Stat 525 Homework 5

Kenny Flagg

October 10, 2016

1. *The text defined multiplicative interaction in terms of RR on pages 148 and 149. Jewell points out on page 149 that a lack of multiplicative interaction based on OR implies*

$$\frac{p_{11}/(1-p_{11})}{p_{01}/(1-p_{01})} = \frac{p_{10}/(1-p_{10})}{p_{00}/(1-p_{00})}$$

Pencil through the algebra to show this, and comment on what it means.

By definition,

$$OR_{11} = \frac{p_{11}/(1-p_{11})}{p_{00}/(1-p_{00})}.$$

If we assume no multiplicative interaction,

$$\begin{aligned}
OR_{11} &= OR_{10} \times OR_{01} \\
\frac{p_{11}/(1-p_{11})}{p_{00}/(1-p_{00})} &= \frac{p_{10}/(1-p_{10})}{p_{00}/(1-p_{00})} \times \frac{p_{01}/(1-p_{01})}{p_{00}/(1-p_{00})} \\
\frac{p_{11}/(1-p_{11})}{p_{00}/(1-p_{00})} \times \frac{p_{00}/(1-p_{00})}{p_{01}/(1-p_{01})} &= \frac{p_{10}/(1-p_{10})}{p_{00}/(1-p_{00})} \\
\frac{p_{11}/(1-p_{11})}{p_{01}/(1-p_{01})} &= \frac{p_{10}/(1-p_{10})}{p_{00}/(1-p_{00})}
\end{aligned}$$

which means that the first exposure variable has the same association with the odds of disease regardless of the level of the second exposure variable.

2. *Researchers in Georgia collected data on death penalty sentencing in cases with black defendants. The data included the race of the victim (White or Black), and types of murder (Aggravation Level from 1 to 6). Aggravation level of 1 was unaggravated (e.g. barroom brawl) and 6 was most aggravated (e.g. vicious crimes involving torture). The data are shown below. I have modified the original data a bit to avoid a numerical issue (we may come back to that if we have time but the change does not impact the overall conclusion). Is there evidence that black murderers of white victims are more likely to receive the death penalty than black murderers of black victims after accounting for the aggravation level of the crime?*

```
# The data
death.CMH <- array(c(2, 1, 60, 181, 2, 1, 15, 21, 6, 2, 7, 9, 9, 2, 3, 4, 9, 4, 1, 3, 17, 4, 1, 1),
                   dim = c(2,2,6), dimnames = list(c("White","Black"), c("Death","NoDeath"),
                     c("AgLevel 1","AgLevel 2","AgLevel 3","AgLevel 4","AgLevel 5","AgLevel 6")))
```

(a) *Pool over Aggravation Level and estimate the the ratio of the odds of a white-victim murderer receiving the death penalty relative to the odds of a black victim murder receiving the death penalty along with an approximate 95% confidence interval.*

```
death.victim <- apply(death.CMH, c(1, 2), sum)
twoby2(death.victim)

2 by 2 table analysis:
------------------------------------------------------
Outcome   : Death
Comparing : White vs. Black

      Death NoDeath    P(Death) 95% conf. interval
White    45      87      0.3409    0.2652   0.4257
Black    14     219      0.0601    0.0359   0.0989

                                   95% conf. interval
             Relative Risk: 5.6737    3.2392   9.9378
         Sample Odds Ratio: 8.0911    4.2274  15.4864
Conditional MLE Odds Ratio: 8.0383    4.0876  16.7027
    Probability difference: 0.2808    0.1963   0.3685

             Exact P-value: 0
        Asymptotic P-value: 0
------------------------------------------------------
```

The odds of receiving the death penalty for a black murderer of a white victim are estimated to be 8.09 times larger than the odds of receiving the death penalty for a black murderer of a black victim, with a 95% confidence interval of 4.23 to 15.5.

(b) *Give an estimate of the Odds Ratio adjusted for Aggravation Level along with an approximate 95% confidence interval. Use the Cochran-Mantel-Haenszel method.*

```
WoolfCMH(death.CMH, alpha = 0.05)


Woolf's Method for Estimating OR, RR, and ER Across Strata
=========================================================
Woolf OR Chi-squared= 0.418 OR p-value = 0.995
Woolf OR = 3.881
Odds Ratio CI = 1.672 Odds Ratio CI = 9.009
Woolf RR Chi-squared= 3.569 RR p-value = 0.613
Woolf RR = 1.474
Relative Risk CI = 1.05 Relative Risk CI = 2.069
Woolf ER Chi-squared= 7.112 ER p-value = 0.212
Woolf ER = 0.0415
Excess Risk CI = -0.0012 Excess Risk CI = 0.0843
=========================================================
CMH Method for Estimating OR, RR, and ER Across Strata
=========================================================
CMH OR = 4.671
Odds Ratio CI = 1.798 Odds Ratio CI = 12.135
CMH RR = 1.836
Relative Risk CI = 1.223 Relative Risk CI = 2.757
CMH ER = 0.0973
Excess Risk CI = 0.0369 Excess Risk CI = 0.1576
```

