



Página www

Página de Abertura

Contenido



Página 1 de 28

Regresar

Full Screen

Cerrar

Abandonar

Estadística Bayesiana: Clase 3

Juan Carlos Correa

3 de marzo de 2021

En la clase anterior el diálogo de la elicitación de la proporción de médicos no abstemios que prefiere la cerveza al aguardiente produjo la siguiente tabla:

Distribución Subjetiva del Cliente

Percentil	Valor del Percentil
0.0625	0.36
0.125	0.42
0.25	0.50
0.50	0.60
0.75	0.68
0.875	0.75
0.9375	0.80

Vamos a graficar la densidad (histograma) en R :

[Página www](#)

[Página de Abertura](#)

[Contenido](#)



[Página 3 de 28](#)

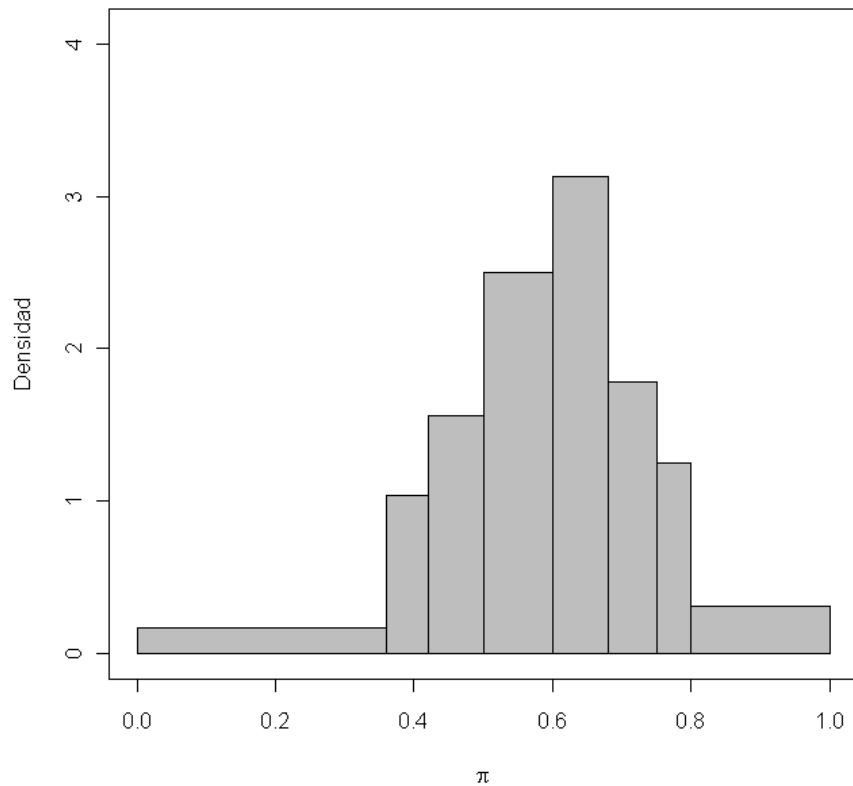
[Regresar](#)

[Full Screen](#)

[Cerrar](#)

[Abandonar](#)

Distribución Subjetiva sobre Médicos Alcohólicos



Los comandos son:

```
temp<-scan()  
0.0625  0.36  
0.125  0.42  
0.25   0.50  
0.50   0.60  
0.75   0.68  
0.875  0.75  
0.9375 0.80
```

```
temp<-matrix(temp,ncol=2,byrow=T)  
temp<-rbind(c(0,0),temp,c(1,1))  
Area<-temp[2:9,1]-temp[1:8,1]  
base<-temp[2:9,2]-temp[1:8,2]  
Altura<-Area/base
```

```
plot(1,1,ylim=c(0,max(Altura)*1.3),xlim=c(0,1),type='n',  
      ylab='',xlab='')  
title(ylab='Densidad',xlab=substitute(pi))  
title(main='Distribución Subjetiva sobre Médicos Alcohólicos')  
for(i in 1:length(Altura)){  
  rect(temp[i,2],0,temp[i+1,2],Altura[i],col='grey')  
}
```

