

# **Ajuste de Curvas de Probabilidad en R para el Evento de Incendio Forestal de Acuerdo con la Información del Atlas Nacional de Riesgos.**

Jorge Uriel Barragan Pozos

08/09/21

Actualice los montos de pérdida para valuar todos en el año 2015 considerando incrementos anuales de la inflación.

### 1)Cargamos las librerías que vamos a necesitar

```
library(knitr)
library(dplyr) #Manipulación de Dataframes
library(readxl) #Importar Datos de Excel
library(MASS) #Distribucion de Probabilidad
library(actuar) #Distribuciones Adicionales
library(fitdistrplus) #Ajuste de Curvas de Probabilidad
library(goftest) #Pruebas de Bondad de Ajuste
library(ggplot2) #Gráficas
```

### 2)Importamos los datos de año y costo del siniestro y las tasa de interes anuales entre el año 2000 y 2015

Encontramos los datos de la tasa de inflación anual en la sig. página de internet: [https://www.proyectosmexico.gob.mx/por-que-invertir-en-mexico/economia-solida/politica-monetaria/sd\\_tasas-de-inflacion-historicas/](https://www.proyectosmexico.gob.mx/por-que-invertir-en-mexico/economia-solida/politica-monetaria/sd_tasas-de-inflacion-historicas/) y, en excel, creamos una nueva columna “factor\_ajuste” que es el producto de  $(1 + i)$  donde  $i$  es la tasa de inflación de los periodos que queremos llevar al 2015.

```
#Importamos los datos de año y costo del siniestro
incendios <- read_excel("AP Ajuste de Curvas de Probabilidad - Atlas de Riesgos.xlsx",
                      sheet = "IF")
#Importamos una tabla con los datos de la tasa de interés anual
inflacion <- read_excel("INFLACION.xlsx")%>%
#nos quedamos con la información entre 2000 y 2015
  filter(Año >= 2000 & Año <= 2015)
```

### 3) Observamos las tablas importadas

```
#Tabla de incendios contiene 464 onservaciones de siniestros con los datos de año y costo.  
#Resumen de los Datos de Incendios Forestales  
kable(  
  top_n(incendios, 20)#Mostramos las primeras 20 observaciones  
)
```

```
## Selecting by Incendio Forestal (mdp)
```

Año	Incendio Forestal (mdp)
2002	88.10000
2002	41.00000
2003	75.62200
2003	72.33700
2003	55.59000
2003	49.87500
2003	37.86000
2005	38.73025
2006	53.61850
2009	71.85466
2009	42.35000
2011	87.92005
2011	79.02220
2011	63.23700
2011	42.27850
2011	42.00540
2012	75.20200
2012	51.90261
2012	51.62643
2014	49.78109

```
#Tabla de Inflación contiene los datos de las tasas anuales en mexico  
#La columna factor_ajuste nos permiten llevar las cantidades a su valor en 2015  
kable(inflacion)
```

Año	tasa_inf	tasa porcentual	factor_ajuste
2000	0.0896	8.96	1.966356
2001	0.0440	4.40	1.804658
2002	0.0570	5.70	1.728600
2003	0.0398	3.98	1.635383
2004	0.0519	5.19	1.572786
2005	0.0333	3.33	1.495186
2006	0.0405	4.05	1.447001
2007	0.0376	3.76	1.390678
2008	0.0653	6.53	1.340284
2009	0.0357	3.57	1.258128
2010	0.0440	4.40	1.214761
2011	0.0382	3.82	1.163564

Año	tasa_inf	tasa porcentual	factor_ajuste
2012	0.0357	3.57	1.120751
2013	0.0397	3.97	1.082120
2014	0.0408	4.08	1.040800
2015	0.0213	2.13	1.000000

```
#Juntamos la tablas de incendios e inflación en una sola tabla incendio_ajustado
incendio_ajustado <- incendios%>%
  inner_join(inflacion, by = "Año")%>%
#También, creamos una nueva columna que es el costo valuado en 2015
mutate(costo_real = `Incendio Forestal (mdp)`*factor_ajuste)
```

Obtenemos la siguiente tabla:

```
kable(top_n(incendio_ajustado,25)) #Mostramos las primeras 25 observaciones
```

