Probabilidad: MIACD

Héctor de la Torre Gutiérrez

Fecha de modificación: 25-10-2025

Contents

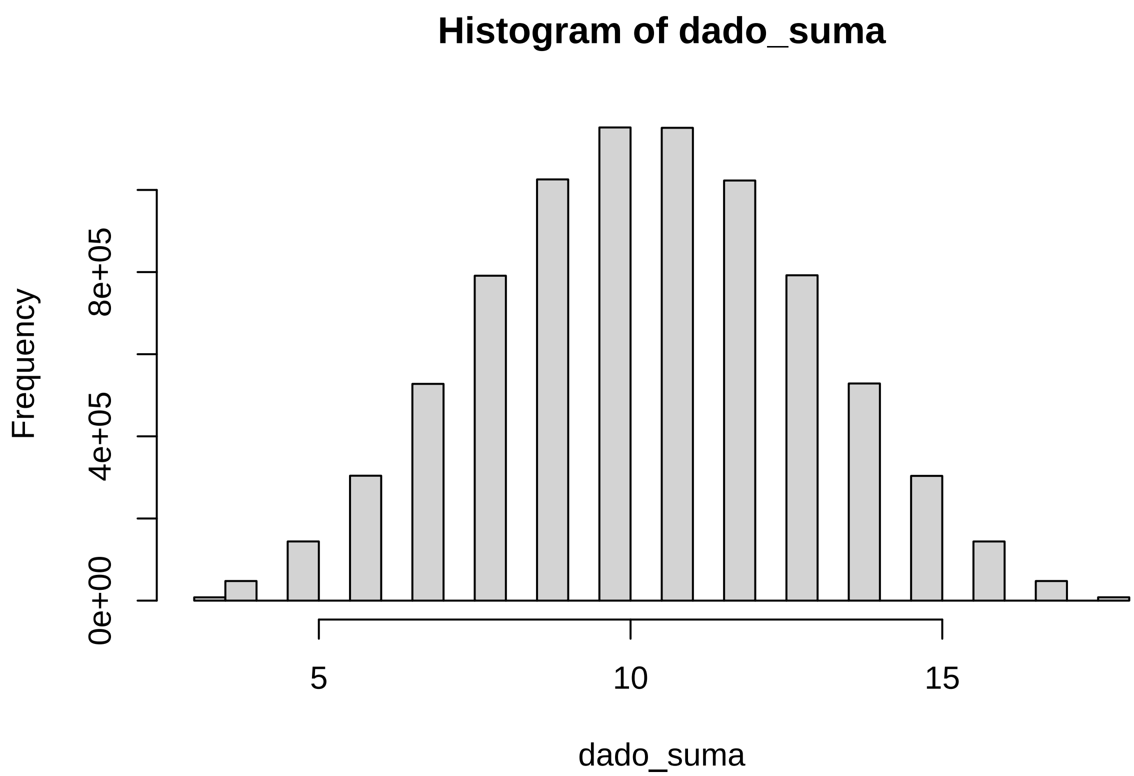
DISTRIBUCIONES DE PROBABILIDAD

Ejemplo de simulación para determinar probabilidades: Simular el experimento del lanzamiento de uno, dos, tres dados. Se analiza la suma.

```
set.seed(1234)
experi<-8000000
dado1<- round(runif(experi,1,6),0) #Dado uno, lanzado X veces
dado2<- round(runif(experi,1,6),0) #Dado dos, lanzado X veces
dado3<- round(runif(experi,1,6),0) #Dado tres, lanzado X veces
dado_suma<-dado1+dado2+dado3
(table(dado_suma)/experi)*100
```

```
## dado_suma
##          3          4          5          6          7          8          9
##  0.0996500  0.5970125  1.8010375  3.8026875  6.5977875  9.8901625 12.8212625
##         10         11         12         13         14         15         16
## 14.4028500 14.3910500 12.7880750  9.9028875  6.6098500  3.7968750  1.8003250
##         17         18
##  0.5966250  0.1018625
```

```
hist(dado_suma)
```



1) DISTRIBUCIONES DE PROBABILIDAD PARA VARIABLES ALEATORIAS DE TIPO DISCRETAS

1.1) DISTRIBUCION BINOMIAL

Tomando en cuenta el ejemplo de la población fumadora dentro de una muestra de 50 personas, se pide determinar:
$P(10 \leq x \leq 40) = P(x \leq 40) - P(x \leq 9)$
$P(x = 10)$

```
proba_binom<-rep(NA,31)
for(i in 10:40){
proba_binom[i-9]<-choose(50,i)*(.3^i)*(.7^(50-i)) }
sum(proba_binom)
```

```
## [1] 0.9597684
```

```
proba_binom<-rep(NA,10)
for(i in 0:9){
proba_binom[i+1]<-choose(50,i)*(.3^i)*(.7^(50-i)) }
sum(proba_binom)
```

```
## [1] 0.04023163
```

```
pbinom(10,50,0.3,lower.tail = TRUE)-pbinom(9,50,0.3,lower.tail = TRUE)
```

```
## [1] 0.03861899
```

```
choose(50,10)*(.3^10)*(.7^(50-10))
```

```
## [1] 0.03861899
```

```r
pbinom(40,50,.3,lower.tail = TRUE)-pbinom(9,50,.3,lower.tail = TRUE)
```

```
## [1] 0.9597684
```

```r
#var_Esp_binom<-50*.3
#var_Esp_binom
```

1) Probabilidad de hasta 15 fumadores en la muestra de 50 personas.

