Homework #6

Dustin Leatherman

5/15/2019

12

Reconsider the data in Display 18.2 from the randomized experiment on 818 volunteers. Let Yi represent the number of individuals with colds in group i , out of mi . Let Xi be an indicator variable that takes on the value 1 for the placebo group and 0 for the vitamin C group. Then the observed values for the two groups are:

a

Using a computer package, obtain an estimate of and a confidence interval for beta_1.

It is estimated that $\beta_1 = 0.4269$. 95% C.I. for $\beta_1 = [0.0948, 0.7628]$

```
cold <- data.frame(Group=c(1,2), Colds=c(335,302), Size=c(411,407), isPlacebo=c(1,0))
lrm <- glm(Colds/Size ~ isPlacebo, weights = Size, family = "binomial", data = cold)
summary(lrm)
```

```
##
## Call:
## glm(formula = Colds/Size ~ isPlacebo, family = "binomial", data = cold,
##     weights = Size)
##
## Deviance Residuals:
## [1]  0  0
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.0565     0.1133   9.325   <2e-16 ***
## isPlacebo     0.4269     0.1702   2.508   0.0121 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6.3573  on 1  degrees of freedom
## Residual deviance: 0.0000  on 0  degrees of freedom
## AIC: 16.162
##
## Number of Fisher Scoring iterations: 3
```

```
confint(lrm)
```

```
## Waiting for profiling to be done...
```

```
##                  2.5 %    97.5 %
## (Intercept) 0.83816004 1.2827497
## isPlacebo   0.09477341 0.7628385
```

b

Interpret the results in terms of the odds of cold for the placebo group relative to the odds of a cold for the vitamin C group.

```
summary(lrm)$coefficients %>% exp
```

```
##             Estimate Std. Error  z value Pr(>|z|)
## (Intercept) 2.876190   1.119959 11216.81 1.000000
## isPlacebo   1.532546   1.185577    12.28 1.012217
```

The odds of catching a cold in the placebo group is 1.5325 times the odds of catching a cold in the Vitamin C group.

c

How does the answer to part (b) compare to the answer in Display 18.9?

It is the same.