Tarea # 1: Elicitación de la edad de una persona

En la próxima diapositiva aparece la foto de una persona (algunos de uds. la conocen). Por favor corran los comandos en R para imprimir la plantilla de elicitación que aparece abajo. Pídale a una persona que le colabore realizando la elicitación de la edad. Dele a su experto todas las indicaciones necesarias para que entienda lo que debe hacer (para cada año de edad debe seleccionar la altura sobre el eje Y que represente su creencia de acuerdo a las frases de incertidumbre que aparecen en la plantilla), también recuerde que debe proporcionarle todo el tiempo que necesite. Sobre la plantilla grafica los puntos y los une con un color que resalte.



Página [www](#)

Página de Abertura

Contenido



Página 6 de 28

Regresar

Full Screen

Cerrar

Abandonar



Página [www](#)

Página de Abertura

Contenido



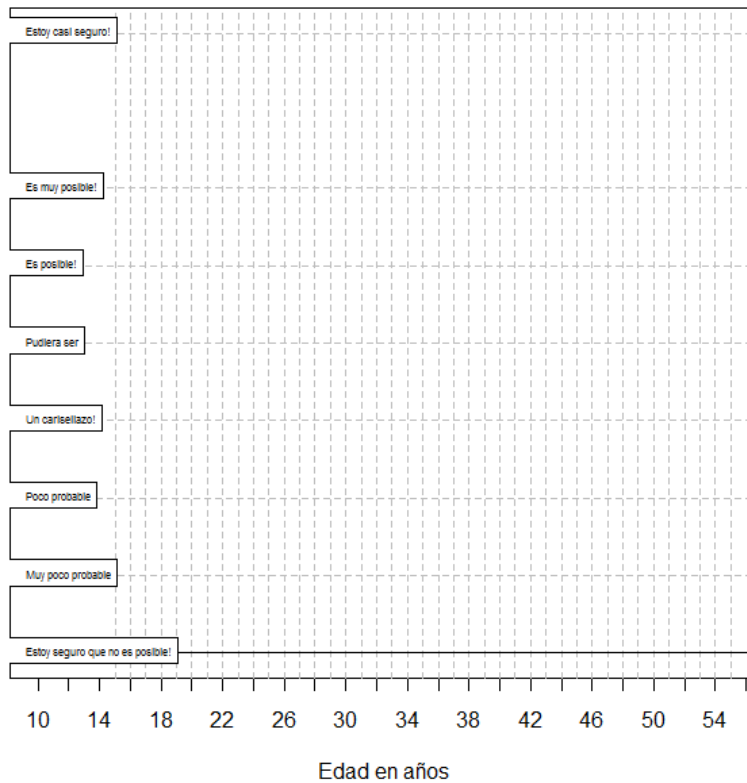
Página 7 de 28

Regresar

Full Screen

Cerrar

Abandonar



Código en *R* para elaborar la plantilla

```
par(lab=c(30,10,10))
par(yaxt='n')
plot(c(10,55),c(0,40),type='n',ylab='',xlab='Edad en años')
abline(v=c(15:55),lty=2,col='grey')
abline(h=0)
abline(h=c(5,10,15,20,25,30,40),lty=2,col='grey')
legend(8,1,'Estoy seguro que no es posible!',cex=0.5,bg='white')
legend(8,6,'Muy poco probable',cex=0.5,bg='white')
legend(8,11,'Poco probable',cex=0.5,bg='white')
legend(8,41,'Estoy casi seguro!',cex=0.5,bg='white')
legend(8,31,'Es muy posible!',cex=0.5,bg='white')
legend(8,26,'Es posible!',cex=0.5,bg='white')
legend(8,21,'Pudiera ser',cex=0.5,bg='white')
legend(8,16,'Un carisellazo!',cex=0.5,bg='white')
```


