

Práctica 7 Simulación de Sistemas

Mónica Sofía Restrepo León

1000181289

grupo 2

1. Comparar el histograma de frecuencias obtenido con las funciones de R con uno obtenido con el método descrito en el libro de ejercicios.

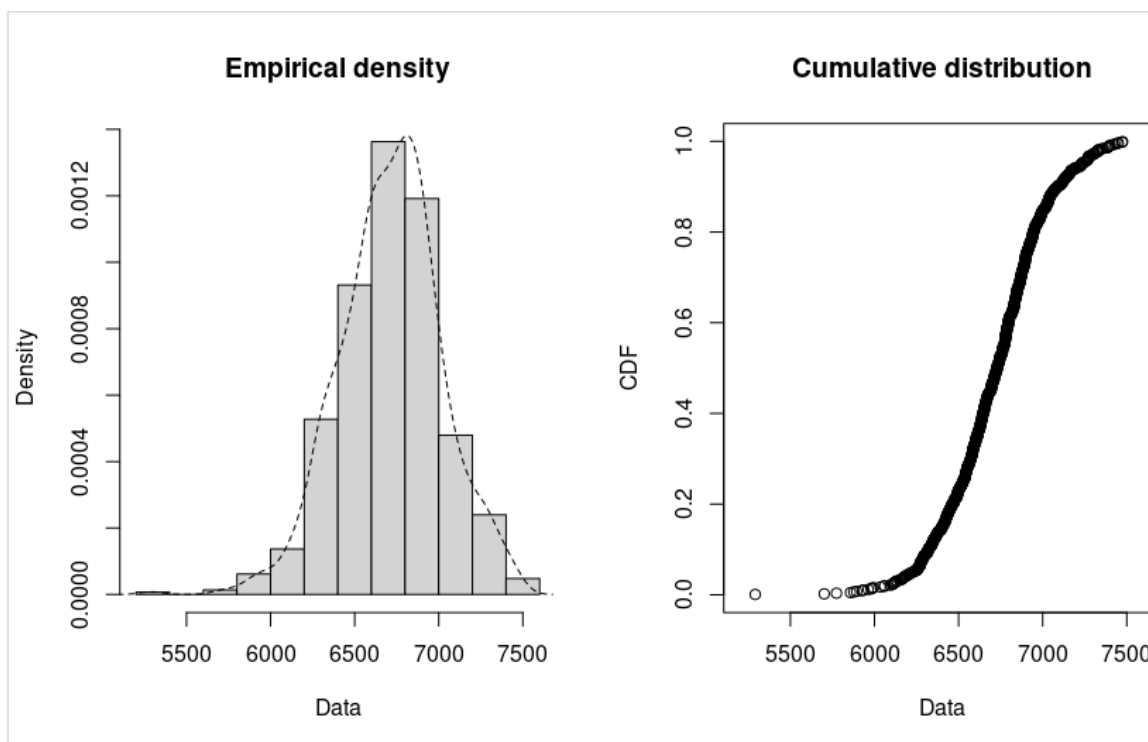
Para el archivo “demcom” los siguientes comandos:

```
demcom = read.table("demcom.txt")
```

```
x<-demcom$V1
```

```
plotdist(x, histo =TRUE, demp = TRUE)
```

Veremos el gráfico de la distribución observada:

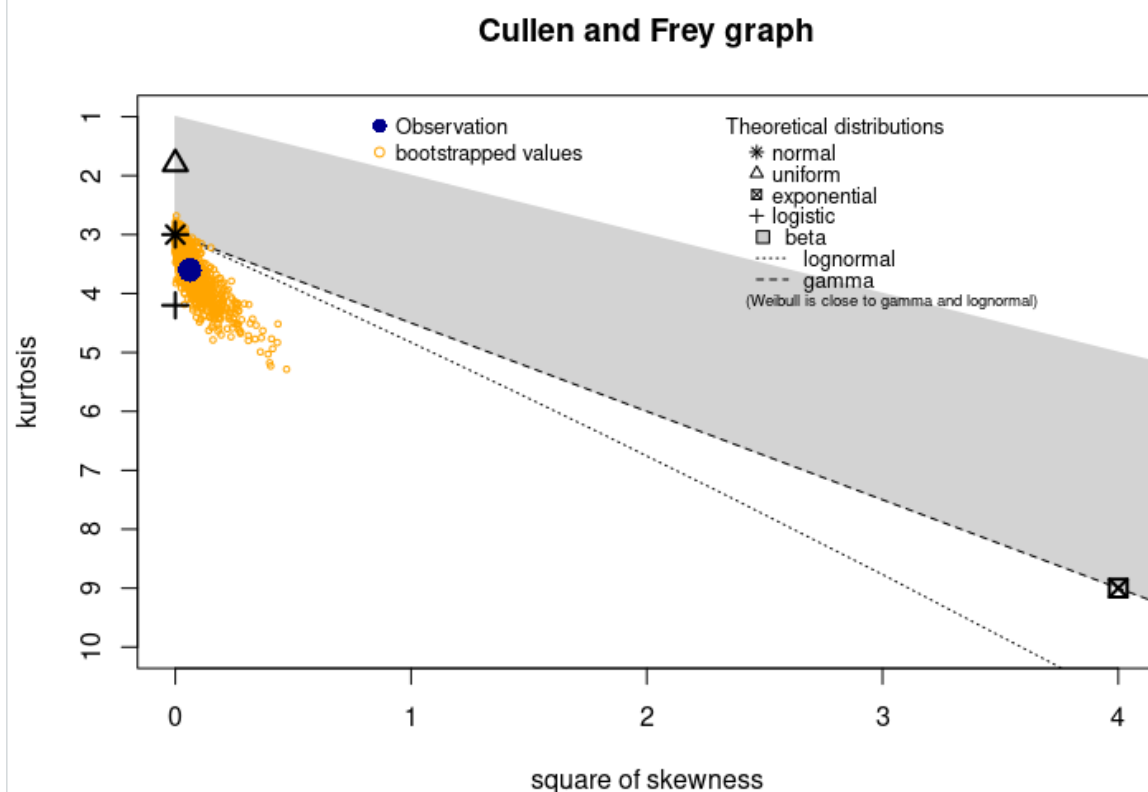


Procedemos con un análisis:

```
> summary(x)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  5291   6527   6729   6714   6902   7475

> descdist(x)
summary statistics
-----
min: 5290.831  max: 7474.839
median: 6728.713
mean: 6713.851
estimated sd: 303.0255
estimated skewness: -0.2479502
estimated kurtosis: 3.603348
> descdist(x,boot=1000)
summary statistics
-----
min: 5290.831  max: 7474.839
median: 6728.713
mean: 6713.851
estimated sd: 303.0255
estimated skewness: -0.2479502
estimated kurtosis: 3.603348
```

ahora, `descdist(x,boot=1000)` para calcular los momentos en muestreo bootstrap, retorna el siguiente ploteo:



Con la información de este gráfico podemos proceder al punto 2.

2. Siguiendo el procedimiento descrito en la clase 7 y el libro de R "¿Cómo ajustar datos de entrada? Proponer y ajustar modelos para representar la muestra de datos suministrada.

Según los datos del gráfico Cullen y Frey, nuestras observaciones son más cercanas en distancia a las distribuciones normal y logística, ahora usaremos diferentes comandos para decidir entre ambas distribuciones.

