Practica\_Clasificación\_1

Valentina Díaz Torres

## Introducción

El objetivo de esta práctica es realizar los ejercicios 4 y 5 del fichero de Prácticas de la Semana 1, sobre regresión logística y el ejercicio 4, los apartados (i) y (iii) del ejercicio 4, con respecto a la semana dos, relacionada con análisis discriminante.

library(foreign)  
library(nnet) # multinomial  
library(stargazer)  
library(ISLR)  
library(Deducer)  
library(pscl)  
library(wooldridge)  
library(MASS)  
library(ISLR)  
library(DAAG)  
library(pastecs)  
library(klaR)  
library(corrplot)

## Prácticas Semana 1

### Ejercicio 4

1. La encuesta Wage del paquete ISLR contiene información sobre salarios y otras variables para un grupo de 3000 trabajadores hombres de la región Mid-Atlantic de USA.

**(i)Construir un modelo de regresión logística para explicar la variable health de las variables age, race, education y logwage**

#Construcción del modelo de regresión logística  
  
  
reg.logs <- glm(health ~ age + race + education + logwage, data=Wage,   
 family=binomial(logit))  
  
# Datos residuales del modelo de regresión lineal simple (diferencia entre los datos observados de la variable dependiente y y los valores ajustados ŷ)  
  
head(reg.logs$residuals,20)

## 231655 86582 161300 155159 11443 376662 450601 377954   
## -4.329341 1.207812 -3.806592 1.231090 -2.173620 1.344103 1.409162 -4.453001   
## 228963 81404 302778 305706 8690 153561 449654 447660   
## 1.351297 1.552778 -3.529053 1.450687 1.426741 1.192407 1.568257 1.621238   
## 160191 230312 301585 153682   
## 1.335697 1.145579 -3.749377 1.350546

#Medida de bondad de ajuste de un modelo lineal generalizado.   
  
reg.logs$deviance

## [1] 3378.41

#La desviación nula muestra qué tan bien se predice la variable de respuesta mediante un modelo que incluye solo la intersección  
  
head(reg.logs$null.deviance,20)

## [1] 3591.187

head(reg.logs$coefficients,20)

## (Intercept) age   
## -2.19222715 -0.03514899   
## race2. Black race3. Asian   
## -0.03604936 -0.17160044   
## race4. Other education2. HS Grad   
## -0.41408867 0.08208608   
## education3. Some College education4. College Grad   
## 0.25867828 0.63781151   
## education5. Advanced Degree logwage   
## 0.87664718 0.93275222

head(reg.logs$fitted.values,20)

## 231655 86582 161300 155159 11443 376662 450601 377954   
## 0.7690179 0.8279437 0.7372978 0.8122882 0.5399379 0.7439908 0.7096416 0.7754323   
## 228963 81404 302778 305706 8690 153561 449654 447660   
## 0.7400300 0.6440072 0.7166378 0.6893285 0.7008979 0.8386398 0.6376506 0.6168127   
## 160191 230312 301585 153682   
## 0.7486730 0.8729211 0.7332890 0.7404415

summary(reg.logs) #para resumir sus datos, valores mínimos, máximos, cuantiles, p-values, nivel de significancia, etc.

##   
## Call:  
## glm(formula = health ~ age + race + education + logwage, family = binomial(logit),   
## data = Wage)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.3091 -1.2063 0.6590 0.8436 1.7613   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.192227 0.621535 -3.527 0.00042 \*\*\*  
## age -0.035149 0.003718 -9.455 < 2e-16 \*\*\*  
## race2. Black -0.036049 0.138152 -0.261 0.79414   
## race3. Asian -0.171600 0.180124 -0.953 0.34075   
## race4. Other -0.414089 0.351048 -1.180 0.23817   
## education2. HS Grad 0.082086 0.146255 0.561 0.57462   
## education3. Some College 0.258678 0.159104 1.626 0.10398   
## education4. College Grad 0.637812 0.166661 3.827 0.00013 \*\*\*  
## education5. Advanced Degree 0.876647 0.198000 4.428 9.53e-06 \*\*\*  
## logwage 0.932752 0.141471 6.593 4.30e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 3591.2 on 2999 degrees of freedom  
## Residual deviance: 3378.4 on 2990 degrees of freedom  
## AIC: 3398.4  
##   
## Number of Fisher Scoring iterations: 4

