



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

TAREA 4

Integrantes

Castillo Suarez Ignacio 316656601

Reyes Rivera Carlos 316159638

Vaquera Herrera Celine Valeria 315032781

Zárate Pérez Juan Carlos 316244709

NOMBRE DEL PROFESOR:

Christian Gabriel Miranda Ruíz



MATERIA:

ADMINISTRACIÓN DE RIESGOS
FINANCIEROS

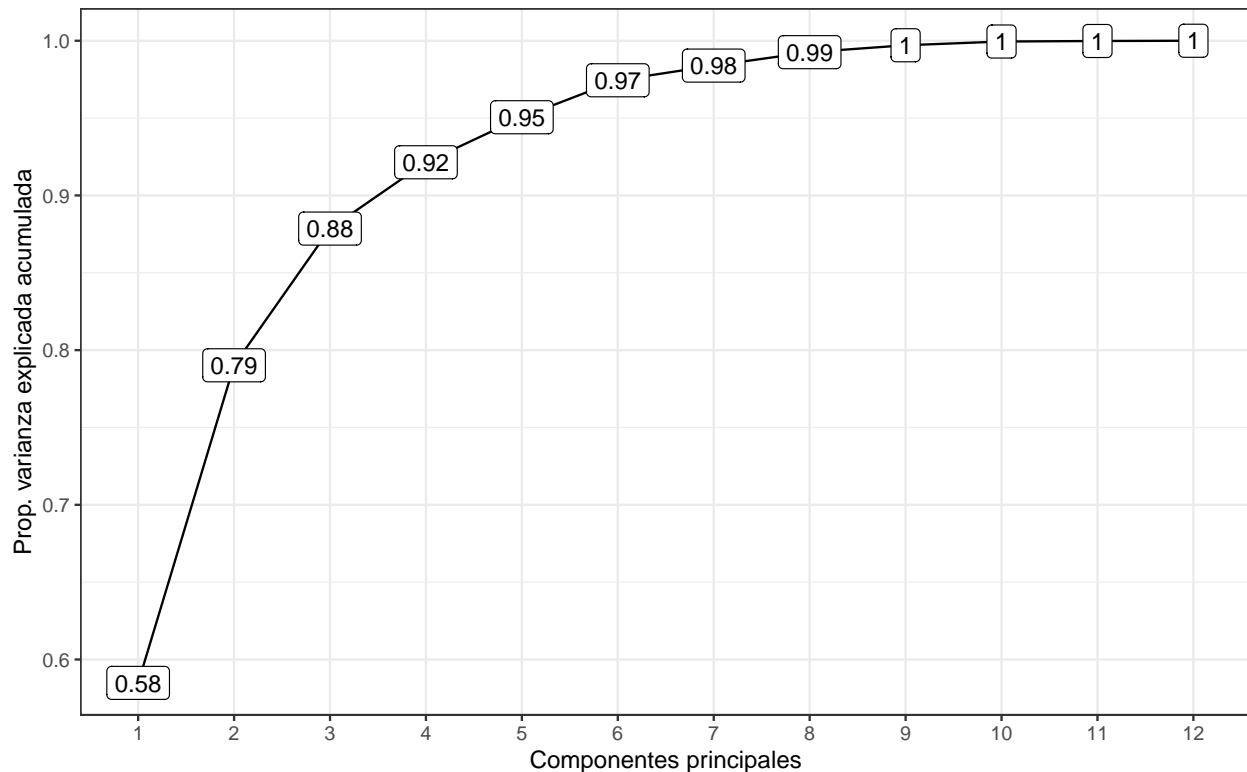
Fecha de entrega: 30.Enero.2021

Ejercicio 1

Paso 1

Se hará un análisis de componentes principales con base en la probabilidad de varianza explicada acumulada obtenida, y así tener las variables más significativas.

```
library(ggplot2)
tabla2 <- read.csv("Libro4.csv", header = TRUE, encoding = "UTF-8")
tabla3 <- tabla2[, -1]
tabla4 <- tabla3[, -1]
pca <- prcomp(tabla4)
# Cálculo de la varianza explicada acumulada
prop_varianza <- pca$sdev^2/sum(pca$sdev^2)
prop_varianza_acum <- cumsum(prop_varianza)
ggplot(data = data.frame(prop_varianza_acum, pc = factor(1:12)),
       aes(x = pc, y = prop_varianza_acum, group = 1)) +
  geom_point() +
  geom_line() +
  geom_label(aes(label = round(prop_varianza_acum, 2))) +
  theme_bw() +
  labs(x = "Componentes principales",
       y = "Prop. varianza explicada acumulada")
```



```
pca$rotation[, 1:7]
```

