

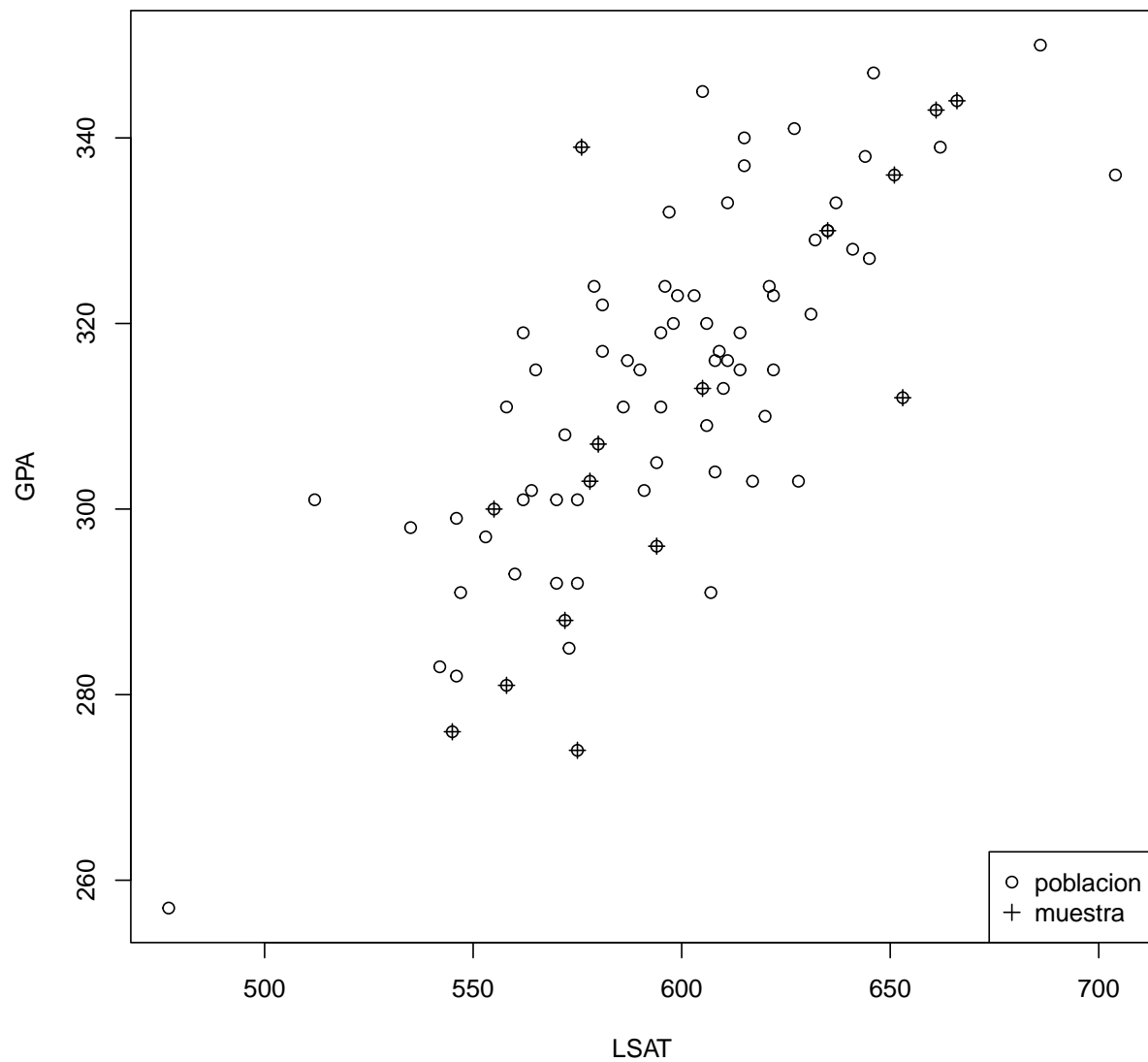
Tema 3. Conceptos relacionados con la Distribución Empírica

Ejemplo sobre estudios de máster

Se toma el ejemplo de las universidades con máster en leyes que está incluido en el libro de Efron y Tibshirani (1993).

```
library(bootstrap)

with(law82, plot(100 * GPA ~ LSAT, ylab = "GPA"))
with(law, points(100 * GPA ~ LSAT, pch = 3))
legend("bottomright", c("poblacion", "muestra"), pch = c(1, 3))
```



Calculo la correlación entre GPA (la puntuación media en los cursos de grado) y LSAT (calificación de admisión).