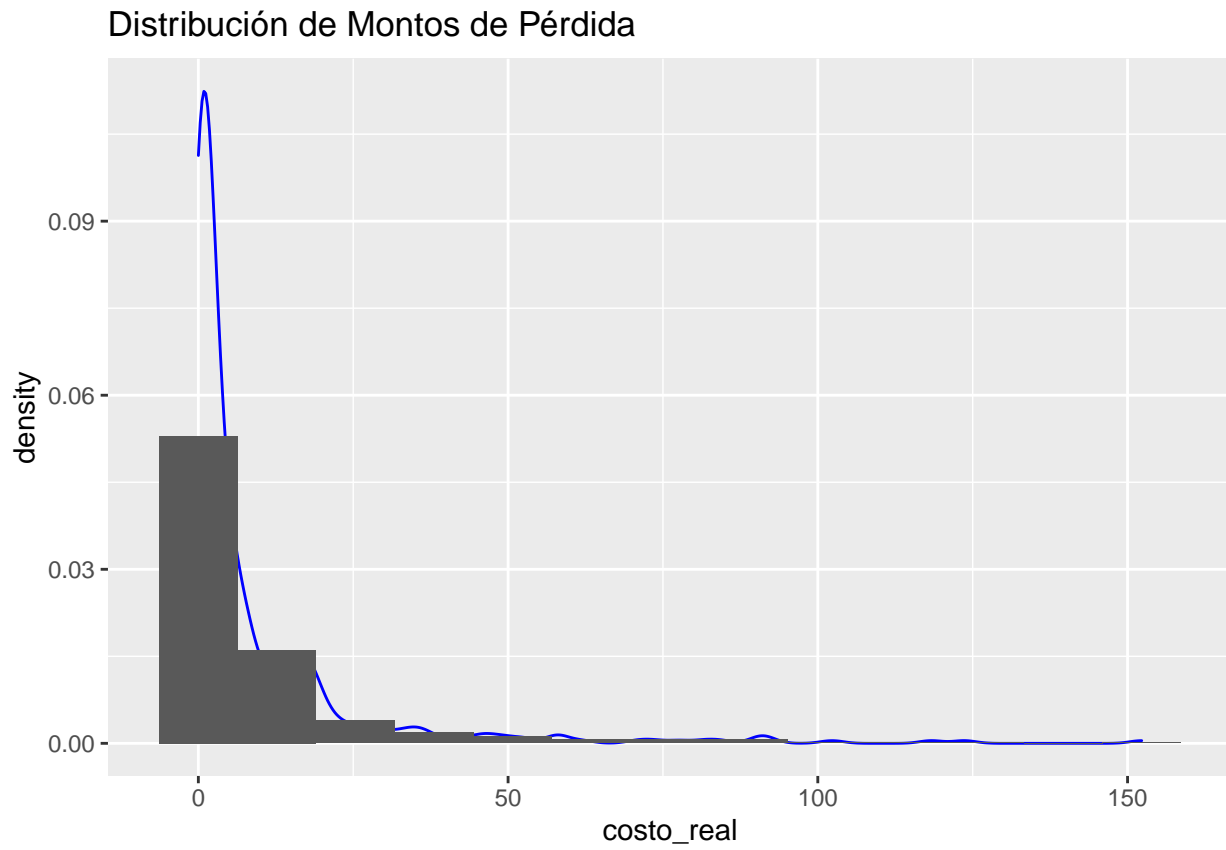
```
## Selecting by costo_real
```

Año	Incendio Forestal (mdp)	tasa_inf	tasa porcentual	factor_ajuste	costo_real
2002	88.10000	0.0570	5.70	1.728600	152.28964
2002	41.00000	0.0570	5.70	1.728600	70.87259
2003	75.62200	0.0398	3.98	1.635383	123.67093
2003	72.33700	0.0398	3.98	1.635383	118.29870
2003	55.59000	0.0398	3.98	1.635383	90.91094
2003	49.87500	0.0398	3.98	1.635383	81.56473
2003	37.86000	0.0398	3.98	1.635383	61.91560
2003	28.26700	0.0398	3.98	1.635383	46.22737
2003	27.60000	0.0398	3.98	1.635383	45.13657
2005	38.73025	0.0333	3.33	1.495186	57.90893
2005	30.93800	0.0333	3.33	1.495186	46.25806
2005	28.40700	0.0333	3.33	1.495186	42.47375
2006	53.61850	0.0405	4.05	1.447001	77.58601
2007	29.68600	0.0376	3.76	1.390678	41.28368
2009	71.85466	0.0357	3.57	1.258128	90.40235
2009	42.35000	0.0357	3.57	1.258128	53.28172
2011	87.92005	0.0382	3.82	1.163564	102.30062
2011	79.02220	0.0382	3.82	1.163564	91.94740
2011	63.23700	0.0382	3.82	1.163564	73.58031
2011	42.27850	0.0382	3.82	1.163564	49.19375
2011	42.00540	0.0382	3.82	1.163564	48.87598
2012	75.20200	0.0357	3.57	1.120751	84.28275
2012	51.90261	0.0357	3.57	1.120751	58.16992
2012	51.62643	0.0357	3.57	1.120751	57.86040
2014	49.78109	0.0408	4.08	1.040800	51.81216