2) Probabilidad de encontrar más de 20 fumadores en la muestra.

```r
pbinom(15,50,.3,lower.tail = TRUE)
```

```
## [1] 0.5691784
```

```r
1-pbinom(20,50,.3,lower.tail = TRUE) # igual a lo siguiente
```

```
## [1] 0.04776384
```

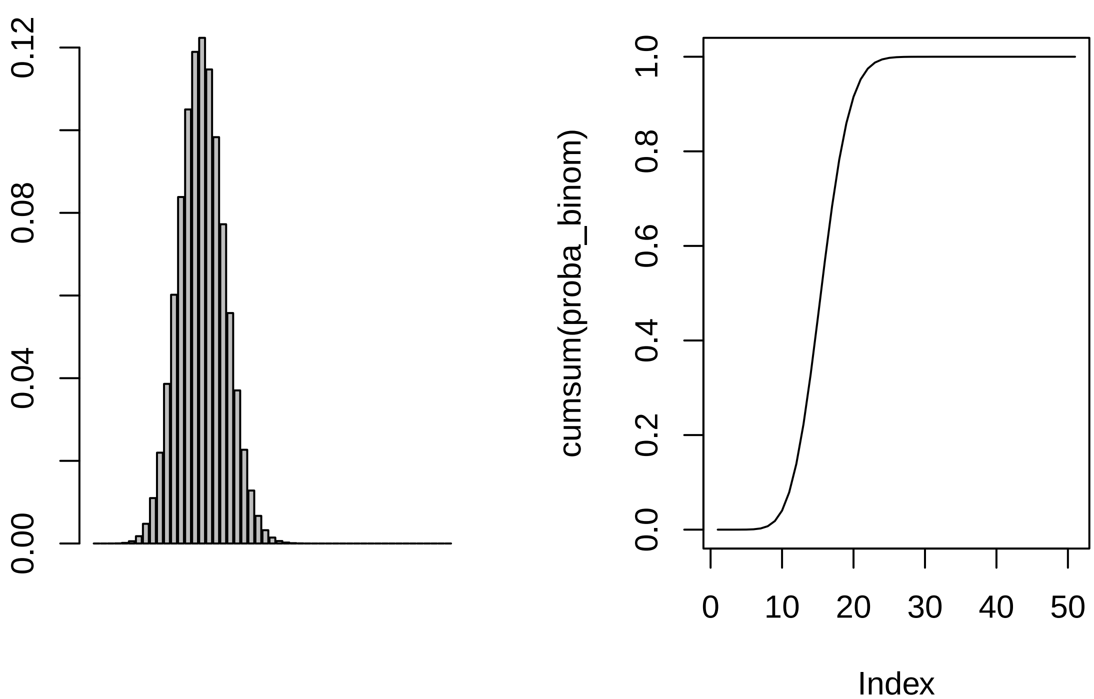
```r
pbinom(20,50,.3,lower.tail = FALSE) #igual a lo anterior
```

```
## [1] 0.04776384
```

```r
pbinom(15,50,.3,lower.tail = TRUE)-pbinom(14,50,.3,lower.tail = TRUE)
```

```
## [1] 0.1223469
```

```r
proba_binom<-rep(NA,51)
for(i in 0:50){
proba_binom[i+1]<-choose(50,i)*(.3^i)*(.7^(50-i)) }
par(mfrow=c(1,2))
barplot(proba_binom)
plot(cumsum(proba_binom),type="l")
```



1.2) DISTRIBUCIÓN POISSON

Se estudia el número de ocurrencias (accidentes) en 20 cruceros de la ciudad de Aguascalientes el día 10 de septiembre (24 horas), los datos son los siguientes:

