HW5

Ningyuan Wang

10/16/2020

1

a. With LRT, the test statistic is 12.845, and the p-value is 0.0016. Since p-value < .05, reject the null hypothesis. With Wald test, the test statistic is 12.4, and the p-value is 0.0021. Similar to the LRT, we also reject the null hypothesis at alpha = 0.05.

b. With LRT, the test statistic is 0.96205, and the p-value is 0.3267. Since p-value > .05, do not reject the null hypothesis. With Wald test, the test statistic is 0.95, and the p-value is 0.33. We also do not reject the null hypothesis at alpha = 0.05.

c. With the covariance matrix of the model, we know that Var(beta_race2) = 0.2054273885, Var(beta_race3) = 0.2412413604, and Cov(beta_race2, beta_race3) = 0.0945459565. Also, based on the model, beta_race2 = 0.91630, and beta_race3 = 0.42271. Se(beta_race2 - beta_race3) = sqrt(Var(beta_race2 - beta_race3)) = 0.5075203. So the 95% confidence interval is (-0.501 1.488). Since the CI includes 0, we do not reject null hypothesis. There is no significant difference in the odds of death comparing black and the other race, while fixed other variables.

d. The residual deviance cannot be used as a goodness of fit statistic, becauase the model has a binary response that we cannot find a saturated model for comparing and constructing a chi-squared distribution.

```
icu = read.csv("icu.csv")
m1 = glm(sta ~ age + can + cpr + inf + factor(race), family = binomial, data =
icu)
summary(m1)

##
## Call:
## glm(formula = sta ~ age + can + cpr + inf + factor(race), family = binomial,
##     data = icu)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2955  -0.6821  -0.4901  -0.2828   2.6817
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -4.10749    0.86008  -4.776 1.79e-06 ***
## age             0.02840    0.01152   2.466   0.0137 *
## can             0.26069    0.61811   0.422   0.6732
## cpr             1.53940    0.62061   2.480   0.0131 *
## inf             0.87996    0.39584   2.223   0.0262 *
## factor(race)2   0.91630    0.45324   2.022   0.0432 *
## factor(race)3   0.42271    0.49116   0.861   0.3894
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 200.16  on 199  degrees of freedom
## Residual deviance: 176.29  on 193  degrees of freedom
## AIC: 190.29
##
## Number of Fisher Scoring iterations: 5

# a.
# LRT
m0 = glm(sta ~ age + can + factor(race), family = binomial, data = icu)
anova(m0, m1)

## Analysis of Deviance Table
##
## Model 1: sta ~ age + can + factor(race)
## Model 2: sta ~ age + can + cpr + inf + factor(race)
##   Resid. Df Resid. Dev Df Deviance
## 1       195     189.13
## 2       193     176.29  2   12.845

1-pchisq(12.845, 2)

## [1] 0.00162459

# Wald
library(aod)
wald.test(b = coef(m1), Sigma = vcov(m1), Terms = c(4,5))

## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 12.4, df = 2, P(> X2) = 0.0021

# b.
# LRT
x = model.matrix(m1)[,-1]
newdat = data.frame(sta =m1$y, x)
m00 = glm(sta ~ age + can + cpr + inf + I(factor.race.2 + factor.race.3), family
= binomial, data=newdat)
anova(m00, m1)

## Analysis of Deviance Table
##
## Model 1: sta ~ age + can + cpr + inf + I(factor.race.2 + factor.race.3)
## Model 2: sta ~ age + can + cpr + inf + factor(race)
##   Resid. Df Resid. Dev Df Deviance
## 1       194     177.25
## 2       193     176.29  1  0.96205
```

```r
1-pchisq(0.96205, 1)

## [1] 0.3266709

# Wald
lc = cbind(0,0,0, 0, 0, 1, -1)
wald.test(b = coef(m1), Sigma = vcov(m1), L = lc)

## Wald test:
## ----------
##
## Chi-squared test:
## X2 = 0.95, df = 1, P(> X2) = 0.33

# c.
vcov(m1)

##                  (Intercept)           age           can           cpr
## (Intercept)    0.739740760 -0.0087221309 -0.0949432092 -0.0673667138
## age           -0.008722131  0.0001326462  0.0006255517  0.0007921354
## can           -0.094943209  0.0006255517  0.3820572207  0.0136495377
## cpr           -0.067366714  0.0007921354  0.0136495377  0.3851541190
## inf           -0.094921781 -0.0001699830  0.0177550426 -0.0254992448
## factor(race)2 -0.129836109  0.0001678859  0.0150420266 -0.0117587927
## factor(race)3 -0.140311649  0.0006496309 -0.0084158970 -0.0326481707
##                        inf factor(race)2 factor(race)3
## (Intercept)   -0.094921781 -0.1298361087 -0.1403116494
## age           -0.000169983  0.0001678859  0.0006496309
## can            0.017755043  0.0150420266 -0.0084158970
## cpr           -0.025499245 -0.0117587927 -0.0326481707
## inf            0.156685697  0.0455245092  0.0205277263
## factor(race)2  0.045524509  0.2054273885  0.0945459565
## factor(race)3  0.020527726  0.0945459565  0.2412413604

summary(m1)

##
## Call:
## glm(formula = sta ~ age + can + cpr + inf + factor(race), family = binomial,
##     data = icu)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2955  -0.6821  -0.4901  -0.2828   2.6817
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -4.10749    0.86008  -4.776 1.79e-06 ***
## age            0.02840    0.01152   2.466   0.0137 *
## can            0.26069    0.61811   0.422   0.6732
## cpr            1.53940    0.62061   2.480   0.0131 *
## inf            0.87996    0.39584   2.223   0.0262 *
## factor(race)2  0.91630    0.45324   2.022   0.0432 *
## factor(race)3  0.42271    0.49116   0.861   0.3894
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 200.16  on 199  degrees of freedom
## Residual deviance: 176.29  on 193  degrees of freedom
## AIC: 190.29
##
## Number of Fisher Scoring iterations: 5
```

```
v = 0.2054273885 + 0.2412413604 - 2*0.0945459565
sqrt(v)
```

```
## [1] 0.5075203
```

```
(0.91630 - 0.42271) + c(-1, 1)*1.96*(0.5075203)
```

```
## [1] -0.5011498  1.4883298
```