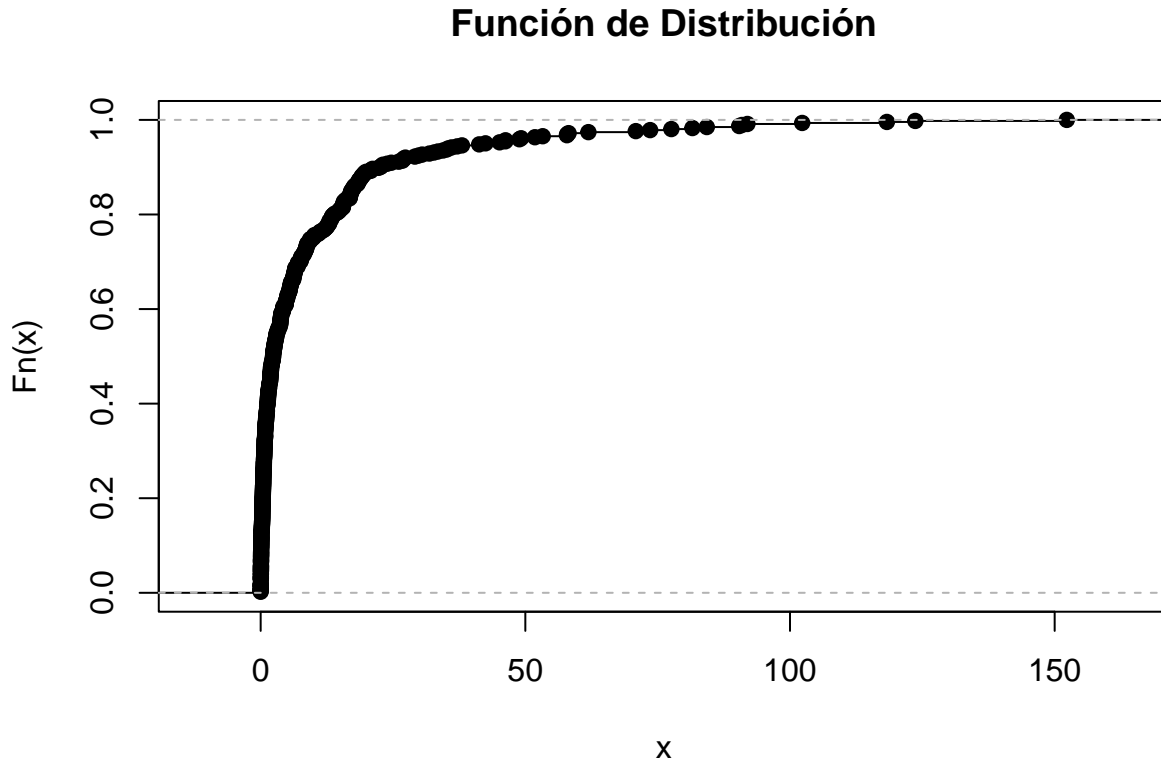
# Análisis de Estadística Descriptiva de los Datos

1) Veamos el histograma de los costos reales.

```
ggplot(incendio_ajustado, aes(costo_real)) + geom_density(color = 'blue')+  
  geom_histogram(aes(y = ..density..), bins = 13)+ labs(title = "Distribución de Montos de Pérdida")
```



```
#Función de Distribución
plot(ecdf(incendio_ajustado$costo_real), main= "Función de Distribución")
```



## 2)Cambiaremos la escala de los datos

Con el objetivo de suavizar el histograma, creamos una columna nueva “logdata” que será  $[\ln(\text{costo\_real}) + 7]$ , sumamos 7 unidades para recorrer el histograma a la derecha y tener valores mayores que 0 para poder ajustar las curvas de probabilidad.

```
datos <- incendio_ajustado %>%
  mutate(logdata = log(costo_real) + 7 )

kable(top_n(datos, 8))#Primeros 8 obs. de tabla actualizada
```