**(ii)Bondad de ajuste mediante test LR y pseudoR2 de McFadden**

#Bondad de ajuste LR test  
  
#Generalizamos el modelo de regresión lineal con las variables health, age, race, education y logwage  
  
model\_all <- glm(health ~ age + race + education + logwage, data=Wage, family = binomial(link = logit))  
  
summary(model\_all)

##   
## Call:  
## glm(formula = health ~ age + race + education + logwage, family = binomial(link = logit),   
## data = Wage)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -2.3091 -1.2063 0.6590 0.8436 1.7613   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.192227 0.621535 -3.527 0.00042 \*\*\*  
## age -0.035149 0.003718 -9.455 < 2e-16 \*\*\*  
## race2. Black -0.036049 0.138152 -0.261 0.79414   
## race3. Asian -0.171600 0.180124 -0.953 0.34075   
## race4. Other -0.414089 0.351048 -1.180 0.23817   
## education2. HS Grad 0.082086 0.146255 0.561 0.57462   
## education3. Some College 0.258678 0.159104 1.626 0.10398   
## education4. College Grad 0.637812 0.166661 3.827 0.00013 \*\*\*  
## education5. Advanced Degree 0.876647 0.198000 4.428 9.53e-06 \*\*\*  
## logwage 0.932752 0.141471 6.593 4.30e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 3591.2 on 2999 degrees of freedom  
## Residual deviance: 3378.4 on 2990 degrees of freedom  
## AIC: 3398.4  
##   
## Number of Fisher Scoring iterations: 4

# Modelo de regresión lineal para la variable health  
  
model\_health <- glm(health ~ 1, data=Wage, family = binomial(link = logit))  
  
#test anova  
anova(reg.logs,test="Chisq")

## Analysis of Deviance Table  
##   
## Model: binomial, link: logit  
##   
## Response: health  
##   
## Terms added sequentially (first to last)  
##   
##   
## Df Deviance Resid. Df Resid. Dev Pr(>Chi)   
## NULL 2999 3591.2   
## age 1 58.041 2998 3533.1 2.567e-14 \*\*\*  
## race 3 5.909 2995 3527.2 0.1161   
## education 4 104.289 2991 3422.9 < 2.2e-16 \*\*\*  
## logwage 1 44.539 2990 3378.4 2.494e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# Odds  
  
exp(coef(model\_all))

## (Intercept) age   
## 0.1116678 0.9654616   
## race2. Black race3. Asian   
## 0.9645927 0.8423157   
## race4. Other education2. HS Grad   
## 0.6609423 1.0855493   
## education3. Some College education4. College Grad   
## 1.2952170 1.8923350   
## education5. Advanced Degree logwage   
## 2.4028299 2.5414943

# Intervalos de confianza  
  
confint(model\_all)

## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) -3.41500441 -0.97719004  
## age -0.04247619 -0.02789839  
## race2. Black -0.30410307 0.23795441  
## race3. Asian -0.51903665 0.18838458  
## race4. Other -1.09411041 0.29252909  
## education2. HS Grad -0.20621174 0.36753703  
## education3. Some College -0.05426872 0.56981917  
## education4. College Grad 0.31063209 0.96436564  
## education5. Advanced Degree 0.49054822 1.26736804  
## logwage 0.65687962 1.21177242

# Pseudo R2 McFadden   
pR2(model\_all)

## fitting null model for pseudo-r2

## llh llhNull G2 McFadden r2ML   
## -1.689205e+03 -1.795594e+03 2.127775e+02 5.924990e-02 6.846901e-02   
## r2CU   
## 9.810443e-02

