Biostatistics 513

Homework 4 Solutions

1)

a) OR = exp(-.60 - 1.5(1))/exp(-.60 – 1.5(2)) = exp(1.5) = 4.48, so the exposure increases the risk of disease

b) For 0/1 coding we have
β_0 = log(odds for unexposed)
β_1 = log(OR)

so

β_0 = log(odds for unexposed) = log(exp(-.60 – 1.5(2))) = log(exp(-3.6)) = -3.6
β_1 = log(OR) = log(4.48) = 1.5

2)

a) Let $\pi(X)$ = Pr(CHD | X). Then the logistic regression model considering NEWIRON only is

$$\text{logit } \pi(X) = \beta_0 + \beta_1 \text{ NEWIRON}$$

b) The coefficient of NEWIRON in this model, β_1, may be interpreted as the difference between the log of the odds of CHD among individuals with high (> 350 mg/month) levels of iron intake and the log of the odds of CHD among individuals with low (≤ 350 mg/month) levels of iron intake. Since we presume that CHD is a relatively rare condition, this difference should approximately the log of the relative risk.

c) Here's my stata output

```
. logistic case newiron

Logistic regression                                     Number of obs   =        908
                                                        LR chi2(1)      =       6.83
                                                        Prob > chi2     =     0.0090
Log likelihood = -595.99383                             Pseudo R2       =     0.0057

------------------------------------------------------------------------------
        case | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     newiron |      1.475   .2187347     2.62   0.009     1.102969    1.972517
------------------------------------------------------------------------------
```

As we found in homework 2, the unadjusted odds ratio, 1.475, is significantly different from 1 (P = .009)

d) The regression model is

$$\text{logit } \pi(X) = \beta_0 + \beta_1 \text{ NEWIRON} + \beta_2 \text{ AGE(2)} + \beta_3 \text{ AGE(3)} + \beta_4 \text{ AGE(4)} + \beta_5 \text{ FEMALE}$$

e) The coefficient of NEWIRON in this model, β_1, may be interpreted as the difference between the log of the odds of CHD among individuals with high (> 350 mg/month) levels of iron intake and the log of the odds of CHD among individuals with low (≤ 350 mg/month) levels of iron intake, after adjusting for (or controlling for, or holding constant) age and gender. (could also state in terms of the log of the odds ratio). We suspect it will be particularly important to control gender since in homework 2 we found that gender is predictive of CHD and that the proportion of females is quite different in the high and low iron groups.

f) The coefficient of AGE(2) in this model, β_2, may be interpreted as the difference between the log of the odds of CHD among individuals aged 50 – 59 years and the log of the odds of CHD among individuals aged ≤49 years (the reference age group), after adjusting for (or controlling for, or holding constant) iron intake level (within the groups >350mg/day and ≤350 mg/day) and gender. (could also state in terms of the log of the odds ratio).

g) The coefficient of AGE(4) in this model, β_4, may be interpreted as the difference between the log of the odds of CHD among individuals aged $\geq$70 years and the log of the odds of CHD among individuals aged $\leq$49 years (the reference age group), after adjusting for (or controlling for, or holding constant) iron intake level (within the groups >350mg/day and $\leq$350 mg/day) and gender. (could also state in terms of the log of the odds ratio).

h) The odds for an individual with (NEWIRON=0, FEMALE=0, AGE=4) is $\exp(\beta_0 + \beta_4)$. The odds for an individual with (NEWIRON=0, FEMALE=0, AGE=2) is $\exp(\beta_0 + \beta_2)$. Thus the odds ratio for comparing these two individuals (or groups) would be (AGE = 4 over AGE = 2) equal to $\exp(\beta_4 - \beta_2)$.

i) Here's my stata output

```
. xi:logit case female i.age
i.age             _Iage_1-4           (naturally coded; _Iage_1 omitted)

Logit estimates                                   Number of obs   =        908
                                                  LR chi2(4)      =     120.26
                                                  Prob > chi2     =     0.0000
Log likelihood = -539.27623                       Pseudo R2       =     0.1003

------------------------------------------------------------------------------
        case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -1.632189   .1838605    -8.88   0.000    -1.992549   -1.271829
     _Iage_2 |   .6817876   .2027506     3.36   0.001     .2844036    1.079172
     _Iage_3 |   .8536424   .1990066     4.29   0.000     .4635965    1.243688
     _Iage_4 |   -.022819   .2911035    -0.08   0.938    -.5933713    .5477333
       _cons |  -.6085052   .1579415    -3.85   0.000    -.9180648   -.2989455
------------------------------------------------------------------------------

. estimates store model0

. xi:logit case female i.age newiron
i.age             _Iage_1-4           (naturally coded; _Iage_1 omitted)

Logit estimates                                   Number of obs   =        908
                                                  LR chi2(5)      =     120.58
                                                  Prob > chi2     =     0.0000
Log likelihood = -539.11767                       Pseudo R2       =     0.1006

------------------------------------------------------------------------------
        case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |  -1.612586   .1870517    -8.62   0.000    -1.979201   -1.245972
     _Iage_2 |   .6850382   .2028911     3.38   0.001     .2873789    1.082698
     _Iage_3 |   .8589874   .1993287     4.31   0.000     .4683103    1.249664
     _Iage_4 |  -.0143397   .2916048    -0.05   0.961    -.5858747    .5571953
     newiron |   .0898209   .1593785     0.56   0.573    -.2225551     .402197
       _cons |  -.6455109   .1712649    -3.77   0.000    -.9811841   -.3098378
------------------------------------------------------------------------------

. estimates store model1

. lrtest model0 model1

likelihood-ratio test                                 LR chi2(1)  =      0.32
(Assumption: model0 nested in model1)                 Prob > chi2 =    0.5733
```

