

PROYECTO FINAL ACTUARIAL III

Michelle García Vazquez 143751

Luz Gardida 141657

Fecha entrega: 1 de junio de 2019

En el presente documento se propondrá el análisis de tres modelos para describir tanto la frecuencia como la severidad individual de siniestros, posteriormente, seleccionaremos el modelo que tenga una mejor caracterización de nuestros datos para definir propuestas de agregación de riesgos y generar proyecciones anuales.

Con el propósito de este análisis, utilizaremos el archivo llamado “DanishInsurance MultivarData Full.csv” que contiene información respecto los siniestros ocurridos de tipo building, contents y profits para los años 1980 a 1990. Nuestro enfoque toma como base los años 1980, 1981, 1982, y 1983 para la predicción de los años 1985 y 1986.

Definir nuestro modelo para la frecuencia de siniestros

Proponemos a las distribuciones binomial, poisson y binomial negativa para observar como se comporta nuestros datos respecto a la frecuencia de siniestros. De primera instancia, intuimos que el modelo menos eficiente sería el binomial dado que usando esta distribución asumimos que únicamente es posible un siniestro por póliza, a diferencia de las otras dos distribuciones que permiten más de un siniestro por asegurado.

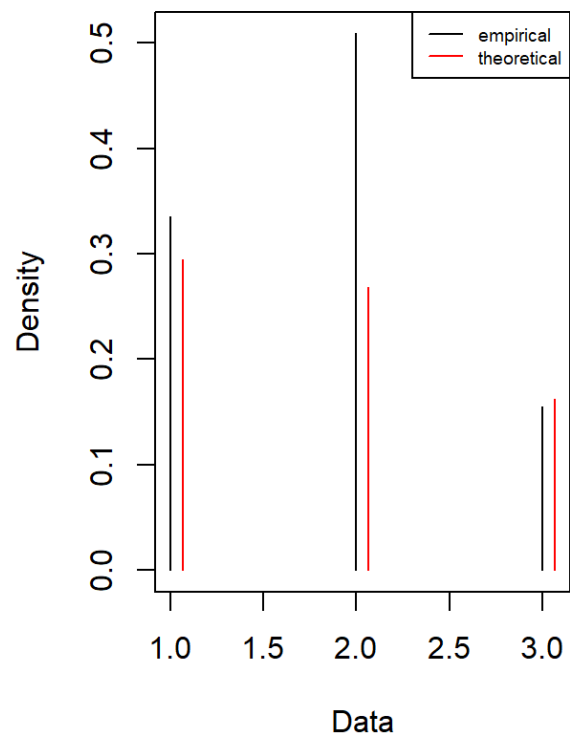
Podemos ver que si graficamos las distribuciones propuestas respecto a nuestros datos base para ver su comportamiento, no nos da una significancia en cuanto a los datos, razón por la cual, usaremos la prueba de bondad de ajuste para mejor caracterización.

```
## Fitting of the distribution 'nbinom' by maximum likelihood
## Parameters :
##           estimate Std. Error
## size 1.033083e+07 7.16591958
## mu    1.819650e+00 0.05211777
## Loglikelihood: -912.1245   AIC: 1828.249   BIC: 1837.263
## Correlation matrix:
##           size          mu
## size 1.000000e+00 1.061471e-08
## mu    1.061471e-08 1.000000e+00
```