# McFadden = 0.05925  
  
# Calculo directo McFadden  
  
1-model\_all$deviance/model\_health$deviance

## [1] 0.0592499

**(iii)Matriz de confusión y curva ROC**

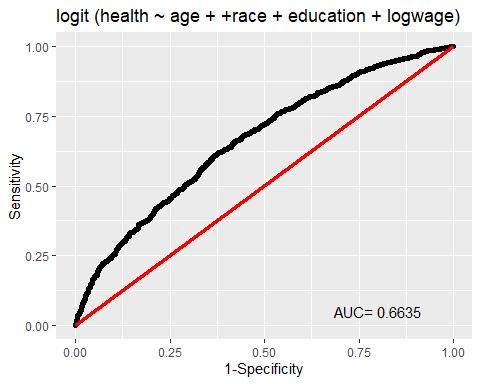
#Matriz de confusión   
  
fit.pred <- ifelse(model\_all$fitted.values>0.5,1,0)  
tabla<-table(fit.pred, Wage$health)  
tabla

##   
## fit.pred 1. <=Good 2. >=Very Good  
## 0 88 73  
## 1 770 2069

(tabla[1,1]+tabla[2,2])/sum(tabla)

## [1] 0.719

#curva ROC  
  
model2 <- glm(health ~ age++race+education+logwage, data=Wage, family = binomial(link = logit))  
  
model2 <- glm(formula=health ~ age++race+education+logwage,data=Wage, family = binomial(link = logit), na.action=na.omit)  
rocplot(model2)



#la curva aparece representada en negro

## Práctica 5

El conjunto de datos happines, del paquete wooldridge, proporciona información sobre el nivel de felicidad de una muestra 17131 individuos encuestados entre 1994 y 2006. Los datos también incluyen información acerca de una serie de características sociodemográfica de los encuestados.

**(i)Estimar el modelo de regresión lineal incluye como variable dependiente a la variable binaria vhappy, que toma valor 1 si el encuestado afirma ser muy feliz y 0 en el caso contrario, y como independientes la variables años de educación (educ), ingreso (income), mujer(female) y desempleado (unem10).**

#modelo de regresión lineal  
  
modelo\_RL <- glm(vhappy ~ educ + income + female + unem10, data = happiness,family=binomial)  
  
summary(modelo\_RL)#resumen de los datos del modelo

##   
## Call:  
## glm(formula = vhappy ~ educ + income + female + unem10, family = binomial,   
## data = happiness)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.0065 -0.9478 -0.7661 1.3935 2.2115   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.930041 0.229074 -4.060 4.91e-05 \*\*\*  
## educ 0.015733 0.008054 1.953 0.050762 .   
## income$1000 to 2999 -0.364266 0.302788 -1.203 0.228961   
## income$3000 to 3999 -0.653166 0.338111 -1.932 0.053383 .   
## income$4000 to 4999 -0.639649 0.338067 -1.892 0.058481 .   
## income$5000 to 5999 -0.303040 0.290866 -1.042 0.297479   
## income$6000 to 6999 -1.117017 0.336860 -3.316 0.000913 \*\*\*  
## income$7000 to 7999 -0.663001 0.299159 -2.216 0.026677 \*   
## income$8000 to 9999 -0.524492 0.261885 -2.003 0.045203 \*   
## income$10000 - 14999 -0.273140 0.222475 -1.228 0.219547   
## income$15000 - 19999 -0.419422 0.225573 -1.859 0.062976 .   
## income$20000 - 24999 -0.108122 0.221141 -0.489 0.624893   
## income$25000 or more 0.173741 0.209284 0.830 0.406443   
## female 0.025440 0.044800 0.568 0.570126   
## unem10 -0.480592 0.049717 -9.666 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 12196 on 9962 degrees of freedom  
## Residual deviance: 11906 on 9948 degrees of freedom  
## (7174 observations deleted due to missingness)  
## AIC: 11936  
##   
## Number of Fisher Scoring iterations: 4