0, 5, 3, 4, 1, 6, 0, 2, 4, 9, 1, 0, 0, 4, 7, 3, 0, 1, 0, 1

Dada una taza de accidentes,
- 1. ¿Cuál es la probabilidad de que haya accidentes en un crucero? - 2. ¿Cuál es la probabilidad de no haya accidentes? - 3. ¿Haya más de tres accidentes en 24 horas? - 4. ¿Cuál es la probabilidad de que haya 2 o más accidentes cuando se observa el fenómenos de 'accidentes en un crucero' por 48 horas?

```
accidentes<-c(0, 5, 3, 4, 1, 6, 0, 2, 4, 9, 1, 0, 0, 4, 7, 3, 0, 1, 0, 1)
theta_pos<-mean(accidentes)
```

```
ppois(0,theta_pos,lower.tail = TRUE) #Probabilidad de que no haya accidentes
```

```
## [1] 0.07808167
```

```
1-ppois(0,theta_pos,lower.tail = TRUE) #Probabilidad de que haya accidentes
```

```
## [1] 0.9219183
```

```
ppois(0,theta_pos,lower.tail = FALSE) #Probabilidad de que haya accidentes
```

```
## [1] 0.9219183
```

```
1-ppois(3,theta_pos,lower.tail = TRUE)
```

```
## [1] 0.2531635
```

```
1-ppois(1,theta_pos*2,lower.tail = TRUE)
```

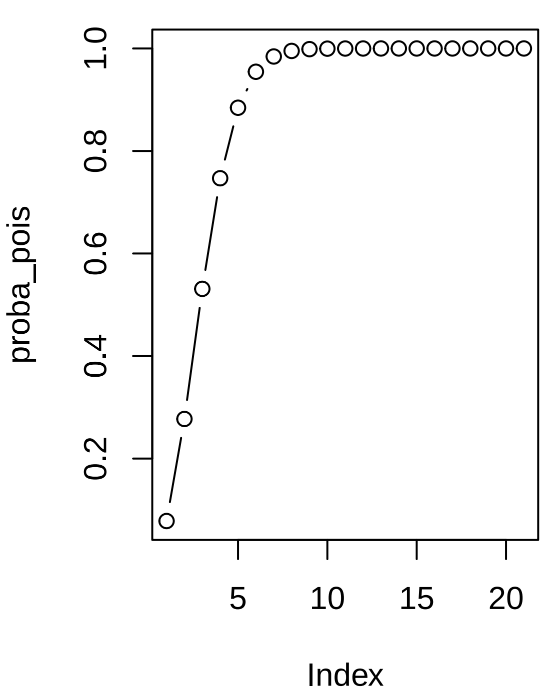
```
## [1] 0.9628098
```

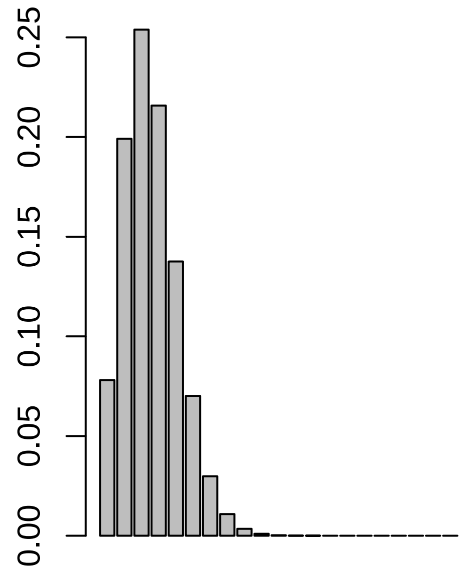
```
ppois(1,theta_pos*2,lower.tail = FALSE)
```

```
## [1] 0.9628098
```

```
par(mfrow=c(1,2))
proba_pois<-rep(NA,21)
for(i in 0:20){
proba_pois[i+1]<-ppois(i,theta_pos)}
plot(proba_pois,type="b")

proba_pois2<-rep(NA,21)
for(i in 0:20){
proba_pois2[i+1]<-((exp(-theta_pos)*theta_pos^i))/factorial(i) }
barplot(proba_pois2)
```





1.3) EJEMPLO INTEGRADOR DE DISTRIBUCIONES DISCRETAS

Utilizar la distribución Binomal para calcular 1) y 2) del ejercicio anterior. Repensar el problema, ahora se estudiará cada semáforo como un evento Bernoulli, y buscaremos saber:
1) en todos los semáforos no haya accidentes $P(X = 0|p = 0.70)$ y
2) en solo un semáforo haya accidentes $P(X = 1|p = 0.70)$. También, que haya mínimo cinco semáforos con choques. ¿Es posible pensar que haya más de 10 semáforos con choques dada la proporción de *p = 0.70* y *n= 20* ?

```
accidentes2<-ifelse(accidentes==0,0,1)
pe_acci<-sum(accidentes2)/20
20*pe_acci #Valor esperado
```

```
## [1] 14
```

```
#Se tiene una flotilla que puede atender hasta cinco accidentes al día
#P() Probabilidad de que cubran el trabajo demandado P(X<=5|n=20,p=0.7)
pbinom(0,20,0.7,lower.tail = TRUE)
```

```
## [1] 3.486784e-11
```

```
pbinom(0,20,0.7,lower.tail = FALSE)
```