The null hypothesis is Ho: $\beta_1 = 0$. The likelihood ratio chi-square test (with one degree of freedom) is equal to .32. I conclude that, after adjusting for age and gender, there is no evidence of a significant association between NEWIRON and CHD (p = .57). I also note that this is identical to the conclusion reached by the Wald test.

j) In the logistic regression analysis the adjusted OR for newiron is exp(.0898) = 1.09 (95% CI 0.80 – 1.49). In homework 2 the adjusted OR from the MH analysis (prob 3) was 1.10 (95% CI 0.81 – 1.51). These two estimates have the same interpretation. They represent the odds ratio for CHD between high and low iron groups, controlling for gender and age.

k) With AGE (and FEMALE) in the model the estimated logodds for NEWIRON is 0.090 (s.e. .159). Without AGE in the model the estimated logodds for NEWIRON is 0.072 (s.e. .157). Thus, inclusion of age causes no change in our interpretation of the the logodds for NEWIRON. You can also do a likelihood ratio test to see if AGE is a significant effect modifier of NEWIRON. The log-likelihood with AGE, NEWIRON and FEMALE in the model is –539.1. If you add AGE*NEWIRON interactions the log-likelihood increases to –537.5. The resulting chi-square statistic is 3.13 with 3 df. The p-value is .37 so there is no evidence of significant effect modelification of AGE on NEWIRON. Thus, it is not particularly important to include AGE in the model for accessing the effect of NEWIRON. It is worth mentioning, however, that AGE is significantly related to risk of CHD. You can do a likelihood ratio test, comparing the model with NEWIRON, FEMALE and AGE in it to the model with NEWIRON and FEMALE only. The resulting value of the likelihood ratio chi-square is 25.99 (3 degrees of freedom. The hypothesis being tested is Ho: $\beta_2 = \beta_3 = \beta_4 = 0$. One would reject this null hypothesis ($p < .001$) and conclude that AGE is an important predictor of the log-odds of CHD.

l) It is very important to include FEMALE in the model, however. Without FEMALE in the model the estimated coefficient for NEWIRON is 0.401 (95% CI 0.106 – 0.697). With FEMALE in the model the coefficient is 0.090 (95% CI -0.222 – 0.402). Thus, our entire interpretation of the effect of NEWIRON changes depending on whether FEMALE is included in the model. FEMALE is a confounder for NEWIRON and it must be included in the model.

m) A model that allows the effect of NEWIRON to vary across gender is

$$\text{logit}\,\pi(X) = \beta_0 + \beta_1\,\text{NEWIRON} + \beta_2\,\text{FEMALE} + \beta_3\,\text{NEWIRON*FEMALE}$$

(we argued in part l that AGE need not be included, though it is okay if you do). The hypothesis of interest is Ho: $\beta_3 = 0$. Here is the stata output:

```
. xi:logit case i.newiron*i.female
i.newiron         _Inewiron_0-1       (naturally coded; _Inewiron_0 omitted)
i.female          _Ifemale_0-1        (naturally coded; _Ifemale_0 omitted)
i.new~n*i.fem~e   _InewXfem_#_#       (coded as above)

Logit estimates                                   Number of obs   =        908
                                                  LR chi2(3)      =      95.80
                                                  Prob > chi2     =     0.0000
Log likelihood = -551.50939                       Pseudo R2       =     0.0799

------------------------------------------------------------------------------
        case |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 _Inewiron_1 |   .0070404   .1668664     0.04   0.966    -.3200117    .3340925
  _Ifemale_1 |  -1.679273   .2068512    -8.12   0.000    -2.084694   -1.273852
_InewXfem_~1 |   .5069054   .4501735     1.13   0.260    -.3754185    1.389229
       _cons |  -.1031842   .1016698    -1.01   0.310    -.3024533    .0960849
------------------------------------------------------------------------------
```

Based on the Wald test, the interaction between female and newiron is not significant ($p = .260$) (you could also do a likelihood ratio test and come to the same conclusion). We conclude that the effect of NEWIRON on CHD is not significantly different between men and women.