**Comandos R para Modelos de Regresión Logística y Modelos de Elección Discreta**

**Lectura de datos**

datos=read.table*(“clipboard",* header=TRUE, sep=“ ")

datos=read.table*(“clipboard",* header=TRUE, sep="\t")

datos=read.csv*(“dirección archivo CSV",* header=TRUE, sep=”\t”)

datos=read.csv2*(“dirección archivo CSV",* header=TRUE, sep=”\t”)

**Exploración de datos**

dim(datos)

head(datos)

summary(datos)

**Estimación modelo regresión logística**

fit1 = glm(compra ~ edad, data = datos, family = binomial(link = "logit"))

fit2 = glm(compra ~ sexo, data = datos, family = binomial(link = "logit"))

fit3 = glm(compra ~ edad + sexo, data = datos, family = binomial(link="logit"))

summary(fit1)

summary(fit2)

summary(fit3)

**Predicción probabilidad según variable edad**

datosp=datos

m=apply(datos,2,mean)

for(i in 1: dim(datos)[1]) {datosp[i,]=m}

datosp$compra=datos$compra

datosp$edad=datos$edad

datosp$predict=predict(fit3,type="response",datosp)

ggplot(datosp)+aes(y=compra,x=edad)+

geom\_point(size=2,alpha=0.25)+ geom\_line(aes(y=predict),color="steelblue",size=2)+

labs(y="Probabilidad",x="Edad")

**Matriz de confusión**

# Instala paquete ROCR

install.packages("ROCR")

library(ROCR)

# Suponga que el data.frame “data” tiene la variable dependiente Y

# y la predicción de probabilidad PROB. La matriz de confusion require # # el siguiente código

**Estimación modelo de elección discreta**

# Instala paquete “mlogit”

install.packages("mlogit")

library(mlogit)

# El data.frame **data** tiene la forma larga en la cual cada fila es una combinación: (individuo,

# alternativa), la variable dependiente es la elección que corresponde a una variable 0 ó 1,

# y las variables independientes X1, X2, X3 son específicas a las alternativas (para el mismo

# individuo toman distinto valor en cada alternativa).

# Estimación del modelo que se almacena en **fit**

fit=mlogit(Y~X1+X2+X3,data)

summary(fit)

# Estimación del modelo sin constantes de las alternativas

fit=mlogit(Y~X1+X2+X3|0,data)

# Si una variable Z es específica al individuo y no varía según las alternativas, se debe indicar

# después del signo **|**

fit=mlogit(Y~X1+X2+X3|Z,data)

# Predicciones ajustadas

fitted(fit,type="probabilities")