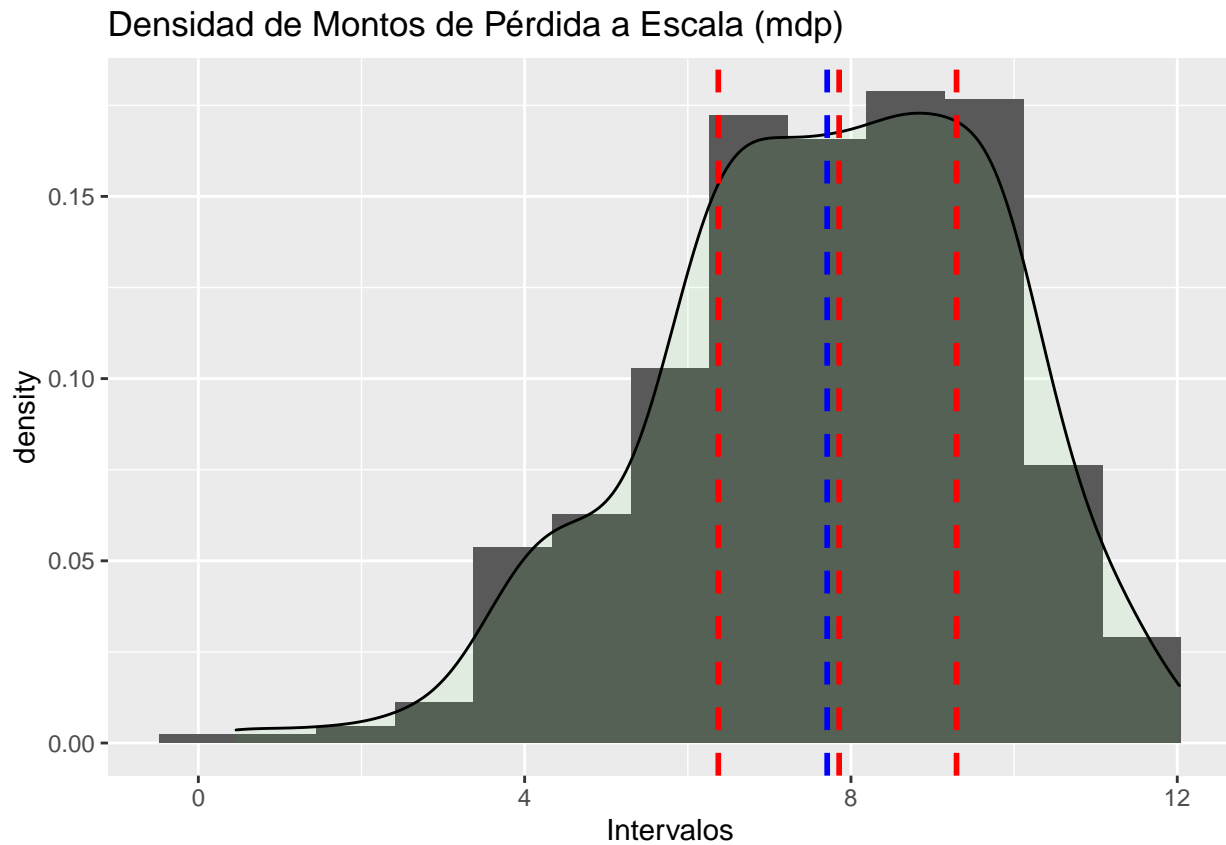
## Selecting by logdata

Año	Incendio Forestal (mdp)	tasa_inf	tasa porcentual	factor_ajuste	costo_real	logdata
2002	88.10000	0.0570	5.70	1.728600	152.28964	12.02578
2003	75.62200	0.0398	3.98	1.635383	123.67093	11.81762
2003	72.33700	0.0398	3.98	1.635383	118.29870	11.77321
2003	55.59000	0.0398	3.98	1.635383	90.91094	11.50988
2009	71.85466	0.0357	3.57	1.258128	90.40235	11.50427
2011	87.92005	0.0382	3.82	1.163564	102.30062	11.62792
2011	79.02220	0.0382	3.82	1.163564	91.94740	11.52122
2012	75.20200	0.0357	3.57	1.120751	84.28275	11.43418

### 3) Veamos el histograma en escala logaritmica recorrido 7 unidades

```
#Datos Estadística Descriptiva
est_des <- summary(datos$logdata)
ggplot(datos, aes(logdata)) + geom_histogram(aes(y = ..density..), bins = 13)+ geom_density(alpha =
  labs(x= "Intervalos", title = "Densidad de Montos de Pérdida a Escala (mdp)"+

  geom_vline(aes(xintercept= est_des[4] ),#Media
    color="blue", linetype="dashed", size=1) +
  geom_vline(aes(xintercept= est_des[2] ),#Primer Cuartil
    color="red", linetype="dashed", size=1) +
  geom_vline(aes(xintercept= est_des[3] ), #Mediana
    color="red", linetype="dashed", size=1) +
  geom_vline(aes(xintercept= est_des[5] ),#Tercer Cuartil
    color="red", linetype="dashed", size=1)
```