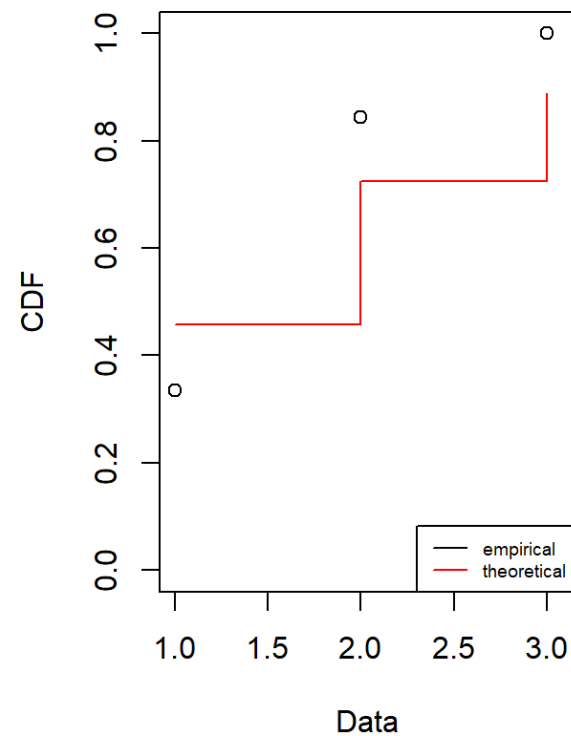
```
## Fitting of the distribution 'pois' by maximum likelihood
## Parameters :
##           estimate Std. Error
## lambda 1.819403    0.0521107
## Loglikelihood: -912.1244   AIC: 1826.249   BIC: 1830.756
```

```
## Fitting of the distribution ' binom ' by maximum likelihood
## Parameters :
##      estimate   Std. Error
## prob 0.003660206 9.604212e-05
## Fixed parameters:
##      value
## size  500
## Loglikelihood:  -911.2327   AIC:  1824.465   BIC:  1828.973
```

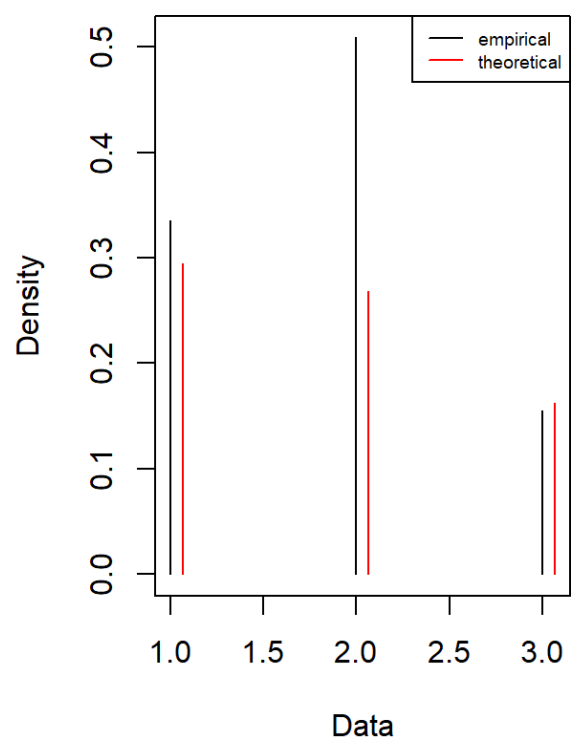
Emp. and theo. distr.