**(ii) Estimar el modelo del apartado anterior por medio de regresión logística. Con las variables anteriores, seleccionar el mejor modelo mediantes stepAIC**

#modelo de regresión y estimación AIC  
modelo\_RL <- glm(vhappy ~ educ + income + female + unem10, data = happiness,  
 family = binomial(logit))  
#estudio de AIC  
modelo\_RL\_AIC <- stepAIC(modelo\_RL, trace = TRUE)

## Start: AIC=11936.46  
## vhappy ~ educ + income + female + unem10  
##   
## Df Deviance AIC  
## - female 1 11907 11935  
## <none> 11906 11936  
## - educ 1 11910 11938  
## - unem10 1 12003 12031  
## - income 11 12031 12039  
##   
## Step: AIC=11934.78  
## vhappy ~ educ + income + unem10  
##   
## Df Deviance AIC  
## <none> 11907 11935  
## - educ 1 11911 11937  
## - unem10 1 12004 12030  
## - income 11 12031 12037

#test anova  
modelo\_RL\_AIC$anova

## Stepwise Model Path   
## Analysis of Deviance Table  
##   
## Initial Model:  
## vhappy ~ educ + income + female + unem10  
##   
## Final Model:  
## vhappy ~ educ + income + unem10  
##   
##   
## Step Df Deviance Resid. Df Resid. Dev AIC  
## 1 9948 11906.46 11936.46  
## 2 - female 1 0.3225805 9949 11906.78 11934.78

#resumen de los datos obtenidos  
summary(modelo\_RL\_AIC)

##   
## Call:  
## glm(formula = vhappy ~ educ + income + unem10, family = binomial(logit),   
## data = happiness)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.0019 -0.9527 -0.7688 1.3919 2.2108   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -0.91469 0.22746 -4.021 5.78e-05 \*\*\*  
## educ 0.01583 0.00805 1.967 0.04923 \*   
## income$1000 to 2999 -0.36518 0.30280 -1.206 0.22780   
## income$3000 to 3999 -0.65074 0.33808 -1.925 0.05425 .   
## income$4000 to 4999 -0.63922 0.33807 -1.891 0.05865 .   
## income$5000 to 5999 -0.30158 0.29086 -1.037 0.29980   
## income$6000 to 6999 -1.11528 0.33684 -3.311 0.00093 \*\*\*  
## income$7000 to 7999 -0.66114 0.29913 -2.210 0.02709 \*   
## income$8000 to 9999 -0.52434 0.26188 -2.002 0.04526 \*   
## income$10000 - 14999 -0.27254 0.22247 -1.225 0.22056   
## income$15000 - 19999 -0.42041 0.22556 -1.864 0.06235 .   
## income$20000 - 24999 -0.11002 0.22112 -0.498 0.61879   
## income$25000 or more 0.17002 0.20918 0.813 0.41635   
## unem10 -0.48129 0.04970 -9.684 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 12196 on 9962 degrees of freedom  
## Residual deviance: 11907 on 9949 degrees of freedom  
## (7174 observations deleted due to missingness)  
## AIC: 11935  
##   
## Number of Fisher Scoring iterations: 4

**(iii) Predecir con el modelo de regresión logística y con el modelo de regresión lineal la probabilidad de ser muy feliz de una mujer con 18 años de educación, que trabaja y tieen un intreso superior a 25000$ anuales (income=$25000 or more)**

#Añadimos los parámetros que indica el enunciado para realizar el filtrado y guardarlos en una nueva variable.  
  
predict\_RL <- data.frame(educ = 18, income = '$25000 or more', female = 1, unem10 = 0)  
  
predict <- predict(modelo\_RL, newdata = predict\_RL, type = 'response')  
  
predict

## 1   
## 0.3899168

## Prácticas Semana 2

1. El fichero de datos spam7. Los datos consisten en 4601 elementos de correo electŕonico,de los cuales 1813 elementos se identificaron como spam. El fichero contiene las siguientes variables:

* crl.tot total length of words in capitals.
* dollar number of occurrences of the $ symbol.
* bang number of occurrences of the ! symbol.
* money number of occurrences of the word ”money”.
* n000 number of occurrences of the string ”000”.
* make number of occurrences of the word ”make”.
* yesno outcome variable, a factor with levels n not spam, y spam.