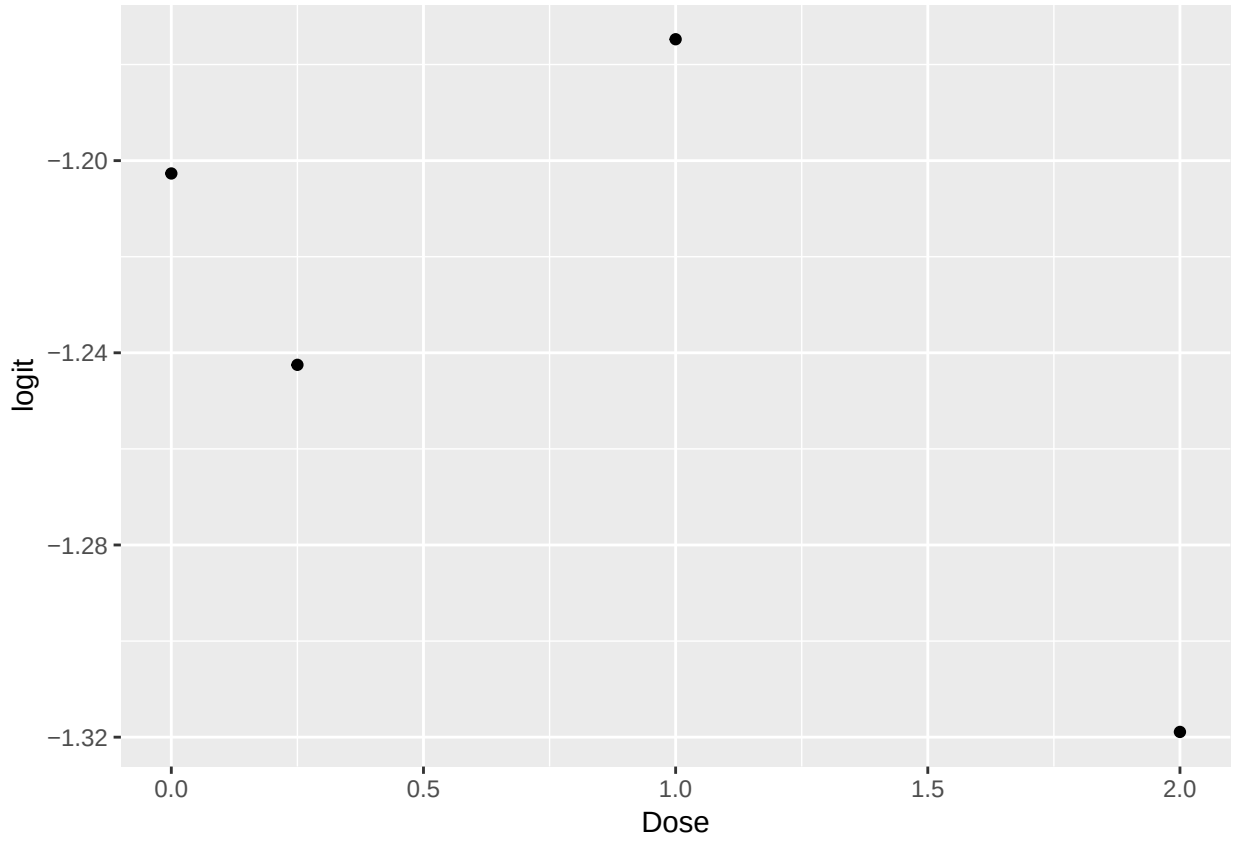
13

Between December 1972 and February 1973, a large number of volunteers participated in a randomized experiment to assess the effect of large doses of vitamin C on the incidence of colds. (Data from T. W. Anderson, G. Suranyi, and G. H. Beaton, "The Effect on Winter Illness of Large Doses of Vitamin C," Canadian Medical Association Journal 111 (1974): 31–36.) The subjects were given tablets to take daily, but neither the subjects nor the doctors who evaluated them were aware of the dose of vitamin C contained in the tablets. Shown in Display 21.18 are the proportion of subjects in each of the four dose categories who did not report any illnesses during the study period.

a

For each of the four dose groups, calculate the logit of the estimated proportion. Plot the logit versus the dose of vitamin C.

```
ex2113 %>%
  mutate(logit = log(ProportionWithout / (1 - ProportionWithout))) %>%
  qplot(Dose, logit, data = .)
```



b

Fit the logistic regression model, $logit(\pi) = \beta_0 + \beta_1 dose$. Report the estimated coefficients and their standard errors. Report the p-value from the deviance goodness-of-fit test. Report the p-value for a Wald's test that β_1 is 0. Report the p-value for a drop-in-deviance test for the hypothesis that β_1 is 0.

```
doses <- glm(WithoutIllness/Number ~ Dose, family = "binomial", weights = Number, data = ex2113)
summary(doses)
```

```
##
## Call:
## glm(formula = WithoutIllness/Number ~ Dose, family = "binomial",
##     data = ex2113, weights = Number)
##
## Deviance Residuals:
##        1         2         3         4
## -0.06857  -0.27405   0.57021  -0.35303
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.20031    0.06167 -19.464   <2e-16 ***
## Dose        -0.03465    0.07113  -0.487    0.626
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 0.76803  on 3  degrees of freedom
## Residual deviance: 0.52957  on 2  degrees of freedom
## AIC: 29.801
##
## Number of Fisher Scoring iterations: 3
```

```
dev_gof <- 1 - pchisq(doses$deviance, doses$df.residual)
dev_null_gof <- 1 - pchisq(doses$null.deviance, doses$df.null)
wald_gof <- tidy(doses)$p.value[2]

dev_gof
```

```
## [1] 0.7673694
```

```
wald_gof
```

```
## [1] 0.626169
```

```
dev_null_gof
```

```
## [1] 0.8570984
```

Deviance Goodness-of-fit Test p-value: 0.7674 Wald Goodness-of-fit Test p-value: 0.6262 Drop-in-Deviance Test for H_0 p-value: 0.8571

c

What can be concluded about the adequacy of the binomial logistic regression model? What evidence is there that the odds of a cold are associated wih the dose of vitamin C?