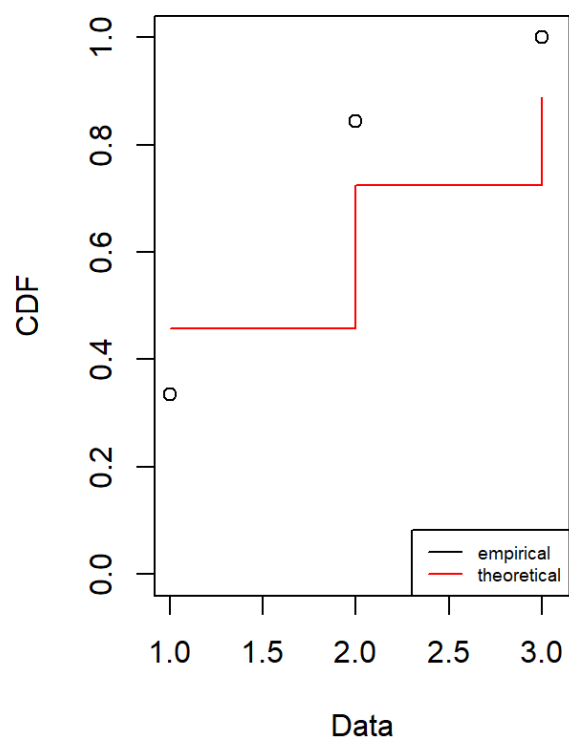
Emp. and theo. CDFs

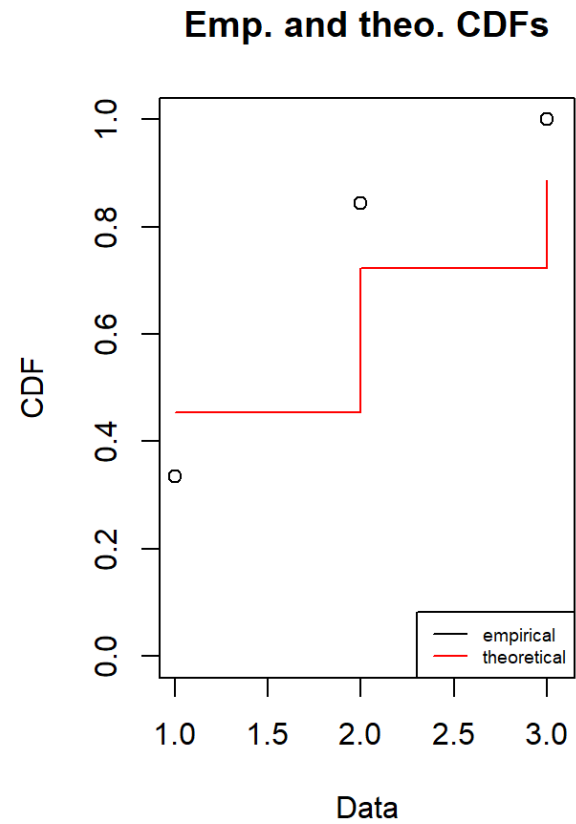
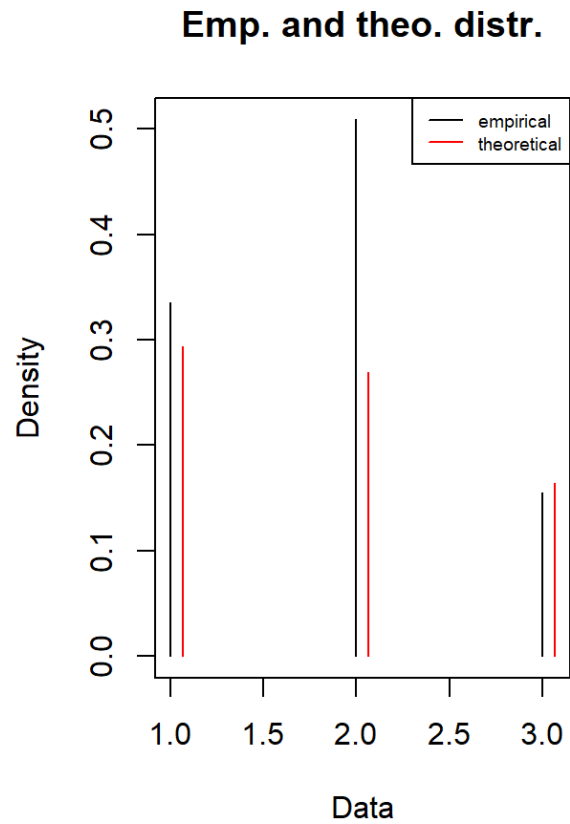


Emp. and theo. distr.



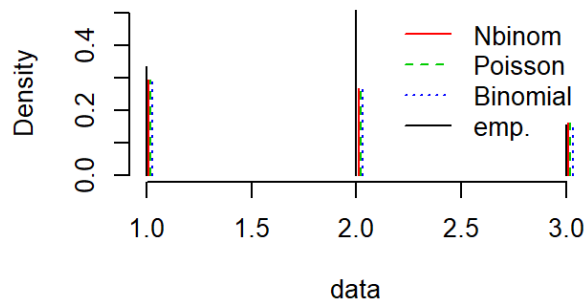
Emp. and theo. CDFs



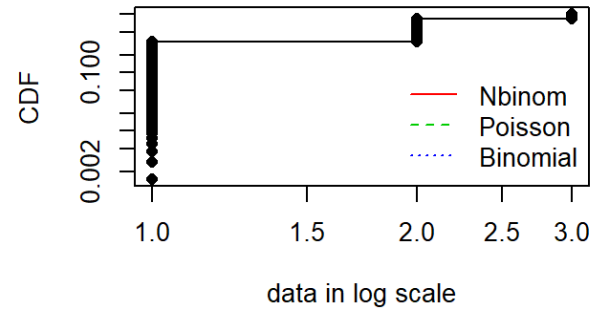


La prueba de bondad de ajuste es un método estadístico que consiste en determinar si los datos de cierta muestra corresponden a cierta distribución, nos permite ver cuál ajusta mejor. Resulta una metodología útil para ver la distribución teórica contra empírica.

