

logistic regression

Amira Ibrahim

September 5, 2019

```
library(readr)
Logestic_Data <- read.csv(file.choose(),header=TRUE)
```

make sure that categorical variables read as a factor :

```
Logestic_Data$outcome <- as.factor(Logestic_Data$outcome)
Logestic_Data$Treatment <- as.factor(Logestic_Data$Treatment)
Logestic_Data$SEX <- as.factor(Logestic_Data$SEX)
```

```
library(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
```

```
logistic_regression <- Logestic_Data %>%
```

```
mutate( Outcome_dummy = (ifelse(outcome == "Success" , 1 , 0)), Treatment_dummy = (ifelse(Treatment == "Control" , 0 , 1)))
```

```
attach(logistic_regression)
```

Check normality

```
library(moments)
skewness(HDL_LEVEL)
```

```
## [1] 0.2515518
```

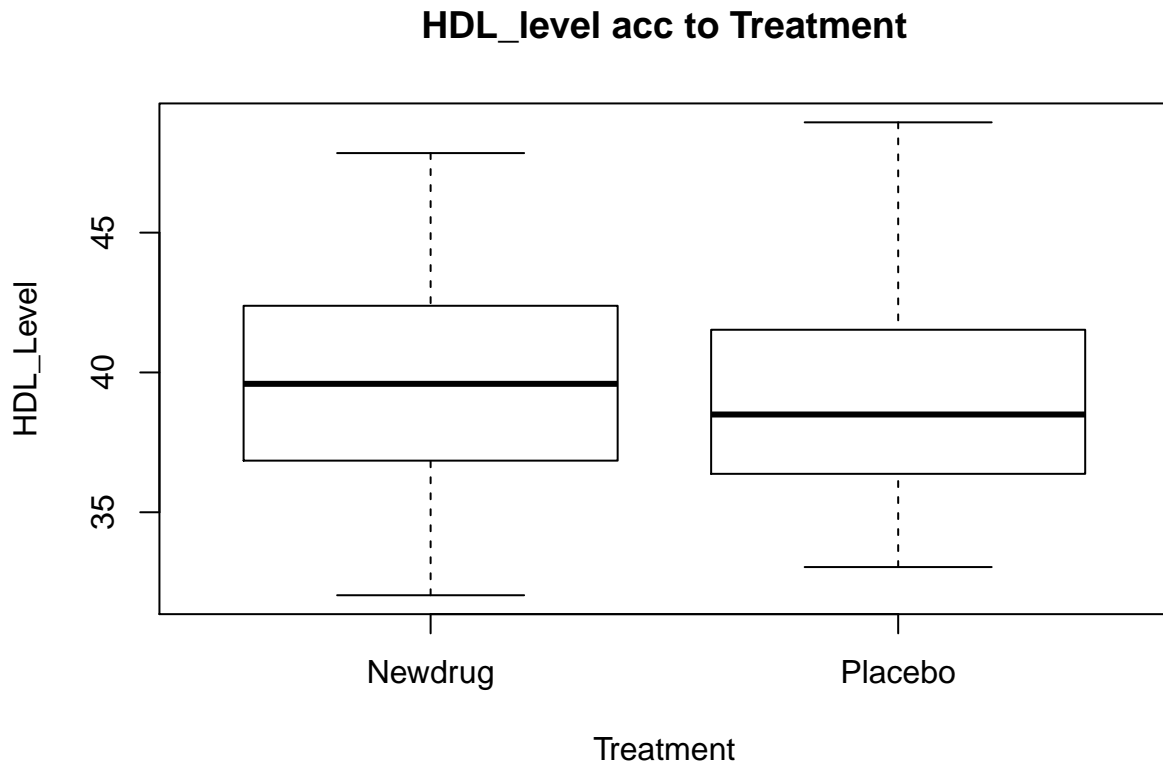
```
accepted range from -1 to +1
```

```
kurtosis(HDL_LEVEL)
```

```
## [1] 2.523675
```

```
accepted range from -2 to +2 , may to +3
```

```
boxplot(HDL_LEVEL~Treatment , xlab = "Treatment" , ylab = "HDL_Level" , main = "HDL_level acc to Treatment")
```



