Usage: When the dependant variable (Y variable) is binary

Example: Develop a model to predict the number of visits of family to a vacation resort based on the salient characteristics of the families. The data collected from 30 households is given in Resort\_Visit.csv

BINARY LOGISTIC REGRESSION

Indian Statistical Institute

1.Reading the file and variables

> mydata = Resort\_Visit

> visit = mydata$Resort\_Visit

> income = mydata$Family\_Income

> attitude = mydata$Attitude.Towards.Travel

> importance = mydata$Importance\_Vacation

> size = mydata$House\_Size > age = mydata$Age.\_Head

2. Converting response variable to discrete

> visit = factor(visit)

Correlation Matrix

> cor(mydata)

Interpretation: Correlation between X variables should be low

3.Converting response variable to discrete

> visit = factor(visit)

5. Checking relation between Xs and Y

> aggregate(income ~visit, FUN = mean)

> aggregate(attitude ~visit, FUN = mean)

> aggregate(importance ~visit, FUN = mean)

> aggregate(size ~visit, FUN = mean)

> aggregate(age ~visit, FUN = mean)

Higher the difference in means, stronger will be the relation to response variable

6. Checking relation between Xs and Y – box plot

> boxplot(income ~ visit)

> boxplot(attitude ~ visit)

> boxplot(importance ~ visit)

> boxplot(size ~ visit)

> boxplot(age ~ visit)

6. Perform Logistic regression

> model = glm(visit ~ income + attitude + importance + size + age, family = binomial(logit))

> summary(model)

6. Perform Logistic regression - Anova > anova(model,test = 'Chisq')

> summary(model)

Since p value < 0.05 for Income, Importance Vacation & Size, redo the modelling with important factors only

7. Perform Logistic regression – Modified

Since p value < 0.05 for both factors, Income & Size, the response variable can be modelled in terms of those two factors

8. Fitted Values and residuals

> predict(model,type = 'response')

> residuals(model,type = 'deviance')

> predclass = ifelse(predict(model, type ='response')>0.5,"1","0")

9. Model Evaluation

> mytable = table(visit, predclass)

> mytable

> prop.table(mytable)

Exercise 2: A car rental company wants to develop a model for brand loyalty. The data was collected from 30 customers, 15 of whom are brand loyal (indicated by 1) and 15 of whom are not (indicated by 0). The company also measured attitude towards the brand (Brand), attitude towards the type of vehicle (vehicle) and attitude toward availing rent a car service (Service), all on a 1 (unfavorable) to 7 (favorable) scale. The data is given in brand.csv file.