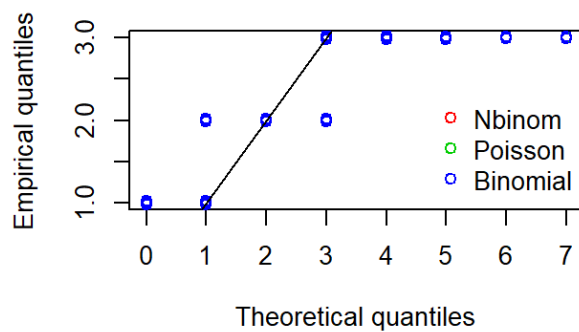
Histogram and theoretical densities



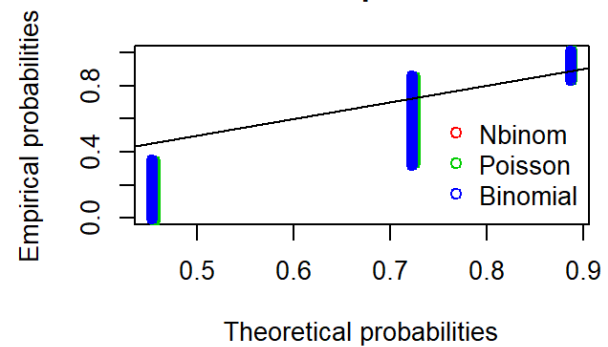
Empirical and theoretical CDFs



Q-Q plot



P-P plot



```
## Chi-squared statistic: 241.2907 241.2961 239.8258
## Degree of freedom of the Chi-squared distribution: 1 2 2
## Chi-squared p-value: 2.057177e-54 4.010775e-53 8.365545e-53
## Chi-squared table:
##      obscounts theo Nbinom theo Poisson theo Binomial
## <= 1      225    306.20070    306.24942    303.8405
## <= 2      341    179.78627    179.78189    180.3224
## <= 3      104    109.04934    109.03190    109.9654
## > 3         0     74.96369     74.93679     75.8717
##
## Goodness-of-fit criteria
##
##               Nbinom Poisson Binomial
## Akaike's Information Criterion 1828.249 1826.249 1824.465
## Bayesian Information Criterion 1837.263 1830.756 1828.973
```

Después de hacer las pruebas de bondad de ajuste, a través de las gráficas observamos que no nos proporcionan información para definir cuál es el mejor modelo para la frecuencia de siniestros, concluimos que las gráficas no nos sirven para definición del modelo.

Recordamos que el valor p nos indica que tan compatible es un modelo a nuestros datos muestra, nos interesa encontrar el que más se aproxime a cero. Para nuestras tres distribuciones propuestas obtenemos el valor p y nos damos cuenta que el menor se obtiene con la binomial negativa con un valor de $2.057177e-54$, por lo que, definimos a ésta como nuestro mejor modelo para la distribución de frecuencias de siniestros.

Definir nuestro modelo para la severidad individual de siniestros

Proponemos a las distribuciones normal, lognormal y gamma para observar como se comporta nuestros datos respecto a la severidad individual de siniestros.

