UNIVERSIDAD NACIONAL DE COLOMBIA

Ajuste de las carteras C2 y C15 sobre el número de siniestros en seguros de autos

Presentado por:
Alejandro Arias Garzón

Presentado a:
José Alfredo Jiménez Moscoso

Bogotá, Colombia, 16 de febrero de 2024

Cartera C2

La cartera C2 se compone de la siguiente manera:

Siniestros	**C-2**
0	20592
1	2651
2	297
3	41
4	7
5	0
6	1
≥ 7	0
Total pólizas	23589

Cálculo índice de dispersión

Hacemos el respectivo cálculo de $\bar{n}$ y de s_n^2.

$$\bar{n} = \frac{\sum \text{Siniestros} \times \text{C-2}}{\text{Total pólizas}} = \frac{3402}{23589} = 0{,}1442198$$

$$s_n^2 = 0{,}1846623 - (0{,}1442198)^2 = 0{,}163863$$

Decimos que d es el índice de dispersión muestral, luego

$$d = \frac{s_n^2}{\bar{n}} = 1{,}136204$$

Como $d > 1$, existe sobredispersión. En este caso, se sugiere usar la pmf binomial negativa. Sin embargo, entre las distribuciones dadas se buscará la pmf que mejor ajuste los datos.

Con esto, determinamos un intervalo de confianza al 95 % para I_n.

$$IC(I_n) = (1{,}107197, 1{,}16521)$$

Estimación de parámetros

Distribución Poisson

Por el método de momentos y por máxima verosimilitud, se obtiene que

$$\bar{\lambda} = \bar{n} = 0{,}1442198$$

Distribución binomial negativa

Por el método de momentos, se obtiene que

$$\widehat{q} = 1 - \frac{\bar{n}}{s_n^2} \approx 0{,}119876$$

$$\widehat{r} = \frac{\bar{n}^2}{s_n^2 - \bar{n}} \approx 1{,}058855$$

Y, por el método de máxima verosimilitud, se obtiene lo siguiente

$$\widehat{q} = 0{,}1142683$$

$$\widehat{r} = 1{,}117895$$

Distribución Panjer

Para determinar los parámetros de la familia de Panjer se usan las expresiones dadas, luego

$$\alpha = 1 - \frac{\bar{n}}{s_n^2} = 0{,}119876$$

$$\beta = (1-\alpha)\bar{n} - \alpha = 0{,}007055291$$

Síntesis

En resumen tendríamos

Distribuciones teóricas	**Parámetros estimados**	
	λ, r, α	q, β
Poisson	0.1442198	-
Binomial negativa	1.058855	0.1198760
Panjer	0.1198760	0.00705529

Aproximaciones con cada distribución

A continuación, en la tabla se presentan las aproximaciones obtenidas con cada una de las diferentes distribuciones trabajadas.

Siniestros (k)	n_k	D. Poisson	D. BN1	D. BN2	D. Panjer
0	20592	20420.938	20605.803	20596.759	20605.803
1	2651	2945.103	2615.521	2631.03	2615.521
2	297	212.371	322.765	318.366	322.765
3	41	10.209	39.451	37.809	39.451
4	7	0.368	4.799	4.448	4.799
5	0	0.011	0.582	0.520	0.582
6	1	0	0.07	0.061	0.07
$7 \geq$	0	0	0.009	0.007	0.009
Total pólizas	23589	23792.88	23592.79	23592.6	23592.79
Test χ^2	-	203.881	3.7896	3.6003	3.7896
v	-	2	2	2	2

[1]Método de momentos
[2]Método máxima verosimilitud

Los grados de libertad (v) se calculan según la definición y se agrupa según el criterio para usar la bondad de ajuste χ^2. Así, Para 2 grados de libertad y $\alpha = 0.05$, se obtiene que el límite de significancia de χ^2 es

$$\chi^2_{2,\alpha} = 5{,}99146 \tag{1}$$

En este caso, se rechaza H_0 si $\chi^2_{2,\alpha} > 5{,}99146$. Luego para los valores menores a 5.99146 no hay evidencia para rechazar los modelos supuestos y, por ende, se puede decir que, con una confianza del 5 %, la distribución Panjer y la binomial negativa proporcionan un ajuste razonable de los datos.