La correlación poblacional es

```
with(law82, cor(GPA, LSAT))
```

```
[1] 0.7599979
```

La correlación muestral (estimador *plug-in*) es

```
with(law, cor(GPA, LSAT))
```

```
[1] 0.7763745
```

Función de distribución empírica

```
# Simulo datos de calificaciones
```

```
mu = 6.5
```

```
sigma = 0.5
```

```
y = rnorm(n = 20, mean = mu, sd = sigma)
```

```
y = round(y, 3)
```

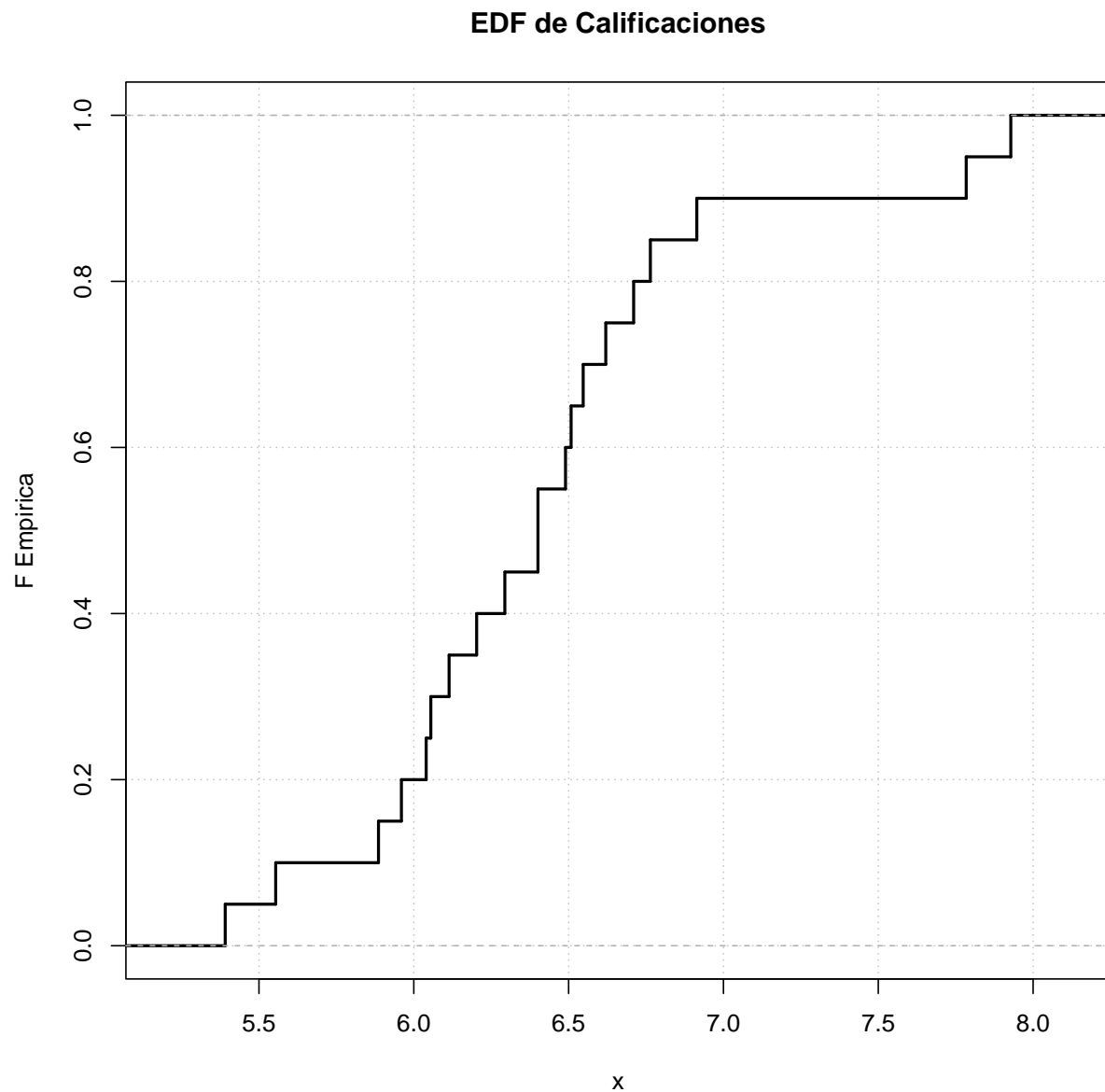
```
t = mean(y)
```

```
cat("La muestra ordenada es", sort(y), "\n y se obtiene una media muestral igual a ",  
    t, "\n")
```

```
La muestra ordenada es 5.391 5.554 5.886 5.96 6.04 6.055 6.114 6.203 6.294 6.401 6.401  
↪ 6.49 6.508 6.547 6.62 6.71 6.764 6.914 7.784 7.928  
y se obtiene una media muestral igual a 6.4282
```