Dadas las gráficas, concluimos que los datos muestra se ajustan mejor a la distribución gamma, por lo que, definimos a ésta como nuestro mejor modelo para la distribución de severidad individual de siniestros.

```
##  
## Begin fitting distributions -----
```

```
## * fitting normal distribution ... OK
```

```
## * fitting Cauchy distribution ... OK
```

```
## * fitting logistic distribution ... OK
```

```
## * fitting beta distribution ... failed
```

```
## * fitting exponential distribution ... OK
```

```
## * fitting chi-square distribution ... OK
```

```
## * fitting uniform distribution ... OK
```

```
## * fitting gamma distribution ... OK
```

```
## * fitting lognormal distribution ... OK
```

```
## * fitting Weibull distribution ... OK
```

```
## * fitting F-distribution ... OK
```

```
## * fitting Student's t-distribution ... OK
```

```
## * fitting Gompertz distribution ... OK
```

```
## * fitting triangular distribution ... failed
```

```
## End fitting distributions -----
```

```
##          logL      AIC      BIC Chisq(value) Chisq(p) AD(value)
## Normal      -2674.89 5353.79  5362.8      6274.69      0      Inf
## Cauchy       -1377.2  2758.4 2767.41      294.40      0     61.89
## Logistic    -1935.57 3875.15 3884.16     1331.88      0      Inf
## Exponential -1656.09 3314.18 3318.68      836.57      0      Inf
## Chi-square  -1685.76 3373.52 3378.02      607.45      0      Inf
## Uniform      NULL     NULL     NULL     8766.78      0      Inf
## Gamma       -1641.88 3287.75 3296.76      733.86      0      Inf
## Lognormal   -1388.75 2781.5 2790.52      329.94      0     28.98
## Weibull     -1652.5  3309 3318.01      859.14      0      Inf
## F           -1668.15 3340.29 3349.31      833.07      0    213.13
## Student     -2280.67 4563.33 4567.84     3199.06      0    623.55
## Gompertz    -4428.46 8860.92 8869.94    2719271.28      0      Inf
##          H(AD) KS(value)   H(KS)
## Normal    rejected      0.41 rejected
## Cauchy     rejected      0.20 rejected
## Logistic   rejected      0.30 rejected
## Exponential rejected      0.26 rejected
## Chi-square  NULL        0.19 rejected
## Uniform     NULL        0.43 rejected
## Gamma      rejected      0.21 rejected
## Lognormal   rejected      0.15 rejected
## Weibull     rejected      0.28 rejected
## F           NULL        0.49 rejected
## Student     NULL        0.78 rejected
## Gompertz    NULL        0.74 rejected
```

```
##
## Chosen continuous distribution is: Gamma (gamma)
## Fitted parameters are:
##      shape      rate
## 1.3128117 0.3013355
```

Propuestas de agregación de riesgos y predicción

Para el modelo de agregación de riesgos, tomamos el número de siniestros de los años 1980, 1981, 1982, 1983, obtenemos la media y la usamos como medida para nuestra simulación ($n=168$). Simularemos 1000 observaciones donde para la frecuencia estimamos a través de la binomial negativa y para la severidad individual a través de la gamma.

Hacemos una comprobación empírica contra teórica para ver que tan aproximado resulta nuestro modelo.

```
## [1] 729.79
```

```
## [1] 610.2132
```

Con estos resultados, tenemos que nuestros datos agregados tienden a una normal.

```
dist_2 <- fit.cont(muestra_s) ### Normal
```

```
##  
## Begin fitting distributions -----
```

```
## * fitting normal distribution ... OK
```

```
## * fitting Cauchy distribution ... OK
```

```
## * fitting logistic distribution ... OK
```

```
## * fitting beta distribution ... failed
```

```
## * fitting exponential distribution ... failed
```

```
## * fitting chi-square distribution ... OK
```

```
## * fitting uniform distribution ... OK
```

```
## * fitting gamma distribution ... OK
```

```
## * fitting lognormal distribution ... OK
```

```
## * fitting Weibull distribution ... OK
```



```
## * fitting F-distribution ... OK
```

```
## * fitting Student's t-distribution ... OK
```

```
## * fitting Gompertz distribution ... failed
```

```
## * fitting triangular distribution ... failed
```

```
## End fitting distributions -----
```

