

Ciencia de Datos: Fundamentos Estadísticos

Resolucion Entrega 27/05

1.

$$\mathbb{E}(X) = 1/\lambda, \mathbb{V}(X) = 1/\lambda^2.$$

2.

Como $q_{0.9}$ es necesariamente positivo, se cumple que

$$1 - e^{-\lambda q_{0.9}} = 0.90.$$

Despejando resulta que $q_{0.90} = \frac{-\log(1-0.9)}{\lambda}.$

3.

$$\lambda = 1/10$$

4.

Se puede calcular usando el comando **qexp(prob,rate)**.

```
tita<-qexp(0.90,1/10)
tita
```

```
## [1] 23.02585
```

```
## comparo con el de la formula que halle
lambda<- 1/10
- log(1-0.9)/lambda
```

```
## [1] 23.02585
```

```
## dan igualitos!!
```

5.

No depende de los datos ya que esta cantidad corresponde a la distribución poblacional.

Para simplificar mi notación llamo de ahora en más $T1 = \hat{q}_{0.9} = \text{quantile}((X_1, \dots, X_n), 0.9)$ y

$$T2 = \tilde{q}_{0.9} = \frac{-\log(1-0.9)}{\hat{\lambda}}.$$

#6. T1

```
set.seed(123)
```

#7.

```
Nrep<- 1000
muchas_exponenciales <- rexp(n=Nrep,rate=1/10)
```

#8.

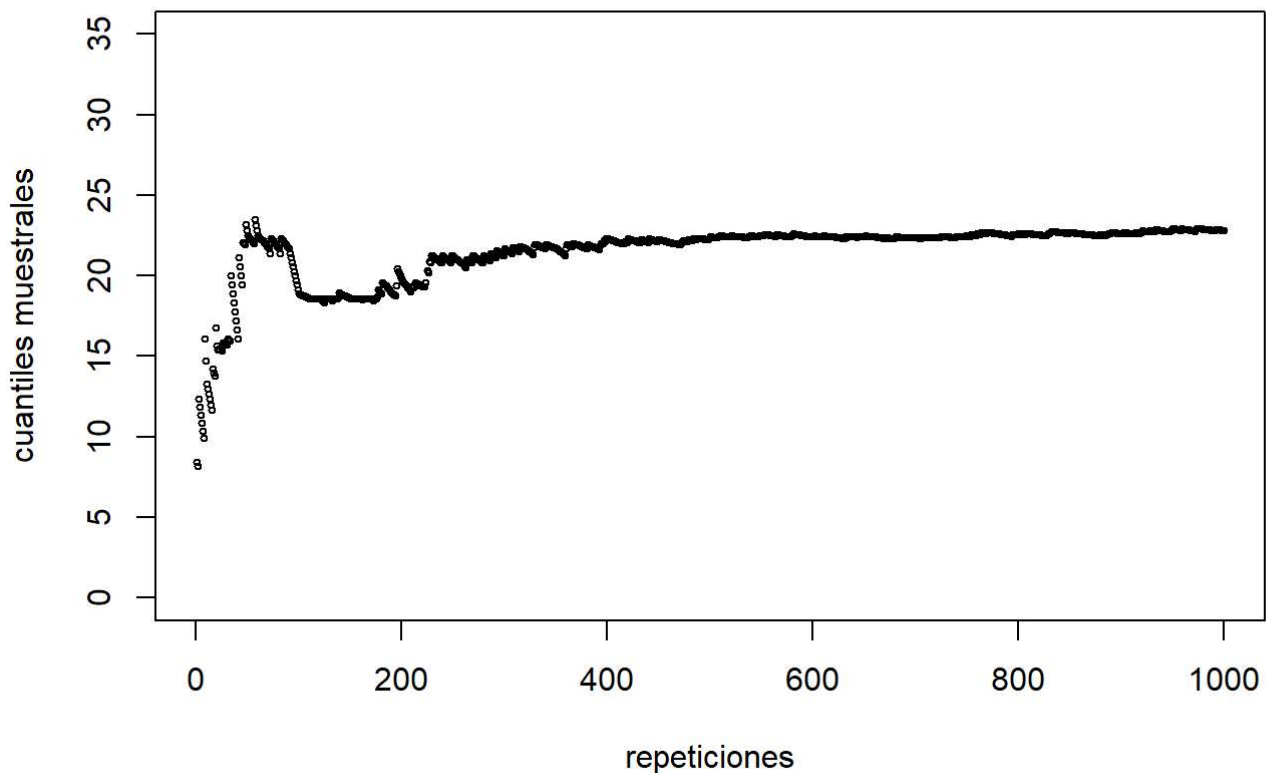
Calculo el cuantil de las primeras n para $n = 1, \dots, Nrep$.

```
muchos_cuantiles_exponenciales<- rep(NA, Nrep)
for(n in 1:Nrep)
{
  muchos_cuantiles_exponenciales[n] <- quantile(x=muchas_exponenciales[1:n], probs = 0.9)
}
```

#9. Grafico n (en el eje x) versus el cuantil empírico de los primeros n datos.

```
plot(1:Nrep, muchos_cuantiles_exponenciales,cex=0.5, xlab = "repeticiones", ylab="cuantiles m
uestrales", main="Simulación Cuantiles T1 - Dist Exponencial",ylim=c(0,35))
```

Simulación Cuantiles T1 - Dist Exponencial

#10 y 11 Repito el grafico $N_{gen} = 10$ veces superponiendo.