Cartera C15

La cartera C15 se compone de la siguiente manera:

Siniestros	**C-15**
0	371481
1	26784
2	2118
3	174
4	18
5	2
6	2
7	0
≥ 8	0
Total pólizas	400579

Cálculo índice de dispersión

Hacemos el respectivo cálculo de $\bar{n}$ y de s_n^2.

$$\bar{n} = \frac{\sum \text{Siniestros} \times \text{C-15}}{\text{Total pólizas}} = \frac{31636}{400579} = 0{,}07897568$$

$$s_n^2 = 0{,}09294546 - (0{,}07897568)^2 = 0{,}0867083$$

Decimos que d es el índice de dispersión muestral, luego

$$d = \frac{s_n^2}{\bar{n}} = 1{,}097911$$

Como $d > 1$, existe sobredispersión. En este caso, se sugiere usar la pmf binomial negativa. Sin embargo, entre las distribuciones dadas se buscará la pmf que mejor ajuste los datos.

Con esto, determinamos un intervalo de confianza al 95 % para I_n.

$$IC(I_n) = (1{,}090627, 1{,}105196)$$

Estimación de parámetros

Distribución Poisson

Por los métodos de los momentos y máxima verosimilitud, se obtiene que

$$\bar{\lambda} = \bar{n} = 0{,}07897568$$

Distribución binomial negativa

Por el método de momentos, se obtiene que

$$\widehat{q} = 1 - \frac{\bar{n}}{s_n^2} \approx 0{,}0891797$$

$$\widehat{r} = \frac{\bar{n}^2}{s_n^2 - \bar{n}} \approx 0{,}8066035$$

Y, por el método de máxima verosimilitud, se obtiene lo siguiente

$$\widehat{q} = 0{,}08791084$$

$$\widehat{r} = 0{,}8193855$$

Distribución Panjer

Para determinar los parámetros de la familia de Panjer se usan las expresiones dadas, luego

$$\alpha = 1 - \frac{\bar{n}}{s_n^2} = 0{,}0891797$$

$$\beta = (1-\alpha)\bar{n} - \alpha = -0{,}01724704$$

Síntesis

En resumen tendríamos

Distribuciones teóricas	**Parámetros estimados** λ, r, α	q, β
Poisson	0.07897568	-
Binomial negativa	0.8066035	0.0891797
Panjer	0.0891797	-0.0172470

Aproximaciones con cada distribución

A continuación, en la tabla se presentan las aproximaciones obtenidas con cada una de las diferentes distribuciones trabajadas.

Siniestros (k)	n_k	D. Poisson	D. BN1	D. BN2	D. Panjer
0	371481	370159.99	371506.532	371486.741	371506.532
1	26784	29233.638	26723.451	26759.254	26723.451
2	2118	1154.373	2152.739	2139.987	2152.739
3	174	30.389	179.604	176.802	179.604
4	18	0.6	15.243	14.841	15.243
5	2	0.009	1.307	1.258	1.307
6	2	0	0.113	0.107	0.113
7	0	0	0.01	0.009	0.01
$8 \geq$	0	0	0.001	0.001	0.001
Total pólizas	400579	402471.7	400581.6	400581.4	400581.6
Test χ^2	-	1892.685	2.576	2.356	2.576
v	-	2	2	2	2

[1]Método de momentos
[2]Método máxima verosimilitud

Los grados de libertad (v) se calculan según la definición y se agrupa según el criterio para usar la bondad de ajuste χ^2. Así, Para 2 grados de libertad y $\alpha = 0.05$, se obtiene que el límite de significancia de χ^2 es

$$\chi^2_{2,\alpha} = 5{,}99146 \tag{2}$$

En este caso, se rechaza H_0 si $\chi^2_{2,\alpha} > 5{,}99146$. Luego para los valores menores a 5.99146 no hay evidencia para rechazar los modelos supuestos y, por ende, se puede decir que, con una confianza del 5 %, la distribución Panjer y la binomial negativa proporcionan un ajuste razonable de los datos.