##	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)
## Normal	-5331.41	10666.82	10676.63	24.57	0.49	0.47
## Cauchy	-5502.53	11009.07	11018.88	224.87	0.00	13.33
## Logistic	-5336.2	10676.4	10686.21	24.59	0.49	0.74
## Chi-square	-5495.54	10993.09	10997.99	676.60	0.00	40.81
## Uniform	NULL	NULL	NULL	Inf	0.00	Inf
## Gamma	-5329.35	10662.69	10672.51	20.76	0.71	0.17
## Lognormal	-5330.08	10664.16	10673.98	21.17	0.68	0.26
## Weibull	-5399.86	10803.73	10813.55	117.61	0.00	10.95
## F	-9483.61	18971.23	18981.04	74963.46	0.00	509.69
## Student	-10193.3	20388.6	20393.5	153454.36	0.00	945.94
##	H(AD) KS(value)		H(KS)			
## Normal	not rejected		0.02	not rejected		
## Cauchy	rejected		0.08	rejected		
## Logistic	rejected		0.02	not rejected		
## Chi-square	NULL		0.09	rejected		
## Uniform	NULL		0.07	rejected		
## Gamma	not rejected		0.01	not rejected		
## Lognormal	not rejected		0.02	not rejected		
## Weibull	rejected		0.06	rejected		
## F	NULL		0.66	rejected		
## Student	NULL		0.82	rejected		

```
##
## Chosen continuous distribution is: Normal (norm)
## Fitted parameters are:
##      mean      sd
## 610.21319  50.02233
```

Validación de nuestra propuesta a través de una predicción anual del riesgo agregado para los años 1985 y 1986

```
##  
## Begin fitting distributions -----
```

```
## * fitting normal distribution ... OK
```

```
## * fitting Cauchy distribution ... OK
```

```
## * fitting logistic distribution ... OK
```

```
## * fitting beta distribution ... failed
```

```
## * fitting exponential distribution ... failed
```

```
## * fitting chi-square distribution ... OK
```

```
## * fitting uniform distribution ... OK
```

```
## * fitting gamma distribution ... OK
```

```
## * fitting lognormal distribution ... OK
```

```
## * fitting Weibull distribution ... OK
```

```
## * fitting F-distribution ... OK
```

```
## * fitting Student's t-distribution ... OK
```

```
## * fitting Gompertz distribution ... failed
```

```
## * fitting triangular distribution ... failed
```

```
## End fitting distributions -----
```

```
##          logL      AIC      BIC Chisq(value) Chisq(p) AD(value)
## Normal      -5331.41 10666.82 10676.63         24.57    0.49    0.47
## Cauchy      -5502.53 11009.07 11018.88        224.87    0.00   13.33
## Logistic    -5336.2  10676.4 10686.21         24.59    0.49    0.74
## Chi-square  -5495.54 10993.09 10997.99        676.60    0.00   40.81
## Uniform      NULL      NULL      NULL          Inf    0.00    Inf
## Gamma       -5329.35 10662.69 10672.51         20.76    0.71    0.17
## Lognormal   -5330.08 10664.16 10673.98         21.17    0.68    0.26
## Weibull     -5399.86 10803.73 10813.55        117.61    0.00   10.95
## F           -9483.61 18971.23 18981.04       74963.46    0.00  509.69
## Student     -10193.3 20388.6  20393.5      153454.36    0.00  945.94
##              H(AD) KS(value)          H(KS)
## Normal      not rejected    0.02 not rejected
## Cauchy       rejected      0.08  rejected
## Logistic     rejected      0.02 not rejected
## Chi-square   NULL          0.09  rejected
## Uniform      NULL          0.07  rejected
## Gamma        not rejected    0.01 not rejected
## Lognormal    not rejected    0.02 not rejected
## Weibull      rejected      0.06  rejected
## F            NULL          0.66  rejected
## Student      NULL          0.82  rejected
```

```
##
## Chosen continuous distribution is: Normal (norm)
## Fitted parameters are:
##      mean      sd
## 610.21319  50.02233
```

```
## [1] 741.6257
```

```
## [1] 606.3709
```

Cálculo de primas

La prima de riesgo es el monto que un asegurado paga por la cobertura parcial o total contra un riesgo.