**(i) Realizar un an ́alisis descriptivo de las variables explicativas**

#Lectura de datos del fichero de datos spam7  
  
spam7 <- spam7  
  
str(spam7)

## 'data.frame': 4601 obs. of 7 variables:  
## $ crl.tot: num 278 1028 2259 191 191 ...  
## $ dollar : num 0 0.18 0.184 0 0 0 0.054 0 0.203 0.081 ...  
## $ bang : num 0.778 0.372 0.276 0.137 0.135 0 0.164 0 0.181 0.244 ...  
## $ money : num 0 0.43 0.06 0 0 0 0 0 0.15 0 ...  
## $ n000 : num 0 0.43 1.16 0 0 0 0 0 0 0.19 ...  
## $ make : num 0 0.21 0.06 0 0 0 0 0 0.15 0.06 ...  
## $ yesno : Factor w/ 2 levels "n","y": 2 2 2 2 2 2 2 2 2 2 ...

head(spam7)

## crl.tot dollar bang money n000 make yesno  
## 1 278 0.000 0.778 0.00 0.00 0.00 y  
## 2 1028 0.180 0.372 0.43 0.43 0.21 y  
## 3 2259 0.184 0.276 0.06 1.16 0.06 y  
## 4 191 0.000 0.137 0.00 0.00 0.00 y  
## 5 191 0.000 0.135 0.00 0.00 0.00 y  
## 6 54 0.000 0.000 0.00 0.00 0.00 y

colnames(spam7) #el nombre de todas las columnas

## [1] "crl.tot" "dollar" "bang" "money" "n000" "make" "yesno"

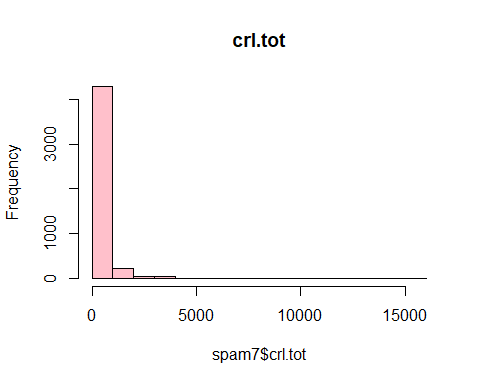
attach(spam7)  
#resumen de los datos y de sus variables.  
summary(spam7)

## crl.tot dollar bang money   
## Min. : 1.0 Min. :0.00000 Min. : 0.0000 Min. : 0.00000   
## 1st Qu.: 35.0 1st Qu.:0.00000 1st Qu.: 0.0000 1st Qu.: 0.00000   
## Median : 95.0 Median :0.00000 Median : 0.0000 Median : 0.00000   
## Mean : 283.3 Mean :0.07581 Mean : 0.2691 Mean : 0.09427   
## 3rd Qu.: 266.0 3rd Qu.:0.05200 3rd Qu.: 0.3150 3rd Qu.: 0.00000   
## Max. :15841.0 Max. :6.00300 Max. :32.4780 Max. :12.50000   
## n000 make yesno   
## Min. :0.0000 Min. :0.0000 n:2788   
## 1st Qu.:0.0000 1st Qu.:0.0000 y:1813   
## Median :0.0000 Median :0.0000   
## Mean :0.1016 Mean :0.1046   
## 3rd Qu.:0.0000 3rd Qu.:0.0000   
## Max. :5.4500 Max. :4.5400

