Applied Logistic Regression - Exercise Week 2

Yago Durán Cid

24/05/2015

WEEK 2

Exercise 1:

Use the Myopia Study (MYOPIA-fixed.dta)
a. Using the results of the output from R, assess the significance of the slope coefficient for SPHEQ using the likelihood ratio test and the Wald test. What assumptions are needed for the p-values computed for each of these tests to be valid? Are the results of these tests consistent with one another? What is the value of the deviance for the fitted model?

We keep the basic model we used in homework 1. This is $\pi(x) = E(y|x) = \frac{e^{(\beta_0+\beta_1 x)}}{1+e^{(\beta_0+\beta_1 x)}}$ which gives foolowing results:

```
##
## Call:
## glm(formula = MYOPIC ~ SPHEQ, family = "binomial", data = data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6435  -0.4533  -0.2681  -0.1029   3.1602
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.05397    0.20675   0.261    0.794
## SPHEQ       -3.83310    0.41837  -9.162   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 480.08  on 617  degrees of freedom
## Residual deviance: 337.34  on 616  degrees of freedom
## AIC: 341.34
##
## Number of Fisher Scoring iterations: 6
```

as per data above, deviance of the model is 337.3448797compared to a deviance of the model with only a constant variable of 480.0770169
We can test the overall significance of the variables (both constant β_0 and SPHEQ β_1) using the likelihood ratio test.

$G = -2ln[\frac{likelihood\,model\,without\,\beta_1}{likelihood\,model\,with\,\beta_1}]$

Function used in R does not show the likelihood ratio, but offers instead the deviance which is directly related to the likelihood ratio: $log\,likelihood = \frac{Deviance}{-2}$

Therefore, we can rewrite the likelihod ratio test based on the deviance as follows:

$G = -2ln[\frac{likelihood\,model\,without\,\beta_1}{likelihood\,model\,with\,\beta_1}] = -2log\,likelihood_{model\,without\,\beta_1} + 2log\,likelihood_{model\,with\,\beta_1} = Deviance_{model\,without\,\beta_1} - Deviance_{model\,with\,\beta_1} = 480.077 - 337.345 = 142.732$

We know that in the case $\beta_1 = 0$ the likelihood ratio follows a χ^2 distribution. That allow us to estimate the probability of $\beta_1 = 0$. Thus, p-value in our case is 6.7266405×10^{-33}so we can assume the model including SPHEQ is appropriate.

The Wald test is easier witht he data proovided by R since it is automatically reported in R. Following this test we can say that SPHEQ is significant while the constant is not. This is consistent with the likelihood ratio test since the likelihood ratio test.
If we want to use likelihood ratio test to asses the significance of the constant, we only have to compare the likelihood of the model including the constant and SPHEQ along with the model not including the constant abut including SPHEQ.

The p-value of $\beta_0 = 0$ is 6.7266405×10^{-33} which is consistent with the Wald statistic.

Exercise 2:

Use the ICU study (icu.dta) Using the results of the output from the logistic regression package used for problem 2 part (d) of week 1, assess the significance of the slope coefficient for AGE using the likelihood ratio test and the Wald test. What assumptions are needed for the p-values computed for each of these tests to be valid? Are the results of these tests consistent with one another? What is the value of the deviance for the fitted model?

As in previous section we show the results of model estimated in frist week homework

```
##
## Call:
## glm(formula = STA ~ AGE, family = "binomial", data = icu)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -0.9536  -0.7391  -0.6145  -0.3905   2.2854
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.05851    0.69608  -4.394 1.11e-05 ***
## AGE          0.02754    0.01056   2.607  0.00913 **
## ---
## 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: 192.31  on 198  degrees of freedom
## AIC: 196.31
##
## Number of Fisher Scoring iterations: 4
```

The p-value of the likelihood ratio test to test that AGE slope is different tahn zero is 0.0050692 Being less than 0.05 we accept the slope of AGE as significant. This is consistent with the Wald test shown in R output.