Espero que se acerque al valor teórico que calculé: 23.02585. Agrego una linea horizontal en ese valor para ver si se acerca.

```
# Para que me de lo mismo fijo la semilla otra vez.
# Hago el primero separado, para dar inicio al plot.

set.seed(123)
muchas_exponenciales <- rexp(n=Nrep,rate=1/10)

muchos_cuantiles_exponenciales<- rep(NA, Nrep)
for(n in 1:Nrep)
{
muchos_cuantiles_exponenciales[n] <- quantile(x=muchas_exponenciales[1:n], probs = 0.9)
}
plot(1:Nrep, muchos_cuantiles_exponenciales,cex=0.5, xlab = "repeticiones", ylab="cuantiles m
uestrales", main="Simulación Cuantiles T1 - Dist Exponencial",ylim=c(0,35))

# Los proximos los hago con un for, y los puntos se agregan al plot inicial con el comando
point

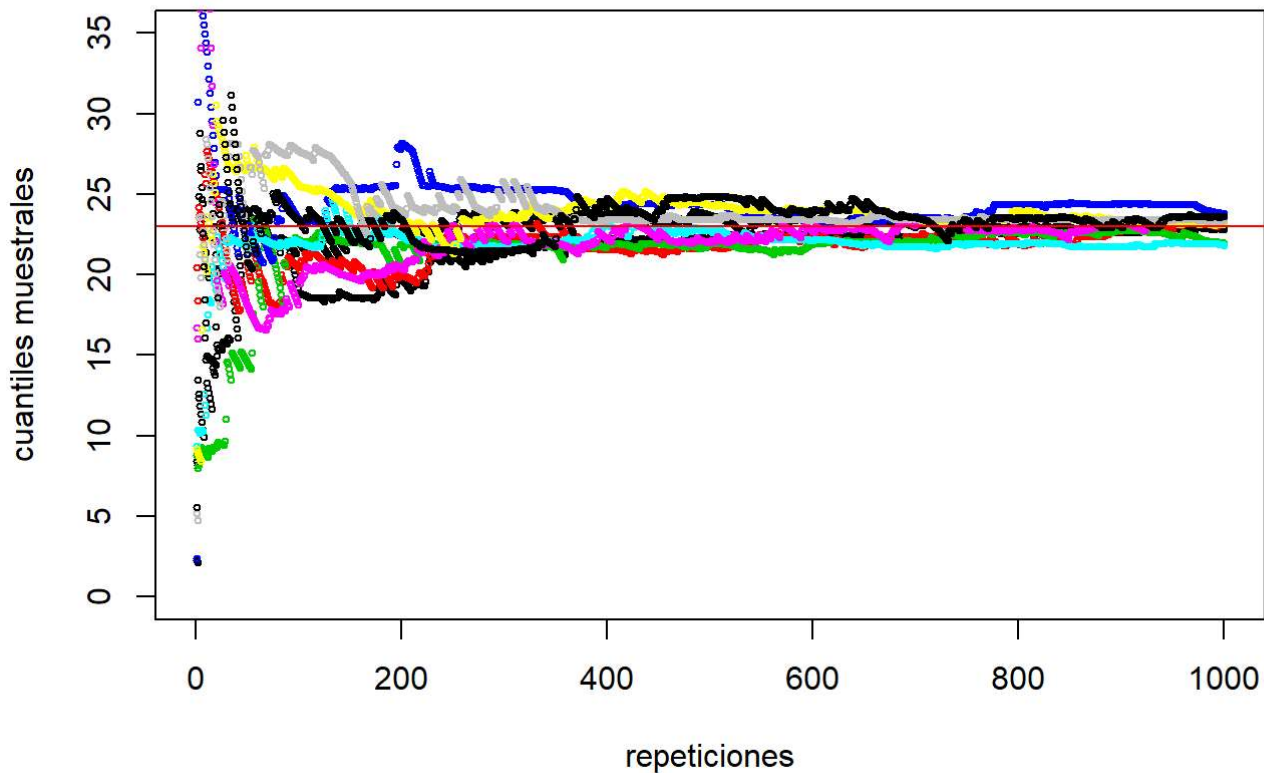
Ngen <- 10 # simula ser la cantidad de alumnos que repite la experiencia
for( i in 1:(Ngen-1))
{
muchas_exponenciales <- rexp(n=Nrep,rate=1/10)

muchos_cuantiles_exponenciales<- rep(NA, Nrep)
for(n in 1:Nrep)
{
muchos_cuantiles_exponenciales[n] <- quantile(x=muchas_exponenciales[1:n], probs = 0.9)
}

points(1:Nrep, muchos_cuantiles_exponenciales, cex=0.51,col=i)
}

abline(h=23.02585,col="red")
```

Simulación Cuantiles T1 - Dist Exponencial



#12.

