Logistic Regression

Johnny Chiu

11/7/2017

## Table of content

* Logistics Regression
  + The idea behind logistics regression model
  + The interpretaion of in logistics regression
  + Estimating the regression coefficients
* Multinomial Logistics Regression
  + Logistics Regression for Nominal Response
  + Bayes Classification
  + Logistics Regression for Ordinal Response

## Logistics Regression

### • The idea behind logistics regression model

Logistic Regression is useful when the outcome of our regression model is categorical. Instead of using y, which we assume the value of it is categorical like {0,1}, {yes, no}, directly as our response variable, we models the probability that y belongs to a particular category. In other words, we build up a linear model on the log-odds (or logit). In the following paragraph, we denote the probability that y belongs to a particular category as

Since is the probability that y belongs to a particular category, the value of it should be between 0 and 1. Therefore, we must model p(X) using a function that gives outputs between 0 and 1. Many functions meet this description. In logistic regression, we use the logistic function:

We can easily see that the output value of this model is within 0 and 1. When goes to , will be 1; When goes to -, will be 0.

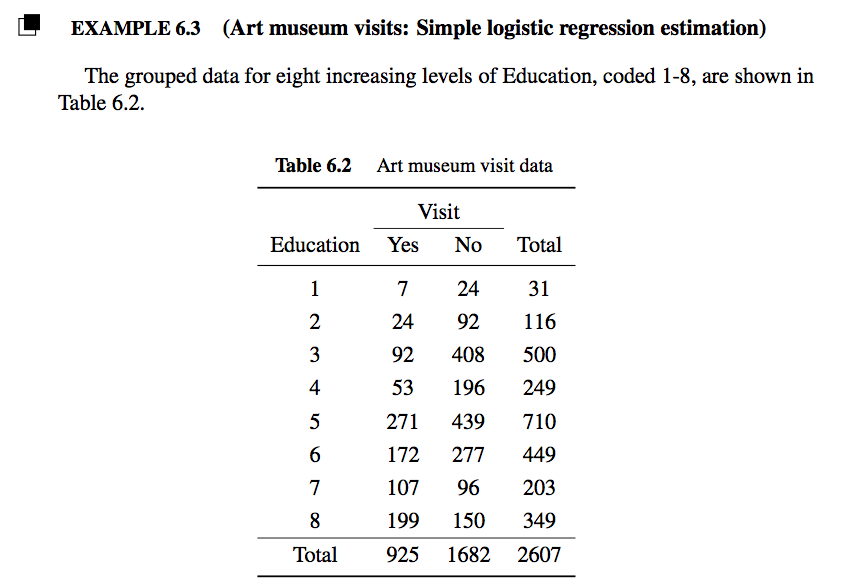
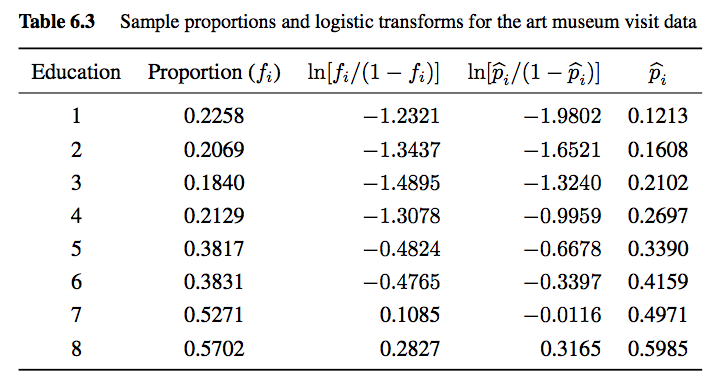
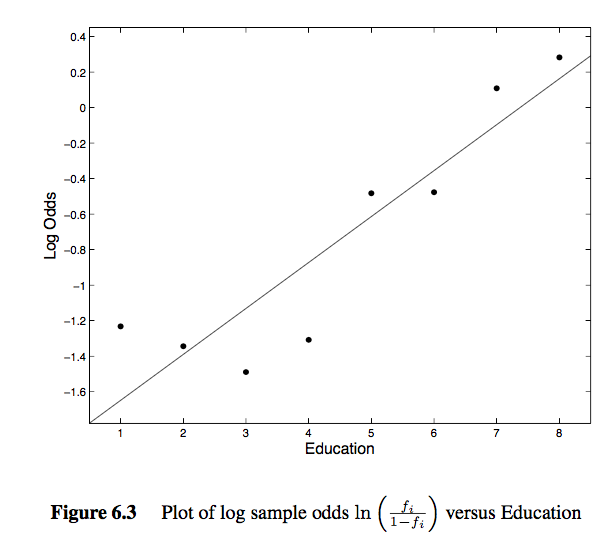
This is the model that we are going to use to build up our regression model on. To make the equation to become more interpretable, after a bit of manipulation, we can get the following one:

In the left hand side of the equation, it is the log odd of the probability that y belongs to a particular category. For example, if the probability of a success event is 0.8, then the odd ratio will be 4. The higher odds we get, the more likely we will get a success event.

The left-hand side is called the log-odds or logit, which is also the link function of logistics regression. We see that the logistic regression has a logit that is linear in .

For the illustration purpose, we can see the Example 6.3 from the textbook *Predictive Analytics: Paramertic Models for Regression and Classificantion*

We can see that the model is fitted using the log odd as y and the Eduation level as x. When eduation level increase from 1 to 2, then the log odd of visit is increased by .

### • The interpretaion of in logistics regression

The above equation also equals:

Here we assume that the we get is 0.3. From the above equation, we can see that when is increased by 1 unit, the odd ratio will increase by =1.3498588. means how much the log odd ratio will increase if our predictor is increased by 1 unit. In other words, if is increased by 1 unit, then the odds ratio will increase by . We can also get the increased probability of success from the increased odd ratio we get. For example, if the original is 0.8, then our odds ratio is 4. When we increased x by 1, the odds ratio increased by =1.3498588. Therefore, the increased odd ratio is 5.3994352. We can then get that our increased is 0.8437362.

### • Estimating the regression coefficients

***Why can't we use least square method to estimate the coefficient for logistic regression?***

***What is Maximum Likelihood Estimation(MLE)?***

To explain MLE intuitively, we can think of it as this way: According to the observed fact, what is the process that will maximize the probability for us to get this outcome? There is one explaination from Quora that explain it very well:

*You only get to see what the nature wants you to see. Things you see are facts. These facts have an underlying process that generated it. These process are hidden, unknown, needs to be discovered. Then the question is: Given the observed fact, what is the likelihood that process P1 generated it? What is the likelihood that process P2 generated it? And so on... One of these likelihoods is going to be max of all. MLE is a function that extracts that max likelihood.*

*Think of a coin toss; coin is biased. No one knows the degree of bias. It could range from o(all tails) to 1 (all heads). A fair coin will be 0.5 (head/tail equally likely). When you do 10 tosses, and you observe 7 Heads, then the MLE estimator for the prob of head of this coin is 0.7.*

*Think of a stock price of say, British Petroleum. BP was selling at $59.88 on April 23, 2010. By June 25, 2010 the price was down to $27.02. There could be several reasons for this fall. But the most likely reason could be the BP oil spill and the public sentiment. Stock price is the observed fact. The MLE will estimate the most likely underlying reason.*

***How to use MLE to estimate the coefficient of logistic regression?***

We are trying to find estimtaes for and , when we plug in these estimators into the our first equation above, yields a number close to 1 for all the observed y equals to 1; close to 0 for all observed y equals to 0. We can formulized the likelehood function as following:

By taking log from both side

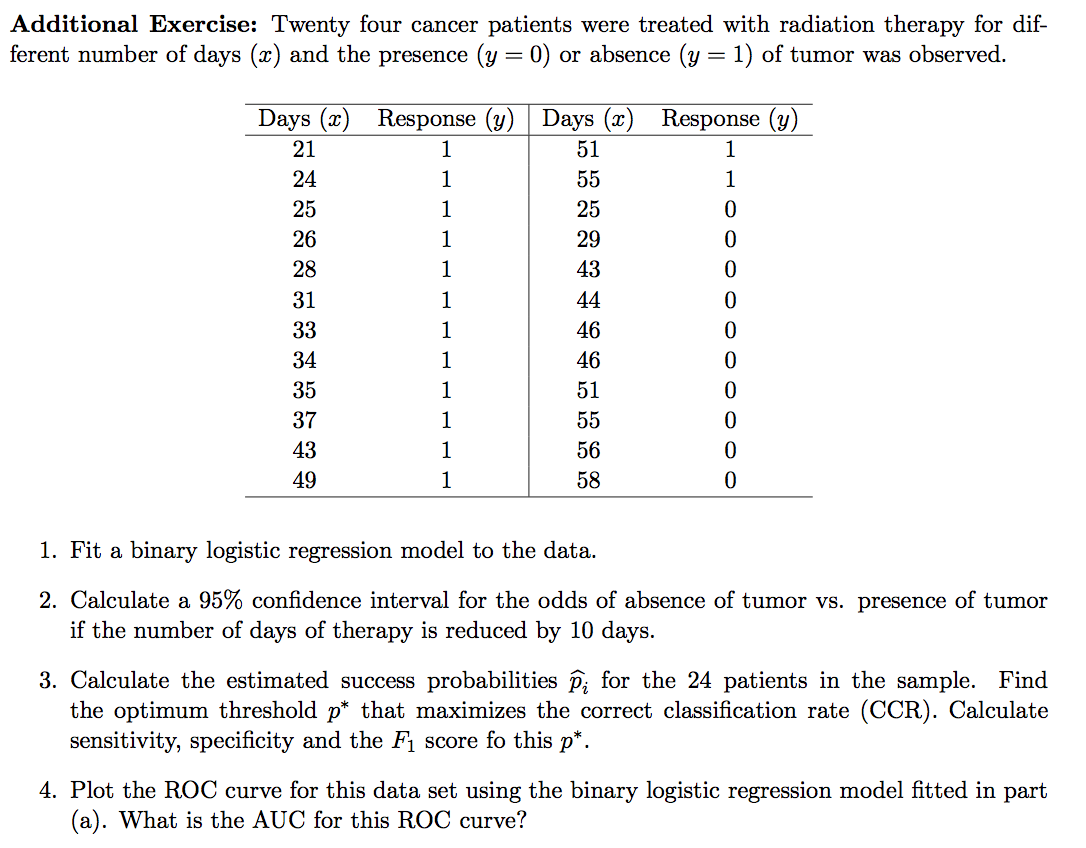
The maximizing values can be found by setting the partial derivatives of with respect to and .

### • Evaluating the correctness of the model

***What are sensitivity, specifity, recall, precision and what is the rationle behind it?***

***What is ROC curve?***

### • An example of building binary logistic regression

 ***Answer(1)***

# Make Dataframe  
df = data.frame(days=c(21,24,25,26,28,31,33,34,35,37,43,49,51,55,25,29,43,44,46,46,51,55,56,58),  
 response=c(rep(1,14),rep(0,10)))  
  
# Fit binary logistic regression  
fit = glm(response ~ days, family=binomial, data=df)  
summary(fit)

##   
## Call:  
## glm(formula = response ~ days, family = binomial, data = df)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.9141 -0.9818 0.5792 0.8411 1.5923   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 3.81944 1.83518 2.081 0.0374 \*  
## days -0.08648 0.04322 -2.001 0.0454 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 32.601 on 23 degrees of freedom  
## Residual deviance: 27.788 on 22 degrees of freedom  
## AIC: 31.788  
##   
## Number of Fisher Scoring iterations: 4

