Cameron Chandra

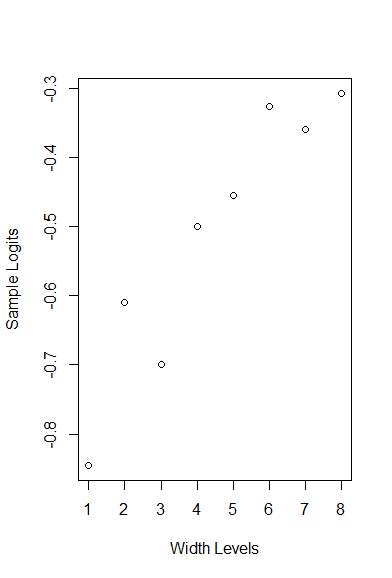
10/11/17

Stats449/Green

Homework 3

1. Sample logit = log [ (yi + 0.5) / (ni + yi + 0.5) ]

|  |  |  |  |
| --- | --- | --- | --- |
| xi | yi | ni | logit |
| 1 | 2 | 15 | -0.845 |
| 2 | 6 | 20 | -0.610 |
| 3 | 7 | 30 | -0.699 |
| 4 | 18 | 40 | -0.500 |
| 5 | 13 | 25 | -0.455 |
| 6 | 22 | 25 | -0.326 |
| 7 | 15 | 20 | -0.360 |
| 8 | 15 | 16 | -0.308 |



There appears to be a linear trend between x and the sample logits.

1. Call:

glm(formula = datmat ~ x, family = binomial)

Residual deviance: 7.1607 on 6 degrees of freedom

G^2 = 7.1607 on 6 degrees of freedom. Because G^2 (the likelihood ratio test statistic) is less than the cutoff point (12.6), we fail to reject the null hypothesis. Therefore, the model fits the data well.

> coefficients(m1)

(Intercept) x

-2.5943388 0.6029013

Estimated log odds ratio = 0.603

Estimated odds ratio = 1.83

Confidence Interval:

e^[ .603 +/- (1.96\*0.099) ] = **(1.51,2.22)**

1. Estimated odds ratio = e^(2.5\*.603) = 4.52

Confidence interval for estimate = e^[ (2.5\*.603) +/- (1.96\*2.5\*0.099) = **(2.78,7.33)**

Code Used:

a.)

> dat <- c(2,6,7,18,13,22,15,15,13,14,23,22,12,3,5,1)

> datmat <- matrix(dat,nrow=8,ncol=2)

> colnames(datmat) <- c("Crabs with Satellites", "Crabs without Satellites")

> rownames(datmat) <- c("22.69","23.83","24.77","25.84","26.79","27.74","28.67","30.41")

> x <- factor(1:8)

> levels(x) <- c("22.69","23.83","24.77","25.84","26.79","27.74","28.67","30.41")

> width <- c(1,2,3,4,5,6,7,8)

> logits <- c(-.845,-.61,-.699,-.5,-.455,-.326,-.360,-.308)

> plot(width,logits, xlab = "Width Levels", ylab = "Sample Logits")

b.)

> x <- 1:8

> m1 <- glm(datmat~x,family=binomial)

> summary(m1)



|  |  |  |
| --- | --- | --- |
| Model | *Df* | *G^2* |
| a | 1 | 0 |
| a + Bi^X | 4 | 3.4409 |
| a + Bk^Z | 3 | 112.57 |
| a + Bi^X + Bk^Z | 2 | 1.8231 |

1. Hypothesis of conditional independence: H0: Θxy1 = Θxy2 = Θxy3 = 1

112.57 – 1.8231 = 110.7469 (test statistic)

3 – 2 = 1 (degrees of freedom)

qchisq(.95,1) = 3.841 (critical value)

110.7469 > 3.841 => reject the null hypothesis.

c. Hypothesis of homogeneity of odds ratios: H0: Θxy1 = Θxy2 = Θxy3

G^2 = 1.8231 (test statistic)

Degrees of freedom = 2

qchisq(.95,2) = 5.99 (critical value)

1.8231 < 5.99 => fail to reject the null hypothesis.

d. Yes, because we failed to reject the null hypothesis of the homogeneity of odds ratios test, this model may have a common odds ratio. The estimated log odds ratio is 0.79409. So, the estimated common odds ratio is 2.212. The confidence interval for the estimated common odds ratio is (1.905,2.569). This common odds ratio implies that, regardless of what center the drug is tested on, patients taking the drug are 2.212 times less likely to be infected. We assert that the true common odds ratio, with 95% confidence, lies within the range (1.905,2.569)

e. The collapsibility condition is satisfied for collapsing the partial X-Y tables into the marginal X-Y table. Because we can find the common odds ratio through the model A+ Bi^x, Y and Z are conditionally independent given X. So, the X-Y association is collapsible over variable Z.

Code Used:

a.

> yes <- c(281,182,252,160,204,114)

> no <- c(319,418,248,340,196,286)

> ymat <- cbind(yes,no)

> x <- rep(c(1,0),3)

> z <- gl(3,2)

> m1 <- glm(ymat ~ z, family= binomial)

> m2 <- glm(ymat ~x+z, family=binomial)

> summary(m1)

> summary(m2)