## 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.

Table 1: Activity Level 0

Diet	Dead	Alive
Low-fat Diet	17	22
Other Diet	75	75

Table 2: Activity Level 1

Diet	Dead	Alive
Low-fat Diet	28	36
Other Diet	40	45

Table 3: Activity Level 2

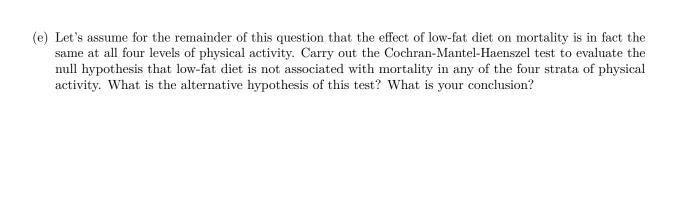
Diet	Dead	Alive
Low-fat Diet	10	37
Other Diet	14	34

Table 4: Activity Level 3

Diet	Dead	Alive
Low-fat Diet	4	24
Other Diet	7	32

(a) Set up a pooled  $2 \times 2$  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.

(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.
(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?



(f)	Calculate an individual $\chi^2$ -statistic for testing independence between low-fat diet and mortality in each stratum. Compare the sum of these four statistics against a $\chi^2$ 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.
(h)	Calculate a Woolf estimate and corresponding $95\%$ confidence interval for the summary odds ratio.

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

(j) Check your calculations for (a), (c), (e), and (g) in R and show your output. You will need to use the by() option for the cs or cc commands to do a stratified analysis.	,