Logistic Regression

Matt Allen

07/09/2018

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 -|-|-|-|-|-|-|-|-|-|-|-

# Create contingency table from anesthesia data  
require(DS705data)

## Loading required package: DS705data

data("anesthesia")  
anesthetic\_nausea <- with(anesthesia,table(anesthetic,nausea))  
addmargins(anesthetic\_nausea)

## nausea  
## anesthetic No Yes Sum  
## A 13 26 39  
## B 23 10 33  
## Sum 36 36 72

chisq.test(anesthetic\_nausea, correct=FALSE)

##   
## Pearson's Chi-squared test  
##   
## data: anesthetic\_nausea  
## X-squared = 9.4545, df = 1, p-value = 0.002106

The p-value for the test is 0.002106. Thus at a 5% level of significance, we can conclude that there is evidence that an anesthetic is associated with nausea.

### 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 -|-|-|-|-|-|-|-|-|-|-|-

# Create and summarize logistic model of nausea as a function of anesthetic  
logistic.model <- glm(nausea~anesthetic,data=anesthesia,family="binomial")  
summary(logistic.model)

##   
## 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 -|-|-|-|-|-|-|-|-|-|-|-

The value of the Wald test statistic from the model summary is the z-value, which is -2.999. The p-value from the standard normal distribution is 0.00271. At a 5% level of significance, we have evidence to support that the coefficient of anesthetic is not zero.

### Part 1d

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

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

# Exponentiate the anesthetic coefficient to get the odds ratio   
exp(-1.5261)

## [1] 0.2173818

#(13\*10)/(26\*23)

The odds of not having nausea with anesthetic B are 78% less than with 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 -|-|-|-|-|-|-|-|-|-|-|-

# Use mosaic to compute the odds ratio for having nausea for anesthetic A vs B.  
anestheticA.oddsRatio <- oddsRatio(anesthetic\_nausea, verbose = FALSE)  
1/anestheticA.oddsRatio

## [1] 0.2173913

### 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 -|-|-|-|-|-|-|-|-|-|-|-

# Compute the reciprocal of odds ratio from part 1 d  
1/exp(1.5261)

## [1] 0.2173818

The odds of not having nausea with anesthetic B is 4.6 times more compared to 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 -|-|-|-|-|-|-|-|-|-|-|-

# Compute Predicted Probaility of nausea when anesthetic A was used.  
newdata <- data.frame(anesthetic='A')  
pred <- predict(logistic.model,newdata,type="response",se.fit = TRUE)  
pred$fit

## 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 -|-|-|-|-|-|-|-|-|-|-|-

# Compute a 95% confidence interval for the predicted probability of nausea when anesthetic A was used.  
C = .95 # define the level of confidence  
crit = qnorm(1-(1-C)/2) # get the appropriate critical value  
lower = exp(pred$fit-crit\*pred$se.fit)/(1+exp(pred$fit-crit\*pred$se.fit))  
upper = exp(pred$fit+crit\*pred$se.fit)/(1+exp(pred$fit+crit\*pred$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 -|-|-|-|-|-|-|-|-|-|-|-

# Calculate model with pain medication as the predictor variable  
logistic.model <- glm(nausea~painmed,data=anesthesia,family="binomial")  
summary(logistic.model)

##   
## 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

The p-value for the test is 2.20e-05. Thus at a 5% level of significance, we can conclude that there is evidence that the amount of pain medication taken is associated with nausea.

### Part 2b

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

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

# Convert the estimate coefficient paind to an odds ratio.  
exp(0.037487)

## [1] 1.038199

The odds of having nauseau increase by 4% for every unit increase in pain medication.

### 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 -|-|-|-|-|-|-|-|-|-|-|-

# Compute probability of having nausea when 50 and 100 units of pain medication are used.  
newdata <- data.frame(painmed=50)  
pred50 <- predict(logistic.model,newdata,type="response",se.fit = TRUE)  
pred50$fit

## 1   
## 0.2335485

newdata <- data.frame(painmed=100)  
pred100 <- predict(logistic.model,newdata,type="response",se.fit = TRUE)  
pred100$fit

## 1   
## 0.6650716

pred100$fit/pred50$fit

## 1   
## 2.847681

Increasing pain medication from 50 units to 100 increases the odds of having nausea by 2.85 times.