Código de R

Cartera C2

```
#########################CARTERA C2##############################3

install.packages("orthopolynom")
install.packages("gtools")
library(orthopolynom)
library(gtools)
### Numero de siniestros
k<-c(0:7)
n_k<-c(20592,2651,297,41,7,0,1,0)
Tcasos<-sum(n_k)
freqrel<-(n_k/Tcasos)
eps=0.000001;
n_bar <- sum(k*n_k)/Tcasos
sum(k*n_k)
(sum(k^2*n_k)/Tcasos)
(n_bar^2)
s_cua <- (sum(k^2*n_k)/Tcasos)-(n_bar^2)
n_bar;s_cua

### Prueba Chicuadrado
Chi_cua<- function(obs, esp,alpha=0.05,m=7,gl=3){
  chi<-sum((obs-esp)^2/esp)
  vchi<-qchisq(1-alpha,df = gl )
  if(chi<vchi){
    return(list("No hay evidencia para rechazar H_0",
                chi,"Valor critico"=vchi))
  }
  if(chi>=vchi){
    return(list("Existe evidencia para rechazar H_0",
                chi, "Valor critico"=vchi))
  }
}
### intervalo de confianza para el índice de dispersión
d <- s_cua/n_bar #índice de dispersión muestral
d
z <- qnorm(0.975, mean = 0, sd = 1)
z
der <- d + z * d * sqrt((1 / Tcasos) * (2 + ((d - 1) * (3 * d - 1)) / s_cua))
izq <- d - z * d * sqrt((1 / Tcasos) * (2 + ((d - 1) * (3 * d - 1)) / s_cua))

izq;der

# Estimación de parámetros

# Distribucion Poisson
lambda<-n_bar
lambda

# Binomial Negativa (método de momentos)
q_bin <- 1-(n_bar/s_cua)
r_bin <- round(n_bar^2/(s_cua-n_bar),10)
q_bin;r_bin
```

```
###Newton

f_r<-function(r,n_bar= n_bar,n_k= n_k){
  as=0;al=0;sol=0
  for(k in 1:(length(n_k)-1)){
    for(m in 0:(k-1)){
      as<-(1/(m+r))+as
    }
    al<- (n_k[k+1]*as)+al
    as=0
  }
  sol<-(log(1+(n_bar/r))*Tcasos)-al
  return(sol)
}
# Funcion f'(r)-----------------
f_der<- function(r,n_bar,n_k){
  as=0;al=0;sol=0
  for(k in 1:(length(n_k)-1)){
    for(m in 0:(k-1)){
      as<-(1/(m+r)^2)+as
    }
    al<-(n_k[k+1]*as)+al
    as=0
  }
  sol <- (-n_bar/(r^2+(n_bar*r)))*Tcasos+al
  return(sol)
}
i=0
r<-r_bin
repeat{
  r1<-r-(f_r(r = r,n_bar = n_bar,n_k = n_k)/
           f_der(r = r,n_bar = n_bar,n_k = n_k))
  i=i+1
  if(abs(r1-r)<eps){break}
  r=r1
  r1=0
}
r1
q<-n_bar/(n_bar+r1)
q

### Panjer (usando (24))

alpha <-  1-(n_bar/s_cua)
alpha
betha <- ((1-alpha)*(n_bar))-alpha
betha

###################### Estimación de probabilidades ##############################3

# Estimación de las probabilidades
Estpois <- dpois(x = k,lambda = lambda)
# Estimacion No. de casos con Poisson
estp<-round(Estpois*Tcasos,3)
data.frame(estp)
sum(estp) # Casos Totales con Poisson
```

```
ajuspois1 <- c(estp[1:3],sum(estp[4:8]))
# Se agrupan los datos
n_ka<-c(n_k[1:3],sum(n_k[4:8]))
# Se agrupa de la misma manera los observados
Chi_p<-sum((n_ka-ajuspois1)^2/ajuspois1)