##		PC1	PC2	PC3	PC4	PC5
##	Liquid	-0.042091309	0.047252634	-0.061982166	-0.15714258	-0.131902660
##	Apalanc	0.075213031	-0.181253734	0.269311044	0.64812620	0.276589426
##	AñosCl	-0.976200811	-0.184146561	0.100970660	0.03175993	0.029283243
##	Rentab	-0.004662393	0.008257103	-0.015789392	-0.01856071	-0.005477467
##	Bancos	0.117331010	-0.241741190	0.857838708	-0.42596777	-0.093063084
##	PCP_CT	0.076224372	-0.110652126	0.196406922	0.38311962	0.175547765
##	Gop_SD	-0.044094248	0.083616790	-0.133628281	-0.28797769	-0.047122258
##	RetrTMP	0.004431585	0.016451155	-0.001974422	0.10012741	-0.287588525
##	PLP_PT	0.003240155	-0.026177456	0.017530030	0.05932998	0.028187696
##	IFAudit	0.008109978	0.012293993	-0.078091455	-0.34199138	0.880944483
##	UQ_VN	-0.003188426	0.004756515	-0.007083552	-0.02831535	-0.009430815
##	AñoOper	-0.133994257	0.923109965	0.338239996	0.10686296	0.062131919
##		PC6	PC7			
##	Liquid	-0.009035069	-0.856161135			
##	Apalanc	0.075115581	-0.434754922			
##	AñosCl	0.010522483	0.025825828			
##	Rentab	0.004772387	-0.014885845			
##	Bancos	0.025226933	-0.005288792			
##	PCP_CT	-0.075557012	0.209419132			
##	Gop_SD	0.011509584	-0.148215427			
##	RetrTMP	0.946466466	0.051415015			
##	PLP_PT	0.029567521	0.086957938			
##	IFAudit	0.301576615	-0.029202670			
##	UQ_VN	0.001968803	-0.013007162			
##	AñoOper	-0.006280699	0.002895003			

Dado el resultados anteriores notamos que la probabilidad de la probabilidad de varianza explicada acumulada en un 98% se acumula en 7 variables , las cuales son las más significativas, estas variables son: Liquid, Apalanc, Rentab, Bancos, PCP_CT, Gop_SD e IFAudit.

Por lo que se procede a hacer los calculos en el excel **Ejer_anal_disc.xls**

Tarea de Riesgo de Crédito

Ejercicio 3

Usando la tabla del archivo “Ejer anal discr.xls” (hoja base) que está en la página del curso, ajustar una regresión logística con las variables más significativas (pueden hacer uso de la hoja frecuencia del archivo “logit y disc datos.xls” y utilizando la prueba de Wald para cada variable), pueden usar el archivo “logitprobit.ipynb” para hacer el análisis, obviamente entregar los coeficientes de regresión, la gráfica de datos observados y ajustados, el nivel de significancia (Devianza) del modelo, es decir la prueba de hipótesis de que la Devianza se distribuye como una $X^2_{n-(m+1)}$, y la interpretación de las variables.

Solución:

Para realizar el ejercicio, primero descargamos el archivo csv en donde se encuentra la base de datos del archivo “Ejer anal discr.xls” (hoja base) e inicializamos las siguientes variables.

```
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

incump = "C:\\Users\\juanm\\OneDrive\\Documentos\\Tareas septimo
semestre\\Admin. Riesgos\\BD-Logit.csv"
data=read.csv(incump, sep=",", quote="/")
print(head(data))

##   Cliente Incump   Liquid   Apalanc   AñosCl       Rentab Bancos
PCP_CT
## 1         1      0 1.418983 0.3678430 5.519115 0.082436996      2
0.9578792
## 2         2      0 1.095459 0.6742535 5.365984 0.050120498      3
0.7856872
## 3         3      0 1.232216 0.8788182 6.579059 0.047376582      1
0.6682592
## 4         4      0 1.112399 0.6111619 3.980561 -0.010216081      2
0.9324931
## 5         5      0 1.495808 0.2262790 5.861760 0.002987486      4
```

```

0.5764472
## 6      6      0 1.106524 0.7106344 6.710337 0.046598640      3
0.7644960
##      Gop_SD RetrTMP      PLP_PT IFAudit      UO_VN AñoOper
## 1 0.9619346      1 0.4255506      1 0.16267763      4
## 2 0.8832112      1 0.3777689      0 0.20572504      4
## 3 1.1103575      1 0.5652263      0 0.23805917      4
## 4 1.0153539      1 0.3970641      1 0.17857237      5
## 5 0.9057852      1 0.4313102      0 0.08164283      6
## 6 1.0523312      0 0.3734093      1 0.24602347      4

n<-nrow(data)-1
m=ncol(data)
x=data[,3:m]
y=data[,2]
xext=as.matrix(cbind(1,x))

```

Realizamos un ajuste de los datos, usando la función glm.

```

ylogr = glm(Incump ~ Liquid + Apalanc + AñosCl + Rentab +Bancos
+ PCP_CT + Gop_SD + RetrTMP + PLP_PT + IFAudit +UO_VN + AñoOper ,
family=binomial("logit"), data = data)

## Warning: glm.fit: algorithm did not converge

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

summary(ylogr)

##
## Call:
## glm(formula = Incump ~ Liquid + Apalanc + AñosCl + Rentab + Bancos +
##      PCP_CT + Gop_SD + RetrTMP + PLP_PT + IFAudit + UO_VN + AñoOper,
##      family = binomial("logit"), data = data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -4.993e-05 -2.100e-08 -2.100e-08  2.100e-08  4.766e-05
##
## Coefficients:
##      Estimate Std. Error z value Pr(>|z|)
## (Intercept)   86.097 345041.658   0.000   1.000
## Liquid        -35.348  83217.419   0.000   1.000
## Apalanc        63.663  74575.329   0.001   0.999
## AñosCl        -7.861 16526.692   0.000   1.000
## Rentab       -395.216 893798.551   0.000   1.000
## Bancos         8.519  48731.575   0.000   1.000
## PCP_CT        47.308 128348.446   0.000   1.000
## Gop_SD       -73.761 111142.606  -0.001   0.999
## RetrTMP       -9.272 112996.503   0.000   1.000
## PLP_PT       -102.053 384818.303   0.000   1.000
## IFAudit        16.038  39775.381   0.000   1.000

```

```
## UO_VN          -124.528 663892.384    0.000    1.000
## AñoOper        -3.374  30617.497    0.000    1.000
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 3.4231e+02  on 246  degrees of freedom
## Residual deviance: 1.3106e-08  on 234  degrees of freedom
## AIC: 26
##
## Number of Fisher Scoring iterations: 25
```

Entonces, los coeficientes de regresión son:

```
print(ylogr$coef)

## (Intercept)      Liquid      Apalanc      AñosCl      Rentab
Bancos
##  86.096523 -35.347612  63.663052  -7.861229 -395.215718
8.518809
##      PCP_CT      Gop_SD      RetrTMP      PLP_PT      IFAudit
UO_VN
##  47.308495 -73.760760  -9.272432 -102.053079  16.038037 -
124.528160
##      AñoOper
##  -3.374335
```

Replicamos la devianza.

```
Llogit=log(prod(((1+exp(-xext%%as.matrix(ylogr$coef)))^(-1))^y*(1-
(1+exp(-xext%%as.matrix(ylogr$coef)))^(-1))^(1-y)))
DLogit=-2*Llogit

print(paste("Devianza replicada = ",DLogit))

## [1] "Devianza replicada =  1.31061188997963e-08"

print(paste("Devianza del modelo lineal generalizado = ", ylogr$dev))

## [1] "Devianza del modelo lineal generalizado =  1.3106114410409e-08"
```