There is no evidence suggesting that the odds of a cold is associated with a dosage of Vitamin C. This binomial logistic regression model is not sufficient in detailing the relationship between cold and Vitamin C Dose.

16

An experiment at the Marine/Freshwater Biomedical Sciences Center at Oregon State University investigated the carcinogenic effects of aflatoxicol, a metabolite of Aflatoxin B1, which is a toxic by-product produced by a mold that infects cottonseed meal, peanuts, and grains. Twenty tanks of rainbow trout embryos were exposed to one of five doses of Aflatoxicol for one hour. The data in Display 21.20 (from George Bailey and Jerry Hendricks) represent the numbers of fish in each tank and the numbers of these that had liver tumors after one year.

Describe the relationship between dose of Aflatoxicol and odds of liver tumor. It is also of interest to determine the dose at which 50% of the fish will get liver tumors. (Note: Tank effects are to be expected, meaning that tanks given the same dose may have slightly different π's. Thus, one should suspect extra-binomial variation.)

```r
tumor <- glm(Tumor/Total ~ Dose, family = "quasibinomial", weights = Total, data = ex2116)
summary(tumor)
```

```
##
## Call:
## glm(formula = Tumor/Total ~ Dose, family = "quasibinomial", data = ex2116,
##     weights = Total)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -7.3577  -4.0473  -0.1515   2.9109   4.7729
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -0.8670     0.2933  -2.956 0.008461 **
## Dose         14.3338     3.5819   4.002 0.000837 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 14.6147)
##
##     Null deviance: 667.20  on 19  degrees of freedom
## Residual deviance: 277.05  on 18  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

There is convincing evidence that the odds having one or more liver tumors is associated with a dosage of Aflatoxicol (Two-tailed Wald Logistic Regression Test. p-value = 0.0008).

$\hat{y} = -0.867 + 14.3338 \times Dose$

$0.50 = -0.867 + 14.3338 \times Dose \rightarrow Dose = \frac{0.50+8.67}{14.3338} = 0.0954$

17

The probability of a male birth in humans is about 0.51. It has previously been noticed that lower proportions of male births are observed when offspring are conceived at times of exposure to smog, floods, or earthquakes. Danish researchers hypothesized that sources of stress associated with severe life events may also have some bearing on the sex ratio. To investigate this theory they obtained the sexes of all 3,072 children who were born in Denmark between January 1, 1980 and December 31, 1992 to women who experienced the following kinds of severe life events in the year of the birth or the year prior to the birth: death or admission to hospital for cancer or heart attack of their partner or of their other children. They also obtained sexes on a sample of 20,337 births for mothers who did not experience these life stress episodes. Shown in Display 21.21 are the percentages of boys among the births, grouped according to when the severe life event took place. Notice that for one group the exposure is listed as taking place during the first trimester of pregnancy. The rationale for this is that the stress associated with the cancer or heart attack of a family member may well have started before the recorded time of death or hospital admission. Analyze the data to investigate the researchers' hypothesis. Write a summary of statistical findings. (Data from D. Hansen et al., "Severe Periconceptional Life Events and the Sex Ratio in Offspring: Follow Up Study Based on Five National Registers," British Medical Journal, 319 (1999): 548–49.)

```r
summary(lrm)
```

```
##
## Call:
## glm(formula = Colds/Size ~ isPlacebo, family = "binomial", data = cold,
##     weights = Size)
##
## Deviance Residuals:
## [1]  0  0
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.0565     0.1133   9.325   <2e-16 ***
## isPlacebo     0.4269     0.1702   2.508   0.0121 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6.3573  on 1  degrees of freedom
## Residual deviance: 0.0000  on 0  degrees of freedom
## AIC: 16.162
##
## Number of Fisher Scoring iterations: 3
```

```r
lrm.additive <- glm(PctBoys/100 ~ Group + Time , data = ex2117, weights = Number, family = "binomial")
```

```
## Warning in eval(family$initialize): non-integer #successes in a binomial
## glm!
```

