Parcial II

Alejandro Salazar Mejía

23/8/2021

```
library('MCMCpack')
## Warning: package 'MCMCpack' was built under R version 4.0.5
## Loading required package: coda
## Warning: package 'coda' was built under R version 4.0.5
## Loading required package: MASS
## ##
## ## Markov Chain Monte Carlo Package (MCMCpack)
## ## Copyright (C) 2003-2021 Andrew D. Martin, Kevin M. Quinn, and Jong Hee Park
## ##
## ## Support provided by the U.S. National Science Foundation
## ## (Grants SES-0350646 and SES-0350613)
## ##
library(hdrcde)
## Warning: package 'hdrcde' was built under R version 4.0.5
## This is hdrcde 3.4
library(KernSmooth)
## Warning: package 'KernSmooth' was built under R version 4.0.5
## KernSmooth 2.23 loaded
## Copyright M. P. Wand 1997-2009
library(knitr)
library(readr)
library(plotrix)
```

Warning: package 'plotrix' was built under R version 4.0.3

library(HH)

```
## Warning: package 'HH' was built under R version 4.0.4
## Loading required package: lattice
## Loading required package: grid
## Loading required package: latticeExtra
## Warning: package 'latticeExtra' was built under R version 4.0.2
## Loading required package: multcomp
## Warning: package 'multcomp' was built under R version 4.0.4
## Loading required package: mvtnorm
## Warning: package 'mvtnorm' was built under R version 4.0.3
## Loading required package: survival
## Loading required package: TH.data
## Warning: package 'TH.data' was built under R version 4.0.4
##
## Attaching package: 'TH.data'
## The following object is masked from 'package:MASS':
##
##
      geyser
## Loading required package: gridExtra
## Warning: package 'gridExtra' was built under R version 4.0.2
##
## Attaching package: 'HH'
## The following object is masked from 'package:coda':
##
##
       acfplot
source('funcionesAux.R')
```

Como se explicó en la sección 3.1, la métodología de elicitación consiste básicamente en elegir al menos dos puntos de diseño, y elicitar la distribución normal de la respuesta en dichos puntos. En esta sección se muestra dicho proceso.

Se escogieron tres puntos de diseño: los modelos 2013, 2015 y 2017. Una vez obtenidos los intervalos en cada año y la frencuencia de la muestra hipotéticas en estos, se hace uso de la función estimaNormal, programada por el profesor Juan Carlos Correa Morales y presentada en el curso de estadística bayesiana en las diapositivas de la clase 10. Esta función nos permite encontrar una distribución normal que mejor ajuste sus probabilidades a las frecuencias relativas correspondientes a cada subintervalo por medio de un proceso de optimización.

A continuación se muestra de manera gráfica el resultado de la elicitación del precio del auto en cada añomodelo, junto con la distribución normal ajustada y su tabla de parámetros estimados.

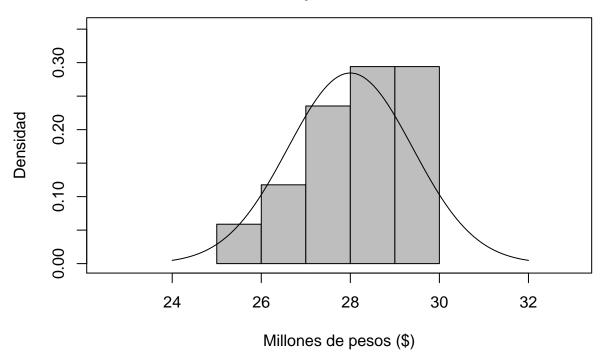
2013:

```
limites 2013 \leftarrow c(25, 26, 27, 28, 29, 30)
frecus2013 <- c(20, 40, 80, 100, 100)
(resu2013 <- optim(c(26,5),estimaNormal,method='L-BFGS-B',</pre>
             lower=c(25,0.01),upper=c(28,50),
             limites=limites2013 ,frecu=frecus2013))
## $par
## [1] 28.000000 1.965467
## $value
## [1] 0.0213493
##
## $counts
## function gradient
##
         16
##
## $convergence
## [1] 0
##
## $message
## [1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"
param.opt2013<-resu2013$par
histograma (frecus 2013, limites 2013,
           main = "Distribución del precio \nen Millones de pesos del modelo 2013",
           xlab = "Millones de pesos ($)",
           ylab = "Densidad")
```

dnorm(xx,mean=param.opt2013[1],sd=sqrt(param.opt2013[2])),type='1')

points(xx<-seq(min(limites2013)-1,max(limites2013)+2,length=100),</pre>

Distribución del precio en Millones de pesos del modelo 2013



```
kable(t(param.opt2013),
    col.names = c("$\\mu_{2013}$", "$\\sigma^2_{2013}$"),
    caption = "Parámetros de la distribución en el año-modelo 2013")
```