Estimamos la prima pura de riesgo básica que se define como la pérdida esperada: Prima pura = 610.2132

```
## [1] 610.2132
```

Estimamos la prima basada en el principio de la varianza con un nivel de $\alpha = .999$. Definida como la pérdida ajustada por un margen de aceptación de riesgo α , es más conservadora que la prima de riesgo pura.

```
## [1] 764.7938
```

```
## [1] 778.6429
```

```
##          gaussian
## Probability  1.0000
## VaR         764.7938
## E.S.        778.6429
```

Por último, estimamos prima de riesgo a través del principio de utilidad cero. Esta prima resulta la más atractiva para la aseguradora pues aparte de que está en función de la utilidad, incorpora todos los momentos del riesgo en S.

```
## [1] 623.5384
```

ANÁLISIS DE SENSIBILIDAD

Queremos ver como cambian nuestras primas de riesgo, si agregamos las siguientes variables:

1.- Deducible del 10%

2.- Límite de cobertura de la siguiente forma: límite total = 20 límite building = 0 límite contents = 0
límite profits = 13

3.- Coaseguro del 95%

```
## # A tibble: 670 x 3
## # Groups:   a, b [577]
##       a         b         c
##   <dbl> <dbl> <dbl>
## 1  1.68    22  1.68
## 2  2.09    22  2.09
## 3  1.73    22  1.73
## 4  1.78    22  1.78
## 5  4.61    22  4.61
## 6  8.73    22  8.73
## 7  7.90    22  7.90
## 8  2.21    22  2.21
## 9  1.49    22  1.49
## 10 2.80    22  2.80
## # ... with 660 more rows
```

```
## # A tibble: 670 x 3
## # Groups:   a, b [452]
##       a      b      c
##   <dbl> <dbl> <dbl>
## 1 1.10      0      0
## 2 1.76      0      0
## 3 1.73      0      0
## 4 0         0      0
## 5 1.24      0      0
## 6 4.45      0      0
## 7 2.49      0      0
## 8 0.776     0      0
## 9 0.813     0      0
## 10 2.37      0      0
## # ... with 660 more rows
```

```
## # A tibble: 670 x 3
## # Groups:   a, b [377]
##       a      b      c
##   <dbl> <dbl> <dbl>
## 1 0.586     0      0
## 2 0.337     0      0
## 3 0         0      0
## 4 1.31      0      0
## 5 3.37      0      0
## 6 4.27      0      0
## 7 3.54      0      0
## 8 0.993     0      0
## 9 0.673     0      0
## 10 0.168     0      0
## # ... with 660 more rows
```

```
## # A tibble: 670 x 3
## # Groups:   a, b [97]
##       a      b      c
##   <dbl> <dbl> <dbl>
## 1 0      13 0
## 2 0      13 0
## 3 0      13 0
## 4 0.474  13 0.474
## 5 0      13 0
## 6 0      13 0
## 7 1.86   13 1.86
## 8 0.439  13 0.439
## 9 0      13 0
## 10 0.256  13 0.256
## # ... with 660 more rows
```

**** Severidad limitada ****

```
##
## Begin fitting distributions -----
```

```
## * fitting normal distribution ... OK
```

```
## * fitting Cauchy distribution ... OK
```

```
## * fitting logistic distribution ... OK
```

```
## * fitting beta distribution ... failed
```

```
## * fitting exponential distribution ... OK
```

```
## * fitting chi-square distribution ... OK
```

```
## * fitting uniform distribution ... OK
```

```
## * fitting gamma distribution ... OK
```

```
## * fitting lognormal distribution ... OK
```

```
## * fitting Weibull distribution ... OK
```

```
## * fitting F-distribution ... OK
```

```
## * fitting Student's t-distribution ... OK
```

```
## * fitting Gompertz distribution ... failed
```

```
## * fitting triangular distribution ... failed
```

```
## End fitting distributions -----
```

##	logL	AIC	BIC	Chisq(value)	Chisq(p)	AD(value)
## Normal	-1786.13	3576.25	3585.27	1789.97	0	99.13
## Cauchy	-1259.01	2522.02	2531.03	294.55	0	61.90
## Logistic	-1576.63	3157.26	3166.28	2172.28	0	58.49
## Exponential	-1442.96	2887.93	2892.44	750.32	0	65.97
## Chi-square	-1374.81	2751.61	2756.12	615.38	0	53.93
## Uniform	NULL	NULL	NULL	Inf	0	Inf
## Gamma	-1368.25	2740.5	2749.52	615.62	0	48.77
## Lognormal	-1234.35	2472.69	2481.71	349.73	0	26.88
## Weibull	-1416.4	2836.79	2845.81	686.26	0	54.39
## F	-1469.06	2942.12	2951.13	640.30	0	187.59
## Student	-2101.48	4204.96	4209.47	2799.67	0	614.85