2) Suppose $Y_1, Y_2, \ldots, Y_n$ is a random sample (iid) from the Bernoulli(p) distribution. Let

$$\hat{p} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

By the Central Limit Theorem

$$\sqrt{n}(\hat{p} - p) \xrightarrow{d} N(0, p(1-p))$$

a) Use the delta method to find the asymptotic distribution of

$$\sqrt{n}\left(\frac{\hat{p}}{1-\hat{p}} - \frac{p}{1-p}\right)$$

b) What is an estimate for

$$Var\left(\frac{\hat{p}}{1-\hat{p}}\right)$$

a. let $g(p) = \frac{p}{1-p}$, $g'(p) = \frac{1}{(1-p)^2}$ $\quad 0<p<1$

with the delta method:

$$\sqrt{n}[g(\hat{p}) - g(p)] \xrightarrow{d} N\left[0,\ p(1-p)\left(\frac{1}{(1-p)^2}\right)^2\right]$$

$$\sim N\left(0, \frac{p}{(1-p)^3}\right)$$

So, the assymptotic distribution is normal with mean $=0$ and variance $= \frac{p}{(1-p)^3}$

b. Since $Var(\sqrt{n}\, g(\hat{p})) = \frac{p}{(1-p)^3}$

$$\widehat{Var}\left(\frac{\hat{p}}{1-\hat{p}}\right) = \frac{\hat{p}}{(1-\hat{p})^3} \cdot \frac{1}{n}$$