Primero asignamos variables al ajuste para las distribuciones más cercanas, normal y logística.

```
fit_logis<-fitdist(x,"logis")
```

```
fit_norm <-fitdist(x,"norm")
```

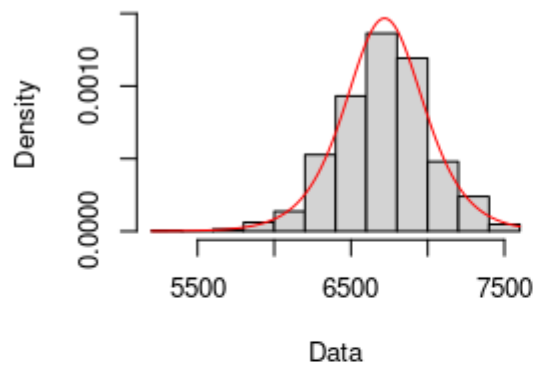
Los comandos:

```
plot(fit_logis)
```

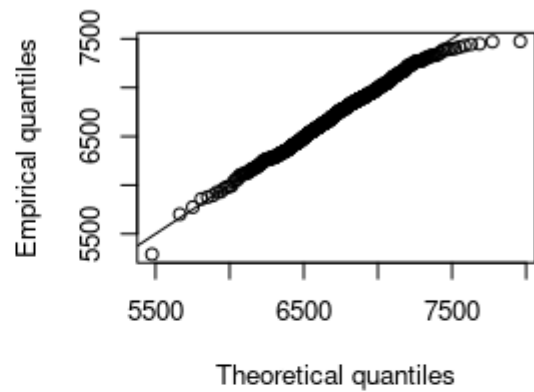
```
summary(fit_ln)
```

Revelan gráficas para ayudarnos a tomar la decisión sobre la distribución logística.

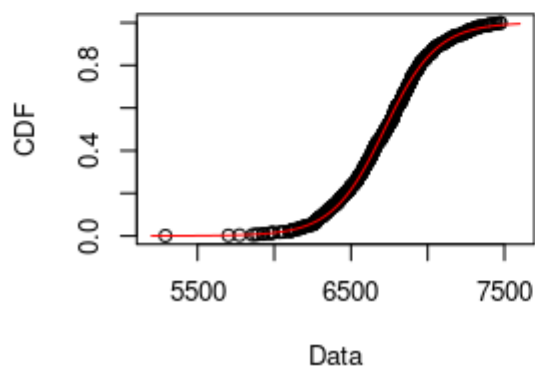
Empirical and theoretical dens.