Histogramas

```
set.seed(123)

Ngen<-100  # simula ser la cantidad de alumnos que repiten la experiencia

Nrep<- 100  # es la cantidad maxima de exponenciales: simula ser la cantidad de mediciones d
e cada alumne

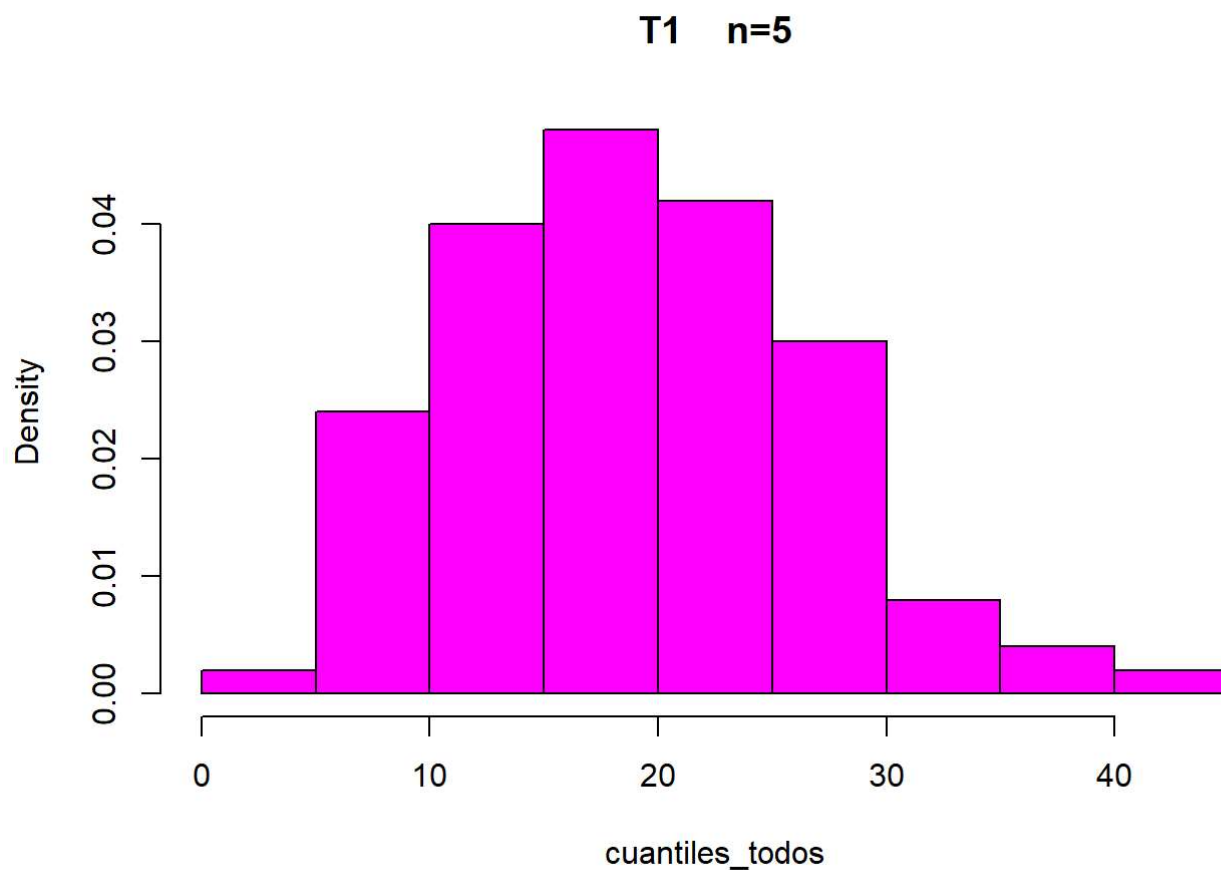
# Voy a generar todo junto ahora

datos_todos <- rexp(n=Ngen*Nrep,rate=1/10)
datos_por_alumno <- matrix(datos_todos,nrow =Ngen ,ncol=Nrep , byrow=TRUE ) #La fila represen
ta al alumno y las columnas las repeticiones de cada uno
```

Cada estudiante mide solo $n = 5$ mediciones y con esas 5 calculo el cuantil. Veamos como es el histograma al considerar los N_{gen} estudiantes.

```
n <- 5

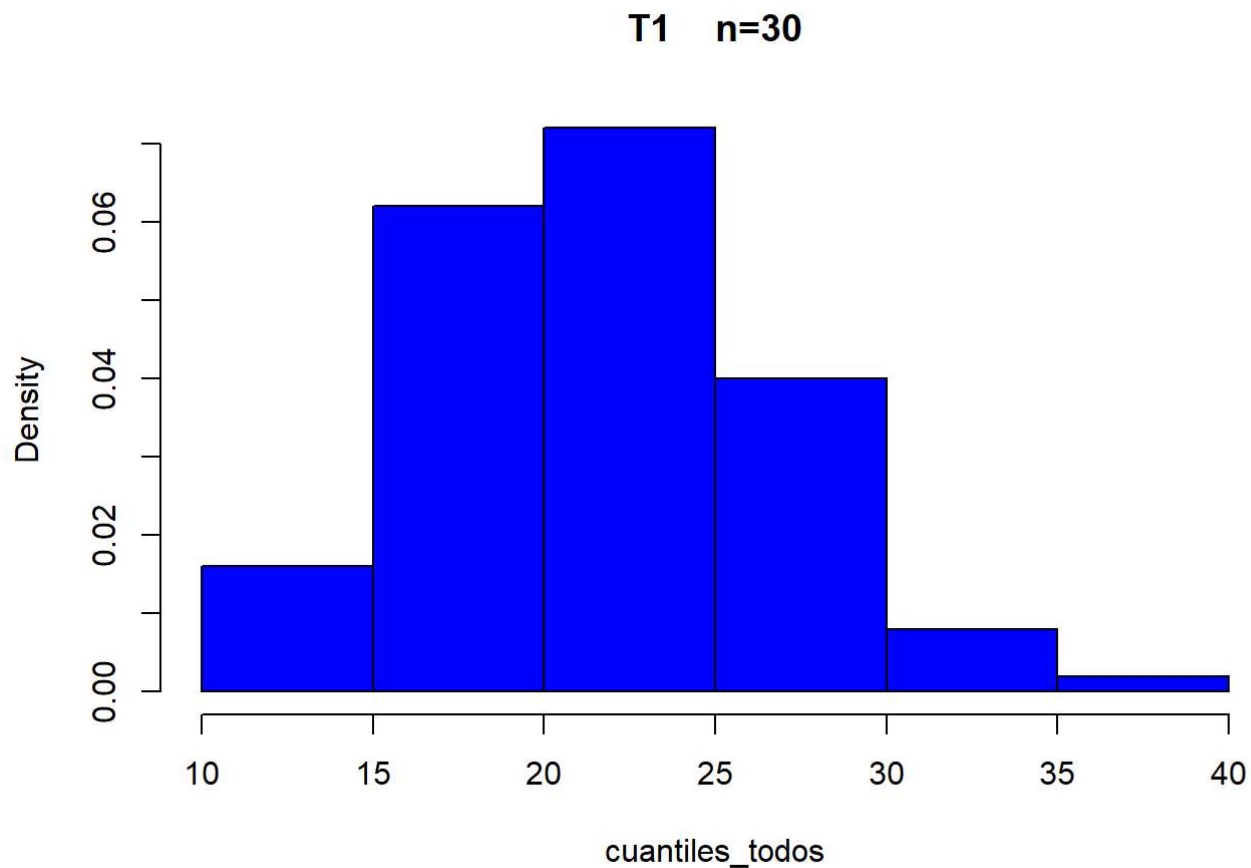
cuantiles_todos <- rep(NA, Ngen)
for(i in 1:Ngen)
{
  cuantiles_todos[i] <- quantile(x=datos_por_alumno[i,1:n], probs = 0.9)
}
hist(cuantiles_todos,probability = TRUE,main="T1    n=5",col="magenta")
```