***Answer(2)*** From the result of our model, we can see that the 95% CI for =

Therefore, a 95% CI on the odds ratio equals the exponential of it, which is

***Answer(3)***

predict\_response = predict(fit, newdata=df, type='response')  
real\_response = df$response  
  
compare\_df = data.frame(real = real\_response, predict = predict\_response)  
compare\_df

## real predict  
## 1 1 0.8811429  
## 2 1 0.8511753  
## 3 1 0.8398841  
## 4 1 0.8279095  
## 5 1 0.8018532  
## 6 1 0.7573984  
## 7 1 0.7242222  
## 8 1 0.7066191  
## 9 1 0.6883758  
## 10 1 0.6501210  
## 11 1 0.5251471  
## 12 1 0.3969417  
## 13 1 0.3563618  
## 14 1 0.2814820  
## 15 0 0.8398841  
## 16 0 0.7877531  
## 17 0 0.5251471  
## 18 0 0.5035475  
## 19 0 0.4603893  
## 20 0 0.4603893  
## 21 0 0.3563618  
## 22 0 0.2814820  
## 23 0 0.2643258  
## 24 0 0.2320858

The predicted value for the 24 patients in the sample is shown in the table above in the predict column

get\_optimal\_p <- function(real\_response, predict\_response, p\_threshold\_list){  
 max\_ccr= 0  
 optimal\_p = 0  
 for (p in p\_threshold\_list){  
 pred = rep(0, 24)  
 pred[predict\_response > p]=1  
 ccr = sum(diag(table(real=real\_response, pred)))/ 24  
   
 if (ccr > max\_ccr){  
 max\_ccr= ccr  
 optimal\_p = p  
 }  
 }  
   
   
 pred = rep(0, 24)  
 pred[predict\_response > optimal\_p]=1  
 confusion\_table=table(real=real\_response, pred)  
  
 confusion\_table=table(real=real\_response, pred)  
 sensitivity=confusion\_table[2,2]/(confusion\_table[2,1]+confusion\_table[2,2])  
 specificity=confusion\_table[1,1]/(confusion\_table[1,1]+confusion\_table[1,2])  
 precision=confusion\_table[2,2]/(confusion\_table[1,2]+confusion\_table[2,2])   
 f1\_score=2\*precision\*sensitivity/(precision+sensitivity)  
   
 return(data.frame(p=optimal\_p,   
 ccr=max\_ccr,  
 sensitivity=sensitivity,   
 specificity=specificity,  
 f1\_score=f1\_score))  
}

The sensitivity, specificity and the F1 score fo the optimal p∗ is shown in the table below

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| p | ccr | sensitivity | specificity | f1\_score |
| 0.51 | 0.75 | 0.7857143 | 0.7 | 0.7857143 |

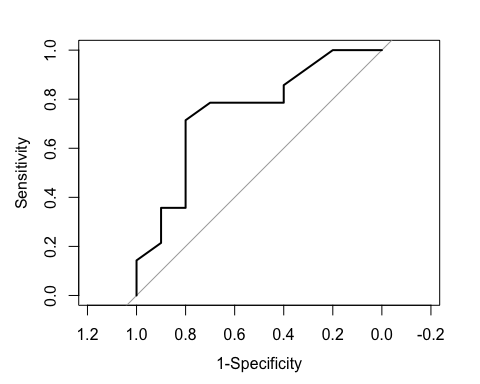
***Answer(4)***

library(pROC)

summary(fit)

##   
## Call:  
## glm(formula = response ~ days, family = binomial, data = df)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.9141 -0.9818 0.5792 0.8411 1.5923   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 3.81944 1.83518 2.081 0.0374 \*  
## days -0.08648 0.04322 -2.001 0.0454 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 32.601 on 23 degrees of freedom  
## Residual deviance: 27.788 on 22 degrees of freedom  
## AIC: 31.788  
##   
## Number of Fisher Scoring iterations: 4

plot.roc(real\_response, fit$fitted.values, xlab="1-Specificity")



my\_auc = auc(real\_response, fit$fitted.values)

The AUC is **0.75**

## Multinomial Logistics Regression

For the data that the outcome is nominal or ordinal, we cannot apply binary logistic regression to our data anymore. One reason is that usually both types of data have more than 2 categories, and ordinal data is a categorical, statistical data type where the variables have natural, ordered categories and the distances between the categories is not known.