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

```
pbinom(10,20,0.7,lower.tail = FALSE)
```

```
## [1] 0.9520381
```

2) DISTRIBUCIONES DE PROBABILIDAD PARA VARIABLES ALEATORIAS CONTINUAS

Una empresa refresquera tiene dos líneas de llenado para las botellas, línea A y B. Ambas, en teoría suministran una cantidad de llenado de 600ml en promedio. Sin embargo, durante la última semana, se ha detectado que en la línea A, dos botellas estaban por encima de los 600 ml. Mientras que en la línea B, todo se encontró en orden. El ingeniero de control de la calidad decide tomar 10 botellas al azar de la línea A y 12 botellas al azar de la línea B. ¿Cuál de las dos máquinas puede cumplir de mejor manera con los requisitos de calidad? (ver tolerancias más adelante). Los resultados se muestran en la tabla siguiente:

```
datos_ejemplo<-matrix(c(601.333,    595.0222,601.744,   606.0491,606.0802,  604.0949,601.2242,  601.951
602.0219,   599.094,599.3259,   598.7811,602.5042,  603.5545,600.9781,  592.3436,
602.7935,   605.3266,592.9382,  600.6886,NA,        601.1702,NA,        601.3226),ncol=2,byrow=T)
colnames(datos_ejemplo)<-c("Linea_A",   "Linea_B")
datos_ejemplo<-as.data.frame(datos_ejemplo)
var(datos_ejemplo$Linea_A,na.rm=T)
```

```
## [1] 11.19482
```

```
datos_ejemplo
```

```
##     Linea_A  Linea_B
## 1  601.3330 595.0222
## 2  601.7440 606.0491
## 3  606.0802 604.0949
## 4  601.2242 601.9510
## 5  602.0219 599.0940
## 6  599.3259 598.7811
## 7  602.5042 603.5545
## 8  600.9781 592.3436
## 9  602.7935 605.3266
## 10 592.9382 600.6886
## 11       NA 601.1702
## 12       NA 601.3226
```

Si se tiene una tolerancia de llenado extra o faltante de 2ml, ¿cuál es la probabilidad de que la Línea A y B cumplan con lo estipulado?

- Proceso A: $P(598 < X < 602|\mu = 601.09, \sigma^2 = 11.19)$
- Proceso B: $P(598 < X < 602|\mu = 600.78, \sigma^2 = 16.37)$

```
mediaA<-mean(datos_ejemplo$Linea_A,na.rm=TRUE)
varA<-var(datos_ejemplo$Linea_A,na.rm=TRUE)
mediaB<-mean(datos_ejemplo$Linea_B,na.rm=TRUE)
varB<-var(datos_ejemplo$Linea_B,na.rm=TRUE)
```

```
pnorm(602,mediaA,varA)-pnorm(598,mediaA,varA) #En este proceso es más probable cumplir con las especifi
```

```
## [1] 0.1411219
```

```
pnorm(602,mediaB,varB)-pnorm(598,mediaB,varB)
```

```
## [1] 0.0971068
```

2) Considere los siguientes datos sobre carga de ruptura (kg/25mm de ancho) para varios tejidos tanto sin cardar como cardados.

```
Sin_cardar<-c(36.4,55.0,51.5,38.7,43.2,48.8,25.6,49.8)
Cardado<-c(28.5,20, 46, 34.5,   36.5,   52.5,   26.5,   46.5)
```