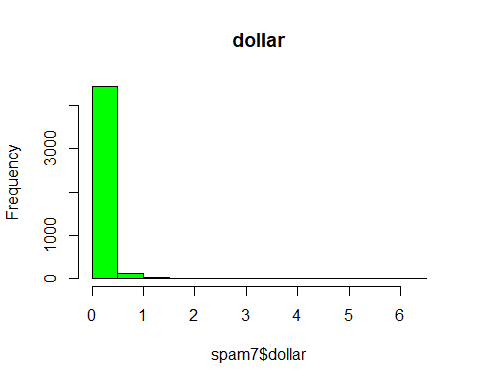
stat.desc(spam7)

## crl.tot dollar bang money n000  
## nbr.val 4.601000e+03 4.601000e+03 4.601000e+03 4.601000e+03 4.601000e+03  
## nbr.null 0.000000e+00 3.201000e+03 2.343000e+03 3.866000e+03 3.922000e+03  
## nbr.na 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00  
## min 1.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00  
## max 1.584100e+04 6.003000e+00 3.247800e+01 1.250000e+01 5.450000e+00  
## range 1.584000e+04 6.003000e+00 3.247800e+01 1.250000e+01 5.450000e+00  
## sum 1.303414e+06 3.488050e+02 1.237995e+03 4.337300e+02 4.676700e+02  
## median 9.500000e+01 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00  
## mean 2.832893e+02 7.581069e-02 2.690709e-01 9.426864e-02 1.016453e-01  
## SE.mean 8.939140e+00 3.624938e-03 1.202512e-02 6.525596e-03 5.164130e-03  
## CI.mean.0.95 1.752500e+01 7.106619e-03 2.357500e-02 1.279330e-02 1.012417e-02  
## var 3.676577e+05 6.045796e-02 6.653202e-01 1.959262e-01 1.227006e-01  
## std.dev 6.063479e+02 2.458820e-01 8.156716e-01 4.426355e-01 3.502864e-01  
## coef.var 2.140384e+00 3.243368e+00 3.031438e+00 4.695470e+00 3.446165e+00  
## make yesno  
## nbr.val 4.601000e+03 NA  
## nbr.null 3.548000e+03 NA  
## nbr.na 0.000000e+00 NA  
## min 0.000000e+00 NA  
## max 4.540000e+00 NA  
## range 4.540000e+00 NA  
## sum 4.810500e+02 NA  
## median 0.000000e+00 NA  
## mean 1.045534e-01 NA  
## SE.mean 4.501762e-03 NA  
## CI.mean.0.95 8.825614e-03 NA  
## var 9.324324e-02 NA  
## std.dev 3.053576e-01 NA  
## coef.var 2.920591e+00 NA

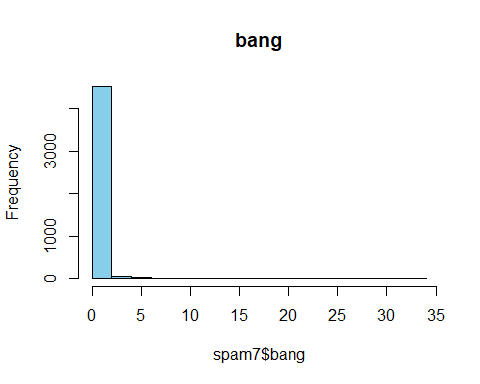
hist(spam7$crl.tot,col = "pink",main = "crl.tot" )



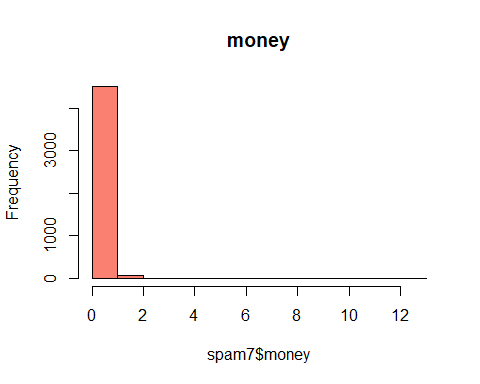
hist(spam7$dollar,col = "green",main = "dollar" )