```r
dev_gof <- 1 - pchisq(lrm.additive$null.deviance, lrm.additive$df.null)

dev_gof
```

```
## [1] 0.1272031
```

```r
#lrm.reduced <- glm(PctBoys/100 ~ Group, data = ex2117, weights = Number, family = "binomial")
```

There is no evidence that Exposure or Time of trauma around death or cancer hospital submission has an effect on the percentage of males born (Deviance Goodness-of-Fit Test. p-value = 0.1272).

18

Researchers in Kenya identified a cohort of more than 1,000 prostitutes who were known to be a major reservoir of sexually transmitted diseases in 1985. It was determined that more than 85% of them were infected with human immunodeficiency virus (HIV) in February 1986. The researchers then identified men who acquired a sexually transmitted disease from this group of women after the men sought treatment at a free clinic. Display 21.22 shows the subset of those men who did not test positive for the HIV on their first visit and who agreed to par- ticipate in the study. The men are categorized according to whether they later

tested positive for HIV during the study period, whether they had one or multiple sexual contacts with the prostitutes, and whether they were circumcised.

Describe how the odds of testing positive are associated with number of contacts and with whether the male was circumcised. (Data from D. W. Cameron et al., "Female to Male Transmission of Human Immunodeficiency Virus Type 1: Risk Factors for Seroconversion in Men," The Lancet (1989): 403–07.)

There is no evidence that the interaction between Contact and Circumcision have any effect on the model (Drop-in-deviance Test. p-value = 0.9616).

There is no evidence that number of contacts had any effect on whether or not a male contracted HIV (Two-Tailed Wald Logistic Regression on a Single Variable. p-value = 0.4685).

There is convincing evidence that Circumcision is associated with whether or not a male contracted HIV (Two-Tailed Wald Logistic Regression on a Single Variable. p-value = 1.94e-06). It is estimated that an Uncircumcised Male is 10.74 times more likely to contract HIV than a Circumcised Male. With 95% confidence, an Uncircumcised Male is between 4.2508 and 31.0129 times more likely to contract HIV than a circumcised Male.

```
ex2118$Circumcised <- relevel(ex2118$Circumcised, ref = "Yes")
lrm.interaction <- glm(HIV/Number ~  Circumcised * Contact, data = ex2118, weights = Number, family = "
summary(lrm.interaction)
```

```
##
## Call:
## glm(formula = HIV/Number ~ Circumcised * Contact, family = "binomial",
##     data = ex2118, weights = Number)
##
## Deviance Residuals:
## [1]  0  0  0  0
##
## Coefficients:
##                             Estimate Std. Error z value Pr(>|z|)
## (Intercept)                -3.48431    0.45402  -7.674 1.66e-14 ***
## CircumcisedNo               2.38570    0.55561   4.294 1.76e-05 ***
## ContactSingle              -0.32235    1.10831  -0.291    0.771
## CircumcisedNo:ContactSingle -0.06064    1.25554  -0.048    0.961
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2.7092e+01  on 3  degrees of freedom
## Residual deviance: 1.5543e-14  on 0  degrees of freedom
## AIC: 20.835
##
## Number of Fisher Scoring iterations: 4
```

```
lrm.additive <- glm(HIV/Number ~  Circumcised + Contact, data = ex2118, weights = Number, family = "bin
summary(lrm.additive)
```

```
##
## Call:
## glm(formula = HIV/Number ~ Circumcised + Contact, family = "binomial",
##     data = ex2118, weights = Number)
```

```
##
## Deviance Residuals:
##        1         2         3         4
##  0.03894  -0.01881  -0.01723   0.01219
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -3.4765     0.4227  -8.225  < 2e-16 ***
## CircumcisedNo   2.3740     0.4988   4.759 1.94e-06 ***
## ContactSingle  -0.3698     0.5218  -0.709    0.479
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 27.0917072  on 3  degrees of freedom
## Residual deviance:  0.0023152  on 1  degrees of freedom
## AIC: 18.837
##
## Number of Fisher Scoring iterations: 4
```

```
dind(lrm.interaction, lrm.additive)
```

```
## [1] 0.9616234
```

```
# Inversing because these are negative values
summary(lrm.additive)$coefficients %>% exp
```