DISTRIBUCION DE PROBABILIDAD DE LA MAGNITUD DE LOS SISMOS

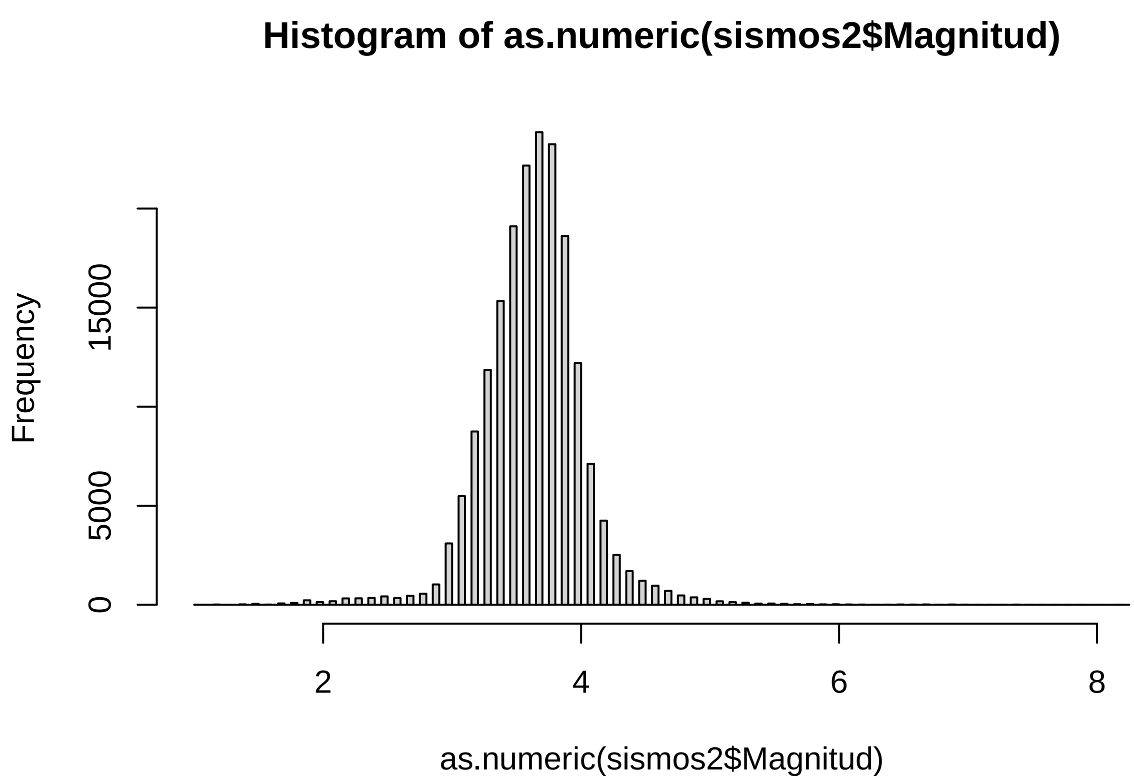
1) Jalar datos junto al histograma de las magnitudes, así como de la densidad
2) ¿Parece ser una distribución normal?
3) ¿Cuál es la probabilidad de que haya un sismo destructor?
4) ¿Y la probabilidad de un sismo cuya magnitud sea mayor a 5?
5) TLC. Pendiente

```
sismos<-read.csv("P:\\UP\\2025\\2025-2\\SSNMX_catalogo_19000101_20251004 (1).csv")
sismos$Magnitud<-as.numeric(sismos$Magnitud)
```

```
## Warning: NAs introduced by coercion
```

```
sismos2<-sismos[which(sismos$Estado2%in%c(" CHIS"," OAX"," JAL")),]
#which(sismos$Estado2%in%c(" CHIS"," OAX"," JAL"))
```

```
sismos2$Magnitud<-as.numeric(sismos2$Magnitud)
hist(as.numeric(sismos2$Magnitud),breaks = seq(1,8.25,0.05))
```




```
#Asumamos que un sismo destructor es aquel cuya magnitud es mayor a 5.0 grados.
#Obtener la probabilidad de dicho tipo de sismo para cada estado y a nivel nacional.
media_estados<-aggregate(Magnitud~Estado2,data=sismos2,FUN=mean,na.rm=TRUE)
sd_estados<-aggregate(Magnitud~Estado2,data=sismos2,FUN=sd,na.rm=TRUE)
media_nac<-mean(sismos$Magnitud,na.rm=TRUE)
sd_nac<-sd(sismos$Magnitud,na.rm=TRUE)
```

```
pnorm(5,media_nac,sd_nac) #Probabilidad de un sismo no-destructo
```

```
## [1] 0.9970959
```

```r
#pnorm(5,media_nac,sd_nac,lower.tail = FALSE) #Probabilidad de un sismo destructo
#1-pnorm(5,media_nac,sd_nac,lower.tail = TRUE) #Similar al anterior
media_nac
```

```
## [1] 3.544628
```

```r
fun_sismo<-function(x){
  media<-mean(x,na.rm=TRUE)
  desvest<-sd(x,na.rm=TRUE)
  probabilidad<-pnorm(5,media,desvest,lower.tail = FALSE)
  #return(data.frame(Media=media,Des_Esta=desvest,Prrobabilidad=probabilidad))}
  return(probabilidad) }
```

```r
1-pnorm(5,media_nac,sd_nac,lower.tail = TRUE) #Similar al anterior
```

```
## [1] 0.002904127
```

```r
1-pnorm(5,media_estados[1,2],sd_estados[1,2]) #Chiapas
```

```
## [1] 0.004558354
```

```r
1-pnorm(5,media_estados[2,2],sd_estados[2,2]) #Jalisco
```

```
## [1] 0.0002066
```

```r
1-pnorm(5,media_estados[3,2],sd_estados[3,2]) #Oaxaca
```

```
## [1] 3.411705e-05
```

```r
cbind(media_estados,sd_estados[,2])
```

```
##   Estado2 Magnitud sd_estados[, 2]
## 1    CHIS 3.780379       0.4677094
## 2     JAL 3.545959       0.4117339
## 3     OAX 3.616808       0.3473296
```

```r
## Calcularlo para los 32 estados y observa aquel donde la probabilidad se la mayor,
# y la menor. Aggregate
```

```r
magnitud<-aggregate(Magnitud~Estado2,data=sismos,FUN=fun_sismo,simplify = TRUE)
mean_magni<-aggregate(Magnitud~Estado2,data=sismos,FUN=mean,simplify = TRUE)
sd_magni<-aggregate(Magnitud~Estado2,data=sismos,FUN=sd,simplify = TRUE)
magnitud[order(magnitud$Magnitud,decreasing = TRUE),]
