PEC3

Paula Corbatón Álvarez

2022-04-06

```
attach(data)
```

Una vez importados los datos, con la misma base de datos y suponiendo que los datos corresponden a una muestra:

PREGUNTA 1

Encontrar un intervalo de confianza para la media de la concentración de partículas de PM10 del año 1999, PM10Concentration1999, con un nivel de confianza del 95% para:
a) (10%) Países con nivel de ingresos bajo, Low income.

```
splitByIncome<- split(PM10Concentration1999, f= IncomeGroup)
str(splitByIncome)
```

```
## List of 4
##  $ High income         : num [1:1095] 33 16 16 19 25 19 11 16 22 22 ...
##  $ Low income          : num [1:99] 46 46 51 60 32 81 91 40 136 97 ...
##  $ Lower middle income: num [1:756] 125 77 120 137 174 86 154 160 151 106 ...
##  $ Upper middle income: num [1:1268] 32 79 87 83 72 78 98 50 73 88 ...
```

```
PM10_LowIncome<- t.test(splitByIncome$'Low income', conf.level=0.95)

PM10_LowIncome$conf.int
```

```
## [1] 59.50163 75.04382
## attr(,"conf.level")
## [1] 0.95
```

Como podemos observar, el intervalo de confianza al 95% para la media en los paises con un nivel de ingresos bajo es (59.50163, 75.04382).

b) (10%) Países con nivel de ingresos alto, High income.

```
PM10_HighIncome<- t.test(splitByIncome$'High income', conf.level=0.95)
PM10_HighIncome$conf.int
```

```
## [1] 28.31483 30.11439
## attr(,"conf.level")
## [1] 0.95
```

Así, el intervalo de confianza al 95% para la media en los paises con un nivel de ingresos alto es (28.31483, 30.11439)

c) (10%) ¿Qué conclusión podemos extraer sobre la concentración de partículas PM10 para los dos niveles de ingresos? (en particular fijaos en las medias y los intervalos de confianza).

```
PM10_LowIncome
```

```
##
##  One Sample t-test
##
## data:  splitByIncome$"Low income"
## t = 17.179, df = 98, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  59.50163 75.04382
## sample estimates:
## mean of x
##  67.27273
```

```
PM10_HighIncome
```

```
##
##  One Sample t-test
##
## data:  splitByIncome$"High income"
## t = 63.708, df = 1094, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  28.31483 30.11439
## sample estimates:
## mean of x
##  29.21461
```

En paises con un alto numero de ingresos la concentracion por metros cúbicos es menor que en paises con pequeño numero de ingresos.
También podemos observar que estos últimos presentan una mayor variabilidad de medidas en comparación con los paises con un alto número de ingresos

d) (20%) Si queremos calcular el intervalo de confianza con un nivel del 95% para la media de la concentración de partículas PM10 del año 1999 para países con ingresos altos, de forma que el margen de error sea inferior a 0.1, calculad el tamaño mínimo de la muestra. Usad la información de la muestra trabajada en el apartado (b) como información previa. En este apartado, usad el R para las operaciones y cálculo del valor crítico.

```
z <- qnorm (0.025, mean = 0, sd = 1, lower.tail = FALSE)
σ <- sd(splitByIncome$'High income')
e <- 0.1

n <- (z*σ/e)^2
print(paste("La muestra tendrá que tener un tamaño de ", ceiling(n)))
```

```
## [1] "La muestra tendrá que tener un tamaño de  88456"
```

PREGUNTA 2

Queremos estudiar la proporción de ciudades con una concentración de PM10 superior a 50μg/m3; para hacerlo:

a) (10%) Calculad un intervalo de confianza del 90% para dicha proporción mediante la función prop.test con la opción correct=FALSE.

```
nrow(data)
```

```
## [1] 3218
```

```
sum(PM10Concentration1999>50)
```

```
## [1] 1182
```

```
prop.test(1182, 3128, conf.level = .95, correct=FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  1182 out of 3128, null probability 0.5
## X-squared = 186.6, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.3610454 0.3950086
## sample estimates:
##         p
## 0.3778772
```

El intervalo de confianza es: (0.3610454, 0.3950086)

b) (10%) Calculad el mismo intervalo siguiendo las fórmulas de las notas de estudio (puede haber pequeñas diferencias en el resultado). Usad R para hacer las operaciones y para calcular el valor crítico.

```
z <- qnorm (0.05, mean = 0, sd = 1, lower.tail = FALSE)
n <- 3128
p <- 1182/n

li <- p-z*sqrt((p*(1-p))/n)
ls <- p+z*sqrt((p*(1-p))/n)
print(paste("el intervalo de confianza para la proporción es:", li, ",", ls))
```

```
## [1] "el intervalo de confianza para la proporción es: 0.363617639385707 , 0.392136
836317618"
```

Si queremos calcular el intervalo de confianza igual que en el apartado anterior (90%) pero con un margen de error inferior a 0.02, calculad cuál tiene que ser el tamaño mínimo de la muestra sí:

c) (15%) Tomamos la muestra trabajada en el apartado anterior, (b) como información previa.

```
e <- 0.02

tamañoMuestra1 <- z^2 * (p*(1-p)/e^2)

print(paste("necesitamos una muestra de tamaño ", ceiling(tamañoMuestra1)))
```

```
## [1] "necesitamos una muestra de tamaño  1591"
```

d) (15%) No tenemos ninguna información previa. Comparad y comentad el resultado con el obtenido en el apartado (c).

Para resolver este problema, nos colocaremos en la situación más desfavorable posible en cuanto a los valores de la proporción poblacional. Por lo tanto, p=q=0.5

```
p <- 0.5

tamañoMuestra2<- z^2 * (p*(1-p)/e^2)
 print(paste("necesitamos una muestra de tamaño ", ceiling(tamañoMuestra2)))
```

```
## [1] "necesitamos una muestra de tamaño  1691"
```

Como podemos comprobar al comparar los resultados, necesitamos una muestra de mayor tamaño en el apartado d. Al establecer p=q=0.5, estamos calculando un tamaño de la muestra para una variable menos precisa y por tanto una misma amplitud necesitará un mayor tamaño de muestra que el realizado con otra p.