Let's see how we can build a regression model on these types of data.

### • Logistics Regression for Nominal Response

Let's assume that we have 3 categorical responses(A,B,C) and 6 predictors including the intercept now. How do we build up a logistic regression model on it? One way to do it is to set one response category as reference and compare the probabilities of the remaining 2 responses to it. Let's denote the probability of response by , and set category as the reference. Then we have the following equations

More generally, we can rewrite it as:

Therefore, in our example, we will need to estimate different 's to build up a model.

***How do we interpret the meaning of all the in the nominal logistic regression model?***

The explanation is quoted from the textbook *Predictive Analytics: Paramertic Models for Regression and Classificantion* explain it very well:

"The interpretation of the coefficient is similar to that of the coefficient for binary logistic regression. It is the change in the log-odds of response ={a or b} relative to that of the regerence category c when the predictor variable is increased by one unit keeping all other variables fixed.

As an example, suppose then . Hence the odds of outcome vesus outcome increased by a factor of if is increased by one unit"

***How to get the predicted probability for different category responses?***

rom the above equation, it follows that

If we denote category {A,B,C} as {1,2,3}, since , we can see that

By solving , we can get

where and

Also, it follows that

To sum up, for given vector, we can then calculate the probability of it for category {A,B,C}, and use the category with the highest predicted probability as the predicted category. We refer to the predicted value we get here as **Maximum probability classifier**

### • Bayes Classification

To go further, we can also use *prior probabilities* as the weight to the predicted probability that we calculate previous to get our bayes predicted probability. Suppose we have available prior probabilities, . Then by using the follow equation derived from *bayes formula*, we can get the *posterior probabilities*

where is the Maximum probability classifier we calculate previously.

In our example, we can get the posterior probabilities as

We refer to the predicted value we get here as **Bayes classifier**. One thing worth mention is that if the prior probabilities are equal, then the Bayes classifier is actually reduce to the maximum probability classifier.

### • Logistics Regression for Ordinal Response

For the categorical response that is ordinal, we build up our linear model on cumulative logits to capture the intrinsically ranked feature of it. The equation is shown as follows:

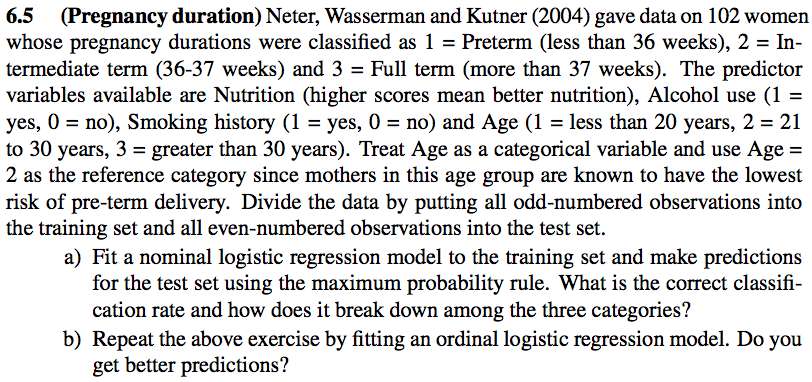
Note that , and the equation can also be rewrite as

If the category response in our preivous example {A,B,C} is ordinal, then the response value of the model we build is

Therefore, we can get that

We can also build up it and apply prior probability to make it into a Bayes classifier.