Table 1: Parámetros de la distribución en el año-modelo 2013

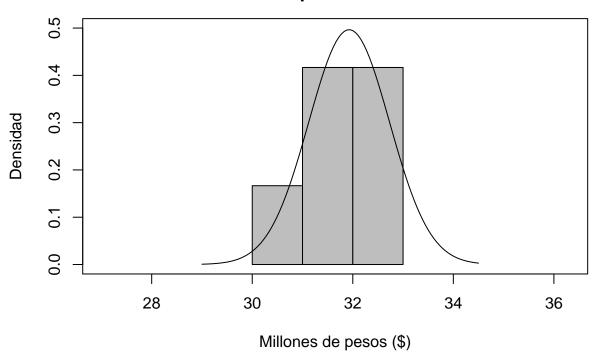
μ_{2013}	σ^2_{2013}
28	1.965467

2015:

\$value

```
## [1] 0.004495004
##
## $counts
## function gradient
##
         29
##
## $convergence
## [1] 0
##
## $message
## [1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"
param.opt2015<-resu2015$par
histograma (frecus 2015, limites 2015,
           main = "Distribución del precio \nen Millones de pesos del modelo 2015",
           xlab = "Millones de pesos ($)",
           ylab = "Densidad")
points(xx<-seq(min(limites2015)-1,max(limites2015)+1.5,length=100),</pre>
       dnorm(xx,mean=param.opt2015[1],sd=sqrt(param.opt2015[2])),type='1')
```

Distribución del precio en Millones de pesos del modelo 2015



```
kable(t(param.opt2015),
    col.names = c("$\\mu_{2015}$", "$\\sigma^2_{2015}$"),
    caption = "Parametros de la distribución en el año-modelo 2015")
```

Table 2: Parámetros de la distribución en el año-modelo 2015

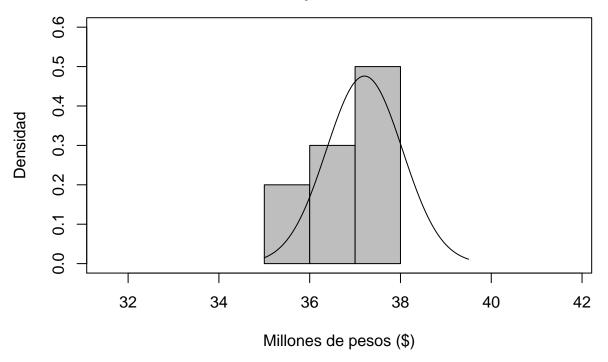
μ_{2015}	σ_{2015}^2
31.92742	0.6453334

2017:

```
limites 2017 < c(35, 36, 37, 38)
frecus2017 \leftarrow c(20, 30, 50)
(resu2017 <- optim(c(36,5),estimaNormal,method='L-BFGS-B',</pre>
             lower=c(35,0.01), upper=c(38,50),
             limites=limites2017 ,frecu=frecus2017))
## $par
## [1] 37.2062508 0.7023848
##
## $value
## [1] 0.02302974
## $counts
## function gradient
##
         22
                   22
##
## $convergence
## [1] 0
##
## $message
## [1] "CONVERGENCE: REL_REDUCTION_OF_F <= FACTR*EPSMCH"
param.opt2017<-resu2017$par
histograma(frecus2017,limites2017,
           main = "Distribución del precio \nen Millones de pesos del modelo 2017",
           xlab = "Millones de pesos ($)",
           ylab = "Densidad")
points(xx<-seq(min(limites2017), max(limites2017)+1.5, length=100),</pre>
```

dnorm(xx,mean=param.opt2017[1],sd=sqrt(param.opt2017[2])),type='1')

Distribución del precio en Millones de pesos del modelo 2017



```
kable(t(param.opt2017),
    col.names = c("$\\mu_{2017}$", "$\\sigma^2_{2017}$"),
    caption = "Parametros de la distribución en el año-modelo 2017")
```

Table 3: Parámetros de la distribución en el año-modelo 2017

μ_{2017}	σ^2_{2017}
37.20625	0.7023848

Una vez obtenidas las tres normales, se simularon 20 datos de cada una (este número busca reflejar el conocimiento que tiene el experto sobre el tema el cual es alto). A dicha muestra se le ajustó un modelo de regresión lineal simple y se guardaron los valores de los tres parámetros de interés. Este proceso se repitió 1000 veces, lo que resulta en una muestra simulada de tamaño 1000 de β_0 , β_1 y σ^2 .