# Se hace la prueba Chi-cuadrado
Chi_cua(obs = n_ka,esp=ajuspois1)

# Estimacion de probabilidades con Binomial
# negativa por el metodo de los momentos
Estbinom1 <- dnbinom(x=k,size=r_bin,prob=(1-q_bin))
# Estimación número de casos con binomial negativa
estbinn1 <- round(Estbinom1*Tcasos,3)
data.frame(estbinn1)
# Casos Totales con la estimación binomial negativa
sum(estbinn1)
ajusbinn11 <- c(estbinn1[1:4],sum(estbinn1[5:8]))
n_ka<-c(n_k[1:4],sum(n_k[5:8]))
# Se agrupa de la misma manera los observados
Chi_binn1<-sum((n_ka-ajusbinn11)^2/ajusbinn11)
# Se hace la prueba Chi-cuadrado
Chi_cua(obs = n_ka,esp=ajusbinn11)

# Estimacion de probabilidades con Binomial
# negativa por el metodo MV
Estbinom2 <- dnbinom(x = k,size = r1,prob = (1-q))
# Estimacion No. de casos con binomial negativa, por MV
estbinn2 <- round(Estbinom2*Tcasos,3)
data.frame(estbinn2)
# Casos Totales con la estimacion binomial negativa por MV
sum(estbinn2)
ajusbinn21 <- c(estbinn2[1:4],sum(estbinn2[5:8]))
# Se agrupa de la misma manera los observados
n_ka<-c(n_k[1:4],sum(n_k[5:8]))
Chi_binn2<-sum((n_ka-ajusbinn21)^2/ajusbinn21)
# Se hace la prueba Chi-cuadrado
Chi_cua(obs = n_ka,esp=ajusbinn21)

# Estimacion de probabilidades con Panjer
ab <- alpha+betha #es mayor que 0 y menor que 1
r <- ab/alpha

distribucionpanjer <- function(r, alpha) {
  x_vals <- 0:7  # Valores de x de 0 a 7
  resultados <- numeric(length(x_vals))  # Vector para almacenar los resultados

  for (i in 1:length(x_vals)) {
    x <- x_vals[i]
    combinacion <- choose(r + x - 1, x)  # Cálculo de la combinación
    resultado <- combinacion * alpha^x * (1 - alpha)^r  # Cálculo del resultado final
    resultados[i] <- resultado
  }

  # Crear un marco de datos con los valores de x y las probabilidades
  df_resultados <- c(probabilidad = round(resultados, digits = 10))
```

```
  return(df_resultados)
}


panjer <- distribucionpanjer(r,alpha)
estpanjer <- round(panjer * Tcasos, 3)
data.frame(estpanjer)
sum(estpanjer) # Casos Totales con Panjer
ajuspanjer1 <- c(estpanjer[1:4], sum(estpanjer[5:8]))
n_ka <- c(n_k[1:4], sum(n_k[5:8]))
Chi_panjer <- sum((n_ka - ajuspanjer1)^2 / ajuspanjer1)
Chi_cua(obs = n_ka, esp = ajuspanjer1)



### Salida Final
test=c("test",".",round(Chi_p,3),round(Chi_binn1,3),round(Chi_binn2,3),round(Chi_panjer,3))
salida=cbind(k,n_k,estp,estbinn1,estbinn2,estpanjer)
salida=rbind(salida,test)
salida
nose<-data.frame(salida)
nose
install.packages("sqldf")
library(sqldf)

total <- sqldf("select sum(n_k) as totalnk, sum(estp) as totalp, sum(estbinn1) as totalb1, sum(estbinn2) a
             from     nose
              ")
```