Visually , Data is normaly distributed

```
shapiro.test(HDL_LEVEL)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  HDL_LEVEL
## W = 0.98428, p-value = 0.2814
P-value > 0.05 , so it is normally distributed
```

test H0: No Change in HDL level according to Treatment :

```
t.test(HDL_LEVEL~Treatment , mu = 0 , paired= FALSE , alternative= "two.sided" ,data=logistic_regression)
```

```
##
##  Welch Two Sample t-test
##
## data:  HDL_LEVEL by Treatment
## t = 0.62698, df = 97.681, p-value = 0.5321
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.016751  1.955911
## sample estimates:
## mean in group Newdrug mean in group Placebo
##           39.56499           39.09541
```

P-value > 0.05 , fail to reject H0

Test H0 : No Difference in outcome Between Placebo & new drug

```
chisq.test(table(outcome,Treatment))
```

```
##  
## Pearson's Chi-squared test with Yates' continuity correction  
##  
## data:  table(outcome, Treatment)  
## X-squared = 25.04, df = 1, p-value = 5.615e-07  
p-value < 0.05 , Reject H0
```

Test H0 : Sex has no effect on outcome

```
chisq.test(table(outcome,SEX))
```

```
## Warning in chisq.test(table(outcome, SEX)): Chi-squared approximation may  
## be incorrect  
##  
## Pearson's Chi-squared test with Yates' continuity correction  
##  
## data:  table(outcome, SEX)  
## X-squared = 0.68114, df = 1, p-value = 0.4092  
P-value > 0.05 , Fail to reject H0
```

Determination of correlation between treatment & outcome :

```
cor(Outcome_dummy , Treatment_dummy , method = "spearman" )
```

```
## [1] 0.5204165
```

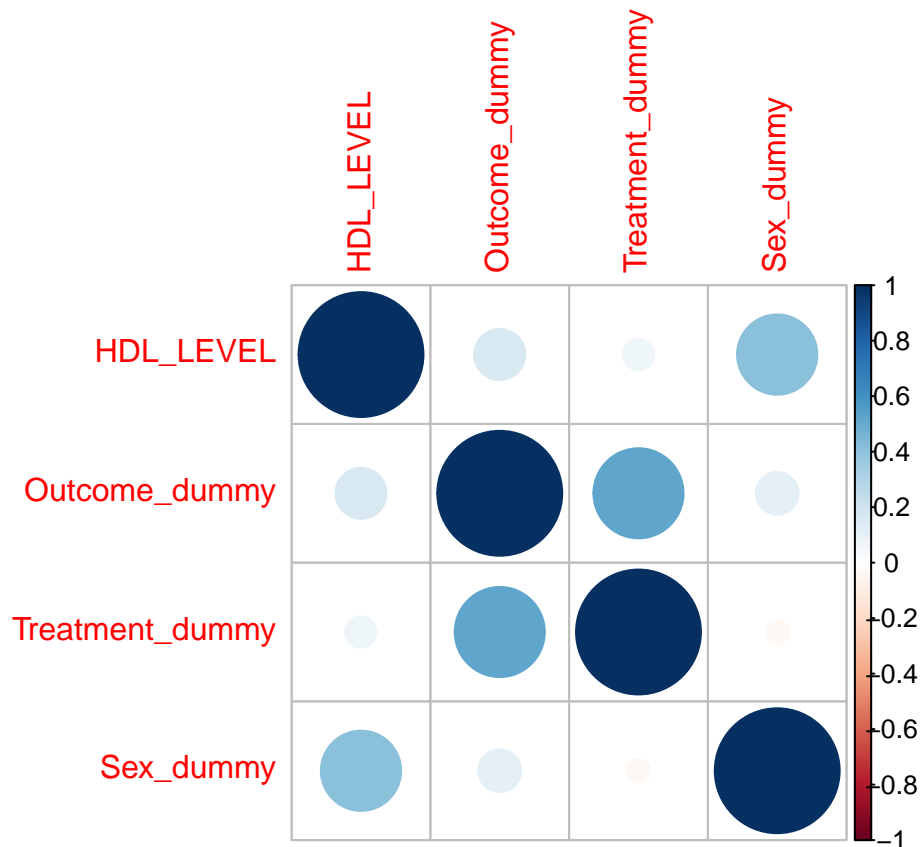
It is +ve moderate correlated

Find correlation between variables :

```
correlations <- cor(logistic_regression[,4:7])  
library(corrplot)
```

```
## Warning: package 'corrplot' was built under R version 3.6.1  
## corrplot 0.84 loaded
```

```
corrplot(correlations , method = "circle")
```



```
correlations
```

```
##           HDL_LEVEL Outcome_dummy Treatment_dummy  Sex_dummy
## HDL_LEVEL      1.00000000      0.1673886      0.06320816  0.41327986
## Outcome_dummy  0.16738857      1.00000000      0.52041650  0.11750194
## Treatment_dummy 0.06320816      0.5204165      1.00000000 -0.03494283
## Sex_dummy      0.41327986      0.1175019     -0.03494283  1.00000000
```