Cada estudiante mide solo $n = 30$ mediciones y con esas 30 calculo el cuantil. Veamos como es el histograma al considerar los N_{gen} estudiantes.

```
n <- 30

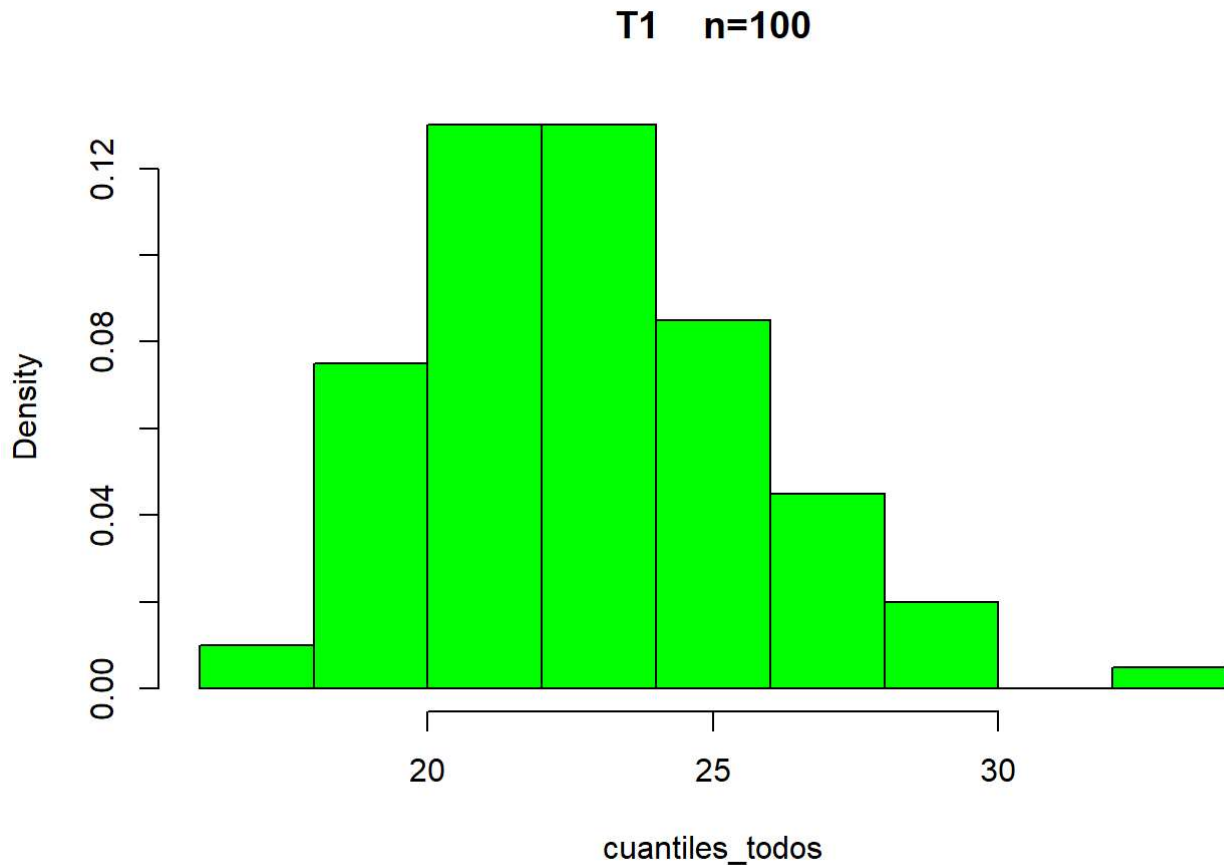
cuantiles_todos <- rep(NA, Ngen)
for(i in 1:Ngen)
{
  cuantiles_todos[i] <- quantile(x=datos_por_alumno[i,1:n], probs = 0.9)
}
hist(cuantiles_todos,probability = TRUE,main="T1    n=30",col="blue")
```



Cada estudiante mide solo $n = 100$ mediciones y con esas 100 calculo el cuantil. Veamos como es el histograma al considerar los N_{gen} estudiantes.

```
n <- 100

cuantiles_todos <- rep(NA, Ngen)
for(i in 1:Ngen)
{
  cuantiles_todos[i] <- quantile(x=datos_por_alumno[i,1:n], probs = 0.9)
}
hist(cuantiles_todos,probability = TRUE,main="T1   n=100",col="green")
```



T2

Notemos que a T2 lo podemos calcular con `qexp(0.9,lambda.hat)` donde `lambda.hat` lo calculo a partir de los datos

#13.

```
set.seed(123)
```

#14.

```
Nrep<- 1000
muchas_exponenciales <- rexp(n=Nrep,rate=1/10)
```

#15.