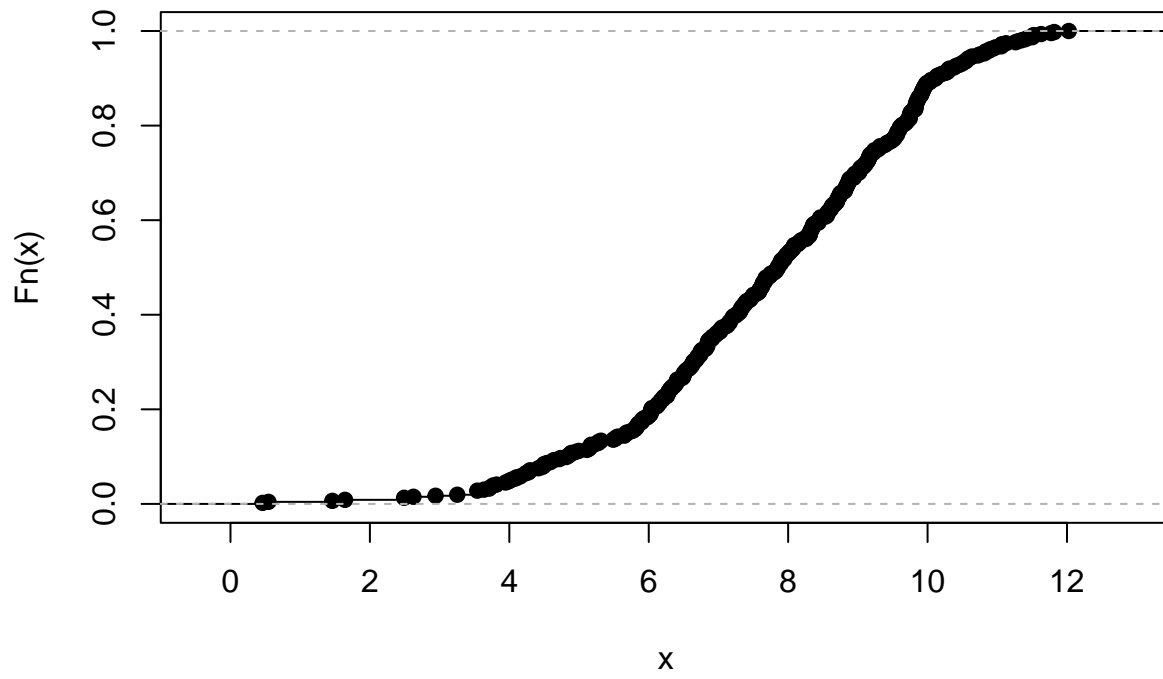


est\_des

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.4617  6.3739   7.8553   7.7089  9.2944  12.0258
```

```
#Función de Distribución
plot(ecdf(datos$logdata), main = "Función de Distribución a Escala")
```

## Función de Distribución a Escala





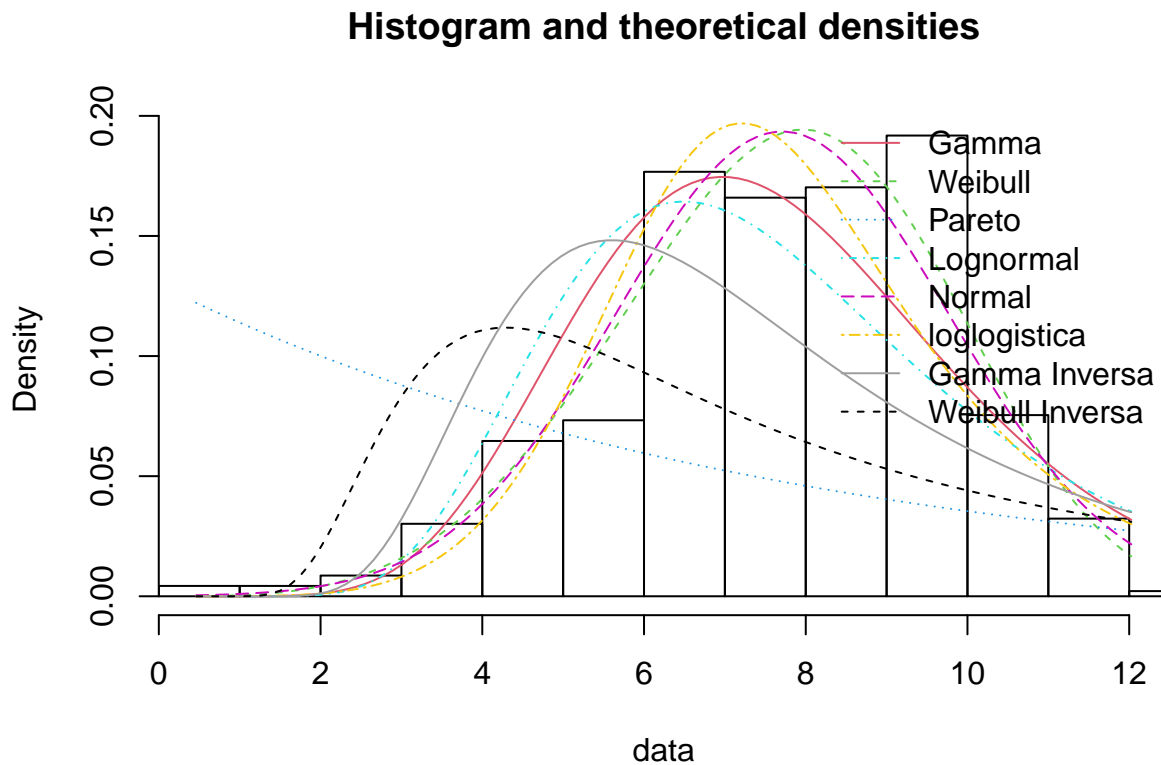
# Ajuste de Curvas de Probabilidad

## Comparativo de Ajustes de Curvas de Probabilidad