Apriori:

```
betas0 <- c()
betas1 <- c()
sigmas2 <- c()

x <- rep(c(13,15,17), each= 20)</pre>
```

```
for (i in 1:1000) {
    y2013 <- rnorm(20, mean = param.opt2013[1], sd = param.opt2013[2])
    y2015 <- rnorm(20, mean = param.opt2015[1], sd = param.opt2015[2])
    y2017 <- rnorm(20, mean = param.opt2017[1], sd = param.opt2017[2])
    y <- c(y2013, y2015, y2017)
    modelo <- lm(y-x)

betas0 <- c(betas0, modelo$coefficients[1])
    betas1 <- c(betas1, modelo$coefficients[2])
    sigmas2 <- c(sigmas2, (summary(modelo)$sigma)**2)
}

taus <- 1/sigmas2</pre>
```

```
# Estimación de parás. de normal bivariada de BetaO y Beta1
betas <- data.frame(betasO, betas1)
b0 <- colMeans(betas)
B0 <- solve(var(betas))

# Estimación de parás. de Gamma inversa por MM
m <- mean(sigmas2)
v <- var(sigmas2)
alpha <- m^2/v+2
beta <- m*(m^2/v+1)</pre>
```

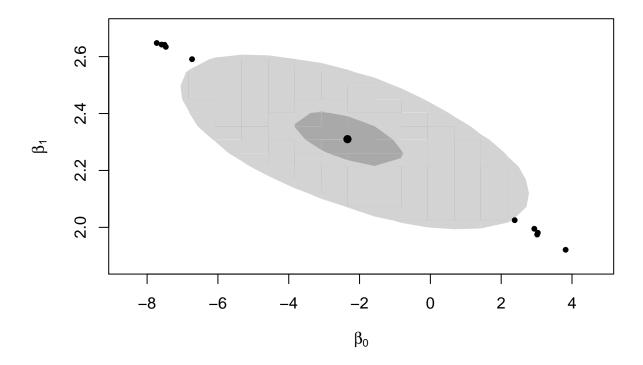
Se busca contruir una distribución apriori con la siguiente p.d.f:

$$\xi(\beta_0, \beta_1, \sigma^2) = \mathcal{N}(\beta_0, \beta_1) \times \mathcal{G}.\mathcal{I}(\sigma^{\epsilon})$$

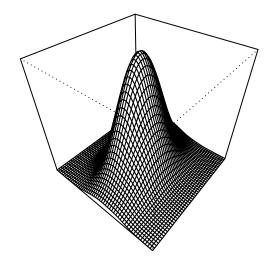
Para esto, vamos a aproximar la distribución apriori del experto al modelo teórico anterior. Se estima entonces el vector de medias y la matriz de var-cov de $\mathcal{N}(\beta_0, \beta_1)$, y los parámetros de escala y forma para $\mathcal{G}.\mathcal{I}(\sigma^{\in})$ usando el método de los momentos.

A continuación se muestra de manera gráfica las distribuciones de $\beta_0, \beta_1, y \sigma^2, y$ su tabla de parámetros estimados necesarios para ajustar el modelo teórico.