Cartera C15

```
    #########################CARTERA C15##############################3

install.packages("orthopolynom")
install.packages("gtools")
library(orthopolynom)
library(gtools)
### Numero de siniestros
k<-c(0:8)
n_k<-c(371481,26784,2118,174,18,2,2,0,0)
Tcasos<-sum(n_k)
freqrel<-(n_k/Tcasos)
eps=0.000001;
n_bar <- sum(k*n_k)/Tcasos
sum(k*n_k)
s_cua <- (sum(k^2*n_k)/Tcasos)-(n_bar^2)
(sum(k^2*n_k)/Tcasos)
n_bar;s_cua

### Prueba Chicuadrado
Chi_cua<- function(obs, esp,alpha=0.05,m=7,gl=3){
  chi<-sum((obs-esp)^2/esp)
  vchi<-qchisq(1-alpha,df = gl )
  if(chi<vchi){
    return(list("No hay evidencia para rechazar H_0",
                chi,"Valor critico"=vchi))
  }
```

```
  if(chi>=vchi){
    return(list("Existe evidencia para rechazar H_0",
                chi, "Valor critico"=vchi))
  }
}
### intervalo de confianza para el índice de dispersión
d <- s_cua/n_bar #índice de dispersión muestral
d
z <- qnorm(0.975, mean = 0, sd = 1)
z
der <- d + z * d * sqrt((1 / Tcasos) * (2 + ((d - 1) * (3 * d - 1)) / s_cua))
izq <- d - z * d * sqrt((1 / Tcasos) * (2 + ((d - 1) * (3 * d - 1)) / s_cua))

izq;der

# Estimación de parámetros

# Distribucion Poisson
lambda<-n_bar
lambda

# Binomial Negativa (método de momentos)
q_bin <- 1-(n_bar/s_cua)
r_bin <- round(n_bar^2/(s_cua-n_bar),10)
q_bin;r_bin

###Newton

f_r<-function(r,n_bar= n_bar,n_k= n_k){
  as=0;al=0;sol=0
  for(k in 1:(length(n_k)-1)){
    for(m in 0:(k-1)){
      as<-(1/(m+r))+as
    }
    al<- (n_k[k+1]*as)+al
    as=0
  }
  sol<-(log(1+(n_bar/r))*Tcasos)-al
  return(sol)
}
# Funcion f'(r)-----------------
f_der<- function(r,n_bar,n_k){
  as=0;al=0;sol=0
  for(k in 1:(length(n_k)-1)){
    for(m in 0:(k-1)){
      as<-(1/(m+r)^2)+as
    }
    al<-(n_k[k+1]*as)+al
    as=0
  }
  sol <- (-n_bar/(r^2+(n_bar*r)))*Tcasos+al
  return(sol)
}
i=0
r<-r_bin
repeat{
```

```
  r1<-r-(f_r(r = r,n_bar = n_bar,n_k = n_k)/
           f_der(r = r,n_bar = n_bar,n_k = n_k))
  i=i+1
  if(abs(r1-r)<eps){break}
  r=r1
  r1=0
}
r1
q<-n_bar/(n_bar+r1)
q

### Panjer (usando (24))

alpha <-  1-(n_bar/s_cua)
alpha
betha <- ((1-alpha)*(n_bar))-alpha
betha

####################### Estimación de probabilidades
##############################

# Estimación de las probabilidades
Estpois <- dpois(x = k,lambda = lambda)
# Estimacion No. de casos con Poisson
estp<-round(Estpois*Tcasos,3)
data.frame(estp)
sum(estp) # Casos Totales con Poisson
ajuspois1 <- c(estp[1:3],sum(estp[4:9]))
# Se agrupan los datos
n_ka<-c(n_k[1:3],sum(n_k[4:9]))
# Se agrupa de la misma manera los observados
Chi_p<-sum((n_ka-ajuspois1)^2/ajuspois1)
# Se hace la prueba Chi-cuadrado
Chi_cua(obs = n_ka,esp=ajuspois1)