```
leyenda <- c("Gamma", "Weibull", "Pareto", "Lognormal","Normal","loglogistica","Gamma Inversa","Weibull
```

### a)Función de Densidad

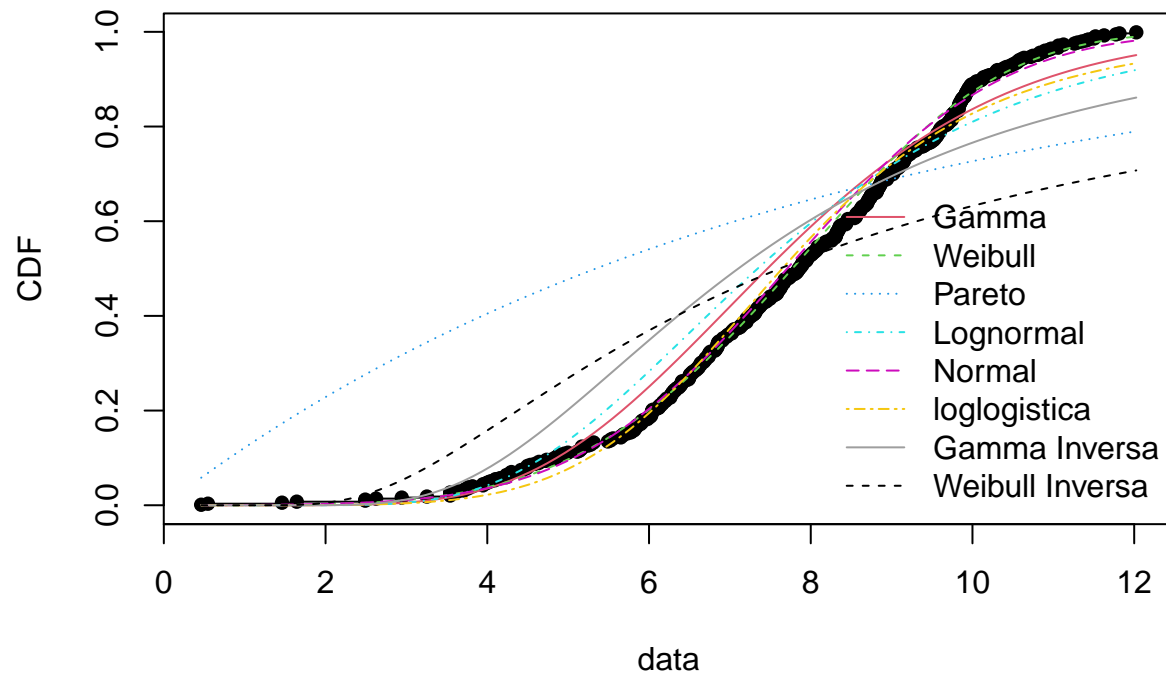
```
denscomp(list(mod1, mod2, mod3, mod4, mod5, mod6, mod7, mod8), legendtext = leyenda)
```