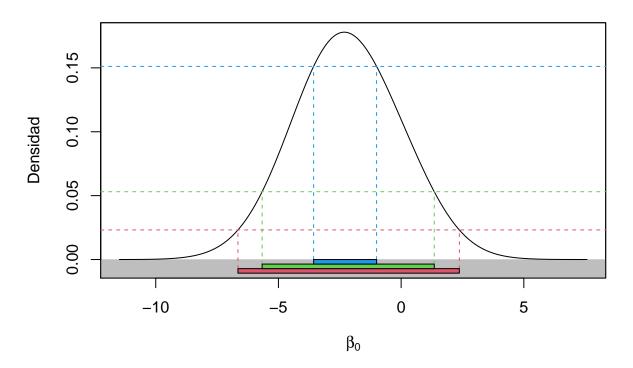
Distribución conjunta de β_0 y β_1



Distribución conjunta de $\,\beta_0\,y\,\beta_1\,$



Distribución Marginal de $\,\beta_0\,$



Distribución Marginal de β_1

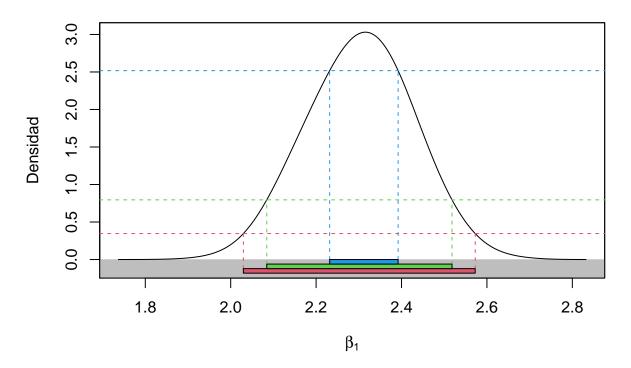


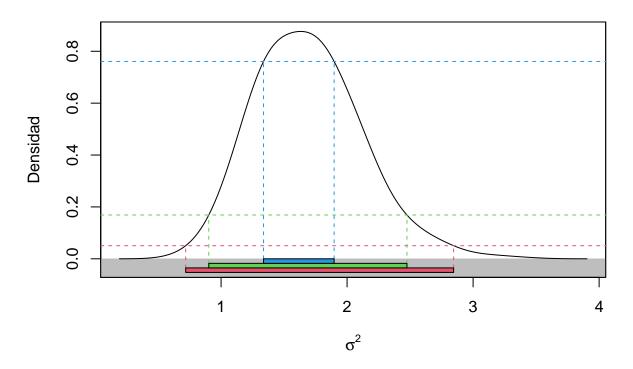
Table 4: Vector de Medias

μ_{eta_0}	μ_{eta_1}
-2.180951	2.303826

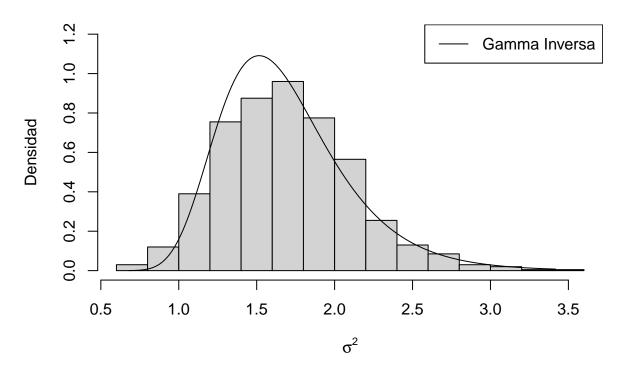
Table 5: Matriz de var-cov

	eta_0	β_1
β_0	83.83875	1341.963
β_1	1341.96265	21556.291

Distribución de σ^2



Distribución de σ^2



```
kable(t(c(alpha, beta)),
    col.names = c("Forma", "Escala"),
    caption = "Parámetros de G.I")
```

Table 6: Parámetros de G.I

Forma	Escala
18.27191	29.20815

Datos Muestrales:

Obtenidos de: https://carros.tucarro.com.co/ https://carros.mercadolibre.com.co/ https://www.olx.com.co/

En particular se buscaron autos cuyo modelo concordaran con los puntos de diseño escogidos, de forma que, más adelante, podamos comparar la información del experto con la muestral.

```
DatosCarros <- read_csv("DatosCarros.csv", col_names = c("modeloAuto", "precio"))

## Parsed with column specification:
## cols(
## modeloAuto = col_double(),</pre>
```

```
## precio = col_double()
## )
```

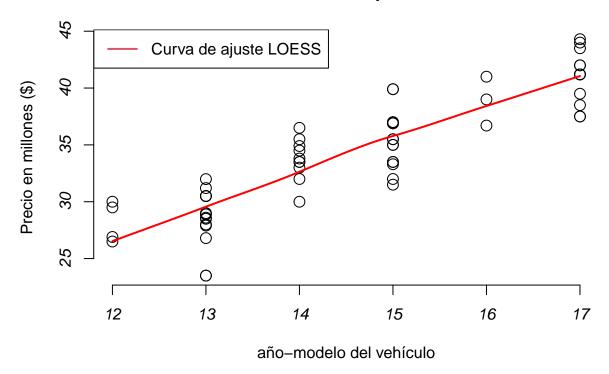
DatosCarros\$modeloAuto <- as.numeric(sub("20","",as.character(DatosCarros\$modeloAuto)))
attach(DatosCarros)</pre>

```
# Table <- data.frame(round(matrix(c(summary(precio),sd(precio)),ncol=7),2))
# names(Table)=c(names(summary(precio)), "sd")
# win.qraph(width = 10)
# layout(rbind(c(1,1),c(2,3)), heights=c(1,3))
# plot.new()
# addtable2plot(x="top", table=Table, xpad=1, ypad=1, bty='o',
                display.rownames = F, hlines = TRUE, vlines = TRUE,
#
                title = "Estadísticos de resumen")
#
# den <- density(precio)</pre>
# hist(precio, breaks="FD", freq=F, xlim=c(min(den$x), max(den$x)),
       main = paste("Histograma del precio en Millones de pesos"),
#
       xlab="Millones de pesos ($)",col="white")
# lines(den,col=2,lwd=2)
# boxplot(precio,boxwex=0.4,xlab="Precio en Millones ($)",
          col="white")
```