##	H(AD)	KS(value)	H(KS)
## Normal	rejected	0.28	rejected
## Cauchy	rejected	0.20	rejected
## Logistic	rejected	0.25	rejected
## Exponential	rejected	0.30	rejected
## Chi-square	NULL	0.20	rejected
## Uniform	NULL	0.33	rejected
## Gamma	rejected	0.19	rejected
## Lognormal	rejected	0.15	rejected
## Weibull	rejected	0.22	rejected
## F	NULL	0.44	rejected
## Student	NULL	0.77	rejected

```
##  
## Chosen continuous distribution is: Log-normal (lnorm)  
## Fitted parameters are:  
## meanlog sdlog  
## 0.8788268 0.6341571
```

Agregación limitada

```
## [1] 412.3151
```

**** Distribución de datos agregados ****

```
##  
## Begin fitting distributions -----
```

```
## * fitting normal distribution ... OK
```

```
## * fitting Cauchy distribution ... OK
```

```
## * fitting logistic distribution ... OK
```

```
## * fitting beta distribution ... failed
```

```
## * fitting exponential distribution ... OK
```

```
## * fitting chi-square distribution ... OK
```

```
## * fitting uniform distribution ... OK
```

```
## * fitting gamma distribution ... OK
```

```
## * fitting lognormal distribution ... OK
```

```
## * fitting Weibull distribution ... OK
```

```
## * fitting F-distribution ... OK
```

```
## * fitting Student's t-distribution ... OK
```

```
## * fitting Gompertz distribution ... failed
```

```
## * fitting triangular distribution ... failed
```

```
## End fitting distributions -----
```



```
##          logL      AIC      BIC Chisq(value) Chisq(p) AD(value)
## Normal      -4782.24  9568.48   9578.3         22.91    0.58    0.52
## Cauchy      -4960.85  9925.7   9935.51        237.46    0.00   12.56
## Logistic    -4793.01  9590.03  9599.84         30.46    0.21    0.85
## Exponential -7021.79 14045.58 14050.48       10018.72    0.00   398.27
## Chi-square  -4780.36  9562.72  9567.62         21.05    0.74    0.34
## Uniform      NULL      NULL      NULL          Inf    0.00    Inf
## Gamma       -4780.33  9564.65  9574.47         20.82    0.70    0.32
## Lognormal   -4780.53  9565.06  9574.87         21.32    0.67    0.38
## Weibull     -4846.12  9696.24  9706.05         105.57    0.00   10.33
## F           -9034.85 18073.69 18083.51       81582.33    0.00  514.04
## Student     -9748.63 19499.26 19504.16      167602.77    0.00  950.29
##          H(AD) KS(value)      H(KS)
## Normal      not rejected    0.02 not rejected
## Cauchy       rejected      0.08  rejected
## Logistic     rejected      0.02 not rejected
## Exponential  rejected      0.56  rejected
## Chi-square   NULL          0.02 not rejected
## Uniform      NULL          0.07  rejected
## Gamma        not rejected    0.02 not rejected
## Lognormal    not rejected    0.02 not rejected
## Weibull      rejected      0.06  rejected
## F            NULL          0.67  rejected
## Student      NULL          0.82  rejected
```

```
##
## Chosen continuous distribution is: Normal (norm)
## Fitted parameters are:
##      mean      sd
## 412.31514 28.88442
```

**** Primas de riesgo con análisis de sensibilidad ****

Prima pura

```
## [1] 412.3151
```

Prima basada en el principio de la varianza

```
##          gaussian
## Probability  1.0000
## VaR         501.5236
## E.S.        509.4905
```

```
## [1] 501.5236
```

```
## [1] 509.4905
```

```
##          gaussian
## Probability  1.0000
## VaR        501.5236
## E.S.       509.4905
```

Prima basada en el principio de utilidad cero

```
## [1] 428.4256
```

Concluimos que nuestros modelos definidos en un principio fueron correctos para la distribución y predicción de los datos.