Elicitación vía muestras hipotéticas

Suponga que deseamos elicitación el número promedio de hijos que tienen las estudiantes de pregrado de primer semestre de la universidad. Para esto procedemos así:

- Le pedimos al experto que tiene una muestra grande ($n = 1000$) de estudiantes mujeres de primer semestre de la universidad.
- A continuación le pedimos que distribuya esta muestra en una tabla como la siguiente:

Nro. de Hijos	0	1	2	3	4	5	6
Nro. de estudiantes							

Suponga que el experto nos entrega los siguientes resultados:

Nro. de Hijos	0	1	2	3	4	5	6
Nro. de estudiantes	880	70	30	17	3	0	0

A continuación...

- Determine el nivel de conocimiento o seguridad que tiene el experto (esto se hace en forma de tamaño muestral equivalente):

Nivel de conocimiento	Tamaño muestral Equivalente
Muy bajo	De 1 a 2
Bajo	De 3 a 5
Algo de conocimiento	De 6 a 10
Bueno	De 10 a 15
Muy bueno	De 16 a 30

- Suponga que el experto califica su nivel de conocimiento con $n_{seguridad}$, para nuestro ejemplo asumamos que con el experto llegamos a la conclusión que $n_{seguridad} = 6$ representa bien su nivel de confianza.
- Ahora SIMULAMOS la distribución apriori del numero promedio de hijos:
 - Fijamos el número de muestras de tamaño $n_{seguridad} = 6$, digamos $NSim = 1000$.
 - A cada muestra simulada le calculamos el promedio y guardamos.
 - Luego cosntruímos el histograma (u otro estimador de densidad). Esta es nuestra distribución apriori (no paramétrica).

```
nro.hijos<-0:6  
frec.experto<-c(880, 70, 30, 17, 3, 0, 0)  
Nsim<-1000  
n.seguridad<-6
```

```
medias<-apply(matrix(  
  sample(nro.hijos,Nsim*n.seguridad,replace=T,prob=frec.experto),  
  ncol=n.seguridad),1,mean)
```

```
hist(medias,freq=F,  
  xlab='Número promedio de hijos',  
  ylab='Densidad',  
  main='Distribución Apriori del Promedio de Hijos')
```

```
summary(medias)  
sd(medias)  
quantile(medias,probs=c(0.05,1:9/10,0.95))  
IQR(medias)
```

[Página www](#)

[Página de Abertura](#)

[Contenido](#)



[Página 13 de 28](#)

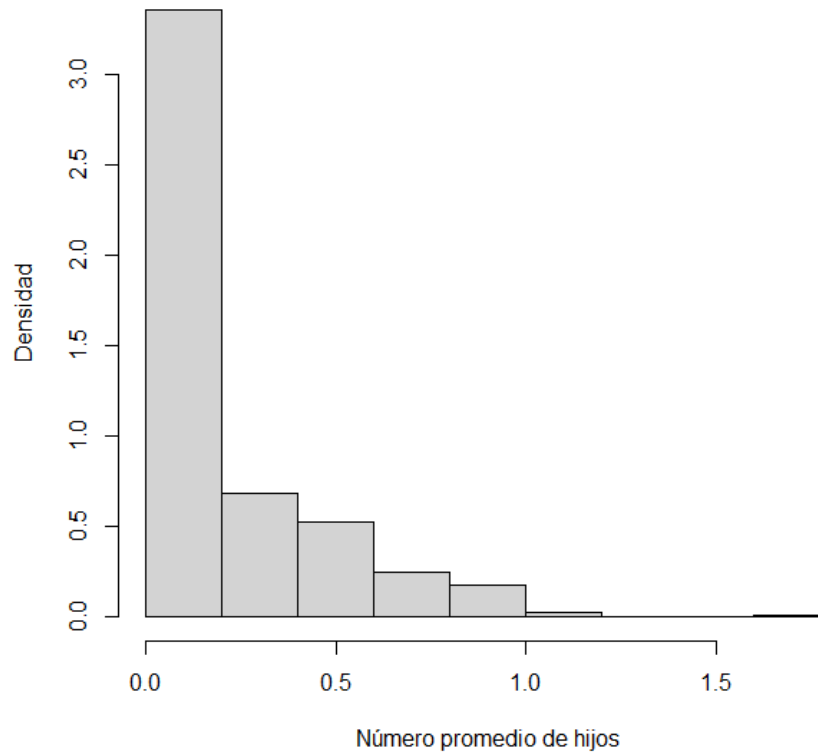
[Regresar](#)

