Lab Assignment: Logistic Regression

Spencer Swartz

April 4, 2017

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

require(DS705data)

## Loading required package: DS705data

data("anesthesia")  
tbl <- table(anesthesia$anesthetic,anesthesia$nausea)  
tbl

##   
## No Yes  
## A 13 26  
## B 23 10

chisq.test(tbl)

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

Based on the -squared test we can reject the null hypothesis that nausea is and anesthetics are not related.

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

nausea.out <- glm(nausea~anesthetic,family="binomial", data = anesthesia)  
summary(nausea.out)

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

Based on a 5% level of significance the coefficient looks to be non-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(-1.5261)

## [1] 0.2173818

The odds of having nausea after surgery after taking anesthesia B are 21.7% as large as the odds of having nausea after surgery after taking anesthesia A, given all other variables are held constant.

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

#install.packages("mosaic")  
require(mosaic)

## Loading required package: mosaic

## Warning: package 'mosaic' was built under R version 3.3.3

## Loading required package: dplyr

## Warning: package 'dplyr' was built under R version 3.3.3

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

## Loading required package: lattice

## Loading required package: ggplot2

## Warning: package 'ggplot2' was built under R version 3.3.3

## Loading required package: mosaicData

## Warning: package 'mosaicData' was built under R version 3.3.3

## Loading required package: Matrix

##   
## The 'mosaic' package masks several functions from core packages in order to add additional features.   
## The original behavior of these functions should not be affected by this.

##   
## Attaching package: 'mosaic'

## The following object is masked from 'package:Matrix':  
##   
## mean

## The following objects are masked from 'package:dplyr':  
##   
## count, do, tally

## The following objects are masked from 'package:stats':  
##   
## binom.test, cor, cov, D, fivenum, IQR, median, prop.test,  
## quantile, sd, t.test, var

## The following objects are masked from 'package:base':  
##   
## max, mean, min, prod, range, sample, sum

oddsRatio(tbl)

## [1] 4.6

4.6 to 1 for nausea if taking anesthesia A as compared to B.

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

exp(1.5261)

## [1] 4.600201

Anestethic A is 4.6 more likely to cause nausea as compared to Anestethic B.

### 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(nausea.out,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(nausea.out, newdata, se.fit=TRUE)  
C = .95 # define the level of confidence  
crit = qnorm(1-(1-C)/2) # get the appropriate critical value  
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   
## 0.5068447 0.7955831

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

nausea.out2 <- glm(nausea~painmed,family="binomial", data = anesthesia)  
summary(nausea.out2)

##   
## 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

Based on a 5% level of significance the coefficient for painmed looks to be non-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(0.0375)

## [1] 1.038212

The odds of having nausea increases by 1.04 for each additional unit of pain meds.

### 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(nausea.out2,newdata2,type="response")

## 1 2   
## 0.2335485 0.6650716

With 50 units of painmeds there is a 23.4% chance of the patient becoming nauseated, and with 100 units of painmeds there is a 66.6% chance of the patient becoming nauseated.