PAC4 Primavera 2023 - Solución

UOC

Las PECs se basarán en una base de datos obtenida a partir del repositorio de microdatos del "Banco Mundial" en https://microdata.worldbank.org/index.php/catalog/424/get-microdata

Contiene indicaciones, entre otros de

1. *City* = Nombre de la ciudad
2. *Country* = País
3. *Population2000* = Población de la ciudad en 2000.
4. *PM10Concentration1999* = "PM10 concentrations (micro gramos por cubic meter) in residential areas of cities larger than 100,000", en 1999
5. *Region* = Clasificación en región geográfica
6. *IncomeGroup* = Clasificación según nivel de ingresos del país.

Para importar los datos podemos usar la siguiente instrucción:

```
dadesPM10<-read.table("AirPollution2000WB_UOC2.csv", header=TRUE,
          sep=";",na.strings="NA",
          fileEncoding = "UTF-8", quote = "\"",
          colClasses=c(rep("character",4),rep("numeric",2),
                       rep("character",2)))
```

Hay que entregar la práctica en forma de fichero pdf (exportando el resultado final a pdf por ejemplo) en esta misma tarea Moodle; no hay que entregarla en el registro de EC.

Os puede ser útil consultar el siguiente material:

1. Módulo de Intervalos de confianza.
2. Actividades resueltas del Reto 3 (Intervalos de confianza).
3. Procurad utilizar las funciones propias de R para hacer los cálculos a no ser que se diga lo contrario.

NOMBRE:

PEC4

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

Pregunta 1 (50%)

a) (10%) Encontrad un intervalo de confianza para la media de la concentración de partículas PM10 del año 1999 *(PM10Concentration1999)* con un nivel del 95% para las ciudades de Canadá.

b) (10%) Encontrad un intervalo de confianza para la media de la concentración de partículas PM10 del año 1999 *(PM10Concentration1999)* con un nivel del 95% para las ciudades que **no** están en Canadá.

c) (10%) ¿Qué conclusión podemos extraer sobre la concentración de partículas PM10 en base a los intervalos encontrados en los apartados anteriores?

d) (20%) Otro estudio afirma que la variable que estamos estudiando sigue una distribución normal con desviación típica 5 microgramos por m^3. Usando este hecho, ¿cuántas ciudades aproximadamente de Canadá tendría que haber en la muestra para tener un intervalo de confianza de la concentración de partículas PM10 del año 1999 *(PM10Concentration1999)* en las ciudades de Canadá con un nivel del 95% y que tenga una longitud de 1 unidad?

Solución

a) Calculamos

```r
options(scipen=999)  # Desactiva la notación científica
attach(dadesPM10)
tuc<-t.test(PM10Concentration1999[Country=="Canada"],
           conf.level = 0.95)
tuc
```

```
##
##   One Sample t-test
##
## data:  PM10Concentration1999[Country == "Canada"]
## t = 29.664, df = 47, p-value < 0.00000000000000022
## alternative hypothesis: true mean is not equal to 0
```

```
## 95 percent confidence interval:
##  19.73118 22.60215
## sample estimates:
## mean of x
##  21.16667
```

El intervalo de confianza pedido será: $(19.7312, 22.6022)$.

b) Calculamos

```
tncan<-t.test(PM10Concentration1999[Country!="Canada"],
          conf.level = 0.95)
tncan
```

```
##
##   One Sample t-test
##
## data:  PM10Concentration1999[Country != "Canada"]
## t = 75.89, df = 3169, p-value < 0.0000000000000022
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  50.22228 52.88624
## sample estimates:
## mean of x
##  51.55426
```

El intervalo de confianza pedido será: $(50.2223, 52.8862)$.

c) En este caso los intervalos son disjuntos y se ve que la concentración de partículas en las ciudades de Canadá es mucho más baja que en resto del mundo (de hecho menos de la mitad).

d) El margen de error tendrá que ser $ME = 0.5$ y podemos aplicar la fórmula

$$n \geq {z_{\alpha/2}}^2 \frac{\sigma^2}{ME^2}$$

Hay que calcular pues

```
za<-abs(qnorm(0.025))
za
```

```
## [1] 1.959964
```

```
ME<-0.5
desv<-5
```

y podemos considerar

$$1.959964^2 \frac{5^2}{0.5^2} = 384.1458821 \approx 385$$

y por lo tanto necesitaríamos aproximadamente ocho veces más ciudades de Canadá que las que tenemos ahora a la muestra.

