NA
##   Exposed2     136 0.6974359      186 0.6098361 0.6780771 0.462563
##           Outcome
## Predictor      upper    p.value
##   Exposed1      NA          NA
##   Exposed2 0.9940021 0.05535248
##
## $measure
## [1] "wald"
##
## $conf.level
## [1] 0.95
##
## $pvalue
## [1] "fisher.exact"
```

- (b) Draw a causal graph to reflect the relationship between low-fat diet, physical activity, and mortality. Based on your graph, is the crude odds ratio you calculated in (a) likely to be a good estimate of the causal odds ratio comparing low-fat diet to other diets?

- (c) For each of the four strata of physical activity, calculate a point estimate for the odds ratio comparing low-fat diet to other diets.

```
epitab(c(17, 22, 75, 75))$tab
```

```
## Outcome
## Predictor Disease1 p0 Disease2 p1 oddsratio lower
## Exposed1 17 0.1847826 22 0.2268041 1.0000000 NA
## Exposed2 75 0.8152174 75 0.7731959 0.7727273 0.3801964
## Outcome
## Predictor upper p.value
## Exposed1 NA NA
## Exposed2 1.570524 0.5899974
```

```
epitab(c(28, 36, 40, 45))$tab
```

```
## Outcome
## Predictor Disease1 p0 Disease2 p1 oddsratio lower
## Exposed1 28 0.4117647 36 0.4444444 1.000 NA
## Exposed2 40 0.5882353 45 0.5555556 0.875 0.4558074
## Outcome
## Predictor upper p.value
## Exposed1 NA NA
## Exposed2 1.679712 0.741056
```

```
epitab(c(10, 37, 14, 34))$tab
```

```
## Outcome
## Predictor Disease1 p0 Disease2 p1 oddsratio lower
## Exposed1 10 0.4166667 37 0.5211268 1.0000000 NA
## Exposed2 14 0.5833333 34 0.4788732 0.6563707 0.2575279
## Outcome
## Predictor upper p.value
## Exposed1 NA NA
## Exposed2 1.672916 0.4800262
```

```
epitab(c(4, 24, 7, 32))$tab
```

```
## Warning in chisq.test(xx, correct = correction): Chi-squared approximation
## may be incorrect
```

```
## Outcome
## Predictor Disease1 p0 Disease2 p1 oddsratio lower
## Exposed1 4 0.3636364 24 0.4285714 1.0000000 NA
## Exposed2 7 0.6363636 32 0.5714286 0.7619048 0.1999751
## Outcome
## Predictor upper p.value
## Exposed1 NA NA
## Exposed2 2.902856 0.7504788
```

- (d) Based on your results in (c), does it seem plausible that the effect of low-fat diet on mortality (as measured on the odds ratio scale) is the same in all four groups of physical activity?

- (e) Let's assume for the remainder of this question that the effect of low-fat diet on mortality is in fact the same at all four levels of physical activity. Carry out the Cochran-Mantel-Haenszel test to evaluate the null hypothesis that low-fat diet is not associated with mortality in any of the four strata of physical activity. What is the alternative hypothesis of this test? What is your conclusion?

```

mantelhaen.test(
array(c(17, 22, 75, 75,
      28, 36, 40, 45,
      10, 37, 14, 34,
      4, 24, 7, 32), dim = c(2, 2, 4)),
correct = FALSE)

##
## Mantel-Haenszel chi-squared test without continuity correction
##
## data:  array(c(17, 22, 75, 75, 28, 36, 40, 45, 10, 37, 14, 34, 4, 24,      7, 32), dim = c(2, 2, 4))
## Mantel-Haenszel X-squared = 1.3534, df = 1, p-value = 0.2447
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.5228808 1.1789039
## sample estimates:
## common odds ratio
##      0.7851281

```

- (f) Calculate an individual χ^2 -statistic for testing independence between low-fat diet and mortality in each stratum. Compare the sum of these four statistics against a χ^2 distribution with four degrees of freedom. What is the alternative hypothesis for the test that you just calculated a p-value for? Compare your p-value to the one you calculated in (e) and explain any difference you might see.

- (g) Calculate a Mantel-Haenszel point estimate for the summary odds ratio.

```

mantelhaen.test(
array(c(17, 22, 75, 75,
      28, 36, 40, 45,
      10, 37, 14, 34,
      4, 24, 7, 32), dim = c(2, 2, 4)),
correct = FALSE)

##
## Mantel-Haenszel chi-squared test without continuity correction
##
## data:  array(c(17, 22, 75, 75, 28, 36, 40, 45, 10, 37, 14, 34, 4, 24,      7, 32), dim = c(2, 2, 4))
## Mantel-Haenszel X-squared = 1.3534, df = 1, p-value = 0.2447
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.5228808 1.1789039
## sample estimates:
## common odds ratio
##      0.7851281

```

(h) Calculate a Woolf estimate and corresponding 95% confidence interval for the summary odds ratio.

(i) Compare your two adjusted estimates in (g) and (h) to the crude estimate in (a). Is the relationship between low-fat diet and mortality confounded?

- (j) Check your calculations for (a), (c), (e), and (g) in R and show your output. You will need to use `epitab()` and `mantelhaen.test()`.