[Full Screen](#)

[Cerrar](#)

[Abandonar](#)

Distribución Apriori del Promedio de Hijos



Sacando medidas descriptivas de resumen...

```
> summary(medias)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.0000	0.1667	0.2075	0.3333	1.6667

```
> sd(medias)
```

```
[1] 0.2534916
```

```
> quantile(medias, probs=c(0.05, 1:9/10, 0.95))
```

5%	10%	20%	30%	40%	50%	60%
0.0000000	0.0000000	0.0000000	0.0000000	0.0000000	0.1666667	0.1666667
80%	90%	95%				
0.3333333	0.5000000	0.6666667				

```
> IQR(medias)
```

```
[1] 0.3333333
```

Gráfico Histograma Decil

```
# Gráfico Histograma Decil
histo.decil<-function(x){
  if(length(x)<20)stop('Datos insuficientes!')
  deciles<-quantile(x,probs=1:9/10)
  xx<-c(min(x),deciles,max(x))
  base<-xx[-1]-xx[-length(xx)]
  area<-0.10
  altura<-area/base
```

```
plot(NA,NA,ylim=c(0,max(altura)),xlim=c(min(x),max(x)),ylab='Densidad')
title(main='Gráfico Histograma Decil')
for(i in 1:length(altura)){
  polygon(c(xx[i],xx[i],xx[i+1],xx[i+1],xx[i]),c(0,altura[i],altura[i],0,altura[i]))
}# Fin for i
}# Fin histo.decil
```

[Página www](#)

[Página de Abertura](#)

[Contenido](#)



[Página 16 de 28](#)

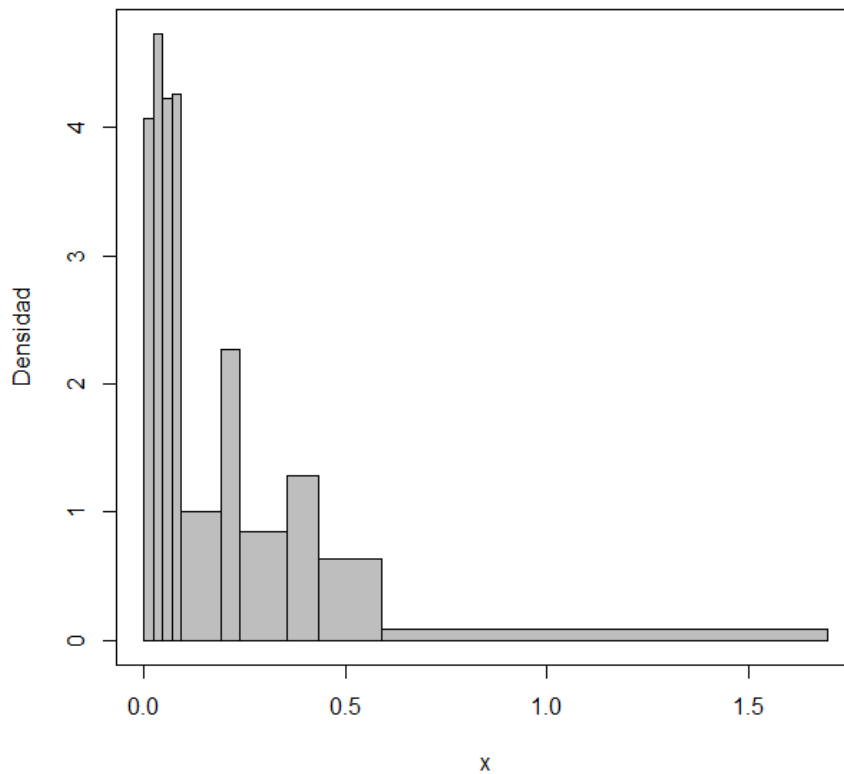
[Regresar](#)

[Full Screen](#)

[Cerrar](#)

[Abandonar](#)

Gráfico Histograma Decil



Ajustando un estimador de densidad tipo kernel

$$\hat{f}_{hn}(x) = \frac{1}{n} \sum_{i=1}^n K_{hn}(x - X_i)$$

es el estimador de Rosenblatt-Parzen de densidades para la densidad marginal de X y donde

$$K_{hn}(u) = \frac{K\left(\frac{u}{h_n}\right)}{h_n}$$

es el kernel con factor de escala h_n .

El kernel es una función real continua, acotada y simétrica K que integra a 1,

$$\int K(u) du = 1.$$

