PEC1 Inferencia Estadistica

Laura Sudupe

16/11/2020

EJERCICIO 1.

a. Las probabilidades de una variable discreta X están representadas mediante uno de los siguientes gráficos:

Las probabilidades de un suceso, deben cumplir los axiomas de Kolmogorov. Uno de los tres axiomas dice que la probabilidad del espacio muestral es 1. En el grafico de la izquiera es 0.09, por lo que el que es correcto es el de la derecha.

- Calcular:

a. $P(X = 6)$ $P = 0.25$
b. $P(X \geq 4)$ $P <= 0.2$
c. $P(6 \leq X < 8)$ $P(6 \leq X < 8) = Fx(8) - Fx(6)$
d. $P(X = 6|X \geq 4)$ $P(X = 6|X \geq 4)$ $P=0.25|P<=0.2$
e. Uno de los siguientes gráficos representa la función de distribución de una variable continua Y, cual y porque:

La correcta es la segunda gráfica, porque la probabilidad acumulada tiene que ser la unidad completa, 1.

a. $P(Y = 0) = 0.35$
b. $P(Y < 25) = P > 0.85$
c. $P(10 < Y < 25) = F(25) - F(10)$ La integral de f(Y) entre 25 y 10, el area comprendida bajo la linea que describe la función.
d. $P(Y > 10|Y < 25)$ $P = 1$

EJERCICIO 2.

Según La revista científica The Lancet que ha publicado un extenso estudio sobre los efectos que la distancia interpersonal, el uso de mascarillas y otros medios de protección contra el coronavirus se describe que la Probabilidad de transmisión del coronavirus en personas que no usan la mascarilla es del 17.8% y en el caso que la usen de del 3.1%

a. ¿Si el 40% de la población no lleva mascarilla? , cual es la probabilidad de infectarse.

$P(A) = P(A \cap Bx) / P(Bx) = 17,8/(40/100)$ $P(A) = 44.8\%$

b. ¿Si uno de los pacientes resulta infectado , cual es la probabilidad que sea de los que no llevaba mascarilla?

$P(Bx) = P(A \cap Bx) / P(A) = 17.8/44.8 = 0.39$

Un 39% de probabilidad

Si 15 sujetos acuden a una fiesta nocturna sin mascarilla. c) ¿Cuál es la probabilidad de que se contagien exactamente 5 ?

```
p_c <- pbinom(5,15,0.178)
p_c
```

```
## [1] 0.9631696
```

d. ¿Cual es la probabilidad de que se contagien más de 5 pacientes?

```
#probabilidad de que se contagien mas de 5 pacientes si 15 acuden sin mascarilla
p_d <- 1-pbinom(4,15,0.178, lower.tail = T)
p_d
```

```
## [1] 0.1124046
```

Genera 1000 muestras aleatorias de 15 pacientes utilizan mascariilla y calcula el intervalo de confianza al 95%. " - Puedes simular los valores observados con el código siguiente para poder replicar los resultados si los ejecutas varias veces, ya que si no fijas la semilla(seed) cada vez que ejecutes los comandos te dará un resultado diferente set.seed (unnumero) #cambia "unnumero" por un número para que siempre genere la misma muestra - rbinom(n, size, prob) # genera n muestras de tamaño "size" con una probabilidad de éxito "prob"

```
#generamos las muestras aleatorias de los 15 pacientes que usan mascarilla

set.seed(9991)
muestra <- rbinom(1000,15,0.031)

#calcular intervalo de confianza al 95%

prop.test(x=15, n=1000, conf.level=0.95, correct=FALSE)
```

