

PEC 3 - MANUEL ROJAS GARCÍA

2023-11-12

Pregunta 1 (resolver con R). (3.5 puntos).

https://rstudio-pubs-static.s3.amazonaws.com/561295_15558c4800f542d3a268a3dd07fc69dc.html

Debido a la crisis energética y sanitaria, la aerolínea que estamos analizando decidió renovar los filtros de aire de los aviones, comprando unos nuevos en Amiexpress. Se ha notificado que estos filtros fallan a veces, pero por suerte solo ocurre en el 40% de las ocasiones. Si la aerolínea solo ha renovado 5 filtros:

¿Qué distribución normal se aproxima a esta distribución binomial?

Para poder obtener $N(\mu, \sigma)$ debemos hacer los siguientes calculos

```
# Filtros
n <- 5
# Probabilidad de fallo.
p <- 0.4

# Fórmula de cálculo de la media n*p
media <- n * p

# Cálculo de la varianza n*p(1 - p)
varianza <- (n * p * (1 - p))

# Cálculo de la desviación
desviacion <- sqrt(varianza)

cat("Tiene una distribución normal, N(", media, ", ", desviacion, ")")
## Tiene una distribución normal, N( 2 , 1.095445 )
```

b) Calcule $P(X < 1)$ mediante la aproximación normal de la binomial.

Para poder calcularla necesitamos tipificar la variable, para convertir nuestra distribución normal anterior en una distribución (0,1) con la siguiente formula. $Z = \frac{X - \mu}{\sigma}$

```
# Filtros
n <- 5
# Probabilidad de fallo.
p <- 0.4
```

```

# Fórmula de cálculo de la media n*p
media <- n * p

# Cálculo de la varianza n*p(1 - p)
varianza <- (n * p * (1 - p))

desviacion <- sqrt(varianza)

X <- 1

z <- (X - media) / desviacion

probabilidad <- pnorm(z)

# Sin cálculos directamente desde pnorm

prob.acu <- pnorm(1, mean =media, sd = desviacion)

cat("Mostramos ambos calculos y coinciden", probabilidad, prob.acu)

## Mostramos ambos calculos y coinciden 0.1806552 0.1806552

```

c) En este caso, ¿la aproximación normal es adecuada? Explique las razones.

Teniendo en cuenta que cuanto mayor sea el numero de n, la variable aleatoria $X(n)$ es aproximadamente normal. Nuestro n no es un valor muy elevado, teniendo en cuenta que 30 es el valor suficiente. Aún así hacemos los calculos.

$$n * p = 0,4 * n \quad n p(1-p) = n(0.4)*(0.6) = 0.24 n$$

Este será el parametro de la variable que se aproxima la distribucion $X(n)$. Así si n fuera lo suficientemente grande, $X(n)$ se comportaría como una $N(0.4n, 0.24n)$.

Si realizamos los calculos con $n = 5$ es correcto, pero no es un valor suficiente.

Otro calculo es cuando $n * p$ y $n(1-p)$ son mayores a 5. Como hemos calculado en otros apartados $n * p = 2$ y $1 - 0.4$ no son mayores a 5.

En conclusión, la aproximacion normal no es adecuada.

Pregunta 2 (resolver con R). (3.5 puntos).

La empresa aZigna-2 publica todos los años información relevante para mejorar los servicios de las aerolíneas. Este año ha publicado que los pasajeros de aviones de tipología Business recorren una distancia media vuelo, distribuida de manera normal, con una $\mu = 1670$ km y una $\sigma = 150$ km. Nuestra aerolínea de estudio trabaja con este tipo de modelo de aviones y opina que la media es mayor de 1670 km.

- a) Se desea realizar un estudio para estimar (con un nivel de confianza del 90%) la distancia media por vuelo de tipología Business con un margen de error de 10 km. Suponiendo que la desviación estándar de los vuelos de tipología Business es $\sigma = 150$ km, ¿qué tamaño de muestra debería utilizar para este estudio? (1.25 puntos).

#Creamos las variables con nuestros datos proporcionados

```
confianza <- 0.90
error <- 10
desviacion <- 150
```

Aplicamos la fórmula del margen de error con los datos conocidos y solo debemos despejar alpha medios.