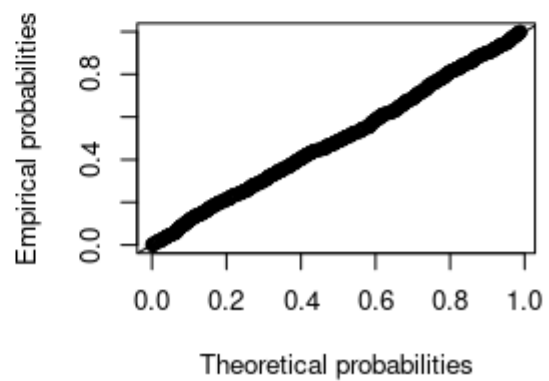
Q-Q plot



Empirical and theoretical CDFs



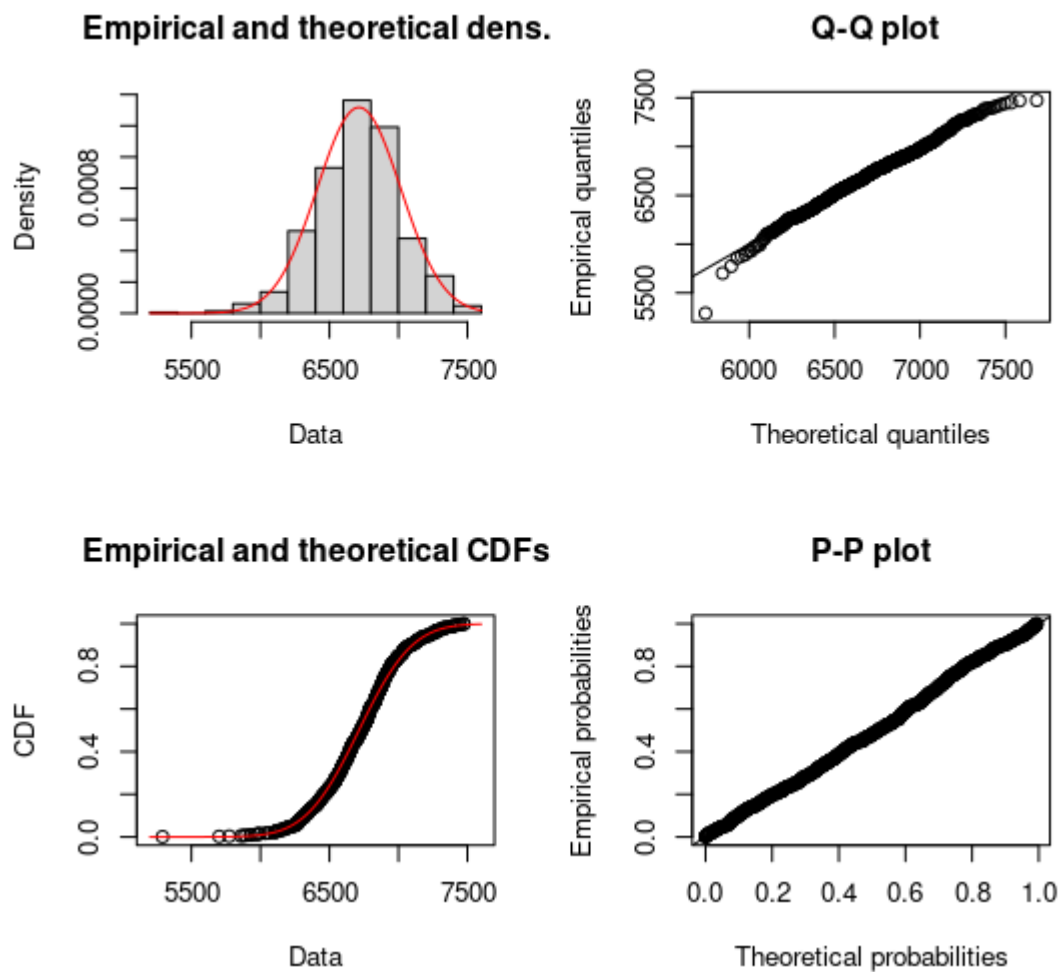
P-P plot



Para ver información similar sobre la distribución normal, usamos los siguientes comandos:

`plot(fit_norm)`

`summary(fit_norm)`

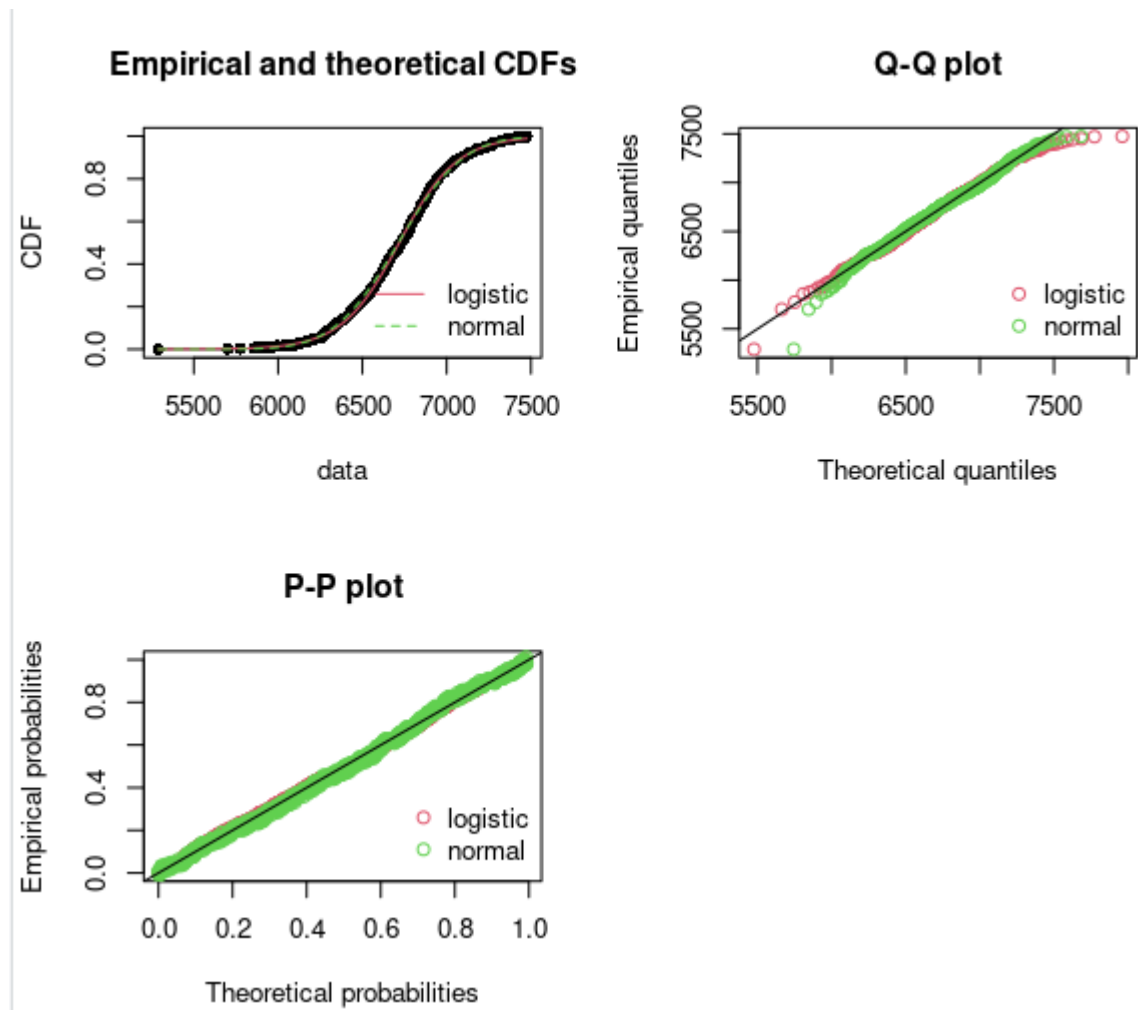


Estas gráficas son muy útiles, pero en este caso solo nos demuestran que es una decisión dura para tomar visualmente.

Procedemos a usar evaluación del ajuste, con una metodología llamada Goodness-of-fit statistic. Los comandos necesarios son los siguientes:

```
gofstat(list(fit_logis, fit_norm))
par(mfrow=c(2,2))
plot.legend<-c("logistic", "normal")
cdfcomp(list(fit_logis, fit_norm), legendtext = plot.legend)
qqcomp(list(fit_logis, fit_norm), legendtext = plot.legend)
ppcomp(list(fit_logis, fit_norm), legendtext = plot.legend)
```

retorna gráficos similares a los anteriores con ambos parámetros al tiempo



, y unas tablas que dan datos mucho más precisos.

```
Goodness-of-fit statistics
1-mle-logis 2-mle-norm
Kolmogorov-Smirnov statistic 0.02960387 0.02774823
Cramer-von Mises statistic 0.08400074 0.12603863
Anderson-Darling statistic 0.57360571 0.73007995

Goodness-of-fit criteria
1-mle-logis 2-mle-norm
Akaike's Information Criterion 10415.58 10416.82
Bayesian Information Criterion 10424.77 10426.01
```

3. Concluir cuál es el modelo que mejor se ajusta a los datos.

Si el valor en la prueba de kolmogorov-smirnov es inferior al estadístico KS (en este caso la $KS(\alpha, n) = KS(0.05, 730)$ es 0.037. Ambos datos que da la prueba Kolmogorov-Smirnov pasan este criterio, así que nos fijamos al valor para cada distribución de las "Akaike's Information Criterion" y "Bayesian Information Criterion" que puede verse al final del punto 2.

Por una unidad, nuestros datos se ajustan más al modelo logístico, osea por el valor que retorna R en la "Akaike's Information Criterion" y "Bayesian Information Criterion", concluyó que es ligeramente más cercano a la distribución normal que a la logística. **Osea, se ajusta mejor al modelo logístico.**