Realizamos la matriz de confusión.

```
ptol=0.5
yest=(1+exp(-xext%%as.matrix(ylogr$coef)))^(-1)

print("Matriz de confusión")

## [1] "Matriz de confusión"

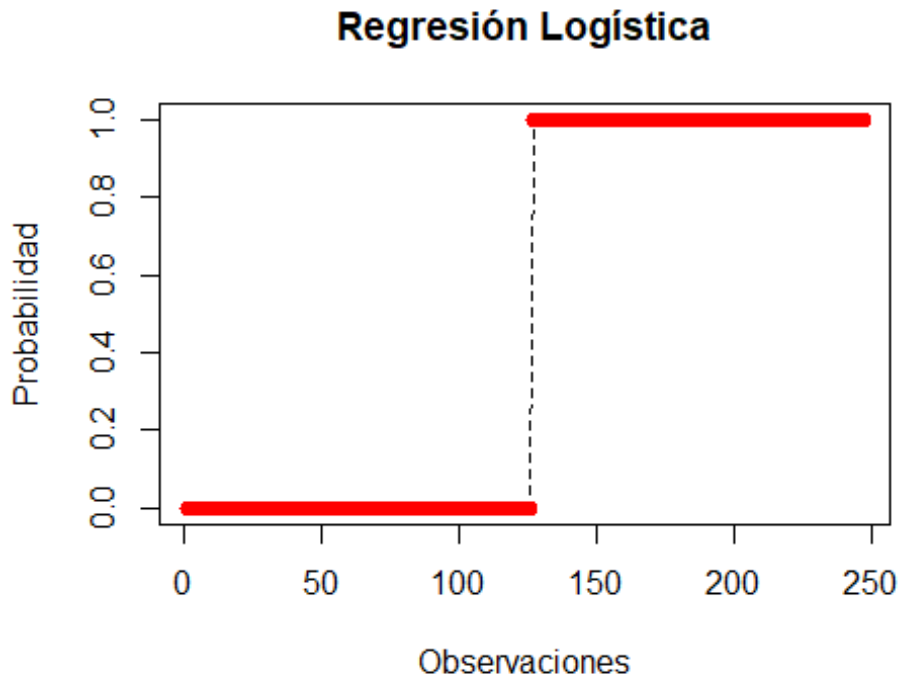
print(table(y,ifelse(yest<0.5,0,1)))

##
## y      0      1
```

```
##    0 126    0
##    1   0 121
```

Generamos la gráfica de datos observados y ajustados.

```
plot(y,type="l", lty=2, xlab="Observaciones",
     ylab="Probabilidad", main="Regresión Logística")
points(yest, lwd=2, col="red")
```



Finalmente, utilizamos la devianza residual para observar si el modelo de regresión logística proporciona un ajuste adecuado para los datos. Esto es posible porque la desviación viene dada por el valor de chi-cuadrado en ciertos grados de libertad. Para probar la significancia, podemos encontrar el p-value asociado.

```
pvalue = 1 - pchisq(5.9299, 234)
pvalue
## [1] 1
```

Con base a los resultados obtenidos, nos damos cuenta que no ha evidencia suficiente para rechazar la hipótesis nula, además que no hay variables que sean significativas para predecir el incumplimiento de los clientes.

Es por eso, que se generará un nuevo modelo utilizando el regresión multiple para obtener mejores resultados.

```
ylogr = glm(Incump ~ Liquid + Apalanc + AñosCl + Rentab +Bancos
+ PCP_CT + Gop_SD + RetrTMP + PLP_PT + IFAudit +UO_VN + AñoOper,
```

```

data = data)
summary(ylogr)

##
## Call:
## glm(formula = Incump ~ Liquid + Apalanc + AñosCl + Rentab + Bancos +
##      PCP_CT + Gop_SD + RetrTMP + PLP_PT + IFAudit + UO_VN + AñoOper,
##      data = data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.61744 -0.10534 -0.01249  0.10217  0.47844
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.311946   0.121002   2.578  0.01055 *
## Liquid      -0.071780   0.039473  -1.818  0.07027 .
## Apalanc      0.307901   0.028870  10.665 < 2e-16 ***
## AñosCl      -0.015446   0.005509  -2.804  0.00548 **
## Rentab     -1.486105   0.301809  -4.924 1.60e-06 ***
## Bancos       0.033695   0.014130   2.385  0.01789 *
## PCP_CT       0.392136   0.042365   9.256 < 2e-16 ***
## Gop_SD      -0.422445   0.052626  -8.027 4.83e-14 ***
## RetrTMP     -0.002731   0.024758  -0.110  0.91225
## PLP_PT       0.040351   0.076964   0.524  0.60057
## IFAudit     -0.007332   0.022477  -0.326  0.74456
## UO_VN       -0.202252   0.208894  -0.968  0.33394
## AñoOper     -0.017458   0.009273  -1.883  0.06099 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.02534142)
##
##      Null deviance: 61.7247  on 246  degrees of freedom
## Residual deviance:  5.9299  on 234  degrees of freedom
## AIC: -192.2
##
## Number of Fisher Scoring iterations: 2