there is -ve weak correlation between sex & treatment

there is +ve weak correlation between HDL level & treatment

there is +ve moderate correlation between outcome & treatment

Bad model in log regrission :

```
badmodel <- glm(outcome~Treatment+SEX+HDL_LEVEL , family = "binomial" , data = logistic_regression)
summary(badmodel)
```

```
##
## Call:
## glm(formula = outcome ~ Treatment + SEX + HDL_LEVEL, family = "binomial",
##      data = logistic_regression)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8707  -0.6796  -0.5550   0.8054   1.8880
```

```
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.61980    2.79087  -0.580    0.562
## TreatmentPlacebo -2.38664    0.48757  -4.895 9.83e-07 ***
## SEXMale        0.90217    0.94601   0.954    0.340
## HDL_LEVEL      0.06643    0.07149   0.929    0.353
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 138.47  on 99  degrees of freedom
## Residual deviance: 106.57  on 96  degrees of freedom
## AIC: 114.57
##
## Number of Fisher Scoring iterations: 4
```

determination of refrence value :

```
logistic_regression <- within(logistic_regression, Treatment<-relevel(Treatment, ref="Placebo"))
goodmodel <- glm(outcome~Treatment+SEX+HDL_LEVEL , family = "binomial" , data = logistic_regression)
summary(goodmodel)
```

```
##
## Call:
## glm(formula = outcome ~ Treatment + SEX + HDL_LEVEL, family = "binomial",
##      data = logistic_regression)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8707  -0.6796  -0.5550   0.8054   1.8880
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -4.00643    2.80069  -1.431    0.153
## TreatmentNewdrug 2.38664    0.48757   4.895 9.83e-07 ***
## SEXMale        0.90217    0.94601   0.954    0.340
## HDL_LEVEL      0.06643    0.07149   0.929    0.353
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 138.47  on 99  degrees of freedom
## Residual deviance: 106.57  on 96  degrees of freedom
## AIC: 114.57
##
## Number of Fisher Scoring iterations: 4
```

```
exp(goodmodel$coefficients)
```

```
##      (Intercept) TreatmentNewdrug      SEXMale      HDL_LEVEL
##      0.01819817      10.87684799      2.46493495      1.06869133
```

```
library(Greg)
```

```
## Warning: package 'Greg' was built under R version 3.6.1
## Loading required package: forestplot
## Warning: package 'forestplot' was built under R version 3.6.1
## Loading required package: grid
## Loading required package: magrittr
## Loading required package: checkmate
## Loading required package: Gmisc
## Warning: package 'Gmisc' was built under R version 3.6.1
## Loading required package: Rcpp
## Warning: package 'Rcpp' was built under R version 3.6.1
## Loading required package: htmlTable
## Registered S3 methods overwritten by 'ggplot2':
##   method      from
##   [.quosures   rlang
##   c.quosures   rlang
##   print.quosures rlang
```

```
printCrudeAndAdjustedModel(goodmodel)[-1,]
```

Crude

Adjusted

Variable

OR

2.5 % to 97.5 %

OR

2.5 % to 97.5 %

Newdrug

10.09

4.15 to 26.41

10.88

4.35 to 29.76

Male

2.33

0.58 to 11.59

2.46

0.41 to 17.85

HDL_LEVEL

1.10

0.99 to 1.23

1.07

0.93 to 1.24

cruded value (the value of OR in the presence of other variables)

the odds of success for patient recieving newdrug taking the considration the sex and HDL level is 10.09

the odds of success for a male patient who is recieving newdrug and has HDL level is 2.33

the odds of success for a patient in the presence of HDL level taking the considration the sex and receiving new drug is 1.10

Adjusted values (the value of OR of variables without the presences of other variables)

the odds of the success for the patient recieving new drug is 10.88

the odds of the success for a male patient is 2.46

the odds of success for a patient with HDL level is 1.07