### • An example of building logistic regression for nomimal and ordinal response



***Answer(a)***

The total *True Positive* number is 29, and we have 51 observations in our test dataset. Therefore, the overall correct classification rate is approximately **0.5686**.

library(mlogit)

# Read Data  
pregnancy\_df = read.csv('\_data/Pregnancy.csv', colClasses=c('factor','integer','factor','factor','factor'))  
  
# Transform Data  
pregnancy\_df$Age <- relevel(pregnancy\_df$Age, ref = "2")  
  
# Make train and test data  
train\_indices <- seq(1, nrow(pregnancy\_df), by=2)  
pregnancy\_train <- pregnancy\_df[train\_indices,]  
pregnancy\_test <- pregnancy\_df[-train\_indices,]  
  
# Build multinominal logistic regression  
pregnancy\_train\_mlogit = mlogit.data(data=pregnancy\_train, choice='Duration', shape='wide')  
model\_65\_1 = mlogit(Duration ~ 0 | Nutrition+Alcohol+Smoking+Age, data=pregnancy\_train\_mlogit)  
  
# Make prediction for pregnancy\_test  
pregnancy\_test\_mlogit = mlogit.data(data=pregnancy\_test, choice='Duration', shape='wide')  
pregnancy\_prediction\_df = data.frame(predict(model\_65\_1, newdata = pregnancy\_test\_mlogit))  
colnames(pregnancy\_prediction\_df) = c(1,2,3)  
pregnancy\_prediction = as.integer(colnames(pregnancy\_prediction\_df)[apply(pregnancy\_prediction\_df,1,which.max)])  
  
# make confusion table  
pregnancy\_acutal = pregnancy\_test$Duration  
print(table(actual = pregnancy\_acutal, prediction = pregnancy\_prediction))

## prediction  
## actual 1 2 3  
## 1 4 6 3  
## 2 4 9 4  
## 3 2 3 16

# calculate the correct classification rate   
sum(diag(table(pregnancy\_prediction, pregnancy\_acutal)))/ dim(pregnancy\_test)[1]

## [1] 0.5686275

From the confusion table above, we can see that the correct classification rate for each category is

|  |  |
| --- | --- |
| Category | Rate |
| 1 | 0.3076923 |
| 2 | 0.5294118 |
| 3 | 0.7619048 |

***Answer(b)***

The total *True Positive* number is 34, and we have 51 observations in our test dataset. Therefore, the overall correct classification rate is approximately **0.6667**, which is higher than the result of multinominal logistic regression.

library(ordinal)

# Transform Data  
pregnancy\_train$Duration = as.ordered(pregnancy\_train$Duration)  
  
# Build ordinal logistic regression  
model\_65\_2 = clm(Duration~ . ,data=pregnancy\_train)  
  
# Make prediction for pregnancy\_test (https://www.rdocumentation.org/packages/ordinal/versions/2015.6-28/topics/predict.clm)  
pregnancy\_prediction\_2 = predict(model\_65\_2, newdata = pregnancy\_test, type="class")$fit  
  
# make confusion table  
pregnancy\_acutal = pregnancy\_test$Duration  
print(table(actual = pregnancy\_acutal, prediction = pregnancy\_prediction\_2))

## prediction  
## actual 1 2 3  
## 1 5 5 3  
## 2 3 10 4  
## 3 2 0 19

# calculate the correct classification rate   
sum(diag(table(actual = pregnancy\_acutal, prediction = pregnancy\_prediction\_2))) / dim(pregnancy\_test)[1]

## [1] 0.6666667

## Reference

* [How do you explain MLE intuitively from Quora](https://www.quora.com/How-do-you-explain-maximum-likelihood-estimation-intuitively)
* [An Introduction to Statistical Learning with Applications in R](http://www-bcf.usc.edu/~gareth/ISL/)
* Predictive Analytics: Paramertic Models for Regression and Classificantion by Ajit C. Tamhane and Edward C. Malthouse