kable(table(modeloAuto))

modeloAuto	Freq
12	4
13	16
14	10
15	16
16	3
17	13

Gráfico de Dispersión



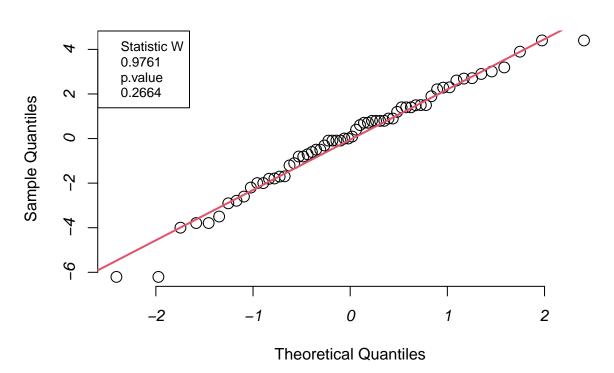
```
modeloVero <- lm(precio~modeloAuto)
anovaModeloVero <- data.frame(anova(modeloVero))
anovaModeloVero[2,c(4,5)] <- c("", "")
kable(anovaModeloVero)</pre>
```

	Df	Sum.Sq	Mean.Sq	F.value	PrF.
modeloAuto Residuals	1 60	1299.4866 329.1919	$1299.486589 \\ 5.486532$	236.85028740771	1.68555407073729e-22

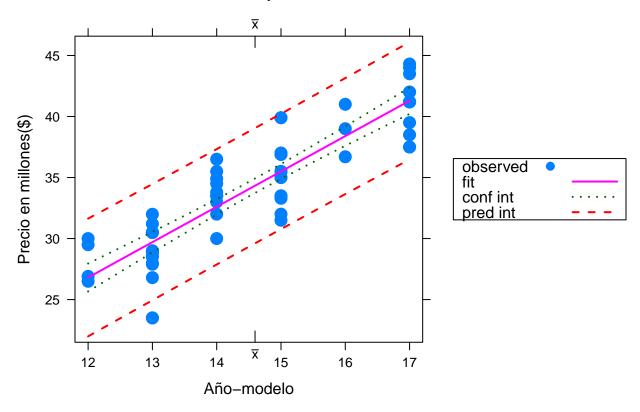
	Estimate	Std. Error	t value	$\Pr(> t)$
β_0	-7.945133	2.7630245	-2.875520	0.0055751
β_1	2.896230	0.1881898	15.389941	0.0000000
σ^2	5.486532	5.4865316	5.486532	5.4865316

```
test <- shapiro.test(residuals(modeloVero))
qqnorm(residuals(modeloVero),cex=1.5,bty="n",font=3,font.main=3)</pre>
```