```
> plot(density(medias.con.ruido,from=0),  
main='Estimador de densidad', ylab='Densidad',xlab='Media')
```

[Página www](#)

[Página de Abertura](#)

[Contenido](#)



[Página 18 de 28](#)

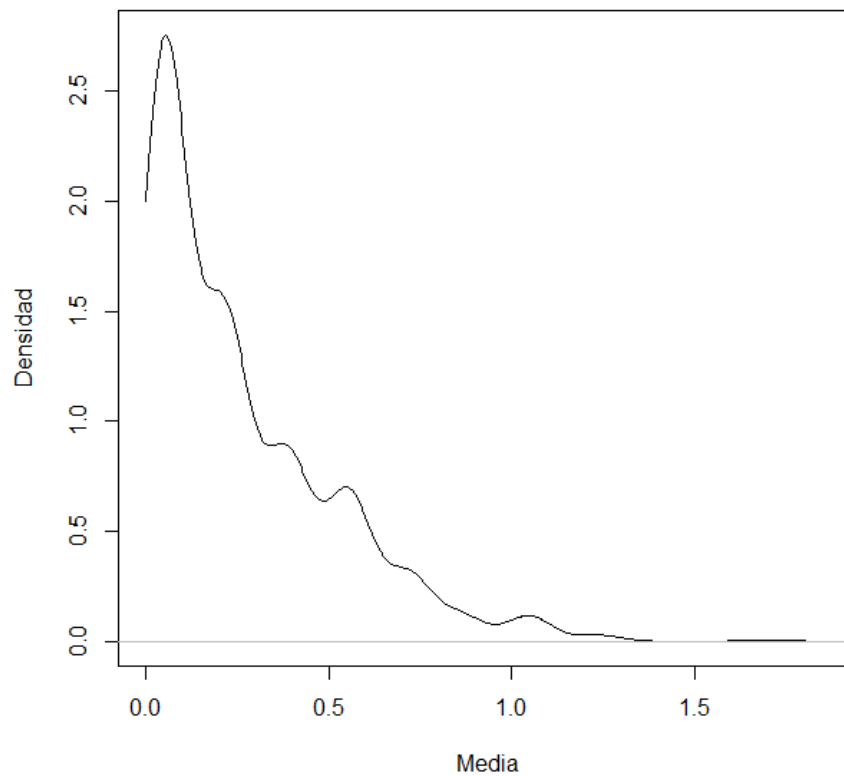
[Regresar](#)

[Full Screen](#)

[Cerrar](#)

[Abandonar](#)

Estimador de densidad



Página www

Página de Abertura

Contenido

◀

▶

◀

▶

Página 19 de 28

Regresar

Full Screen

Cerrar

Abandonar

Estimador de densidad kernel con ancho de ventana o banda=0.1