```

```
##    Estado2     Magnitud
## 23      QR 3.461654e-01
## 32     YUC 4.523902e-02
## 25     SIN 2.256600e-02
## 6     CHIH 1.303216e-02
## 4     CAMP 1.152166e-02
## 3      BCS 6.357967e-03
## 7     CHIS 4.558354e-03
## 18     NAY 2.384860e-03
## 15     MEX 2.164871e-03
## 8     COAH 1.076265e-03
## 26     SLP 1.028104e-03
## 27     SON 1.017773e-03
## 22     PUE 4.138874e-04
```

```
## 28     TAB 2.231477e-04
## 14     JAL 2.066000e-04
## 31     VER 1.143970e-04
## 13     HGO 1.045731e-04
## 11     GRO 6.930489e-05
## 16    MICH 4.894393e-05
## 21     OAX 3.411705e-05
## 10     DGO 3.246049e-05
## 5     CDMX 2.811593e-05
## 17     MOR 1.953568e-05
## 2       BC 1.098871e-05
## 19      NL 8.238268e-06
## 33     ZAC 7.364855e-06
## 29    TAMS 4.604504e-06
## 9      COL 2.171194e-06
## 1      AGS 1.520671e-06
## 24     QRO 1.050707e-06
## 30    TLAX 4.733117e-07
## 12     GTO 7.434940e-09
## 20     OAx           NA
```

```r
magni1<-merge(magnitud,mean_magni,by.x="Estado2",by.y = "Estado2")
magni2<-merge(magni1,sd_magni,by.x="Estado2",by.y = "Estado2")
colnames(magni2)<-c("Estado","Probabilidad","Media","Sd")
magni2
```

```
##    Estado Probabilidad    Media        Sd
## 1     AGS 1.520671e-06 2.791489 0.4731163
## 2      BC 1.098871e-05 3.415207 0.3734379
## 3     BCS 6.357967e-03 2.437123 1.0285957
## 4    CAMP 1.152166e-02 4.213333 0.3461351
## 5    CDMX 2.811593e-05 2.087049 0.7231594
## 6    CHIH 1.303216e-02 3.487569 0.6796673
## 7    CHIS 4.558354e-03 3.780379 0.4677094
## 8    COAH 1.076265e-03 3.815517 0.3860340
## 9     COL 2.171194e-06 3.406098 0.3469315
## 10    DGO 3.246049e-05 3.830588 0.2927796
## 11    GRO 6.930489e-05 3.539758 0.3832017
## 12    GTO 7.434940e-09 3.758929 0.2191522
## 13    HGO 1.045731e-04 2.841137 0.5822639
## 14    JAL 2.066000e-04 3.545959 0.4117339
## 15    MEX 2.164871e-03 2.786099 0.7759683
## 16   MICH 4.894393e-05 3.582820 0.3637743
## 17    MOR 1.953568e-05 3.417722 0.3847108
## 18    NAY 2.384860e-03 3.850877 0.4071744
## 19     NL 8.238268e-06 3.556210 0.3351445
## 20    OAx           NA 3.000000        NA
## 21    OAX 3.411705e-05 3.616808 0.3473296
## 22    PUE 4.138874e-04 3.395629 0.4798712
## 23     QR 3.461654e-01 4.748571 0.6354116
## 24    QRO 1.050707e-06 3.476471 0.3211881
## 25    SIN 2.256600e-02 3.993798 0.5022414
## 26    SLP 1.028104e-03 3.685897 0.4263810
## 27    SON 1.017773e-03 3.746959 0.4061727
## 28    TAB 2.231477e-04 3.901455 0.3128797
```

```
## 29   TAMS 4.604504e-06 3.639462 0.3067758
## 30   TLAX 4.733117e-07 3.513534 0.3032106
## 31    VER 1.143970e-04 3.747764 0.3398291
## 32    YUC 4.523902e-02 3.920000 0.6379655
## 33    ZAC 7.364855e-06 1.730249 0.7546697
```