*# $10 = z \alpha/2 * 150/\sqrt{n}$*

$z \alpha/2 = 1 + 0.90/2 = 0.95$

```
zetamedios = qnorm(0.95)
```

Sustituimos

*# $10 = 1.645 * 150 / \sqrt{n}$*

```
muestra = ((zetamedios * desviacion) / error)^2
```

```
cat("El valor es", muestra, "es decir, necesitamos 609")
```

```
## El valor es 608.7473 es decir, necesitamos 609
```

- b) Finalmente se llevó a cabo el estudio usando los datos proporcionados en el dataset data_pac3, ignorando la recomendación sobre el tamaño de la muestra. Construya un intervalo de confianza del 95% para la distancia media poblacional de los vuelos de tipología Business (Suponga normalidad y que σ es desconocida). (1.25 puntos). Nota: Los km vienen en la variable "Flight.Distance" y la tipología de vuelos Business corresponde con la variable "Class" y el nivel Business.

```
dataset <- read.csv("C:/Users/Manuel/Desktop/UOC/SEMESTRE 3 (Sep 2023 - Feb 2024)/Estadística/PEC 3/data_pac3.csv", sep = ",")
```

```
head(dataset)
```

```
##   ID Type.of.Travel   Class Flight.Distance
## 1  1      Business Business             821
## 2  2      Business Business             821
## 3  3      Business Business             853
```

```
## 4 4      Business Business      1905
## 5 5      Business Business      3470
## 6 6      Business Business      3788

datasetbusiness <- subset(dataset, Class == "Business")
mediamuestral <- mean(datasetbusiness$Flight.Distance)
cat("Media muestral =", mediamuestral, "n =", nrow(datasetbusiness))

## Media muestral = 1676.088 n = 62160
```

Web para encontrar los valores. <https://www.datanovia.com/en/lessons/how-to-do-a-t-test-in-r-calculation-and-reporting/>

#Aplicamos t test para calcular el intervalo de confianza

```
tstudent <- t.test(datasetbusiness$Flight.Distance, conf.level = 0.95)
```

utilizo parameter para comprobar que usa los grados de libertad adecuados y el valor de p

```
libertad <- tstudent$parameter
pvalor <- tstudent$p.value
```

#Los cálculos son correctos.

Para imprimir los valores de tstudent aplicamos los test que se indican en la web superior.

```
cat("Intervalo de confianza del 95% para la distancia media de vuelos
Business:",
    (tstudent$conf.int[1]), "a",
    (tstudent$conf.int[2]), "km.\n")
```

```
## Intervalo de confianza del 95% para la distancia media de vuelos
Business: 1667.153 a 1685.024 km.
```

¿Se podría rechazar la afirmación de la empresa aZigna-2 de que la distancia media de los pasajeros en vuelos con aviones de tipología Business es de 1670 km basándose en el intervalo de confianza que ha construido anteriormente?

La pregunta que indica el enunciado es que la empresa considerará que la media es mayor a 1670km y como se ha demostrado en el apartado anterior, la media está comprendida entre esos dos valores y 1670 está dentro. La media está bastante ajustada a la realidad.

#Pregunta 3 (resolver con R). (3 puntos).

Se están analizando los vuelos con una distancia recorrida estrictamente mayor de 1000 km (variable "Flight.Distance"). a) Encuentre un intervalo de confianza al 99%

para la proporción de vuelos con una distancia recorrida estrictamente mayor de 1000 km. Haga este apartado siguiendo las fórmulas de las notas de estudio, y usando R para hacer los diferentes cálculos. (1.5 puntos).

```
# Filtramos Los datos

dataset1000 <- subset(dataset, Flight.Distance > 1000)

# muestra (n) y La proporción muestral ( $\hat{p}$ )

n <- nrow(dataset)
p <- nrow(dataset1000)/n
z <- qnorm(0.995)

# Calculamos el error estandar

SE <- sqrt(p * (1 - p) / n)

# Calcular Los límites del intervalo de confianza
limiteinferior <- p - z * SE
limitesuperior <- p + z * SE

# Mostrar el resultado
cat("Intervalo de confianza del 99% para la proporción de vuelos con
distancia mayor de 1000 km:", round(limiteinferior, 3), "a",
round(limitesuperior, 3), "\n")

## Intervalo de confianza del 99% para la proporción de vuelos con
distancia mayor de 1000 km: 0.418 a 0.425

# Hacemos Los cálculos automaticamente y vemos que Los valores son
iguales.
Intervalo <- prop.test(x = nrow(dataset1000), n = n, conf.level = 0.99)
head(Intervalo)

## $statistic
## X-squared
## 3180.672
##
## $parameter
## df
## 1
##
## $p.value
## [1] 0
##
## $estimate
## p
## 0.4217508
```

```
##
## $null.value
##   p
## 0.5
##
## $conf.int
## [1] 0.4182214 0.4252883
## attr(,"conf.level")
## [1] 0.99
```

- b) Si queremos calcular el intervalo de confianza del apartado anterior con un nivel de confianza del 99% y con una precisión de 0.001, ¿qué tamaño de muestra necesitamos? (1.5 puntos).

```
confianza <- 0.99
precision <- 0.001
z <- qnorm((1 + confianza) / 2)

proporcionanterior <- p

#Despejamos n de la formula

muestra <- (z^2 * proporcionanterior * (1 - proporcionanterior)) /
precision^2

cat("El tamaño de muestra necesario es:", muestra)

## El tamaño de muestra necesario es: 1618099
```