```
> plot(density(medias.con.ruido,bw=0.1,from=0),  
main='Estimador de densidad',  
ylab='Densidad',  
xlab='Media')  
>
```

[Página www](#)

[Página de Abertura](#)

[Contenido](#)



[Página 20 de 28](#)

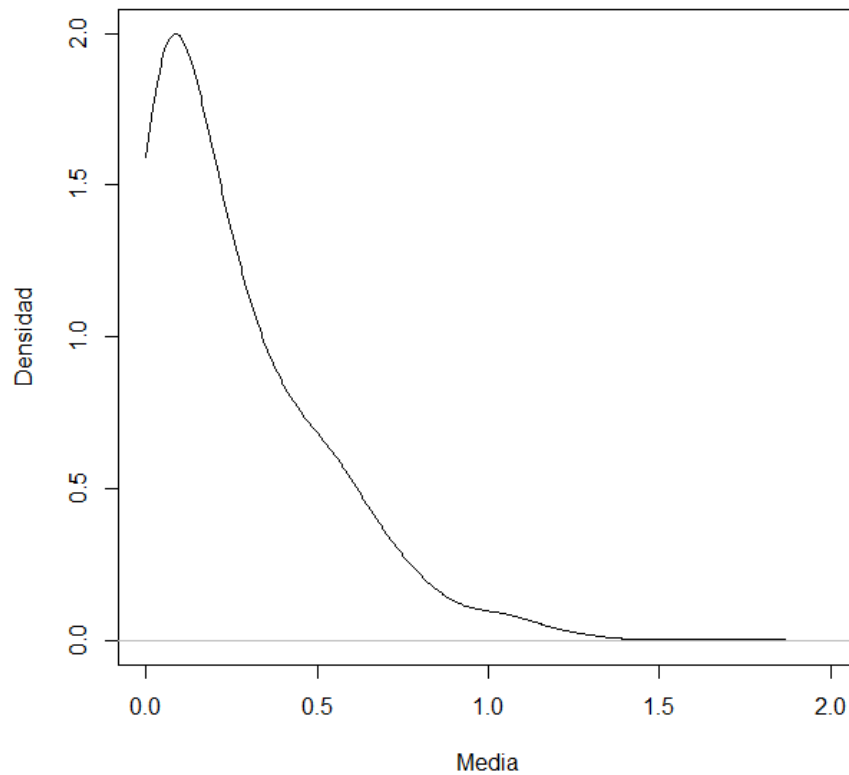
[Regresar](#)

[Full Screen](#)

[Cerrar](#)

[Abandonar](#)

Estimador de densidad



- La forma de las ponderaciones kernel está determinada por K , mientras el tamaño de las ponderaciones es parametrizada por h , que es llamada la ventana o la amplitud de banda.
- La selección de h es un proceso delicado.
- La normalización de las ponderaciones hace posible adaptarse a la intensidad local de las observaciones X , y en adición garantiza que las ponderaciones sumen uno.

Qué pasa si ajustamos una distribución paramétrica como la gamma?

```
library(MASS)
fitdistr(medias,'gamma')
Error in stats::optim(x = c(0, 0.166666666666667, 0.166666666666667)
  valor inicial en 'vmin' no es finito

medias.cor<-ifelse(medias==0.00000,0.000001,medias)

fitdistr(medias.cor,'gamma')
      shape      rate
0.150216507 0.723867053
(0.005063139) (0.063901267)
There were 23 warnings (use warnings() to see them)
```

Página www

Página de Abertura

Contenido



Página 23 de 28

Regresar

Full Screen

Cerrar

Abandonar

Sobrepongamos el gráfico de la gamma sobre el histograma