## b)Función de Distribución

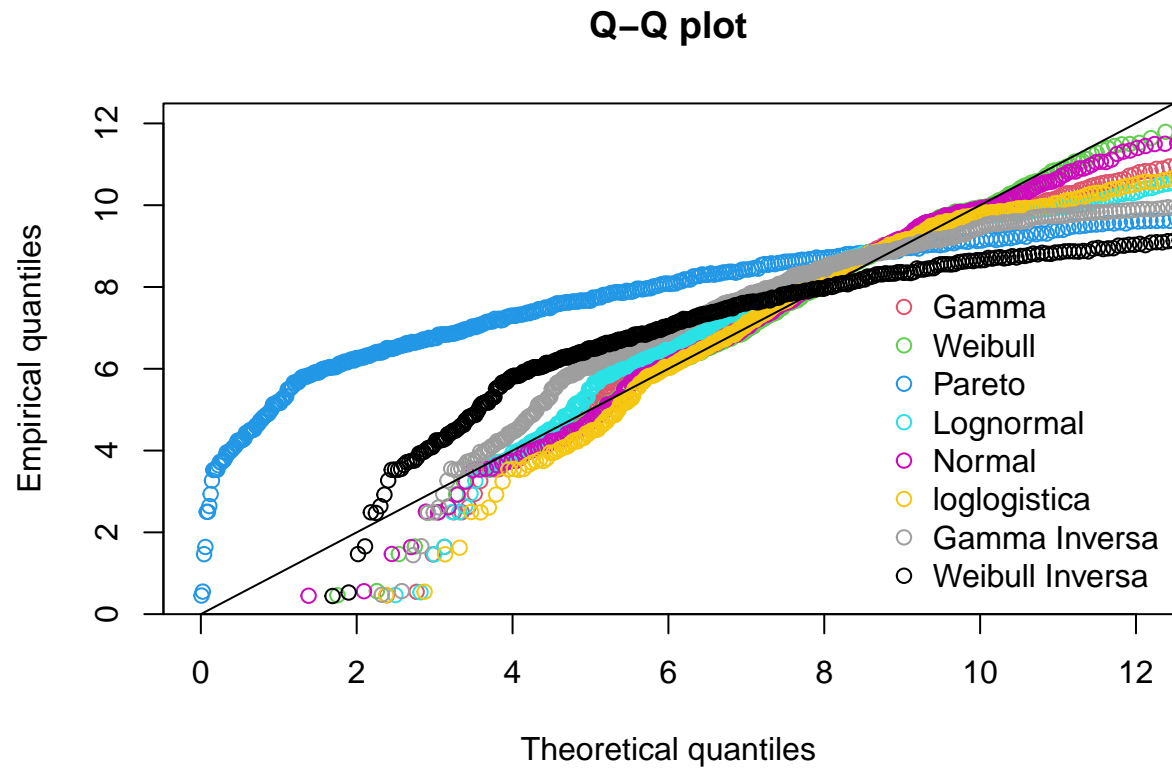
```
cdfcomp(list(mod1, mod2, mod3, mod4, mod5, mod6, mod7, mod8), legendtext = leyenda)
```

### Empirical and theoretical CDFs



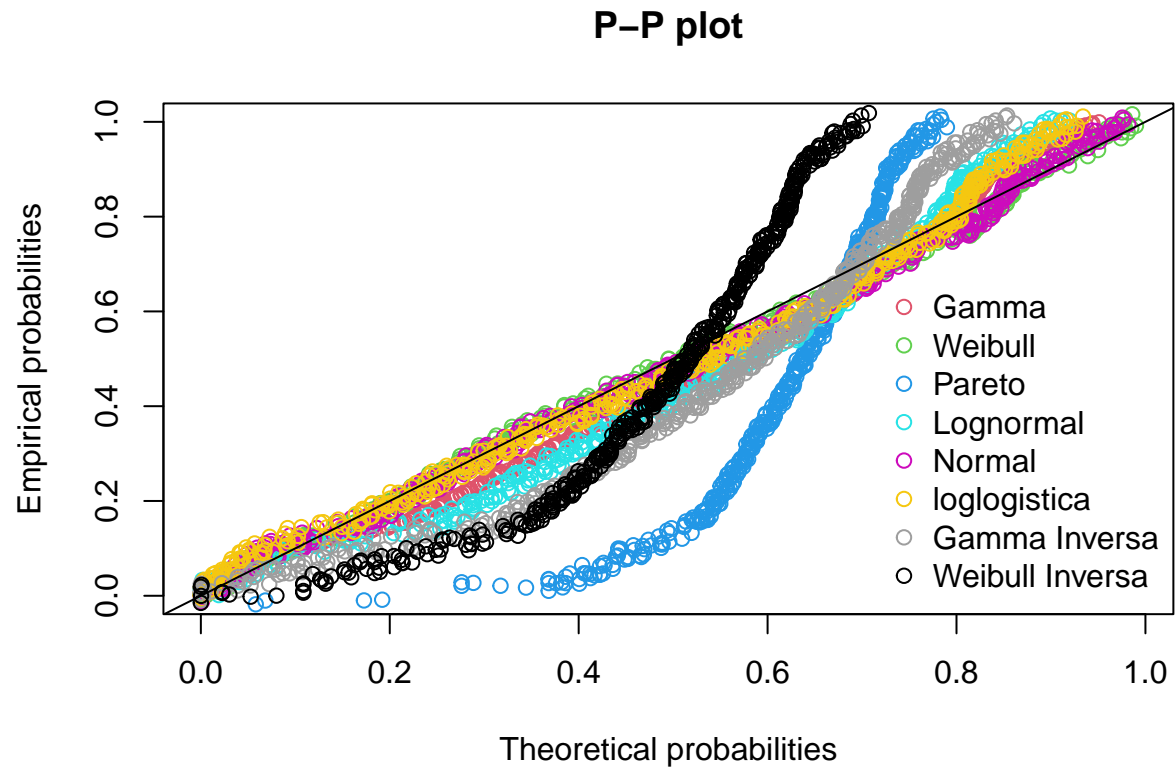
### c) Cuantiles Teóricos vs Empíricos