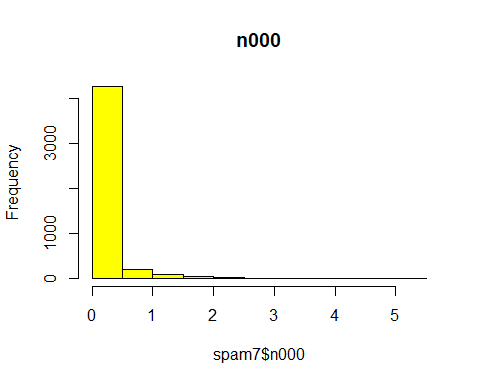
hist(spam7$bang,col = "skyblue",main = "bang" )



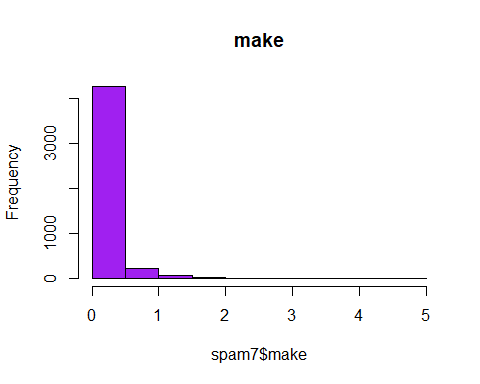
hist(spam7$money,col = "salmon",main = "money" )



hist(spam7$n000,col = "yellow",main = "n000" )



hist(spam7$make,col = "purple",main = "make" )



#Todas las variables se distribuyen de forma muy similar, solo crl.tot muestra una mayor distribución de frecuencia.

**(ii) Comparar los modelos LR, LDA y QDA mediante la matriz de confusíon. Realizar gráficosd partición**

Se realiza los modelos LR, LDA y QDA para las variable dependiente Yesno y el resto. Después estos son utilizados para las matrices de confusión. Por último se crea el gráfico de partición LDA y QDA para representar dicho análisis.

#Modelo LR:  
  
modelo\_LR <- glm(yesno ~., data = spam7, family = binomial(link = logit))

## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

result\_LR <- predict(modelo\_LR, newdata = spam7)  
  
#Modelo LDA:  
  
modelo\_LDA <- lda(yesno ~., data = spam7)   
  
modelo\_LDA

## Call:  
## lda(yesno ~ ., data = spam7)  
##   
## Prior probabilities of groups:  
## n y   
## 0.6059552 0.3940448   
##   
## Group means:  
## crl.tot dollar bang money n000 make  
## n 161.4709 0.01164849 0.1099835 0.01713773 0.007087518 0.0734792  
## y 470.6194 0.17447821 0.5137126 0.21287921 0.247054606 0.1523387  
##   
## Coefficients of linear discriminants:  
## LD1  
## crl.tot 0.0005832625  
## dollar 1.6357112508  
## bang 0.5229084780  
## money 0.8295009511  
## n000 1.5117933586  
## make 0.1509208601

#Modelo QDA:  
  
modelo\_QDA <- qda(yesno ~., data = spam7)   
modelo\_QDA

## Call:  
## qda(yesno ~ ., data = spam7)  
##   
## Prior probabilities of groups:  
## n y   
## 0.6059552 0.3940448   
##   
## Group means:  
## crl.tot dollar bang money n000 make  
## n 161.4709 0.01164849 0.1099835 0.01713773 0.007087518 0.0734792  
## y 470.6194 0.17447821 0.5137126 0.21287921 0.247054606 0.1523387

