Biostatistics 513

Homework 2 Solutions

1.

a) The four counterfactuals for a given woman are the coverage outcomes when she receives the following combination of treatments:

Antenatal	Labor and Delivery
C	C
C	T
T	C
T	T

where T = targeted therapy and C = combined therapy

Using the risk difference as a measure of effect, I will define the treatment effect as RD = P(Y = 1 | C,C) – P(Y = 1 | T,T). This is the difference in coverage if the standard of care is combined therapy versus if the standard of care is targeted therapy. Note that the outcomes (C, T) and (T, C) are a spurious feature of the trial design. These groups would not exist if one treatment or the other is the standard of care over all of Lusaka.

b) The approach I am proposing is to compare the coverage between the group of women who receive combined therapy both antenatally and at LD versus the coverage in women who receive targeted therapy at both their antenatal visit and at LD. Using this approach, women who receive a combination of treatments are not included in the (primary) analysis.

c) The really important assumption here is that women are not choosing their site of antenatal and/or LD care based on the treatments offered. If a woman is able to self-select her treatment then we no longer have a randomized trial. At first glance this may seem like a dubious assumption. However, it is not easy for women to attend a clinic that isn't close to where they live. Also, we have some preliminary data from a previous study that suggests that such migration is not common.

d) It is quite likely that at least some of the variation in coverage might be due to the quality of care received in a given antenatal or LD clinic (for instance, how good was the counseling about the importance of taking the NVP at the onset of labor?). For this reason, responses from women who attend the same antenatal clinic or the same LD clinic are likely to be correlated. The assumption of independence is likely to be violated.

2. a)

```
. cs yn txn [freq=count], or

                 | txn                   |
                 |   Exposed   Unexposed  |     Total
-----------------+-----------------------+----------
           Cases |        55          47  |       102
        Noncases |        75          96  |       171
-----------------+-----------------------+----------
           Total |       130         143  |       273
                 |                       |
            Risk |  .4230769    .3286713  |  .3736264
                 |                       |
                 |      Point estimate    |  [95% Conf. Interval]
                 |------------------------+----------------------
```

```
 Risk difference |         .0944056       | -.0202237    .2090349
      Risk ratio |         1.287234       |  .9455465    1.752396
 Attr. frac. ex. |         .2231405       | -.0575895    .4293526
 Attr. frac. pop |         .1203209       |
      Odds ratio |         1.497872       |   .916455    2.448146  (Cornfield)
                 +-------------------------------------------------
                             chi2(1) =     2.59  Pr>chi2 = 0.1073
```

From the above, we see that the proportion of favorable responses in the drug group is about 42% versus 33% in the control group. The odds ratio is 1.5 indicating that individuals in the drug group are at a higher "risk" for having a favorable response. However, this increase in risk may be due to chance variations (p = .11) and we cannot conclude that the probability of a favorable response in the drug group is greater than the probability of a favorable response in the control group.

b)
. table txn [freq=count], by(clinic) row c(mean yn) f(%5.2f)

clinic	P(favorable)			
	drug	control	RR	OR
1	0.31	0.27	1.13	1.19
2	0.80	0.69	1.16	1.82
3	0.74	0.37	2.00	4.80
4	0.13	0.06	2.12	2.28
5	0.35	0.00	--	--
6	0.09	0.00	--	--
7	0.20	0.11	1.80	2.00
8	0.67	0.86	0.78	.33

We see that the drug treatment has a higher favorable response rate at 7 of the 8 clinics, suggesting that the response is fairly consistent across clinics. We also see that the probability of a favorable response varies substantially across clinics.

c)
The hypothesis is

Ho: common OR = 1
Ha: common OR ≠ 1

```
Mantel-Haenszel estimate controlling for clinic
------------------------------------------------------------------
 Odds Ratio    chi2(1)        P>chi2        [95% Conf. Interval]
------------------------------------------------------------------
   2.134549       6.38        0.0115         1.168685   3.898656
------------------------------------------------------------------
```

The estimate of the common OR is 2.1. Based on the MH chi-square test, this value is significantly different from 1 (p = .01). We conclude that the probability of a favorable response is different between the drug and placebo groups.

d) The hypothesis is

Ho: $OR_1 = OR_2 = \ldots = OR_8$

Ha: at least one not equal

```
Test of homogeneity (B-D)      chi2(7) =     8.00  Pr>chi2 = 0.3330
```

The MH homogeneity test is not significant (p = .33). There is not strong evidence to suggest that the OR relating response to treatment varies significantly across clinics. Note that this is NOT a statement about whether the response rate varies across clinic. It is a statement about the OR.

e) From the output in (c) we see that the estimate of the common OR is 2.1 with a 95% CI of 1.2 – 3.9. In part (a) we found an OR of 1.5, while the OR adjusting for clinic is 2.1. Not only is the adjusted estimate larger, but it is now statistically significantly different from 1.0.