EJERCICIO INTEGRADOR

Una compañía fabrica focos con vida media de 500 horas y desviación estándar de 100. Suponga que los tiempos de vida útil de los focos se distribuyen normalmente, esto es que los tiempos de vida forman una distribución normal.

- a) Encuentre la probabilidad de que cierta cantidad de focos dure menos de 650 horas.
- b) Calcule la probabilidad de que cierta cantidad de focos dure más de 780 horas.
- c) Determine la probabilidad de que cierta cantidad de focos dure entre 650 y 780 horas (ambos inclusive).
- d) Halle el valor de k tal que el 5% de los focos tenga un tiempo de vida mayor que k horas?
- e) Si se eligen 10,000 focos, ¿cuántos tuvieron un tiempo de vida entre 650 y 780 horas (ambos inclusive)?
- f) Si se eligen 1200 focos, ¿cuál es la probabilidad de que al menos 3 duren más de 780 horas?
- g) Si se eligen 20 focos, ¿cuál es la probabilidad de que entre 16 y 19 (ambos inclusive) duren menos de 650 horas?
- h) Si se eligen 20 focos, ¿cuál es la probabilidad de que entre 16 y 19 (ambos inclusive) no duren menos de 650 horas?

```r
pnorm(650,500,100) # a)
```

```
## [1] 0.9331928
```

```r
1-pnorm(650,500,100) # pre-h
```

```
## [1] 0.0668072
```

```r
pnorm(780,500,100,lower.tail = FALSE) #b)
```

```
## [1] 0.00255513
```

```r
pnorm(780,500,100,lower.tail = TRUE)-pnorm(650,500,100,lower.tail = TRUE) #c
```

```
## [1] 0.06425207
```

```r
qnorm(0.95,500,100) #d)
```

```
## [1] 664.4854
```

```r
qnorm(0.05,500,100,lower.tail = FALSE) #d)
```

```
## [1] 664.4854
```

```r
pnorm(664.485362695147,500,100) #d)
```

```
## [1] 0.95
```

```r
(pnorm(780,500,100,lower.tail = TRUE)-pnorm(650,500,100,lower.tail = TRUE))*10000 #e
```

```
## [1] 642.5207
```

```r
1-pbinom(2,size=1200,prob=pnorm(780,500,100,lower.tail = FALSE)) #f)
```

```
## [1] 0.5917659
```

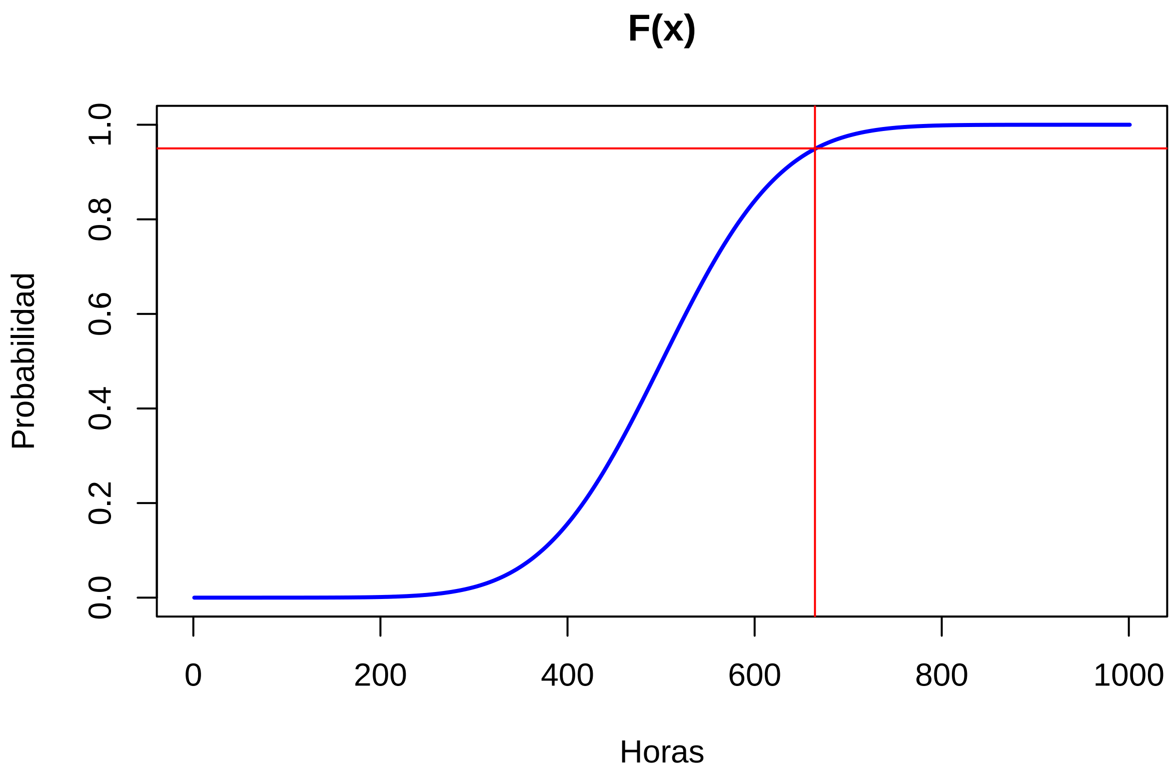
```r
pbinom(19,size=20,prob=pnorm(650,500,100))-pbinom(15,size=20,prob=pnorm(650,500,100)) #g)
```

```
## [1] 0.7403082
```

```r
pbinom(19,size=20,prob=pnorm(650,500,100,lower.tail = FALSE))-pbinom(15,size=20,prob=pnorm(650,500,100,l
```

```
## [1] 6.661338e-16
```

```r
#Gráficamente, ¿cómo se ve la distribución?
plot(pnorm(seq(0,1000),500,100),type="l",lwd=2,col="blue",main="F(x)",xlab="Horas",
ylab="Probabilidad")
abline(v=664.485362695147,col="red")
abline(h=0.95,col="red")
```