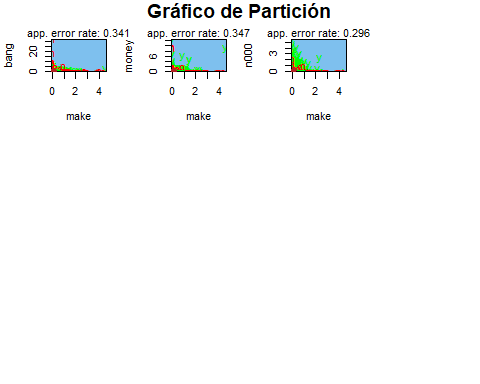
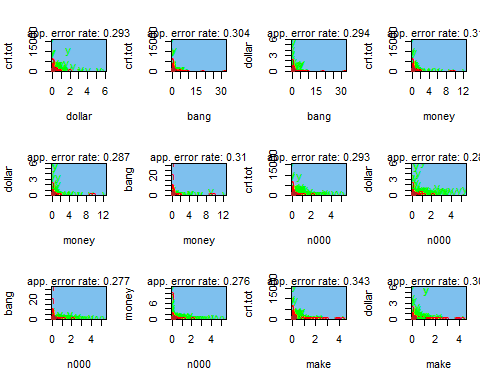
#Matriz de Confusión LDA:  
  
result\_LDA <- predict(modelo\_LDA, newdata = spam7)   
matrizLDA <- table(result\_LDA$class, spam7$yesno)  
matrizLDA

##   
## n y  
## n 2723 1015  
## y 65 798

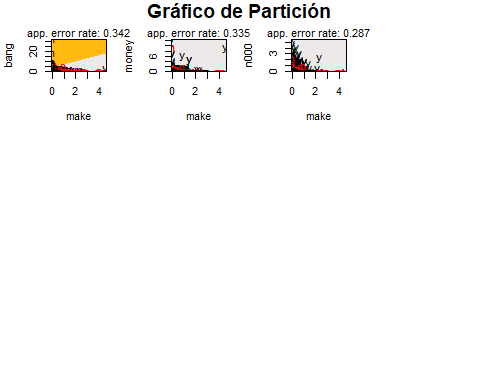
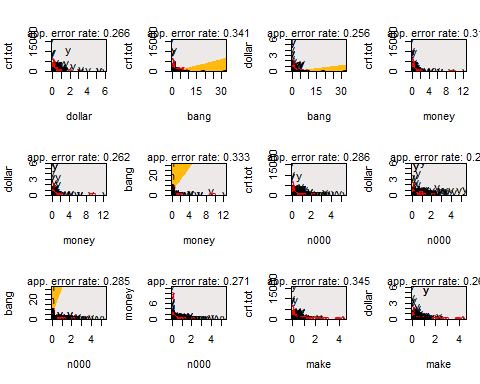
#Matriz de Confusión QDA:  
  
result\_QDA <- predict(modelo\_QDA, newdata = spam7)   
matrizQDA <- table(result\_QDA$class, spam7$yesno)  
matrizQDA

##   
## n y  
## n 2687 1003  
## y 101 810

#Gráfico de Partición LDA  
  
 graf\_PartLDA <- partimat(yesno ~., data = spam7, method="lda",main="Gráfico de Partición", col.correct="green",col.wrong = "red"  
  
, image.colors = c("darkgoldenrod1", "skyblue2"), col.mean = "firebrick")



#Gráfico de Partición QDA  
  
graf\_PartQDA <- partimat(yesno ~., data = spam7, method="qda",image.colors = c("darkgoldenrod1", "snow2"),main="Gráfico de Partición")



# % de precisión del modelo QDA  
sum(diag(matrizQDA))/sum(matrizQDA) #la precisión del modelo QDA. El resultado es de un 76% de precisión.

## [1] 0.7600522

# % de precisión del modelo LDA  
sum(diag(matrizLDA))/sum(matrizLDA) #la precisión del modelo LDA. El resultado obtenido es muy próximo al anterior, 76,52%

## [1] 0.7652684

El color verde de estos gráficos representa el acierto y el rojo el error de cada una de las variables