La gráfica de la función de distribución empírica es

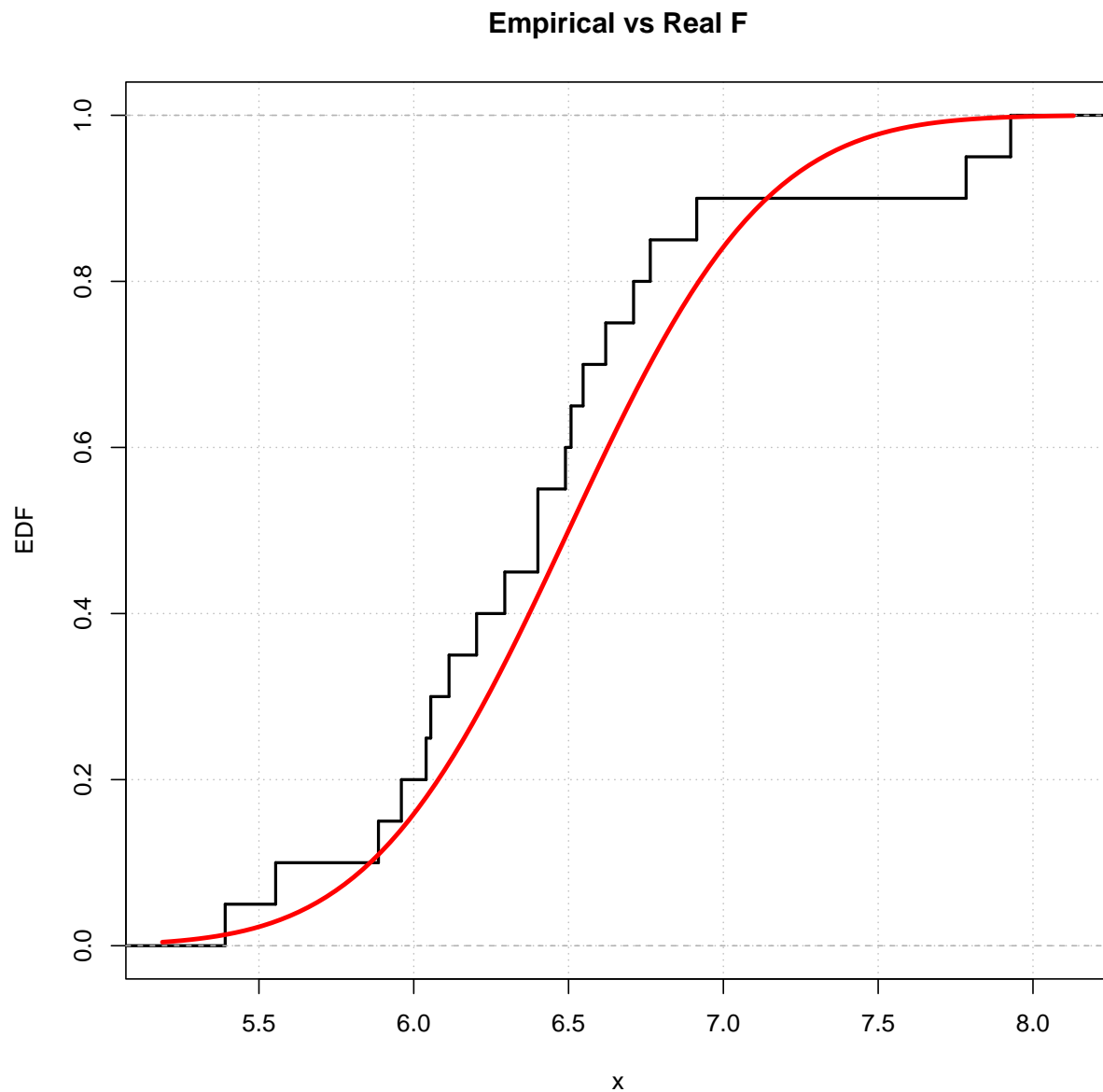
```
plot.ecdf(x = y, verticals = TRUE, do.p = FALSE, main = "EDF de Calificaciones",  
          lwd = 2, panel.first = grid(col = "gray", lty = "dotted"), ylab = "F Empirica")
```