```
> secuencia<-seq(from=0.001,to=1.5,length=100)
> yy<-dgamma(secuencia,0.150216507,rate=0.723867053 )
> points(secuencia,yy,type='l',col='red')
```

[Página www](#)

[Página de Abertura](#)

[Contenido](#)



[Página 24 de 28](#)

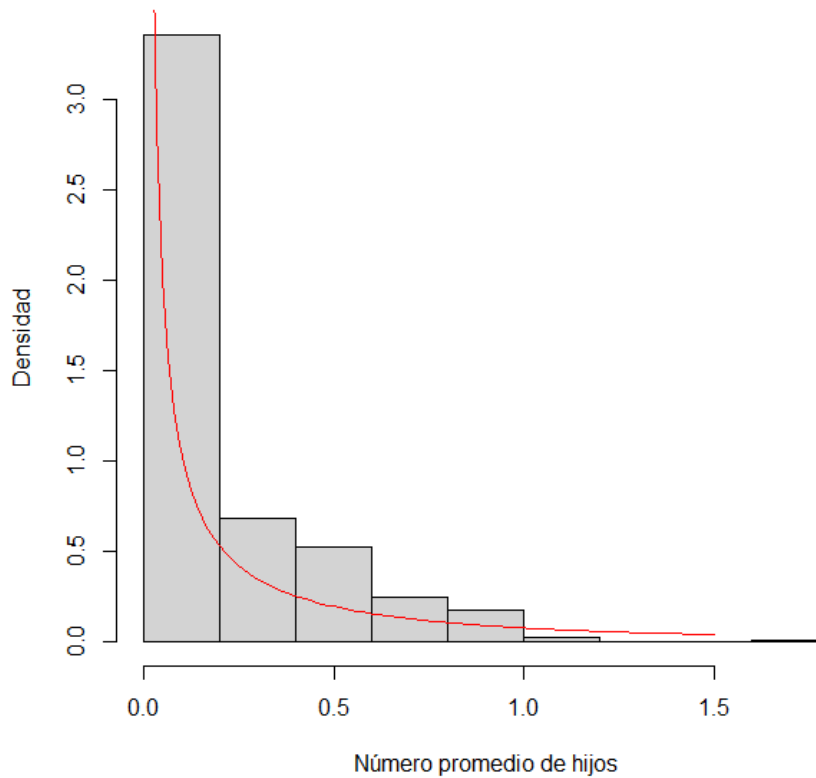
[Regresar](#)

[Full Screen](#)

[Cerrar](#)

[Abandonar](#)

Distribución Apriori del Promedio de Hijos



Qué pasa si ajustamos una distribución paramétrica como la lognormal?

```
library(MASS)
```

```
fitdistr(medias.cor,'lognormal')
```

```
      meanlog      sdlog  
-6.6882522      6.2820160  
( 0.1986548) ( 0.1404701)
```

```
hist(medias,freq=F,  
xlab='Número promedio de hijos',  
ylab='Densidad',  
main='Distribución Apriori del Promedio de Hijos')
```