Normal Q-Q Plot

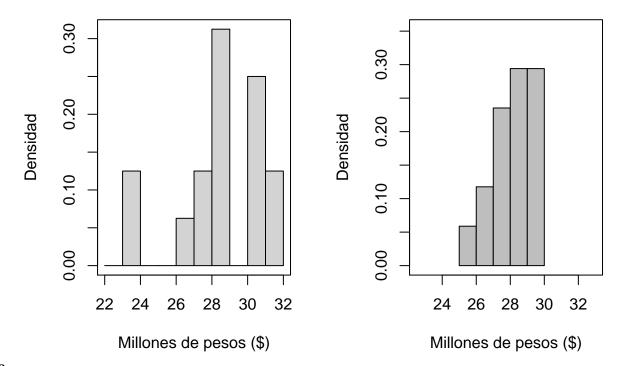


95% confidence and prediction intervals for modeloVero

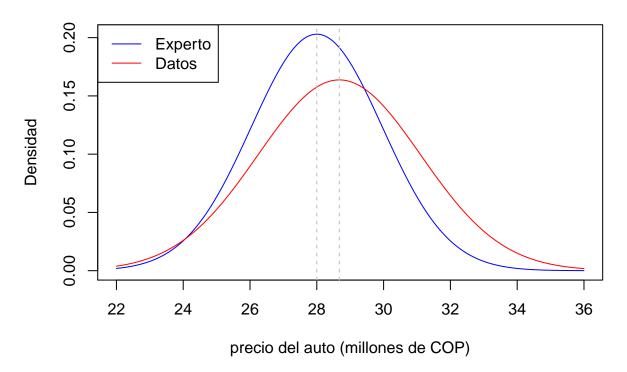


Datos Muestrales vs. Experto

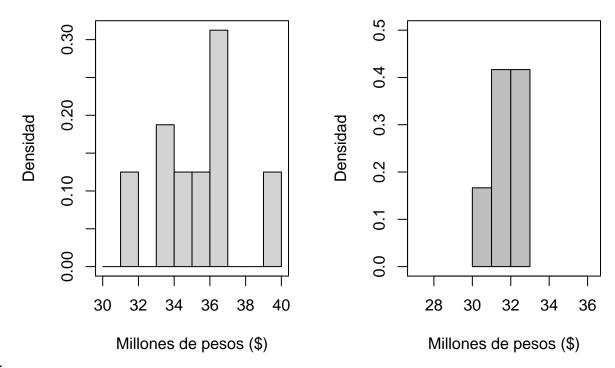
Datos precios vs. elicitación: 2013



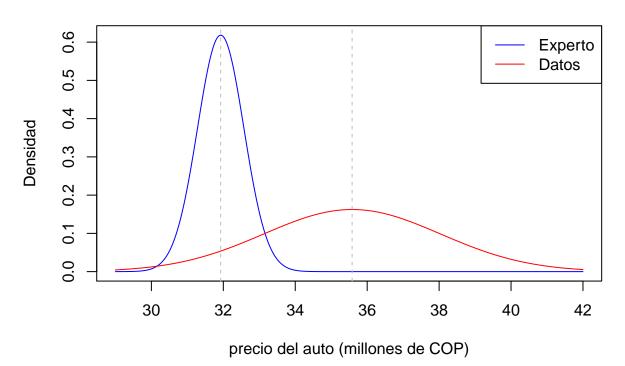
Distribución del precio según experto y datos del modelo 2013



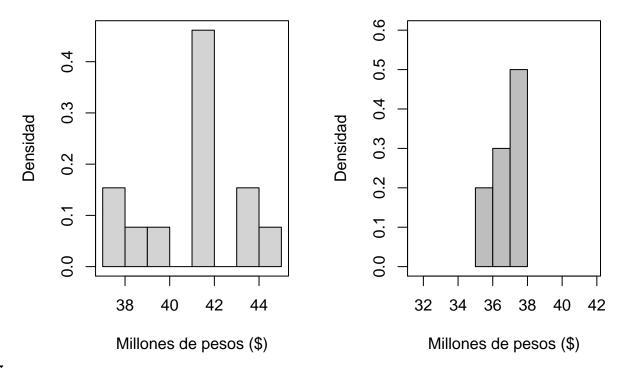
Datos precios vs. elicitación: 2015



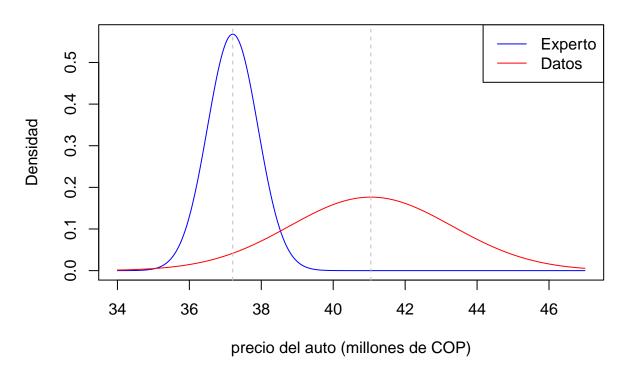
Distribución del precio según experto y datos del modelo 2015



Datos precios vs. elicitación: 2017



Distribución del precio según experto y datos del modelo 2017



Posterior

```
suPoste <- summary(resPoste)
kable(suPoste$statistics, col.names = colnames(suPoste$statistics))</pre>
```

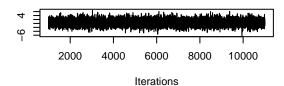
	Mean	SD	Naive SE	Time-series SE
(Intercept) modeloAuto sigma2	0.4466519 2.1528019 8.8325715	$\begin{array}{c} 1.5529481 \\ 0.0973066 \\ 1.3500479 \end{array}$	$\begin{array}{c} 0.0155295 \\ 0.0009731 \\ 0.0135005 \end{array}$	0.0153135 0.0009536 0.0148292

kable(suPoste\$quantiles, col.names = colnames(suPoste\$quantiles))

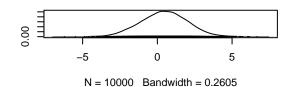
	2.5%	25%	50%	75%	97.5%
(Intercept)	-2.604465	-0.5890419	0.4618165	1.488456	$3.462205 \\ 2.343329 \\ 11.853601$
modeloAuto	1.963473	2.0867092	2.1515699	2.217531	
sigma2	6.584194	7.8767489	8.7080555	9.629242	

plot(resPoste)

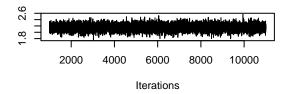
Trace of (Intercept)