The Cochran-Mantel-Haenzel estimate of the odds ratio, accounting for aggravation level, is 4.671 with a 95% confidence interval of 1.80 to 12.1.

(c) *One key assumption of the CMH method is homogeneity or no interaction between race and aggravation level, i.e. the aggravation level specific odds ratios are all equal to one another. Is this assumption reasonable here? Justify your answer. Use the Breslow-Day test to investigate.*

```
breslowday.test(death.CMH)

       AgLevel 1 AgLevel 2 AgLevel 3 AgLevel 4 AgLevel 5 AgLevel 6
log OR  1.797300  1.029619 1.3499267 1.7917595  1.909543  1.446919
Weight  1.049209  1.296672 0.8245267 0.9786154  1.203201  1.615591
       OR      Stat        df    pvalue
4.6707923 0.2863058 5.0000000 0.9978927
```

The Breslow-Day statistic is $\chi^2_5 = 0.2863$ with p-value $= 0.9979$, giving essentially no evidence that the odds ratio differs by aggravation level. It is reasonable to assume homogeneity and use the Cochran-Mantel-Haenszel method.

(d) *Based on these results which Odds Ratio would you report (1) the pooled odds ratio, (2) the Odds Ratio adjusted for Aggravation Level, or (3) the stratum level Odds Ratios? Justify your answer.*

```
death.agg <- t(apply(death.CMH, 2:3, sum))
print(death.agg)
```

```
           Death NoDeath
AgLevel 1      3     241
AgLevel 2      3      36
AgLevel 3      8      16
AgLevel 4     11       7
AgLevel 5     13       4
AgLevel 6     21       2
```

```
chisq.test(death.agg, simulate.p.value = TRUE)
```

```

Pearson's Chi-squared test with simulated p-value (based on 2000
replicates)

data:  death.agg
X-squared = 215.73, df = NA, p-value = 0.0004998
```

We should report the adjusted odds ratio. There is no evidence of an interaction, so we do not need to report separate odds ratios. There is evidence of an association between aggravation level and whether the defendent received the death penalty, so the adjusted odds ratio will have lower variance than the pooled odds ratio.

3. *The table below contains data from a cohort study of smoking and lung cancer. Asbestos exposure was another risk factor of interest. The numbers in the table are deaths from lung cancer per 100000 people, i.e. they are P_{ij} values times 100000. You can work on this scale to answer the questions below.*

	Asbestos No	Asbestos Yes
Cig No	11.3	58.4
Cig Yes	112.6	601.6

(a) *What would you expect the value to be in the lower right cell if the relation between asbestos and smoking were additive?*

The expected number of deaths per 100,000 would be $122.6 + (58.4 - 11.3) = 169.7$.

(b) *What would you expect the value to be in the lower right cell if the relation between asbestos and smoking were multiplicative?*

The expected number of deaths per 100,000 would be $122.6 \times \dfrac{58.4}{11.3} = 633.6$.

(c) *Is the observed value more indicative of an additive interaction or a multiplicative interaction? Is the interaction synergistic or antagonistic and why?*

This is indicative of a multiplicative interaction because the observed value of 601.6 is closer to the expected value under multiplicative effects than to the expected value under additive effects. If we assume effects are additive, the observed asbestos effect for smokers (489.0) is about ten times as large as the observed asbestos effect for nonsmokers (47.1) which seems implausible.

As a multiplicative effect, the interaction is antagonistic because the observed value is smaller than the expected value.