Se puede dibujar la correspondiente función de distribución empírica junto con la curva de la función de distribución real.

```
plot.ecdf(x = y, verticals = TRUE, do.p = FALSE, main = "Empirical vs Real F",
  lwd = 2, xlab = "x", panel.first = grid(nx = NULL, ny = NULL, col = "gray",
    lty = "dotted"), ylab = "EDF")

curve(expr = pnorm(x, mean = mu, sd = sigma), col = "red", add = TRUE, lw = 3)
```



Simulaciones de la función de distribución empírica

Por ejemplo se toma una m.a.s de una distribución de Poisson

```
x = rpois(20, 3)
P = ecdf(x)
P(3)
```

```
[1] 0.65
```

```
acumula.dist = function(muestra, z) {  
  cuento = 0  
  for (t in muestra) {  
    if (t <= z)  
      cuento = cuento + 1  
  }  
  return(cuento/length(muestra))  
}  
  
acumula.dist(x, 3)
```

```
[1] 0.65
```

Para simular de la función de distribución empírica una vez observado vector x , se puede usar la función `sample`.

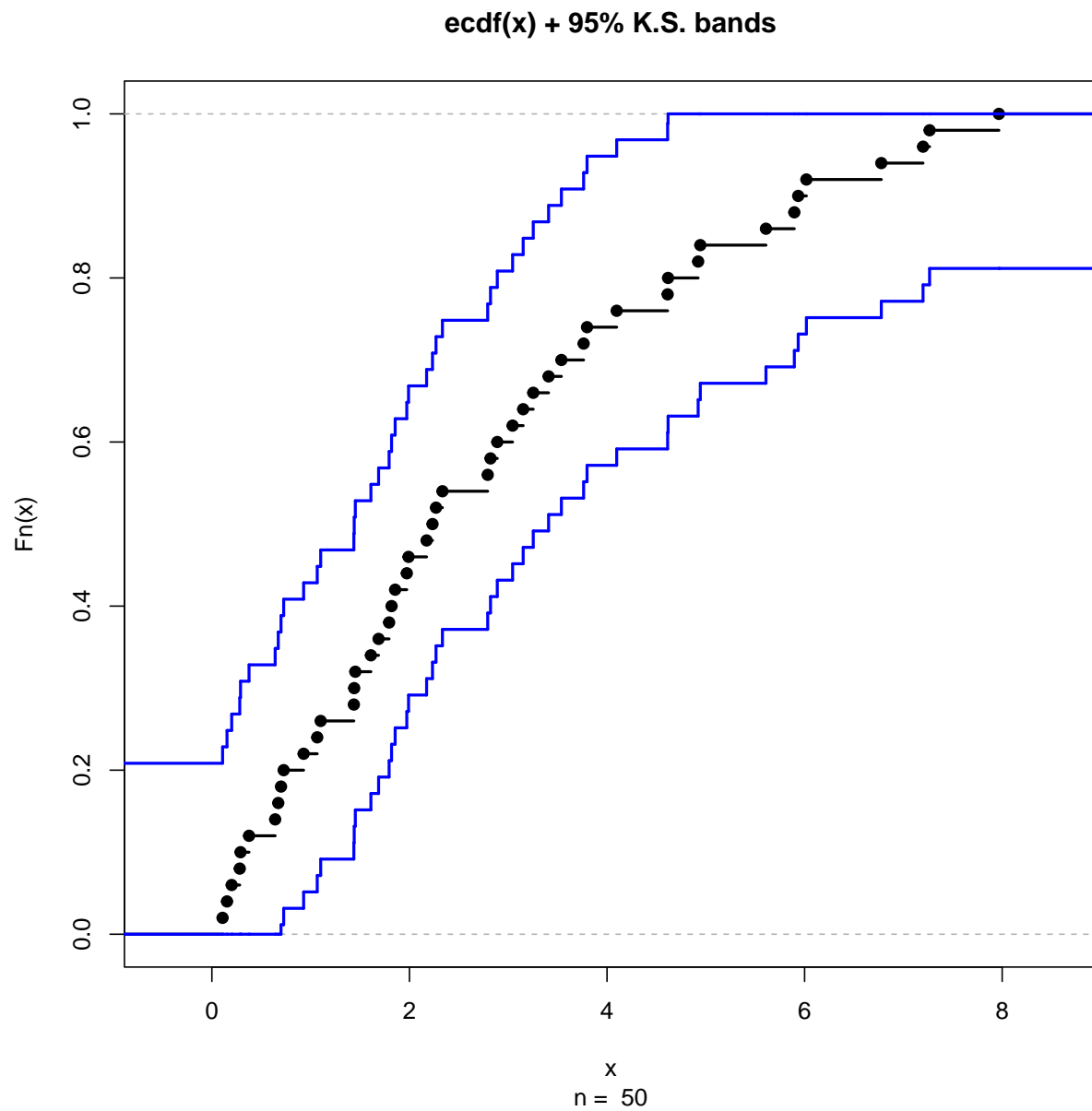
```
sample(x, size = 20, replace = TRUE)
```

```
[1] 2 3 2 4 2 2 1 1 2 2 5 4 2 1 4 2 2 6 6 5
```

Intervalos de confianza basados en la función de distribución empírica

Simulas datos de una v.a. chi cuadrado con 3 g.l.

```
library(sfsmisc)  
  
x = rchisq(50, 3)  
  
ecdf.ksCI(x, ci.col = "blue", lwd = 2)
```



Simulas observaciones de una distribución t de Student

```
datos = rt(20, 3)
```

```
dkw_cota = function(datos, x, alfa) {
  P = ecdf(datos)
  F_boina = P(x)
  epsilon = sqrt(log(2/alfa)/(2 * length(datos)))
}
```

```
    inf_cota = pmax(F_boina - epsilon, 0)
    sup_cota = pmin(F_boina + epsilon, 1)
    return(c(inf_cota, sup_cota))
}

dkw_cota(datos, -0.5, 0.05)
```

```
[1] 0.04631927 0.65368073
```

Calculas los intervalos

```
ecdf.ksCI(datos, ci.col = "pink", lwd = 2)
```


ecdf(datos) + 95% K.S. bands