```
qqcomp(list(mod1, mod2, mod3, mod4, mod5, mod6, mod7, mod8), legendtext = leyenda, xlim = c(0,12))
```



d) Probabilidades Teóricas vs Empíricas

```
ppcomp(list(mod1, mod2, mod3, mod4, mod5, mod6, mod7, mod8), legendtext = leyenda)
```



## Pruebas de Bondad de Ajuste

```
gofstat(list(mod1, mod2, mod3, mod4, mod5, mod6, mod7, mod8))
```

```
## Goodness-of-fit statistics
##               1-mle-gamma 2-mle-weibull 3-mle-pareto 4-mle-lnorm
## Kolmogorov-Smirnov statistic 0.07380089 0.04269195 0.3773036 0.0991657
## Cramer-von Mises statistic 1.13791170 0.16530104 23.4435938 2.0992650
## Anderson-Darling statistic 7.28290738 1.16801873 113.5048724 13.0880721
##               5-mle-norm 6-mle-llogis 7-mle-invgamma
## Kolmogorov-Smirnov statistic 0.05175177 0.0753979 0.1647074
## Cramer-von Mises statistic 0.26067735 0.5511577 4.9655879
## Anderson-Darling statistic 1.81308628 5.7607027 28.7049124
##               8-mle-invweibull
## Kolmogorov-Smirnov statistic 0.3004398
## Cramer-von Mises statistic 11.1957559
## Anderson-Darling statistic 61.1729527
##
## Goodness-of-fit criteria
##               1-mle-gamma 2-mle-weibull 3-mle-pareto
## Akaike's Information Criterion 2095.951 1989.265 2827.327
## Bayesian Information Criterion 2104.231 1997.544 2835.607
##               4-mle-lnorm 5-mle-norm 6-mle-llogis
## Akaike's Information Criterion 2201.189 1992.381 2090.629
## Bayesian Information Criterion 2209.469 2000.661 2098.909
##               7-mle-invgamma 8-mle-invweibull
## Akaike's Information Criterion 2427.044 2692.361
## Bayesian Information Criterion 2435.324 2700.641
```

## Conclusión

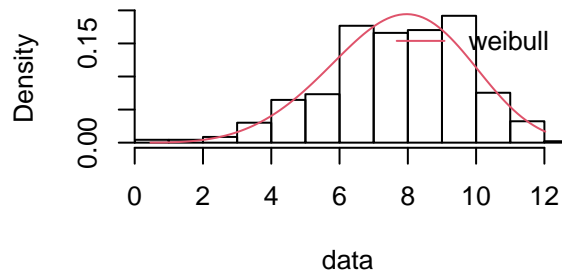
Con base en la información de las pruebas de bondad de ajuste, podemos ver que la distribución con el menor criterio de Akaike y Bayesiano es la distribución Weibull. Así concluimos que la distribución que mejor se ajusta a los datos es la distribución Weibull con parámetros:

```
mod2
```

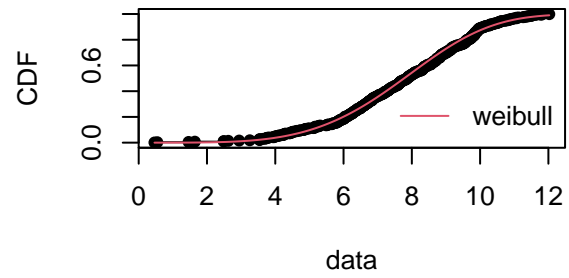
```
## Fitting of the distribution ' weibull ' by maximum likelihood
## Parameters:
##      estimate Std. Error
## shape 4.340739 0.16279276
## scale 8.458998 0.09477672
```

```
par(mfrow = c(2,2))
denscomp(mod2) # Función de Densidad
cdfcomp(mod2) #Función de Distribución
qqcomp(mod2) #cuantiles Teóricos vs Empíricos
ppcomp(mod2) #Probabilidades Teóricas vs Empíricas
```

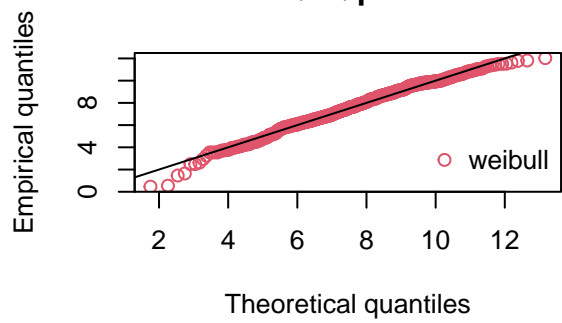
**Histogram and theoretical densities**



**Empirical and theoretical CDFs**



**Q-Q plot**



**P-P plot**

