414HW6

Ningyuan Wang

11/3/2020

1.

a. The odds ratio that compares the odds of approval on the second survey to the odds of approval on the first survey is 0.5733, and the 95% CI is (0.4398, 0.7474). The value of the likelihood ratio test is 17.58 on 1 df, and p-value of the test is p=0. Reject Ho. WIth 95% confidence, the odds ratio taht comapres the odds of approcal on the second survey to the odds of approval on the first survey locates in the interval of 0.4398 and 0.7474.

Recall the value of McNemar's test statistic is 17.4 on 1 df. The Score(logrank) Test statistic is 17.36 on 1 df. The two statistic are pretty close to each other.

```
m1 = clogit(appdapp ~ strata(subj) + srvy)
summary(m1)
```

```
## Call:
## coxph(formula = Surv(rep(1, 3200L), appdapp) ~ strata(subj) +
##     srvy, method = "exact")
##
##   n= 3200, number of events= 1824
##
##          coef exp(coef) se(coef)      z Pr(>|z|)
## srvy -0.5563    0.5733   0.1353 -4.113 3.91e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##      exp(coef) exp(-coef) lower .95 upper .95
## srvy    0.5733      1.744    0.4398    0.7474
##
## Concordance= 0.636  (se = 0.044 )
## Likelihood ratio test= 17.58  on 1 df,   p=3e-05
## Wald test            = 16.92  on 1 df,   p=4e-05
## Score (logrank) test = 17.36  on 1 df,   p=3e-05
```

b. Please check the hand-written part. The results match to the results in (a).

2.

a. Please check the hand-written part.

b. The odds ratio of MI that compares those with diabetes to those without diabetes is 2.313. 95% CI is (1.286， 4.157). WIth 95% confidence, the odds ratio of MI that compares those with diabetes to those without diabetes is located in the interval between 1.286 and 4.157.

```
pairs = rep(c(1:144), each = 2)
y = rep(c(1,0), times = 72)
x = c(rep(c(1, 1), 9),
      rep(c(1, 0), 37),
      rep(c(0, 1), 16),
      rep(c(0, 0), 82))
newdf = data.frame(pairs, y, x)
m2 = clogit(y ~ strata(pairs) + x, data = newdf)
summary(m2)

## Call:
## coxph(formula = Surv(rep(1, 288L), y) ~ strata(pairs) + x, data = newdf,
##     method = "exact")
##
##   n= 288, number of events= 144
##
##     coef exp(coef) se(coef)     z Pr(>|z|)
## x 0.8383    2.3125   0.2992 2.802  0.00508 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##   exp(coef) exp(-coef) lower .95 upper .95
## x     2.313     0.4324     1.286     4.157
##
## Concordance= 0.573  (se = 0.035 )
## Likelihood ratio test= 8.55  on 1 df,   p=0.003
## Wald test            = 7.85  on 1 df,   p=0.005
## Score (logrank) test = 8.32  on 1 df,   p=0.004
```

c. The statistic is 8.551263, and it matches the likelihood ratio statistic produced in the model of part b.

```
l_beta = 37*0.8383- (16+37)* log(2.3125+1)
l0 = -(16+37)* log(2)
2*(l_beta-l0) #  4.749658e-05

## [1] 8.551263
```

d. The score statistic is 8.320755, and it matches to the score(logrank) statistic in the model.

```
d2 = (37 - (16+37)/2)^2
I = (16+37)/4
d2/I

## [1] 8.320755
```

e. Compared to the model in part b, the estmation, standard error and the Wald test are the same. The deviance difference between null deviance and residual deviance is equal to the likelihood ratio test statistic in the part b.

```
index = 1:144*2
newdata = newdf[index-1,] - newdf[index,]
mod = glm(y ~ -1 + x, family = binomial, data = newdata)
summary(mod)

##
## Call:
## glm(formula = y ~ -1 + x, family = binomial, data = newdata)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## 0.8478  0.8478  1.1774  1.1774  1.5477
##
## Coefficients:
##   Estimate Std. Error z value Pr(>|z|)
## x   0.8383     0.2992   2.802  0.00508 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 199.63  on 144  degrees of freedom
## Residual deviance: 191.07  on 143  degrees of freedom
## AIC: 193.07
##
## Number of Fisher Scoring iterations: 4
```

1b.

		Second A	Second D	
First	A	794	150	944
	D	86	570	656
		880	720	1600

$\hat{\theta} = n_{21} | n_{12} = 86 | 150 = 0.5733$

$\log(\hat{\theta}) = \log(0.5733) = -0.5563$

$\hat{se} \log(\hat{\theta}) = \sqrt{1/n_{21} + 1/n_{12}} = \sqrt{0.0183} = 0.1353$

95% CI for θ: $e^{\log(\hat{\theta}) \pm 1.96 \hat{se} \log(\hat{\theta})}$

$= e^{(-0.8215, -0.2911)}$

$= (0.4398, 0.7474)$

2a. Since both response and predictor variables are binary.

$l(\beta) = \log L(\beta) = \log(\exp(\beta)^{n_{21}}) - \log((1+\exp(\beta))^{n_{12}+n_{21}})$

$= n_{21}\beta - (n_{12}+n_{21}) \log(e^{\beta}+1)$

$\frac{\partial l(\beta)}{\partial \beta} = n_{21} - (n_{12}+n_{21}) \frac{1}{e^{\beta}+1} e^{\beta}$

set $\frac{\partial l(\beta)}{\partial \beta} = 0 \Rightarrow \hat{\beta} = \log(\frac{n_{21}}{n_{12}})$

$$\frac{\partial^2 l(\beta)}{\partial \beta^2} = -(n_{12}+n_{21})\frac{e^{\beta}}{(e^{\beta}+1)^2} \quad \text{Since } p = \frac{e^{\beta}}{1+e^{\beta}}$$

$$I(\beta) = -E\left(\frac{\partial^2 l(\beta)}{\partial \beta^2}\right) = E\left((n_{12}+n_{21})\frac{e^{\beta}}{(e^{\beta}+1)^2}\right) \;\&\; \hat{p} = \frac{n_{21}}{n_{12}+n_{21}}$$

$$= (n_{12}+n_{21})\,p(1-p)$$

$$I(\hat{\beta}) = \frac{n_{12}n_{21}}{n_{12}+n_{21}} \Rightarrow Var(\hat{\beta}) = I(\hat{\beta})^{-1} = \frac{n_{12}+n_{21}}{n_{12}n_{21}} = \frac{1}{n_{12}} + \frac{1}{n_{21}}$$

$$\hat{se}(\hat{\beta}) = \sqrt{\frac{1}{n_{12}} + \frac{1}{n_{21}}}$$