Lab Assignment: Logistic Regression

Your name here

Nov. ??, 2015

Create a Word document from this R Markdown file for the following exercises. Submit the R markdown file and resulting Word document via D2L Dropbox.

## Exercise 1

A study was conducted whereby the type of anesthetic (A or B), nausea after the surgery (Yes or No), the amount of pain medication taken during the recovery period, and age for a random sample of 72 patients undergoing reconstructive knee surgery.

The data is in the file anesthesia.rda.

### Part 1a

Use R to create a two-way table with the type of anesthetic defining the rows and nausea after the surgery as the columns and also produce the output for a chi-square test for independence.

Is there an association between these two categorical variables at a 5% level of significance?

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1a -|-|-|-|-|-|-|-|-|-|-|-

library(DS705data)  
data(anesthesia)  
  
  
tbl <- with(anesthesia, table(anesthetic, nausea))  
chisq.test(tbl)

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: tbl  
## X-squared = 8.0559, df = 1, p-value = 0.004535

At , we reject the null hypothesis (p-value = 0.004535). That is to say, the two categorical values appear to be independent.

### Part 1b

Obtain the output from R (including the Wald tests for coefficients - so use "summary" function) for the logistic regression model with nausea as the dependent variable and the type of anesthetic as the predictor variable.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1b -|-|-|-|-|-|-|-|-|-|-|-

fit <- glm(nausea~anesthetic, data=anesthesia, family='binomial')  
summary(fit)

##   
## Call:  
## glm(formula = nausea ~ anesthetic, family = "binomial", data = anesthesia)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.4823 -0.8497 0.0254 0.9005 1.5453   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.6931 0.3397 2.041 0.04129 \*   
## anestheticB -1.5261 0.5088 -2.999 0.00271 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 99.813 on 71 degrees of freedom  
## Residual deviance: 90.133 on 70 degrees of freedom  
## AIC: 94.133  
##   
## Number of Fisher Scoring iterations: 4

### Part 1c

What is the outcome of the hypothesis test that the coefficient of **anesthetic** is "zero" vs "not zero" at a 5% level of significance? (use the Wald test from the R output from the logistic regression you performed)

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1c -|-|-|-|-|-|-|-|-|-|-|-

At , we reject the null hypothesis (p-value = 0.00271). That is to say, the coefficient is not zero.

### Part 1d

Convert the estimated coefficient of **anesthesia** to an odds ratio and interpret it in the context of the problem.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1d -|-|-|-|-|-|-|-|-|-|-|-

exp(confint(fit))

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) 1.04695740 4.0156727  
## anestheticB 0.07728494 0.5743758

With 95% confidence, the odds of feeling nausea when taking anesthetic B is 7.72% to 57.44% that of taking anesthetic A.

### Part 1e

Install the package "mosaic" (if you don't have it installed already), then load it. Use the oddsRatio function to compute the odds ratio for having nausea for anesthetic A vs B. You may have to refer back to Week 8 for details on odds ratios and the oddsRatio function in R.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1e -|-|-|-|-|-|-|-|-|-|-|-

library(mosaic)  
oddsr <- oddsRatio(tbl)

### Part 1f

When logistic regression coefficients are negative, the interpretation sometimes has more impact when we switch the perspective and use the reciprocal of the exponentiated coefficient. Find the odds ratio for having nausea for anesthetic B compared to anesthetic A (i.e. take the reciprocal of the odds ratio you computed in part **1d**).

Interpret this odds ratio in the context of the problem.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1f -|-|-|-|-|-|-|-|-|-|-|-

1 / oddsr

## [1] 0.2173913

The odds of feeling nausea are 78.26% less when anesthetic B is present vs anesthetic A.

### Part 1g

Compute the predicted probability of a reconstructive knee surgery patient having nausea after surgery when anesthetic A was used.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1g -|-|-|-|-|-|-|-|-|-|-|-

newdata <- data.frame(anesthetic='A')  
predict(fit, newdata, type='response')

## 1   
## 0.6666667

### Part 1h

Compute a 95% confidence interval for the predicted probability of a reconstructive knee surgery patient having nausea after surgery when anesthetic A was used.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1h -|-|-|-|-|-|-|-|-|-|-|-

out <- predict(fit, newdata, type='response', se.fit=TRUE)  
C = 0.95  
crit = qnorm(1-(1-C)/2)  
lower = exp(out$fit-crit\*out$se.fit)/(1+exp(out$fit-crit\*out$se.fit))  
upper = exp(out$fit+crit\*out$se.fit)/(1+exp(out$fit+crit\*out$se.fit))  
c(lower,upper)

## 1 1   
## 0.6268481 0.6930920

## Exercise 2

Continue using the anesthesia.rda data set to do the following.

### Part 2a

Obtain the output from R (including the Wald tests for coefficients - so use "summary" function) for the logistic regression model with nausea as the dependent variable and the amount of pain medication taken as the predictor variable.

At , is there a statistically significant relationship between nausea and the amount of pain medication taken?

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2a -|-|-|-|-|-|-|-|-|-|-|-

fit2 <- glm(nausea~painmed, data=anesthesia, family='binomial')  
summary(fit2)

##   
## Call:  
## glm(formula = nausea ~ painmed, family = "binomial", data = anesthesia)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.8555 -0.6167 -0.1072 0.8206 1.7894   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.062742 0.764501 -4.006 6.17e-05 \*\*\*  
## painmed 0.037487 0.008833 4.244 2.20e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 99.813 on 71 degrees of freedom  
## Residual deviance: 68.049 on 70 degrees of freedom  
## AIC: 72.049  
##   
## Number of Fisher Scoring iterations: 5

Replace this text with your written answer here. At , we reject the null hypothesis (p-value = ). That is to say, the coefficient is not zero.

### Part 2b

Convert the estimated coefficient of **painmed** to an odds ratio and interpret it in the context of the problem.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2b -|-|-|-|-|-|-|-|-|-|-|-

exp(fit2$coefficients[2])

## painmed   
## 1.038199

The odds of having nausea increases by 103.81% as the pain medicine increases by 1.

### Part 2c

Compute the predicted probabilities of a reconstructive knee surgery patient having nausea in the recovery time after surgery for when 50 units of pain medication are used and also for when 100 units of pain medication are used.

Comment on these two probabilities.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2c -|-|-|-|-|-|-|-|-|-|-|-

newdata2 <- data.frame(painmed=c(50, 100))  
predict(fit2, newdata2, type='response')

## 1 2   
## 0.2335485 0.6650716

The probability of having nausea is 23.35% when pain medicine is 50 units and 66.50% when pain medicine is 100 units.