f) From 2b we see that clinic is clearly related to the outcome … the overall rate of favorable responses varies significantly across clinics in both the case and control groups. Is clinic related to the "exposure" (drug treatment in this case)? Here's a table:

clinic	P(drug)
1	.49
2	.38
3	.50
4	.48
5	.59
6	.52
7	.36
6	.46

It makes sense that the outcome varies across clinic. It is less clear why there should be variation in the treatment assignment rates across clinic. Typically, when we do a randomized trial we would randomize <u>within</u> clinic to avoid this problem. Most of the clinics seem reasonably balanced … the exceptions are clinics 2 and 7. So there is justification for treating clinic as a confounder. It also makes sense to control for clinic effects a priori. Given all this, together with the fact that our adjusted OR has changed noticeably, it appears as though clinic is confounding the relationship between treatment and response.

g) Overall we see a 33% favorable response for control subjects compared to a 42% favorable response for treatment subjects. However, these rates vary substantially across clinics and we would certainly modify these numbers if we knew which clinic a particular subject was going to attend.

We obtain an OR of 2.13 after stratification meaning that within a given clinic we estimate that the odds of a favorable outcome among treated subjects are about twice the odds for a control subject. We conclude that the drug is effective (p = .01)

3) In order to assess the association between iron (dichotomized at >350mg) we will estimate the odds ratio comparing the odds of disease among subjects with high iron (>350mg) to the odds of disease among subjects with low iron consumption (<=350mg). We will also compute the odds ratio after adjusting for gender and age, as these may be potential confounders.

Univariate summaries: The data contain 570 controls and 338 subjects with CHD. Overall, there are 270/908 = 30% of subjects with iron consumption greater than 350mg/month. Also, 216 (or

24%) of subjects are under 50 years of age, while 291 (32%) are aged 50-59, 302 (33%) are aged 60-69, and 99 subjects (11%) are aged 70 or older. Females comprise only 32% of the subjects (291 women and 617 men).

Bivariate summaries: First we consider associations between disease status and covariates. We find that 35% of the cases have high iron consumption while only 27% of the controls have high iron consumption. Also, the cases tend to be a little older with 48% aged 60 or older as compared to 41% of controls. A small fraction of the cases are women (only 13%) while women comprise 43% of the controls. Thus, we see a large difference between the cases and controls in terms of gender, and a modest difference in terms of iron consumption, and a small difference in ages.

Next we consider associations between the predictor of interest, iron consumption, and the covariates age and gender. The age distribution of subjects that consume low and high levels of iron are not too different with 56% of subjects that consume low levels of iron over 60 compared to 50% of subjects that consume high levels of iron. However, we find that only 16% of subjects with high iron are female as compared to 39% of subjects with low iron. Thus, iron consumption appears to vary greatly with gender, yet appears not to be association with age.

Confirmatory Analysis: The crude odds ratio comparing the odds of CHD among subjects with high iron consumption to the odds of CHD among subjects with low iron consumption is estimated as 1.475 (95% confidence interval: 1.09, 1.99). However, after adjusting for gender we obtain an adjusted odds ratio of 1.07 (95% confidence interval: 0.79, 1.45). If we adjust for both gender and age we obtain an adjusted odds ratio estimate of 1.10 with 95% confidence interval (0.81, 1.51). Thus, although we find a significant crude association between disease and high iron consumption, this association is greatly diminished and is not statistically significant after controlling for gender. As we found in the bivariate analysis above, gender is strongly associated with both the exposure of interest and with CHD, and thus an analysis such as the crude odds ratio that does not control for gender would lead to a biased assessment of the impact of high iron consumption (ie. a crude comparison of NewIron=1 to NewIron=0 is comparing one group with 16% women (NewIron=1) to another group with 39% women (NewIron=0) and thus blurs the effects of gender and iron).

```
*** adjusting for female only

. mhodds case newiron female

Mantel-Haenszel estimate of the odds ratio
Comparing newiron==1 vs. newiron==0, controlling for female

     ------------------------------------------------------------------
      Odds Ratio    chi2(1)        P>chi2        [95% Conf. Interval]
     ------------------------------------------------------------------
        1.073283       0.21        0.6475         0.792519   1.453514
     ------------------------------------------------------------------

*** adjusting for female and age (8 strata)
. mhodds case newiron female age

Mantel-Haenszel estimate of the odds ratio
Comparing newiron==1 vs. newiron==0, controlling for female and age

     ------------------------------------------------------------------
      Odds Ratio    chi2(1)        P>chi2        [95% Conf. Interval]
     ------------------------------------------------------------------
        1.104245       0.39        0.5316         0.809221   1.506829
     ------------------------------------------------------------------
```