```

Observemos que en este caso, si encontramos variables significativas. El modelo muestra que el apalancamiento, la liquidez, los años de ser cliente, la rentabilidad, el número de bancos, el pasivo corto plazo a capital de trabajo, la Generación Operativa y los años operando son significativos para el incumplimiento de los clientes.

Tomando en cuenta las variables mencionadas, generamos un modelo donde unicamente se encuentren estas variables con el fin de dar una mayor precisión a nuestro modelo.

```

data2 = data %>%
  select(Apalanc, AñosCl, Rentab, PCP_CT, Gop_SD)

```



```

xext2=as.matrix(cbind(1,data2))
ylogr2 = glm(Incump ~ Apalanc + AñosCl + Rentab + PCP_CT +
Gop_SD, data = data)
summary(ylogr2)

##
## Call:
## glm(formula = Incump ~ Apalanc + AñosCl + Rentab + PCP_CT + Gop_SD,
##      data = data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.68280  -0.10883   0.00431   0.11333   0.46782
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.251781   0.088174   2.856 0.004671 **
## Apalanc      0.330885   0.028521  11.601 < 2e-16 ***
## AñosCl      -0.018206   0.005451  -3.340 0.000971 ***
## Rentab      -1.718526   0.298939  -5.749 2.71e-08 ***
## PCP_CT       0.414708   0.042167   9.835 < 2e-16 ***
## Gop_SD      -0.454613   0.051644  -8.803 2.64e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.02618935)
##
##      Null deviance: 61.7247  on 246  degrees of freedom
## Residual deviance:  6.3116  on 241  degrees of freedom
## AIC: -190.79
##
## Number of Fisher Scoring iterations: 2

```

Entonces, los coeficientes de regresión son:

```

print(ylogr2$coef)

## (Intercept)      Apalanc      AñosCl      Rentab      PCP_CT
## Gop_SD
##  0.25178090  0.33088544 -0.01820618 -1.71852649  0.41470833 -
## 0.45461331

```

Replicamos la devianza.

```

Llogit2=log(prod(((1+exp(-xext2%%as.matrix(ylogr2$coef)))^(-1))^y*(1-
(1+exp(-xext2%%as.matrix(ylogr2$coef)))^(-1))^(1-y)))
DLogit2=-2*Llogit2

print(paste("Devianza replicada = ",DLogit2))

## [1] "Devianza replicada = 261.62491079492"

```

```
print(paste("Devianza del modelo lineal generalizado = ", ylogr2$dev))
## [1] "Devianza del modelo lineal generalizado = 6.31163348561568"
```

Realizamos la matriz de confusión.

```
ptol=0.5
yest2=(1+exp(-xext2%*%as.matrix(ylogr2$coef)))^(-1)

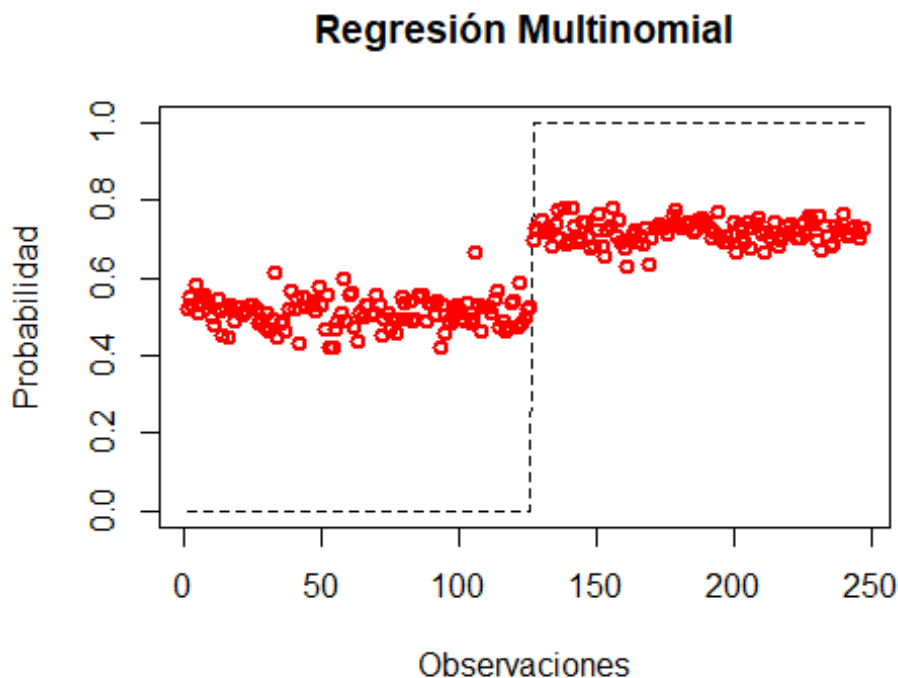
print("Matriz de confusión")
## [1] "Matriz de confusión"

print(table(y,ifelse(yest2<0.5,0,1)))

##
## y      0    1
##   0  48   78
##   1   0  121
```