```
##                  Estimate Std. Error      z value Pr(>|z|)
## (Intercept)    0.03091538   1.526067 2.679702e-04 1.000000
## CircumcisedNo 10.74010144   1.646770 1.166575e+02 1.000002
## ContactSingle  0.69088636   1.684997 4.922781e-01 1.613658
```

```
confint(lrm.additive) %>% exp
```

```
## Waiting for profiling to be done...
```

```
##                    2.5 %      97.5 %
## (Intercept)   0.01200791  0.06475246
## CircumcisedNo 4.25077517 31.01292717
## ContactSingle 0.23006691  1.83114444
```

19

Meta-analysis refers to the analysis of analyses. When the main results of studies can be cast into 2 X 2 tables of counts, it is natural to combine individual odds ratios with a logistic regression model that includes a factor to account for different odds from the different studies. In addition, the odds ratio itself might differ slightly among studies because of different effects on different populations or different research techniques. One approach for dealing with this is to suppose an underlying common odds ratio and to model between-study

variability as extra-binomial variation. Display 21.23 shows the results of 10 separate case–control studies on the association of breast cancer and whether a woman had breast fed children.

How much greater are the odds of breast cancer for those who did not breast feed than for those who did breast feed? (Data gathered from various sources by Karolyn Kolassa as part of a Master's project, Oregon State University.)

```
cancer <- ex2119
cancer$Total <- with(cancer, Cancer + NoCancer)
lrm.additive <- glm(Cancer/Total ~ factor(Study) + Lactate, family = "quasibinomial", weights = Total, d
summary(lrm.additive)
```

```
##
## Call:
## glm(formula = Cancer/Total ~ factor(Study) + Lactate, family = "quasibinomial",
##     data = cancer, weights = Total)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.70217  -0.57823  -0.00853   0.47100   1.43668
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -0.78029    0.07945  -9.821 4.16e-06 ***
## factor(Study)2   0.21757    0.09582   2.271 0.049317 *
## factor(Study)3   0.77526    0.14180   5.467 0.000397 ***
## factor(Study)4   0.91200    0.09573   9.527 5.35e-06 ***
## factor(Study)5  -3.25657    0.08388 -38.824 2.48e-11 ***
## factor(Study)6   0.84493    0.11804   7.158 5.32e-05 ***
## factor(Study)7   0.12387    0.09575   1.294 0.227975
## factor(Study)8   0.83808    0.09789   8.561 1.28e-05 ***
## factor(Study)9   0.81036    0.08211   9.870 3.99e-06 ***
## factor(Study)10  0.78969    0.08251   9.571 5.15e-06 ***
## Lactateyes      -0.10943    0.03130  -3.496 0.006768 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasibinomial family taken to be 1.847086)
##
##     Null deviance: 33218.213  on 19  degrees of freedom
## Residual deviance:    16.659  on  9  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 3
```

```
# get inverse of odds for Lactate since the baseline reference is "yes"
(summary(lrm.additive)$coefficients %>% exp)[11,]^-1
```

```
##   Estimate Std. Error    t value   Pr(>|t|)
##  1.1156387  0.9691822 32.9758793  0.9932545
```

```
(confint(lrm.additive) %>% exp)^-1
```

```
## Waiting for profiling to be done...

##                       2.5 %     97.5 %
## (Intercept)       2.5534106  1.8698516
## factor(Study)2    0.9700760  0.6662422
## factor(Study)3    0.6080944  0.3486788
## factor(Study)4    0.4841907  0.3326614
## factor(Study)5   30.5685254 22.0000010
## factor(Study)6    0.5411183  0.3406276
## factor(Study)7    1.0652503  0.7318151
## factor(Study)8    0.5235847  0.3566917
## factor(Study)9    0.5217148  0.3780967
## factor(Study)10   0.5330353  0.3856931
## Lactateyes        1.1861960  1.0492165
```

There is convincing evidence that breast feeding is associated with contracting breast cancer (Two-tailed Wald Single Value Extra-Binomial Variation Logistic Regression Test. p-value = 0.0068). It is estimated that women who do not breast feed are 1.1156 times more likely to develop breast cancer compared to women who do. With 95% confidence, women who do not breast feed are between 1.0492 and 1.1862 times more likely to develop breast cancer than those who do not.