

PRACTICA 7: intervalos de confianza. Enunciados.

Estadística, Grado en Biología, UAH, 2022/23.

Objetivos

- Determinar las condiciones de aplicabilidad de los intervalos de confianza y el tipo de intervalo a utilizar en cada situación.
- Construir los intervalos de confianza para la media, la varianza y la proporción.
- Interpretar los resultados obtenidos.

Los datos que contiene el fichero `Practica07-ciudades.csv` se refieren (a excepción de la última columna) a 41 ciudades de Estados Unidos y fueron extraídos de distintas revistas publicadas por el gobierno de este país durante los años 1969 y 1970. Las variables que aparecen son:

- CIUDAD (nombre de la ciudad).
- SO2 (contenido de SO2 en el aire medido en microgramos por metro cúbico).
- TEMP (promedio de la temperatura anual en grados Fahrenheit).
- MAN (número de empresas de manufacturación con 20 o más empleados).
- POP (tamaño demográfico en decenas de miles según el censo de 1970).
- VIENTO (promedio de la velocidad del viento anual en millas por hora)
- LLUVIA (promedio de la precipitación anual en pulgadas).
- DIASLLUVIA (promedio del número de días que llueve al año).
- CLIMA (variable categórica que especifica el tipo de clima).
- SO2_tr (la variable SO2 transformada adecuadamente).
- En la última columna se añaden datos ficticios correspondientes a una posible evolución de la variable SO2, en el año 1975; el nombre que recibe esta variable, en el fichero, es SO2tr_75.

Empieza por leer el fichero `Practica07-ciudades.csv` y guardar su contenido en la variable `ciudades`. Puedes usar la función de R `read.table()` (no olvides fijar correctamente la carpeta de trabajo)

```
ciudades = read.table(file = "Practica07-ciudades.csv", sep = ";", header = TRUE, dec = ".")
```

o bien usar el botón **Import dataset** de RStudio.

Visualiza las primeras filas de la tabla

```
head(ciudades, 4)

##          CIUDAD SO2 TEMP MAN POP VIENTO LLUVIA DIASLLUVIA      CLIMA  SO2_tr
## 1      Phoenix  10  70.3 213 582    6.0   7.05         36 continental 2.302585
## 2  LittleRock  13  61.0  91 132    8.2  48.52        100 subtropical 2.564949
## 3 San Francisco 12  56.7 453 716    8.7  20.66         67 subtropical 2.484907
## 4      Denver  17  51.9 454 515    9.0  12.95         86 continental 2.833213
## SO2_tr75
## 1  4.68068
## 2  4.15766
## 3  6.47724
## 4  2.03278
```

Responde de forma concisa y razonadamente a las siguientes preguntas sobre la variable $SO2_{tr}$:

1. Estima por intervalos de confianza al 95% de confianza la media poblacional.
2. Estima por intervalos de confianza al 95% de confianza la desviación típica poblacional.
3. Si quisiéramos estimar el valor medio de $SO2_{tr}$ al nivel de confianza del 95% con una precisión de 0.1, ¿qué tamaño de muestra necesitaríamos?
4. Se considera que valores de $SO2_{tr}$ superiores a 2.90 entrañan un riesgo grave para la salud. ¿Dirías, a un 95% de confianza, que es el caso de las ciudades americanas? ¿Se mantiene la misma conclusión al 99% de confianza?
5. Con un nivel de confianza del 95%, ¿qué porcentaje de ciudades verifica $SO2_{tr} > 2.90$?
6. Por un lado, en el ejercicio 4 se tiene que con un nivel de confianza del 95% el valor medio poblacional de $SO2_{tr}$ supera el umbral de 2.9. Por otro lado, del ejercicio 5 se desprende que entre el 40% y el 70% de las poblaciones superan dicho umbral con el mismo nivel de confianza. ¿Resultan contradictorios ambos resultados?
7. Calcula un intervalo de confianza del 92% para la temperatura media de las ciudades con clima continental húmedo. ¿Cabía esperar, un año cualquiera, que la temperatura media rondara los 50 grados Fahrenheit? ¿Y los 55?
8. En el primer ejercicio hemos averiguado que la estimación puntual de la media de $SO2_{tr}$ es

```
[1] 3.153004
```

Del intervalo de confianza calculado se deduce que la precisión de la estimación (la semi anchura del intervalo) al nivel de confianza del 95% es

```
[1] 0.21497
```

En el ejercicio 3 se pregunta cuánto hay que aumentar el tamaño de la muestra para mejorar la precisión hasta 0.1 unidades. Esa estrategia implica aumentar la muestra desde 200 hasta 190 individuos, lo que implica consumir muchos recursos.

Otra opción es la siguiente: reducir el nivel de confianza reduce el tamaño del intervalo, lo que aumenta la precisión (ojo, a costa de reducir el nivel de certidumbre).

Para explorar esa vía (redondeando valores para que salgan números más limpios) supón que la media muestral de la variable $SO2_{tr}$ es $\bar{X} = 3.3$, $s = 0.7$ y $n = 41$. Considera el intervalo de confianza $IC_\mu = (3.2, 3.4)$ (es decir, tiene precisión 0.1, pero has mantenido la muestra en 41 elementos) ¿cuál es el nivel de confianza de este intervalo?