Generamos la gráfica de datos observados y ajustados.

```
plot(y,type="l", lty=2, xlab="Observaciones",
      ylab="Probabilidad", main="Regresión Multinomial")
points(yest2, lwd=2, col="red")
```



Obtenemos la precisión de nuestro modelo utilizando la diagonal de nuestra matriz de confusión.

```
(48+121)/(48+121+78)
```

```
## [1] 0.6842105  
pvalue2 = 1 - pchisq(6.3116, 241)  
pvalue2  
## [1] 1
```

Entonces con base a este modelo, podemos observar que se tienen varias variables que son significativas para predecir el incumplimiento de los clientes y que se tiene una precisión de 0.6639676 donde no hay evidencia suficiente para rechazar el modelo.

Ejercicio 5

Calcular la correlación de activos, probabilidad media de incumplimiento y probabilidad conjunta de incumplimiento por el método de momento de la información del archivo "Histórico Deudores (modelo de valores de referencia).xls".

Solución

Para el ejercicio se hizo uso de los datos en el archivo sugerido, únicamente con el cambio de ser en formato .csv

Se utilizaron dos librerías: rootSolve, para el cálculo los ceros de la función normal bivariada, y mvtnorm para la implementación de la función bivariada normal.

```
library(rootSolve) #Se carga La Librería para encontrar Las raices de una función multivariada  
library(mvtnorm) #Se carga La Librería para La distribución normal bivariada
```

Posteriormente se leen los datos y se extraen las columnas de nuestro interés.

```
datos = read.csv("C:/ICS/9.Optativas/RiesgosFinancieros/HistoricoDeudores.csv") #Se lee el archivo con los datos a analizar  
D_t = datos$D_t  
N_t = datos$N_t
```

Se calcula la probabilidad media de incumplimiento.

```
PMI = mean(D_t/N_t)
```

Se calcula la probabilidad conjunta de incumplimiento.

```
PCI_t = (D_t*(D_t-1))/(N_t*(N_t-1))  
PCI = mean(PCI_t)
```

Se calcula el valor crítico de referencia

```
VCR = qnorm(PMI)
```

Se hace una estimación de la correlación de activos

```
correl = 0.15
```

Se crea una función que reciba el valor de la correlación y calcule la función de distribución acumulada de una normal estándar bivariada

```
G <- function(x){  
  sigma = matrix(c(1,x,x,1), nrow = 2)  
  pmvnorm(lower = c(-Inf, -Inf), upper=c(VCR,VCR), mean=rep(0, 2), sigma=sigma)/PCI-1  
}
```

Se calcula la raíz de la función, la cual representa la correlación de activos

```
raiz = multiroot(G,correl,positive = TRUE, maxiter=10000)
CA = raiz$root
```

En resumen

```
## [1] "La probabilidad media de incumplimiento es: "  
## [1] 0.001004205  
## [1] "La probabilidad conjunta de incumplimiento es: "  
## [1] 1.543424e-06  
## [1] "El valor crítico de referencia es: "  
## [1] -3.088986  
## [1] "La correlación de activos es: "  
## [1] 0.03886633
```