```
secuencia<-seq(from=0.001,to=1.5,length=100)  
yy2<-dlnorm(secuencia,meanlog = -6.6882522, sdlog = 6.2820160 )  
points(secuencia,yy2,type='l',col='blue')
```

[Página www](#)

[Página de Abertura](#)

[Contenido](#)



[Página 26 de 28](#)

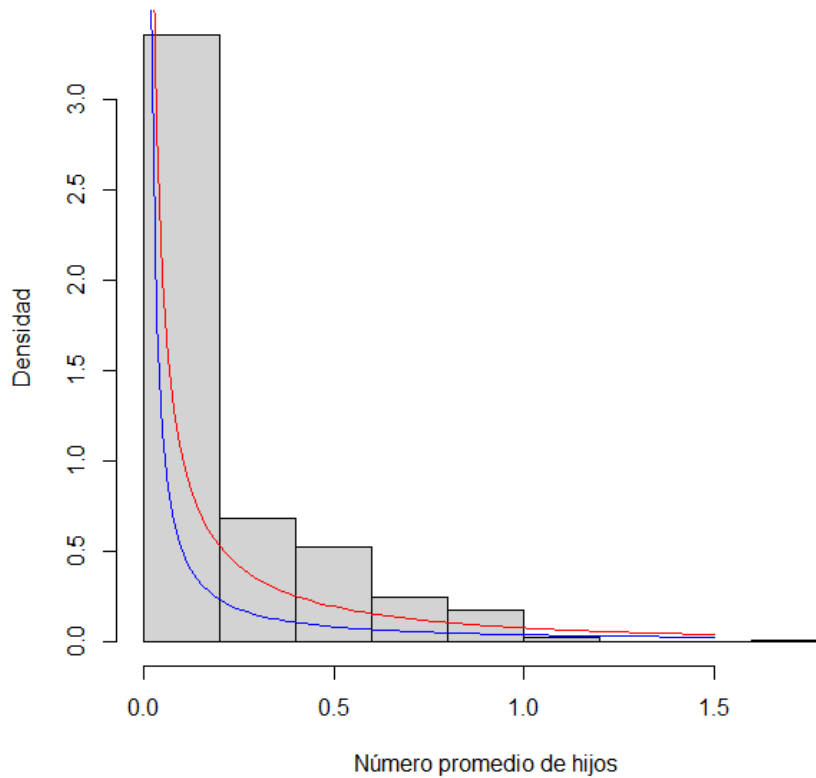
[Regresar](#)

[Full Screen](#)

[Cerrar](#)

[Abandonar](#)

Distribución Apriori del Promedio de Hijos



Reflexiones sobre la elicitación

Kadane y Wolfson (1998) recogiendo la opinión de varios autores concluyen:

- la opinión de los expertos es la más valiosa para elicitar;
- a los expertos se les debe pedir que expresen su opinión sobre cantidades obseables, condicionadas solo en covariables (que también son observables) o en otras cantidades observables;
- a los expertos no se les debe pedir estimaciones de los momentos de una distribución (excepto posiblemente el primer momento); a ellos se les debe pedir que determinen cuantiles o probabilidades de la distribución predictiva;
- hay que proporcionarles retroalimentación frecuentemente durante el proceso de elicitación;
- a los expertos se les debe pedir que den asignaciones tanto incondicionalmente como condicionalmente basados en datos observados hipotéticos.

Garthwaite, Kadane, y O'Hagan (2004) señalan varios aspectos que deben tenerse en cuenta es un proceso de elicitación.

- El objetivo es elicitación una distribución que represente el conocimiento presente del experto, y es útil tener un resumen sobre cuál es la justificación d este conocimiento.
- Cualquier interés personal o financiero que el experto pueda tener tener en las inferencias o decisiones que dependerán (aún marginalmente) en la distribución del experto, debe ser declarado.

Página www

Página de Abertura

Contenido

◀ ▶

◀ ▶

Página 29 de 28

Regresar

Full Screen

Cerrar

Abandonar

- Debe proporcionarse un entrenamiento para familiarizar al experto con las interpretaciones probabilísticas y tanto con sus conceptos y propiedades que se requieran en la elicitación. Es útil correr un ejercicio de elicitación ficticia para proporcionar práctica en el protocolo que el facilitador se propone utilizar.
- Un registro debe llevarse de la elicitación. Este debe idealmente contener todas las preguntas que fueron realizadas por el facilitador junto con las respuestas del experto, también como el proceso con el que se ajustó la distribución de probabilidad con estas respuestas.

Pasos en una elicitación

En el proceso de elicitación tenemos pasos muy claros:

1. seleccionar y motivar el experto
2. entrenar el experto
3. estructurar las preguntas
4. elicitar y documentar los juicios del experto
5. verificar los resultados.