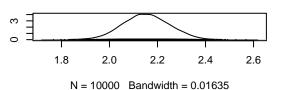
Density of (Intercept)



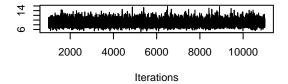
Trace of modeloAuto



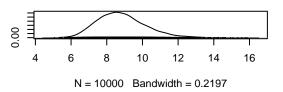
Density of modeloAuto



Trace of sigma2



Density of sigma2



cor(resPoste)

```
## (Intercept) modeloAuto sigma2

## (Intercept) 1.0000000 -0.9975114 -0.1428914

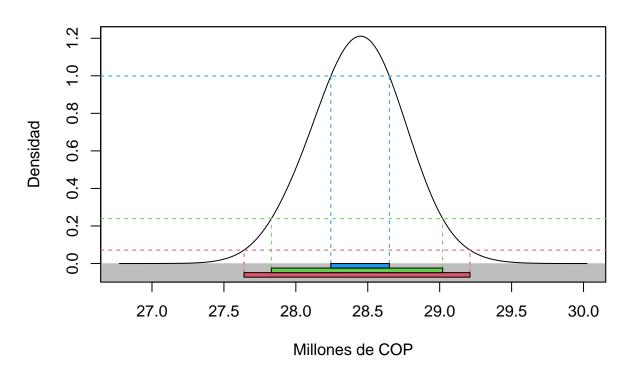
## modeloAuto -0.9975114 1.0000000 0.1241324

## sigma2 -0.1428914 0.1241324 1.0000000
```

Precio del modelo 2013

Distribución del precio medio del modelo 2013

Distribución del precio medio del modelo 2013



```
suM2013 <- summary(preciosMedios2013)
kable(suM2013$statistics, col.names = colnames(suM2013$statistics))</pre>
```

Mean	28.4330767
SD	0.3044668
Naive SE	0.0030447
Time-series SE	0.0031036

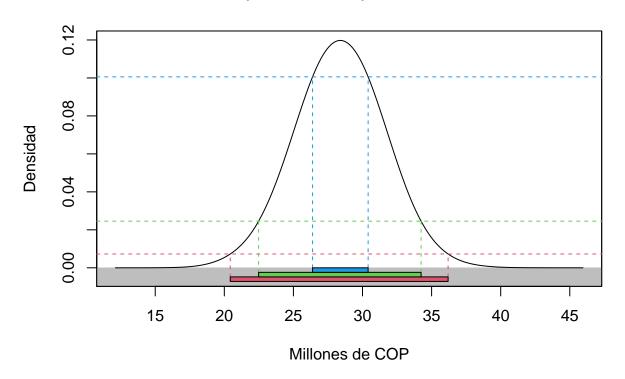
kable(suM2013\$quantiles, col.names = colnames(suM2013\$quantiles))

2.5%	27.83028
25%	28.23036
50%	28.43807
75%	28.63710
97.5%	29.02216

Distribución predictiva del precio del modelo 2013

```
errores <- sapply(resPoste[,3], genera.error)
preciosPredic2013 <- preciosMedios2013 + errores
```

Distribución predictiva del precio del modelo 2013



```
suP2013 <- summary(preciosPredic2013)
kable(suP2013$statistics, col.names = colnames(suP2013$statistics))</pre>
```

Mean	28.3697892
SD	2.9970633
Naive SE	0.0299706
Time-series SE	0.0308643

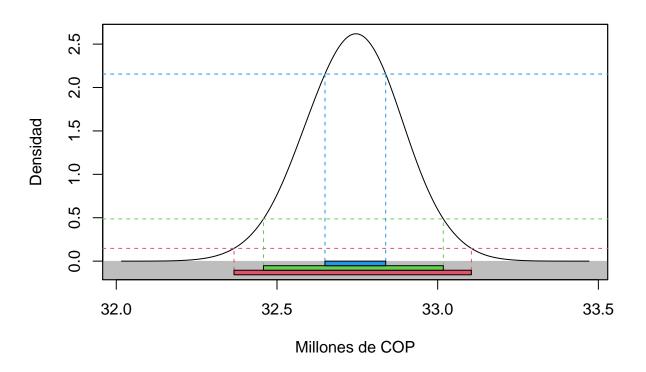
kable(suP2013\$quantiles, col.names = colnames(suP2013\$quantiles))

2.5%	22.46036
25%	26.37350
50%	28.38800
75%	30.39281
97.5%	34.22801

Precio del modelo 2015

Distribución del precio medio del modelo 2015

Distribución del precio medio del modelo 2015



```
suM2015 <- summary(preciosMedios2015)
kable(suM2015$statistics, col.names = colnames(suM2015$statistics))</pre>
```