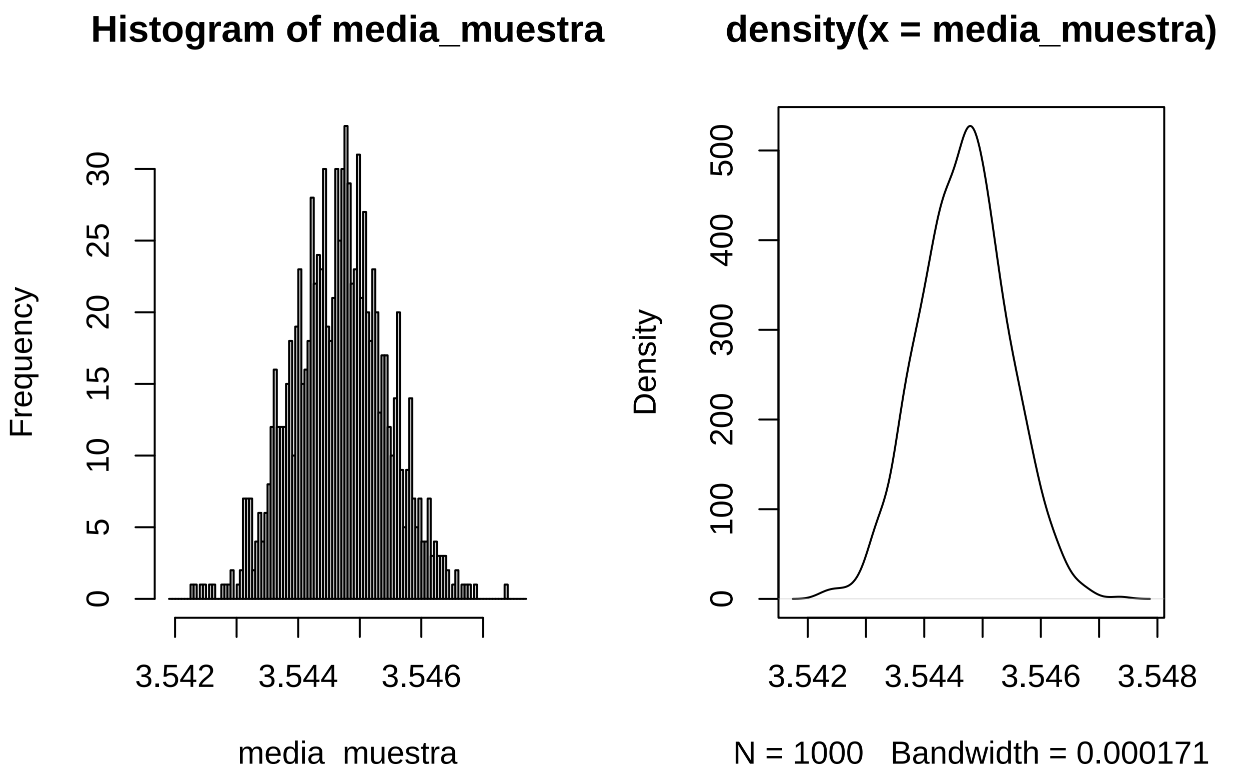


EJEMPLO DE LA APLICACIÓN DE TEOREMA DEL LÍMITE CENTRAL

```r
# 1) Tomar 200,000 muestras
# 2) Calcular la media de esta muestra de tamaño 200,000
# 3) Guardar el valor
# 4) Repertir lo anterior 200 veces

#1000 medias... ¿cómo se observa gráficamente estas 200 medias? Realizar un histograma
# o un gráfico de densidad.
sismos$Magnitud<-as.numeric(sismos$Magnitud)
media_muestra<-rep(NA,1000)
for(i in 1:1000) {
  media_muestra[i]<-mean(sample(sismos$Magnitud,size=200000+i,replace=FALSE),na.rm=TRUE)}
```

```
par(mfrow=c(1,2))
hist(media_muestra,breaks=seq(min(media_muestra)*.9999,max(media_muestra)*1.0001,0.00005))
plot(density(media_muestra))
```




```
media_media<-mean(media_muestra)
sd_media<-sd(media_muestra)/sqrt(1000)
```