Logistic Regression

Josh Jarvey

06/16/2020

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

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

#load the data set  
library(DS705data)  
data("anesthesia")  
  
 #create a table of drug type and sickness outcome.  
outcome = table(anesthesia$anesthetic, anesthesia$nausea)  
addmargins(outcome)

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

#testing independence between drug type and sickness outcome.  
chisq.test(outcome)

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

At a 95% level of significance, there is enough evidence to suggest that the type of Anesthetic used and Nausea experienced are associated for knee surgery patients (p=0.004535).

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

#creating a logistic regression of nausea as a function of anesthetic used.   
model = glm(nausea~anesthetic, data = anesthesia, family = "binomial")  
summary(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 **anestheticB** 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)

*Note: When there is a categorical predictor variable with more than two categories, R will set 1 category as “baseline” (usually the category that comes first alphabetically), also known as the reference category. So in this problem anesthetic A will be the baseline category. The fit model will use two dummy variable anestheticB = 1 for anesthetic B and 0 otherwise. Anesthetic A will be the reference category and is modeled in the output when anestheticB=0.*

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

At a 95% level of significance, there is enough evidence to support the claim that the coefficient for anestheticB is different from 0 (p=0.00271).

### Part 1d

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

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

#exponentiating the coefficients to get the odds ratio.  
exp(model$coefficients[2])

## anestheticB   
## 0.2173913

The odds of having a sickness after knee surgery with the use of anestheticB are 21.74 % as large as the odds of having sickness after knee surgery with the use of anestheticA. That is, they are 78.26% less!

### 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 Lesson 5 for details on odds ratios and the oddsRatio function in R.

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

library(mosaic)

## Loading required package: dplyr

##   
## 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: ggformula

## Loading required package: ggplot2

## Loading required package: ggstance

##   
## Attaching package: 'ggstance'

## The following objects are masked from 'package:ggplot2':  
##   
## geom\_errorbarh, GeomErrorbarh

##   
## New to ggformula? Try the tutorials:   
## learnr::run\_tutorial("introduction", package = "ggformula")  
## learnr::run\_tutorial("refining", package = "ggformula")

## Loading required package: mosaicData

## Loading required package: Matrix

## Registered S3 method overwritten by 'mosaic':  
## method from   
## fortify.SpatialPolygonsDataFrame ggplot2

##   
## 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.  
##   
## Note: If you use the Matrix package, be sure to load it BEFORE loading mosaic.  
##   
## Have you tried the ggformula package for your plots?

##   
## Attaching package: 'mosaic'

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

## The following object is masked from 'package:ggplot2':  
##   
## stat

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

## The following objects are masked from 'package:stats':  
##   
## binom.test, cor, cor.test, cov, 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

#calculating the odds ratio of having nausea for anesthetic A vs. B. OR = 4.6  
oddsRatio(outcome, verbose = TRUE)

##   
## Odds Ratio  
##   
## Proportions  
## Prop. 1: 0.3333   
## Prop. 2: 0.697   
## Rel. Risk: 2.091   
##   
## Odds  
## Odds 1: 0.5   
## Odds 2: 2.3   
## Odds Ratio: 4.6   
##   
## 95 percent confidence interval:  
## 1.271 < RR < 3.439   
## 1.697 < OR < 12.47   
## NULL

## [1] 4.6

### 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 A compared to anesthetic B by taking the reciprocal of the odds ratio that was computed in part **1d**.

Interpret this odds ratio in the context of the problem.

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

#exponentiating the coefficients to get the odds ratio.  
1/exp(model$coefficients)

## (Intercept) anestheticB   
## 0.5 4.6

The odds of having a sickness after knee surgery with the use of anestheticA are 4.6 times more likely than odds of having sickness after knee surgery with the use of anestheticB.

### Part 1g

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

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

#when the use of anesthetic A, the probability of getting sick is 66.67%. This is the same thing as 26 Yes's / 39 total for A.   
predict(model,data.frame(anesthetic = "A"), 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 -|-|-|-|-|-|-|-|-|-|-|-

#allow standard error in the prediction.  
predicted\_prob = predict(model,data.frame(anesthetic = "A"), se.fit = TRUE)  
  
 #set confidence interval to 95%  
C = 0.95  
 #calculate the critical cutoff edges (2.5 and 97.5%)  
crit = qnorm(1-(1-C)/2)  
 #calculate the lower and upper confidence interval bounds based on the fitted value and the standard error, using the desired confidence level  
lower = exp(predicted\_prob$fit-crit\*predicted\_prob$se.fit)/(1+exp(predicted\_prob$fit-crit\*predicted\_prob$se.fit))  
upper = exp(predicted\_prob$fit+crit\*predicted\_prob$se.fit)/(1+exp(predicted\_prob$fit+crit\*predicted\_prob$se.fit))  
 #print bounds  
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 -|-|-|-|-|-|-|-|-|-|-|-

#create a 2nd model that uses pain med amount as the predictor variable. (AIC is better...).  
model2 = glm(nausea~painmed, data = anesthesia, family = "binomial")  
summary(model2)

##   
## 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

At a 95% level of significance, there is enough evidence to claim that the coefficient for the amount of pain medication administered as a predictor of sickness after knee surgery is different than 0 (p=2.20e-05).

### Part 2b

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

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

#converting log odds into normal odds. (take 1 - result, and that is your percentage. If its negative, then thats a decrease for every additional unit, whereas if its positive, then its in increase in probability).   
exp(model2$coefficients[2])

## painmed   
## 1.038199

The odds of having a sickness after knee surgery increase by 3.82% for each additional unit of pain medication administered.

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

#predict the probability of getting sick with 50 and 100 units of pain medication.   
predict(model2,data.frame(painmed = c(50,100)), type = "response")

## 1 2   
## 0.2335485 0.6650716

Holding everything else constant, when a knee surgery patient is administered 50 units of pain medication, they have a 23.35% chance of getting sick post surgery. When a patient is administered 100 units of pain medication, they have a 66.51% chance of getting sick post surgery.