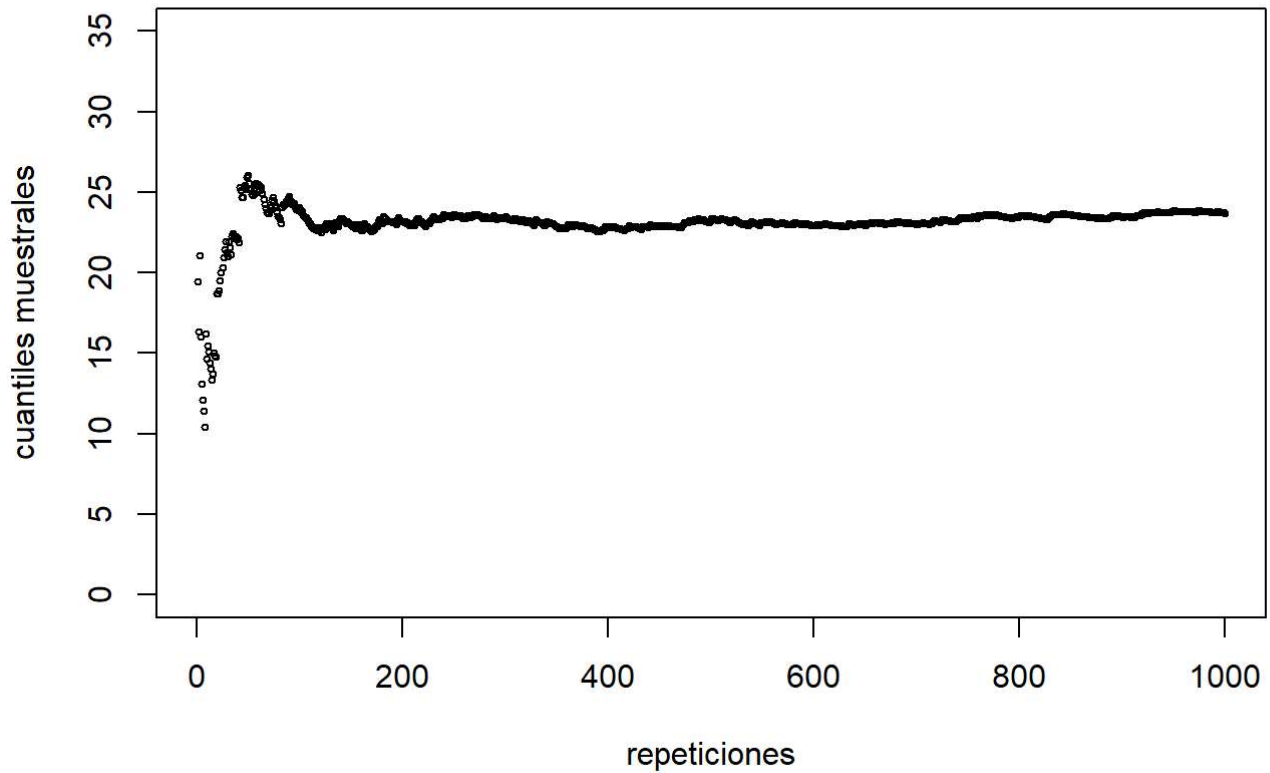
Calculo el cuantil de las primeras n para $n = 1, \dots, Nrep$.

```
muchos_cuantiles_rulo_exponenciales<- rep(NA, Nrep)
for(n in 1:Nrep)
{
  lambda.hat<- 1/mean(muchas_exponenciales[1:n])
  muchos_cuantiles_rulo_exponenciales[n] <- qexp(0.90,lambda.hat)
}
```

#16. Grafico n (en el eje x) versus el cuantil empírico de los primeros n datos.

```
plot(1:Nrep, muchos_cuantiles_rulo_exponenciales,cex=0.5, xlab = "repeticiones", ylab="cuantiles muestrales", main="Simulación Cuantiles T2 - Dist Exponencial",ylim=c(0,35))
```

Simulación Cuantiles T2 - Dist Exponencial



#17 y 18 Repito el grafico $N_{gen} = 10$ veces superponiendo.