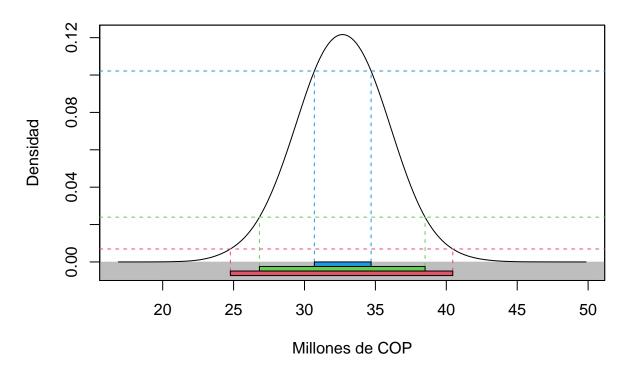
Mean	32.7386805
SD	0.1414064
Naive SE	0.0014141
Time-series SE	0.0015226

```
kable(suM2015$quantiles, col.names = colnames(suM2015$quantiles))
```

2.5%	32.45697
25%	32.64418
50%	32.74013
75%	32.83338
97.5%	33.01537

Distribución predictiva del precio del modelo 2015

Distribución predictiva del precio del modelo 2015



```
suP2015 <- summary(preciosPredic2015)
kable(suP2015$statistics, col.names = colnames(suP2015$statistics))</pre>
```

Mean	32.6753931
SD	2.9846612
Naive SE	0.0298466
${\it Time-series SE}$	0.0307561

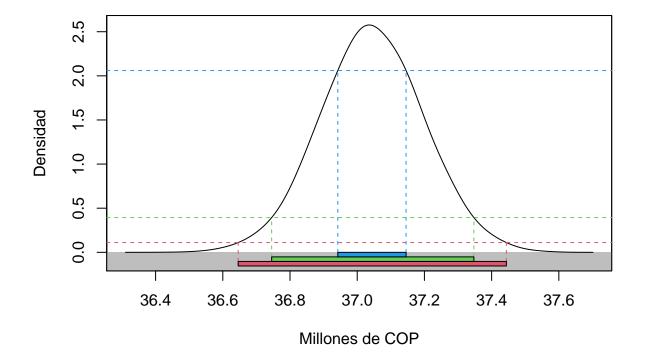
```
kable(suP2015$quantiles, col.names = colnames(suP2015$quantiles))
```

2.5%	26.80926
25%	30.68312
50%	32.68747
75%	34.69018
97.5%	38.48212

Precio del modelo 2017

Distribución del precio medio del modelo 2017

Distribución del precio medio del modelo 2017



```
suM2017 <- summary(preciosMedios2017)
kable(suM2017$statistics, col.names = colnames(suM2017$statistics))</pre>
```

Mean	37.0442843
SD	0.1517895
Naive SE	0.0015179
Time-series SE	0.0015179

```
kable(suM2017$quantiles, col.names = colnames(suM2017$quantiles))
```

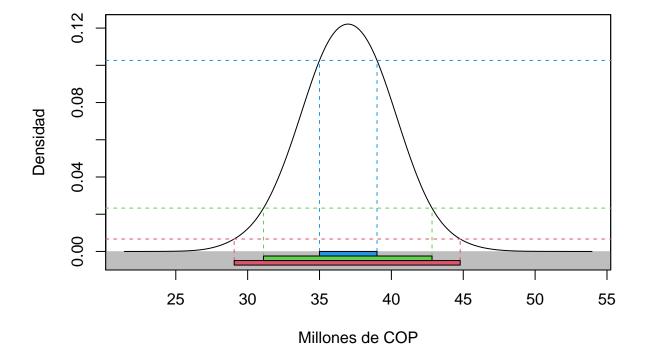
2.5%	36.74423
25%	36.94307
50%	37.04256
75%	37.14640
97.5%	37.34642

Distribución predictiva del precio del modelo 2017

```
preciosPredic2017 <- preciosMedios2017 + errores</pre>
```

```
hdrPreciosPredic2017 <- hdr.den(preciosPredic2017, xlab = "Millones de COP",
ylab = "Densidad",
main = expression("Distribución predictiva del precio del modelo 2017"))
```

Distribución predictiva del precio del modelo 2017



suP2017 <- summary(preciosPredic2017) kable(suP2017\$statistics, col.names = colnames(suP2017\$statistics))</pre>

Mean	36.9809969
SD	2.9849229
Naive SE	0.0298492
Time-series SE	0.0307652

kable(suP2017\$quantiles, col.names = colnames(suP2017\$quantiles))

2.5%	31.09711
25%	34.99406
50%	36.98867
75%	38.99287
97.5%	42.82485