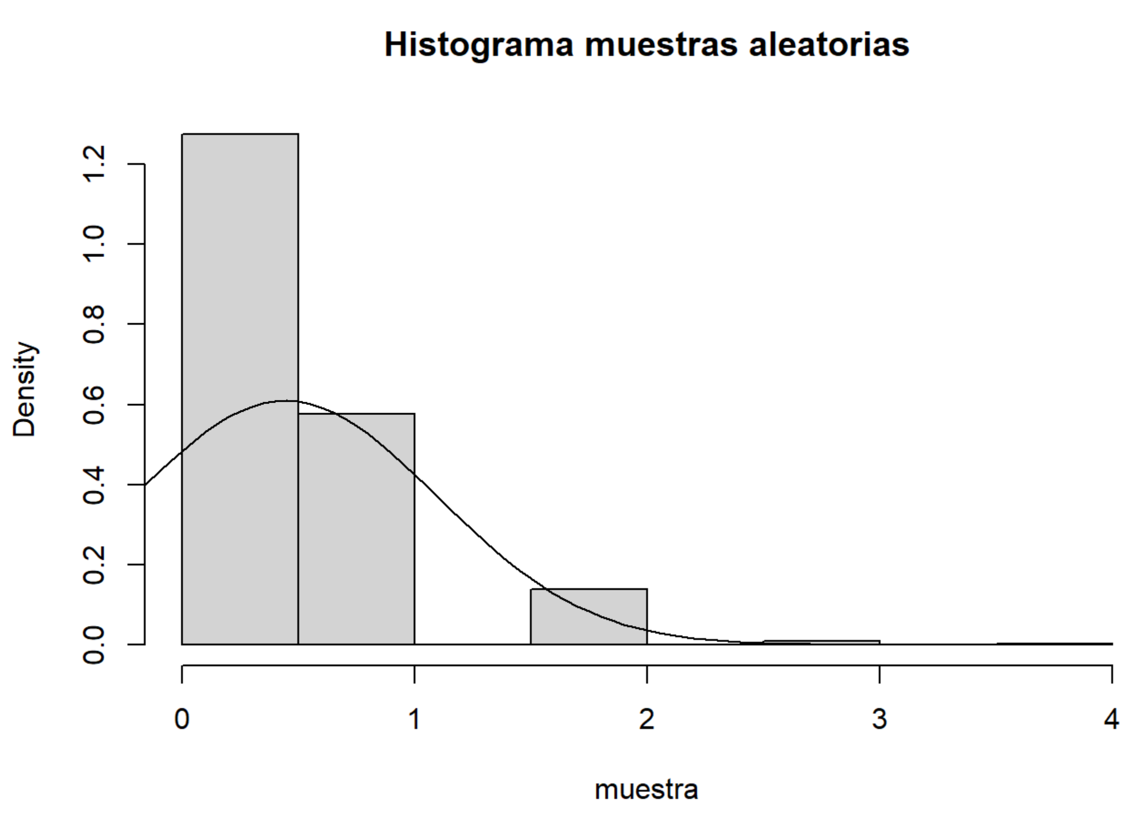
```
##
##  1-sample proportions test without continuity correction
##
## data:  15 out of 1000, null probability 0.5
## X-squared = 940.9, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.009110979 0.024600977
## sample estimates:
##     p
## 0.015
```

e. Cuantos intervalos esperas que contengan y cuantos verdaderamente contienen el verdadero % de transmisión de coronavirus.

El intervalo en el que estara nuestra media un 95% de las veces, es 0.009110979 0.024600977

f. Dibuja un histograma con los valores observados y superpón la curva normal teórica que se esperaría según el teorema central del límite.

```
media <- mean(muestra)
sd <- sd(muestra)
#barplot(table(muestra)/length(muestra),ylim=c(0,1.5))
hist(muestra, freq = FALSE, main= "Histograma muestras aleatorias")
curve(dnorm(x, media, sd),-1,4, add =TRUE)
```