Pregunta 2 (50%)

Queremos estudiar la proporción de ciudades correspondientes a diferentes niveles de ingresos.

a) (10%) Calculad un intervalo de confianza del 85% para la proporción de ciudades que están en países de nivel "Upper middle income" mediante la función *prop.test* con la opción *correct=FALSE*.

b) (10%) Calculad un intervalo de confianza del 85% para la proporción de ciudades que están en países de nivel "High income" mediante la función *prop.test* con la opción *correct=FALSE*.

c) (10%) En base en los apartados anteriores, ¿podemos decir que las proporciones son diferentes? Razonad vuestra respuesta. Comparad también los intervalos obtenidos.

d) (20%) Encontrad ahora un intervalo de confianza del 40% para la proporción de ciudades que están en países de nivel "Upper middle income" mediante la función *prop.test* con la opción *correct=FALSE*. Comparadlo con el obtenido en el apartado a) y explicad cual tiene una longitud mayor y el porqué.

Solución

a) Para usar la instrucción *prop.test*, primero tenemos que calcular el tamaño de la muestra y el número de observaciones de la muestra correspondientes a ciudades "Upper middle income"

```
len<-length(PM10Concentration1999)
len
```

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

```r
lenUM<-length(IncomeGroup[IncomeGroup=="Upper middle income"])
lenUM
```

```
## [1] 1268
```

```r
propUM<-prop.test(lenUM, len, alternative='two.sided', p=0.5,
              conf.level=0.85,correct=FALSE)
propUM
```

```
##
##   1-sample proportions test without continuity correction
##
## data:  lenUM out of len, null probability 0.5
## X-squared = 144.54, df = 1, p-value < 0.00000000000000022
## alternative hypothesis: true p is not equal to 0.5
## 85 percent confidence interval:
##  0.3817056 0.4064979
## sample estimates:
##         p
## 0.3940336
```

El intervalo de confianza pedido será: $(0.3817, 0.4065)$.

b) Como en el apartado anterior

```r
lenHI<-length(PM10Concentration1999[IncomeGroup=="High income"])
lenHI
```

```
## [1] 1095
```

```r
propHI<-prop.test(lenHI, len, alternative='two.sided', p=0.5,
               conf.level=0.85,correct=FALSE)
propHI
```

```
##
##   1-sample proportions test without continuity correction
##
## data:  lenHI out of len, null probability 0.5
## X-squared = 328.4, df = 1, p-value < 0.00000000000000022
## alternative hypothesis: true p is not equal to 0.5
## 85 percent confidence interval:
##  0.3283564 0.3523961
## sample estimates:
##         p
## 0.3402735
```

El intervalo de confianza pedido será: $(0.3284, 0.3524)$.

c) En este caso los intervalos son disjuntos y podemos concluir que la proporción de ciudades correspondientes a ingresos altos es inferior a la de ciudades correspondientes a ingresos medianos-altos.

d) Podemos hacer directamente

```
propUM2<-prop.test(lenUM, len, alternative='two.sided', p=0.5,
              conf.level=0.4,correct=FALSE)
propUM2
```

```
##
##   1-sample proportions test without continuity correction
##
## data:  lenUM out of len, null probability 0.5
## X-squared = 144.54, df = 1, p-value < 0.00000000000000022
## alternative hypothesis: true p is not equal to 0.5
## 40 percent confidence interval:
##  0.3895257 0.3985595
## sample estimates:
##         p
## 0.3940336
```

El intervalo de confianza pedido será: $(0.3895, 0.3986)$ y es más pequeño que en el apartado a) porque la probabilidad que contengan la proporción es también más pequeña. En el apartado a) la longitud del intervalo es 0.0247922 y ahora es 0.0090339 y es aproximadamente la tercera parte.