# Estimacion de probabilidades con Binomial
# negativa por el metodo de los momentos
Estbinom1 <- dnbinom(x=k,size=r_bin,prob=(1-q_bin))
# Estimación número de casos con binomial negativa
estbinn1 <- round(Estbinom1*Tcasos,3)
data.frame(estbinn1)
# Casos Totales con la estimación binomial negativa
sum(estbinn1)
ajusbinn11 <- c(estbinn1[1:4],sum(estbinn1[5:9]))
n_ka<-c(n_k[1:4],sum(n_k[5:9]))
# Se agrupa de la misma manera los observados
Chi_binn1<-sum((n_ka-ajusbinn11)^2/ajusbinn11)
# Se hace la prueba Chi-cuadrado
Chi_cua(obs = n_ka,esp=ajusbinn11)

# Estimacion de probabilidades con Binomial
# negativa por el metodo MV
Estbinom2 <- dnbinom(x = k,size = r1,prob = (1-q))
# Estimacion No. de casos con binomial negativa, por MV
estbinn2 <- round(Estbinom2*Tcasos,3)
data.frame(estbinn2)
```

```
# Casos Totales con la estimacion binomial negativa por MV
sum(estbinn2)
ajusbinn21 <- c(estbinn2[1:4],sum(estbinn2[5:9]))
# Se agrupa de la misma manera los observados
n_ka<-c(n_k[1:4],sum(n_k[5:9]))
Chi_binn2<-sum((n_ka-ajusbinn21)^2/ajusbinn21)
# Se hace la prueba Chi-cuadrado
Chi_cua(obs = n_ka,esp=ajusbinn21)

# Estimacion de probabilidades con Panjer
ab <- alpha+betha #es mayor que 0 y menor que 1
ab
r <- ab/alpha

distribucionpanjer <- function(r, alpha) {
  x_vals <- 0:8  # Valores de x de 0 a 8
  resultados <- numeric(length(x_vals))  # Vector para almacenar los resultados

  for (i in 1:length(x_vals)) {
    x <- x_vals[i]
    combinacion <- choose(r + x - 1, x)  # Cálculo de la combinación
    resultado <- combinacion * alpha^x * (1 - alpha)^r  # Cálculo del resultado final
    resultados[i] <- resultado
  }

  # Crear un marco de datos con los valores de x y las probabilidades
  df_resultados <- c(probabilidad = round(resultados, digits = 10))

  return(df_resultados)
}

data.frame(estpanjer)
panjer <- distribucionpanjer(r,alpha)
estpanjer <- round(panjer * Tcasos, 3)
sum(estpanjer) # Casos Totales con Panjer
ajuspanjer1 <- c(estpanjer[1:4], sum(estpanjer[5:9]))
n_ka <- c(n_k[1:4], sum(n_k[5:9]))
Chi_panjer <- sum((n_ka - ajuspanjer1)^2 / ajuspanjer1)
Chi_cua(obs = n_ka, esp = ajuspanjer1)



### Salida Final
test=c("test",".",round(Chi_p,3),round(Chi_binn1,3),round(Chi_binn2,3),round(Chi_panjer,3))
salida=cbind(k,n_k,estp,estbinn1,estbinn2,estpanjer)
salida=rbind(salida,test)
salida
nose<-data.frame(salida)
nose
install.packages("sqldf")
library(sqldf)

total <- sqldf("select sum(n_k) as totalnk, sum(estp) as totalp, sum(estbinn1) as totalb1, sum(estbinn2) a
             from     nose
              ")
```

Referencias

[Jiménez, J. A. (2022)] Introducción a la teoría estadística del riesgo

[Vanegas, L. H. (2022)] https://r2022.netlify.app/