Espero que se acerque al valor teórico que calculé: 23.02585. Agrego una linea horizontal en ese valor para ver si se acerca.


```
# Para que me de lo mismo fijo la semilla otra vez.
# Hago el primero separado, para dar inicio al plot.

set.seed(123)
muchas_exponenciales <- rexp(n=Nrep,rate=1/10)

muchos_cuantiles_rulo_exponenciales<- rep(NA, Nrep)
for(n in 1:Nrep)
{
  lambda.hat<- 1/mean(muchas_exponenciales[1:n])
  muchos_cuantiles_rulo_exponenciales[n] <- qexp(0.90,lambda.hat)
}
plot(1:Nrep, muchos_cuantiles_rulo_exponenciales,cex=0.5, xlab = "repeticiones", ylab="cuantiles muestrales", main="Simulación Cuantiles T2 - Dist Exponencial",ylim=c(0,35))

# Los proximos los hago con un for, y los puntos se agregan al plot inicial con el comando
point

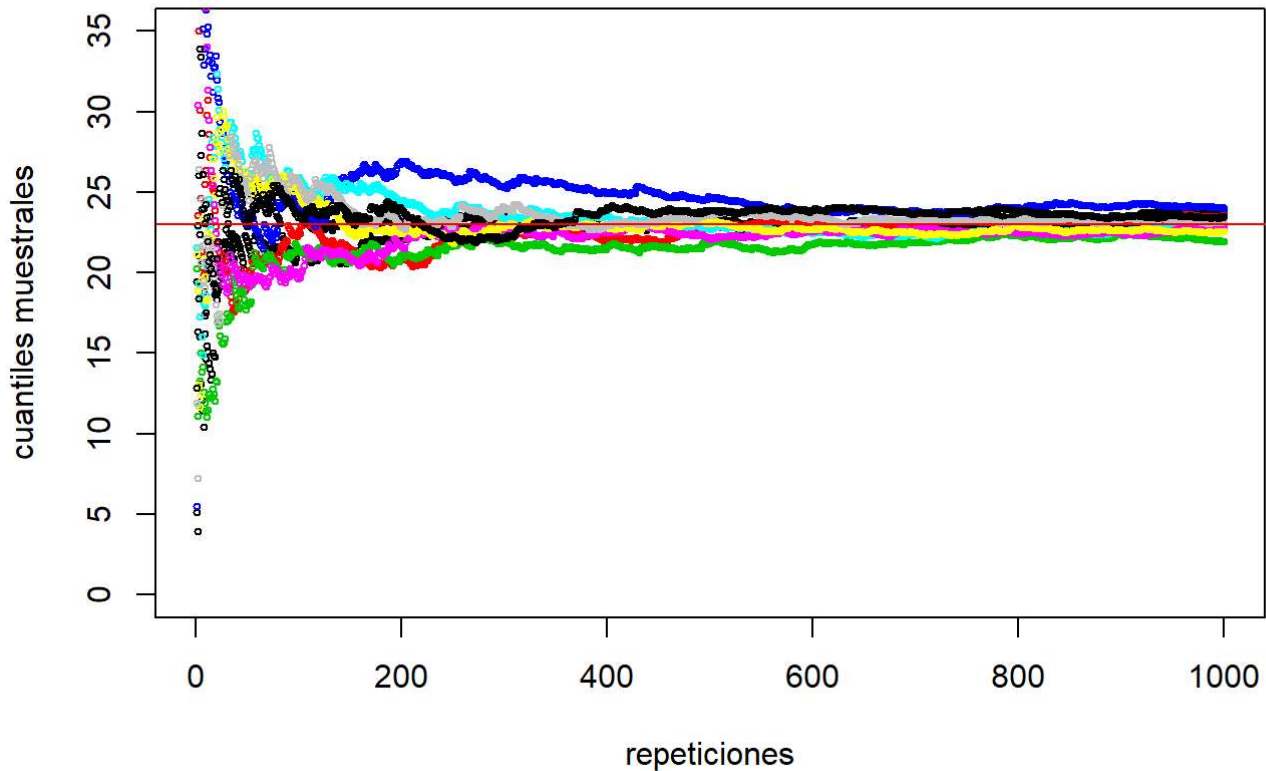
Ngen <- 10 # simula ser la cantidad de gente que repite la experiencia
for( i in 1:(Ngen-1))
{
  muchas_exponenciales <- rexp(n=Nrep,rate=1/10)

  muchos_cuantiles_rulo_exponenciales<- rep(NA, Nrep)
  for(n in 1:Nrep)
  {
    lambda.hat<- 1/mean(muchas_exponenciales[1:n])
    muchos_cuantiles_rulo_exponenciales[n] <- qexp(0.90,lambda.hat)
  }

  points(1:Nrep, muchos_cuantiles_rulo_exponenciales, cex=0.51,col=i)
}

abline(h=23.02585,col="red")
```

Simulación Cuantiles T2 - Dist Exponencial



#19.

Histogramas

```
set.seed(123)

Ngen<-100  # simula ser la cantidad de alumnos que repiten la experiencia

Nrep<- 100  # es la cantidad maxima de exponenciales: simula ser la cantidad de mediciones d
e cada alumne

# Voy a generar todo junto ahora

datos_todos <- rexp(n=Ngen*Nrep,rate=1/10)
datos_por_alumno <- matrix(datos_todos,nrow =Ngen ,ncol=Nrep , byrow=TRUE ) #La fila represen
ta al alumno y las columnas las repeticiones de cada uno
```

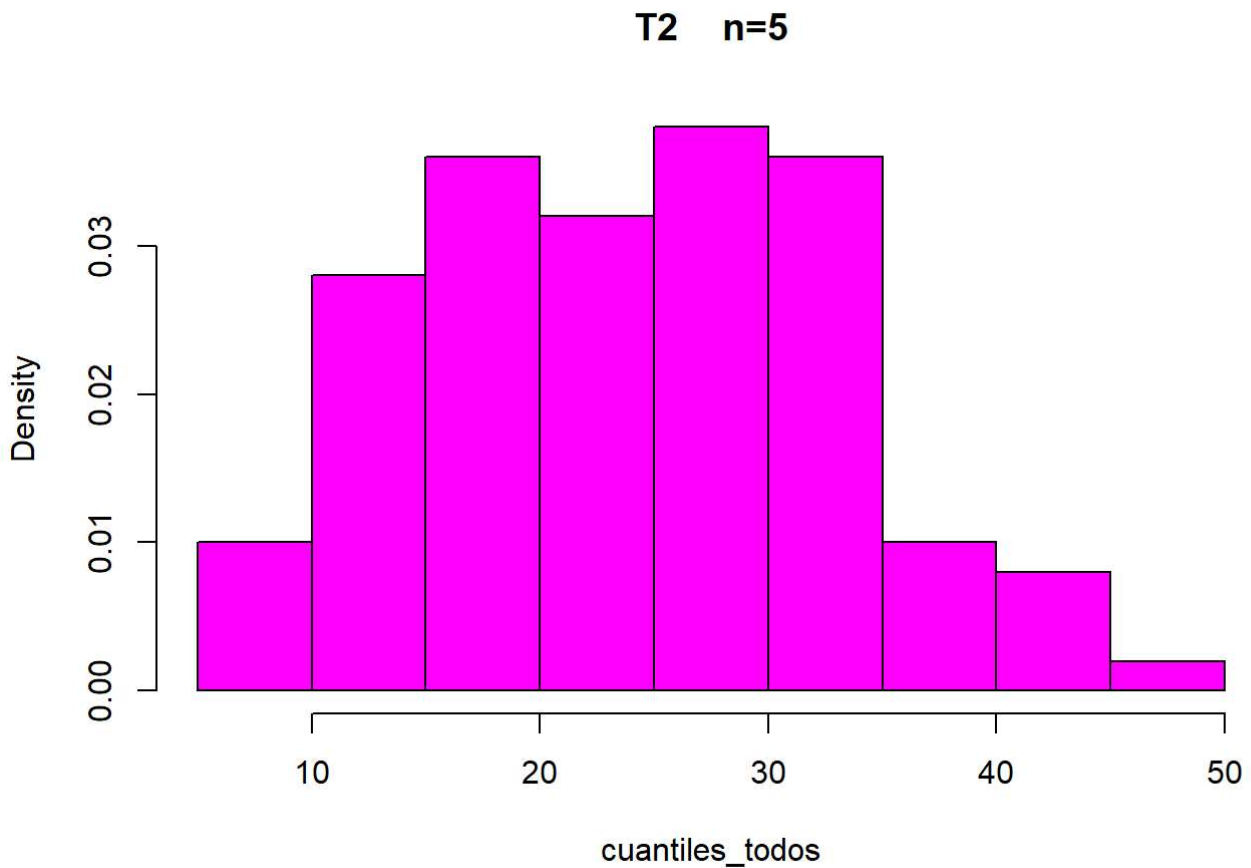
Cada estudiante mide solo $n = 5$ mediciones y con esas 5 calculo el cuantil. Veamos como es el histograma al considerar los N_{gen} estudiantes.

```

n <- 5

cuantiles_todos <- rep(NA, Ngen)
for(i in 1:Ngen)
{
  lambda.hat<- 1/mean(datos_por_alumno[i,1:n])
  cuantiles_todos[i] <- qexp(0.90,lambda.hat)
}
hist(cuantiles_todos,probability = TRUE,main="T2    n=5",col="magenta")