4. *Problem 11.2 on page 177. The way he sets it up is confusing.*

(a) *Ignore the exposure variable and just use the cases and controls. The R code below sets up the data for you.*

```
cases <- c(59, 54, 53, 61, 61)
n <- c(192, 204, 173, 181, 118)
```

Do what he asks in the second paragraph.

```
prop.test(cases, n)


5-sample test for equality of proportions without continuity
correction

data:  cases out of n
X-squared = 23.435, df = 4, p-value = 0.0001037
alternative hypothesis: two.sided
sample estimates:
   prop 1    prop 2    prop 3    prop 4    prop 5
0.3072917 0.2647059 0.3063584 0.3370166 0.5169492
```

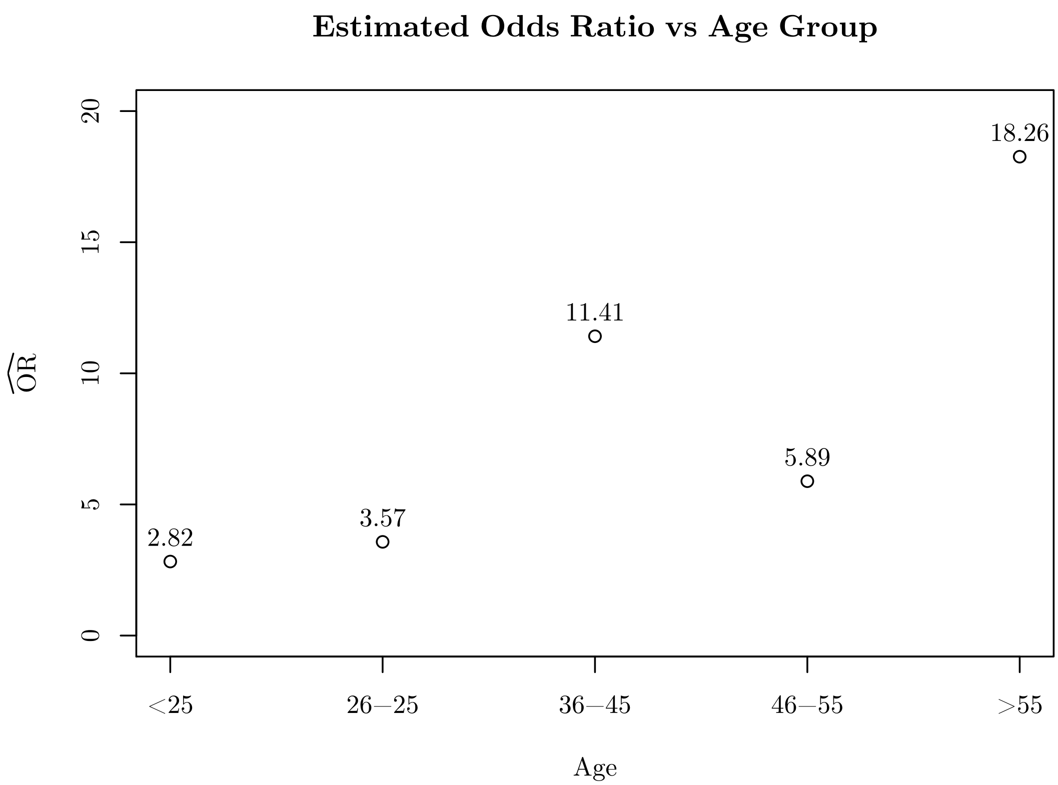
For the overall test of association, we have very strong evidence ($\chi^2_4 = 23.43$, p-value $=$ 0.0001) that tuberculosis prevalence differs among the age groups.

```
prop.trend.test(cases, n)


Chi-squared Test for Trend in Proportions

data:  cases out of n ,
 using scores: 1 2 3 4 5
X-squared = 13.005, df = 1, p-value = 0.0003106
```

There is very strong evidence ($\chi^2_1 = 13.01$, p-value $= 0.0003$) that tuberculosis prevalence has a linear relationship with age.

(b) *Now using all the data in the table do what he asks in the third paragraph. You can just calculate age specific Odds Ratios using the ratio of cross products. Note that he is asking for a qualitative assessment only. You will just be looking at a simple graph and/or table.*

```
prob.11.2 <- data.frame(age = c('<25', '26-25', '36-45', '46-55', '>55'),
                        a = c(8, 7, 12, 8, 15),
                        b = c(7, 6, 3, 3, 1),
                        c = c(51, 47, 41, 53, 46),
                        d = c(126, 144, 117, 117, 56))
prob.11.2$OR.hat <- with(prob.11.2, (a * d) / (b * c))
plot(prob.11.2$OR.hat, main = 'Estimated Odds Ratio vs Age Group',
     ylab = expression(widehat(OR)), xlab = 'Age', xaxt = 'n', ylim = c(0, 20))
axis(1, at = 1:5, labels = prob.11.2$age)
text(prob.11.2$OR.hat ~ I(1:5), labels = round(prob.11.2$OR.hat, 2), pos = 3)
```



There is a clear trend of higher estimated odds ratios with increasing age.