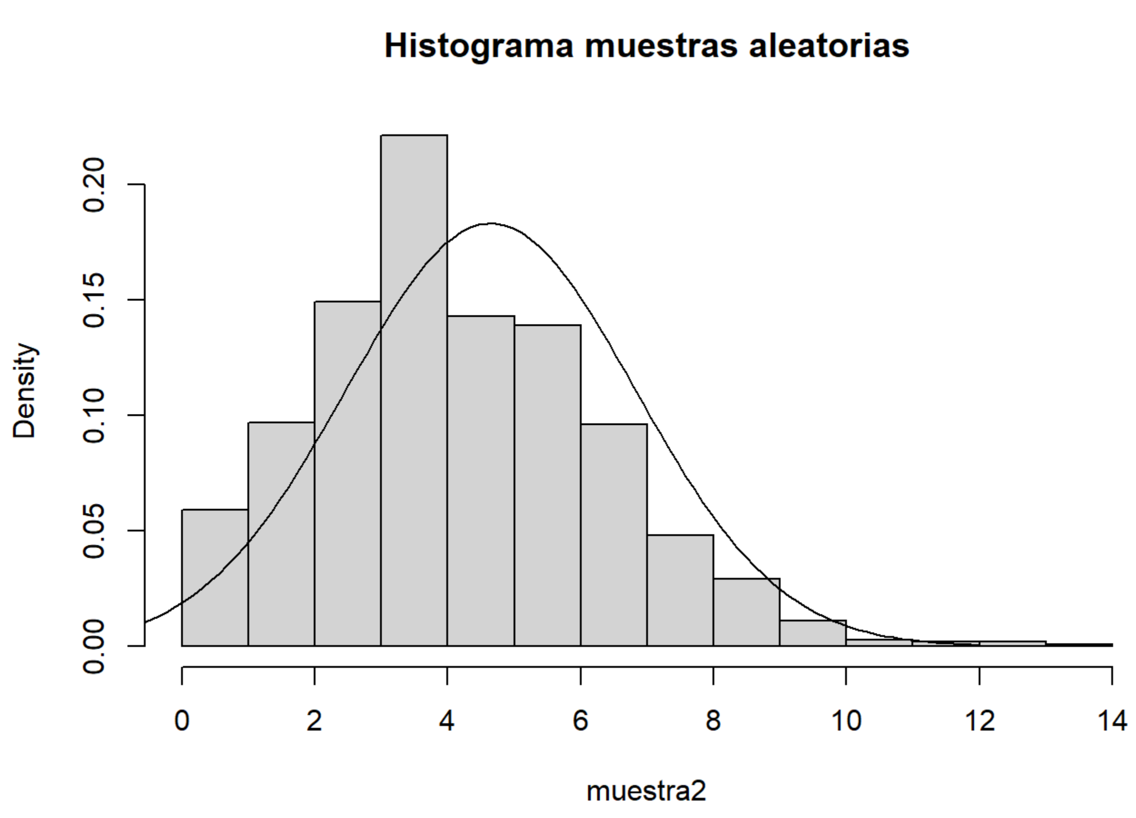
g. Si en lugar de ser la muestra de 15 fuera de 150 que esperarias que ocurri- era con respecto a los apartados anteriores (no es necesario que presentes los cálculos)

```
muestra2 <- rbinom(1000,150,0.031)
media2 <- mean(muestra2)
sd2 <- sd(muestra2)

#calcular intervalo de confianza al 95%
prop.test(x=150, n=1000, conf.level=0.95, correct = FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  150 out of 1000, null probability 0.5
## X-squared = 490, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.1292101 0.1734687
## sample estimates:
##    p
## 0.15
```

```
#histograma y curva normal
hist(muestra2, freq = FALSE, main= "Histograma muestras aleatorias")
curve(dnorm(x, media2, sd2),-1,12, add =TRUE)
```



La distribución binomial tiende a parecerse a una distribución normal cuando aumentamos el N. Esto podemos verlo cuando en el histograma y la curva de la muestra de N = 150.

EJERCICIO 3. El colesterol sérico es un factor de riesgo importante de enfermedad coronaria. Se puede demostrar que el colesterol sérico sigue un a distribución normal , con media = μ = 219mg / dL y desviación estándar = σ = 50mg / dL.

a. Si el rango clínicamente deseable para el colesterol es <200 mg / dL, ¿qué proporción de personas tienen clínicamente niveles deseables de colesterol?

```
#Mediante los valores acumulados de la curva normal, podemos calcular la
#proporción de observaciones que se encuentran antes o despues de la medida
#de una observación.
pnorm(200, mean=219,sd=50, lower.tail = TRUE)
```

```
## [1] 0.3519727
```

b. Algunos investigadores creen que solo el colesterol con niveles superiores a 250 mg / dL indican un riesgo suficientemente alto para justificar el tratamiento. ¿Qué proporción de la población representa este grupo?

```
pnorm(250, mean=219, sd=50, lower.tail = FALSE)
```

```
## [1] 0.2676289
```

c. ¿Qué proporción de la población general tiene niveles límite de colesterol alto, es decir,> 200 pero menors de <250 mg / dL?

```
props <-pnorm(c(200,250), mean=219, sd=50, lower.tail = TRUE)
lim_alto <- props[2] - props[1]
lim_alto
```

```
## [1] 0.3803984
```

d. Genera una muestra de 10 pacientes de la población. Estima la media de la muestra con su intervalo de confianza al 99%. Interpreta los resultados

```
set.seed(9991)
muestra_d <- rnorm(10, mean=219, sd=50)
t.test(muestra_d, conf.level = 0.99)
```

```
##
##  One Sample t-test
##
## data:  muestra_d
## t = 28.282, df = 9, p-value = 4.202e-10
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
##  195.6325 246.4296
## sample estimates:
## mean of x
##   221.031
```

e. Replica el apartado anterior con una muestra de 20.Comenta los resultados y justifica las diferencias con el apartado anterior.

```
muestra_e <- rnorm(20, mean=219, sd=50)
t.test(muestra_e, conf.level = 0.99)
```

```
##
##  One Sample t-test
##
## data:  muestra_e
## t = 24.252, df = 19, p-value = 9.333e-16
## alternative hypothesis: true mean is not equal to 0
## 99 percent confidence interval:
##  195.5007 247.7959
## sample estimates:
## mean of x
##  221.6483
```

Como hemos dicho en el ejercicio anterior, cuantas mayor sea N, mayor se asemejara nuestra función de distribución a una normal. Pero los valores de intervalos y medias obtenidas con 10 y 20 muestras son muy parecidadas.