```



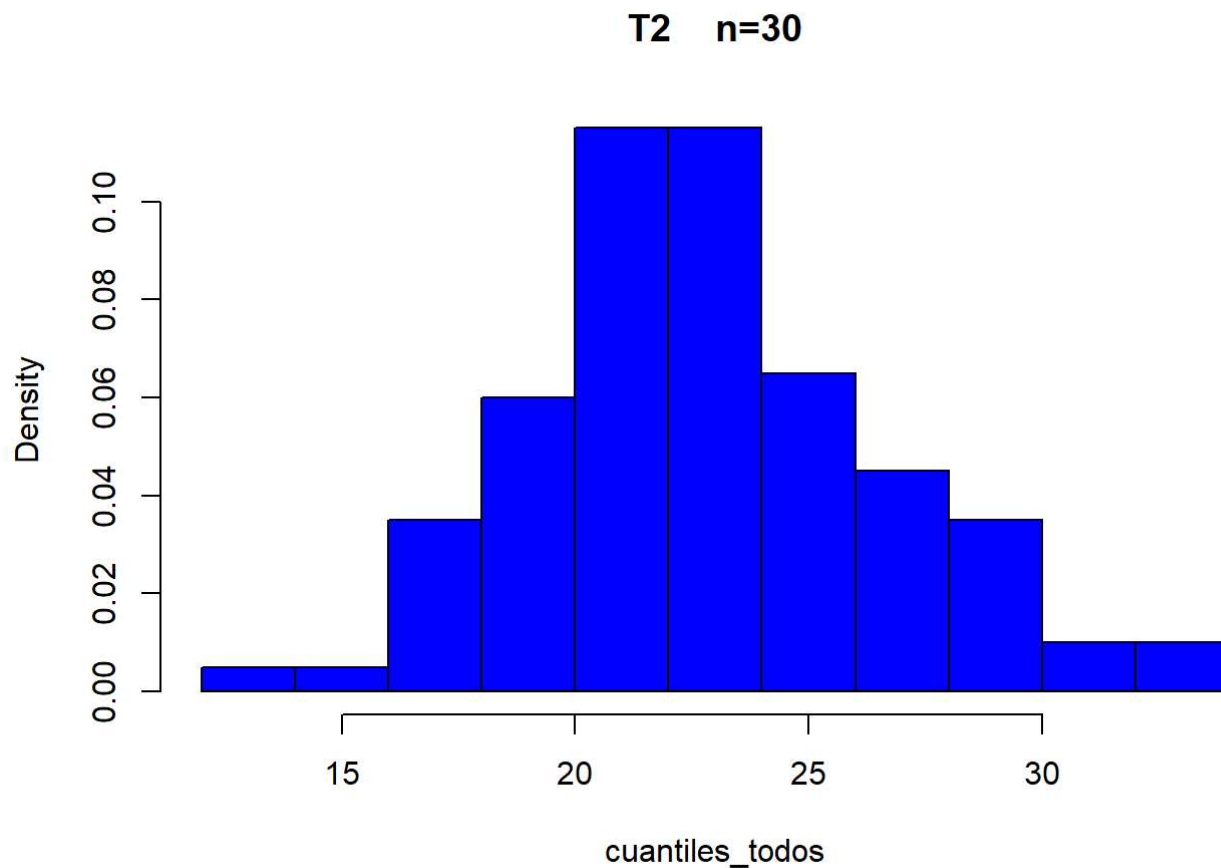
Cada estudiante mide solo $n = 30$ mediciones y con esas 30 calculo el cuantil. Veamos como es el histograma al considerar los N_{gen} estudiantes.

```

n <- 30

cuantiles_todos <- rep(NA, Ngen)
for(i in 1:Ngen)
{
  lambda.hat<- 1/mean(datos_por_alumno[i,1:n])
  cuantiles_todos[i] <- qexp(0.90,lambda.hat)
}
hist(cuantiles_todos,probability = TRUE,main="T2    n=30",col="blue")

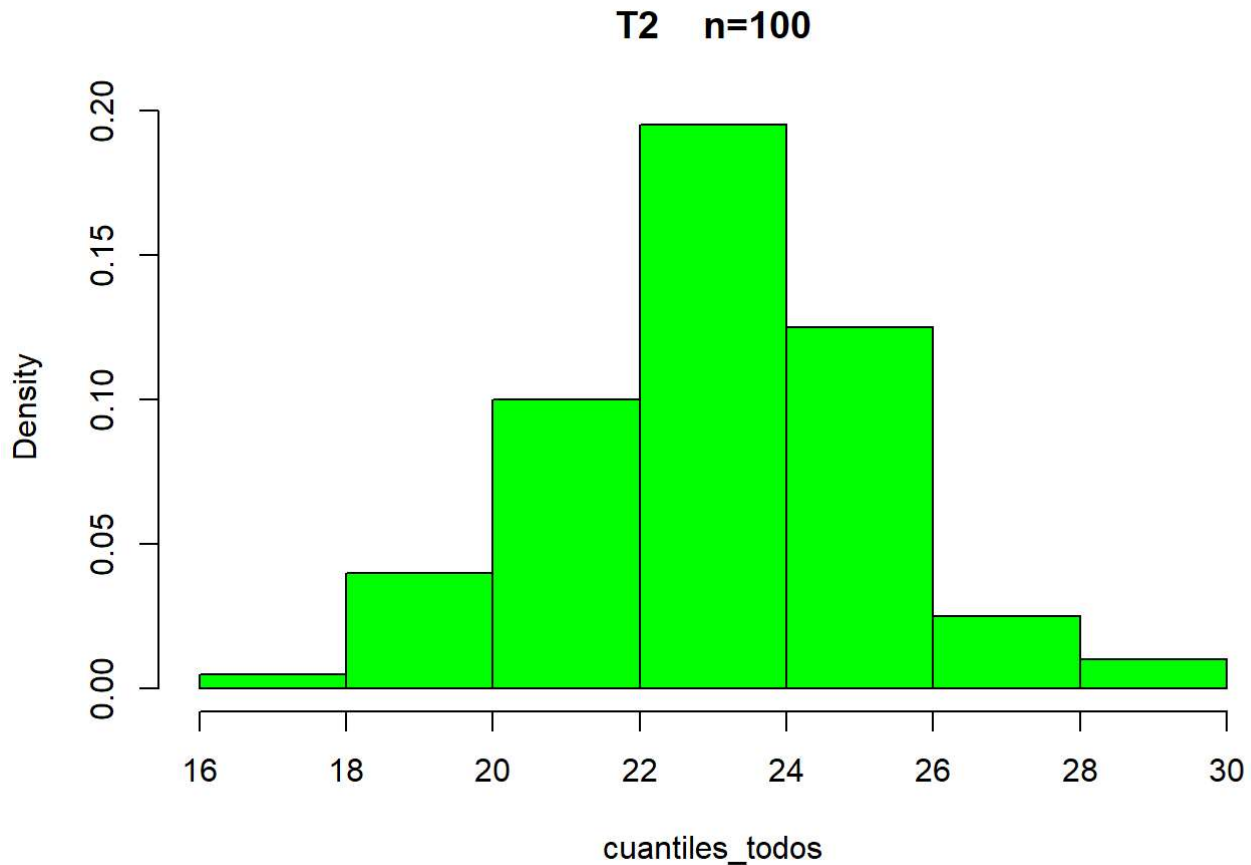
```



