***1. You need to predict the type of program a student is in based on other attributes --> mdata.csv***

***prog: is a categorical variable indicating what type of program a student is in: ?General? (1), ?Academic? (2), or ?Vocational? (3)***

***Ses:?is a categorical variable indicating someone?s socioeconomic class: ?Low? (1), ?Middle? (2), and ?High? (3)***

***read,write,math,science: is their scores on different tests***

***honors: Whether they have enrolled or not***

**Inferences from the Data Set:**

Data Set talks about the predict the type of program a student is in based on other attributes is a binary variable with respect to around 8 Independent variables & 200 observations.

**Data Set Size:**

Data gives 9 columns and 200 rows found to be a binary data for which a multinomial regression can be performed getting deeper into the data analysis and its behavior

Here we creating a model output of the data set is more than two variables is considered as multinomial regression for this we take reference output as **”academic”.**

Summary of the model gives coefficients and standard errs and the coefficients are in negative because the model is log of output converts as odds ratio i.e; exponential of that coefficients gives positive coefficients.

**p- value calculations:**

we run our model using multinom. The multinom package does not include p-value calculation for the regression coefficients, so we calculate p-values using 2-tailed z test we get p values

**p\_value <- (1-pnorm(abs(z),0,1))\*2.**

for this p -values must satisfies the condition that statistically significant p-value < 0.05

**Find the accuracy of the model:**

For finding the accuracy of the model, we have to predict the probabilities.

**prob <- fitted(model)**

**Find the accuracy of the model:**

After predicting the probabilities set the prob. into data frame

**prob <- data.frame(prob)**

and create a column as pred and get the names like academic vocation and general.

**Creating Confusion matrix:**

Create confusion matrix find out the best accuracy of the model / best fit model for the out multinomial regression.

i.e; table of predicted values to actual gives us confusion matrix table

**table(pred\_name,Mdata1$prog)**

pred\_name academic general vocation

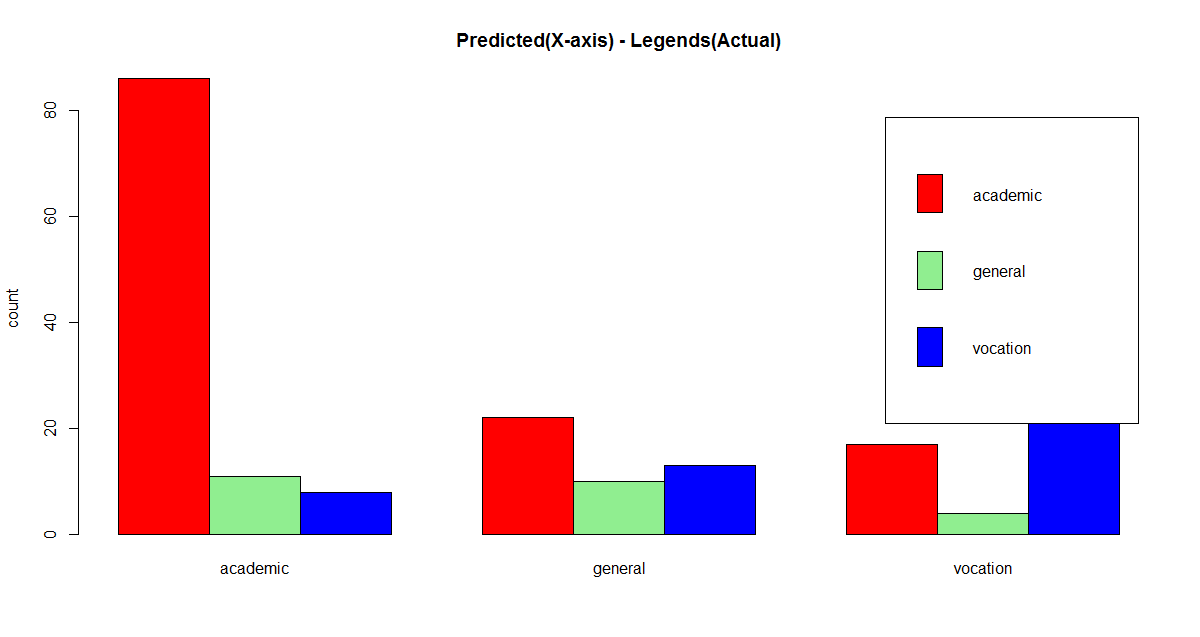
academic 86 22 17

general 11 10 4

vocation 8 13 29

**confusion matrix visualization:**

barplot(table(pred\_name,Mdata1$prog),beside = T,col=c("red","lightgreen","blue"),legend=c("academic","general","vocation"),main = "Predicted(X-axis) - Legends(Actual)",ylab ="count")



By seeing the plot we can understand red color shows as academic lite green shows as general and blue color shows as vocation and frequency of academic is very high compare to the other.

And inference/we say that the student is **Academic** of categorical program indicating.

**Accuracy**

**mean(pred\_name==Mdata1$prog)**

we can say about accuracy of the model by using above equation that mean of comparing predicted name to the actual value of the categorical output give accuracy of the model and this model gives 62%

and applying transformation to the model t get the better accuracy of the model.

**R-code:**

**# Multinomial Logit Model**

**# packages required**

**require('mlogit')**

**require('nnet')**

#In built dataset

data()

Mdata <- read.csv("C:/RAVI/Data science/Assignments/Module 10 Multinomial Regression/Dataset/mdata.csv")

View(Mdata)

attach(Mdata)

head(Mdata)

table(Mdata$prog)

summary(Mdata)

Mdata1 <- Mdata[ ,-c(1,2)]

View(Mdata1)

model <- multinom(prog ~ female + ses + schtyp + read + write + math + science + honors, data=Mdata1)

summary(model)

Mdata1$prog <- relevel(Mdata1$prog, ref= "academic") # change the baseline level

##### Significance of Regression Coefficients###

z <- summary(model)$coefficients / summary(model)$standard.errors

z

#2-tailed z test we get p values

p\_value <- (1-pnorm(abs(z),0,1))\*2

summary(model)$coefficients

p\_value

# odds ratio

exp(coef(model))

# predict probabilities

prob <- fitted(model)

prob

# Find the accuracy of the model

class(prob)

prob <- data.frame(prob)

View(prob)

prob["pred"] <- NULL

# Custom function that returns the predicted value based on probability

get\_names <- function(i){

return (names(which.max(i)))

}

pred\_name <- apply(prob,1,get\_names)

?apply

prob$pred <- pred\_name

View(prob)

# Confusion matrix

table(pred\_name,Mdata1$prog)

# confusion matrix visualization

barplot(table(pred\_name,Mdata1$prog),beside = T,col=c("red","lightgreen","blue"),legend=c("academic","general","vocation"),main = "Predicted(X-axis) - Legends(Actual)",ylab ="count")

# Accuracy

mean(pred\_name==Mdata1$prog)