Cada estudiante mide solo $n = 100$ mediciones y con esas 100 calculo el cuantil. Veamos como es el histograma al considerar los N_{gen} estudiantes.

```
n <- 100

cuantiles_todos <- rep(NA, Ngen)
for(i in 1:Ngen)
{
  lambda.hat<- 1/mean(datos_por_alumno[i,1:n])
  cuantiles_todos[i] <- qexp(0.90,lambda.hat)
}
hist(cuantiles_todos,probability = TRUE,main="T2   n=100",col="green")
```



#20

Voy a hacer todo junto otra vez.

```

set.seed(123)

Nrep<-1000  # numero de replicaciones

n_max<- 500  # Los valores de n=50, 150,200 y 500. Genero todo y en cada caso uso lo que me
             # piden
enes<- c(50,150,200,500) # valores de n que me piden  considerar

# Voy a generar todo junto ahora
datos_todos <- rexp(n=Nrep*n_max,rate=1/10)
datos <- matrix(datos_todos,nrow =Nrep ,ncol=n_max , byrow=TRUE ) #Nrep es el numero de filas

T1<- T2 <- rep(NA,Nrep)

#Guardo ECME
#length(enes) es la cantidad de n que voy a considerar, aca es 4
#matriz de 2 x 4. En filas los estimadores, en columnas los distintos n

ECME<- matrix(rep(NA,2*length(enes)),nrow=2,ncol=length(enes))

```

Ahora calculo

```

for (ik in 1:length(enes)){

  for(i in 1:Nrep){

    T1[i]<- quantile(x=datos[i,1:enes[ik]], probs = 0.9)

    lambda.hat<- 1/mean(datos[i,1:enes[ik]])
    T2[i] <- qexp(0.90,lambda.hat)
  }
  ECME[1,ik]<- mean((T1-tita)^2)
  ECME[2,ik]<- mean((T2-tita)^2)
}

ECME<-round(ECME,4)

```

Imprimo la tabla

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 3.6.3
```

```
ECME.tab<- cbind(c("T1","T2"),ECME)
```

```
kable(ECME.tab,col.names = c("Estimador","n=5","n=150","n=200","n=500"),caption="ECME Nrep=1000")
```

ECME Nrep=1000

Estimador	n=5	n=150	n=200	n=500
T1	15.0264	5.3608	3.7336	1.6954
T2	9.476	3.2254	2.3575	0.9812