Repaso de intervalos y estimación puntual

Curso completo de estadística inferencial con R y Python

Contents

1 Estimación puntual

Sea X_1 y X_2 una muestra aleatoria de dos observaciones independientes de una población con media μ y varianza σ^2. Considerar los siguientes tres estimadores puntuales de μ:

$$\begin{aligned}\overline{X} &= \frac{1}{2}X_1 + \frac{1}{2}X_2\\ \hat{\mu}^{(1)} &= \frac{1}{4}X_1 + \frac{3}{4}X_2\\ \hat{\mu}^{(2)} &= \frac{1}{3}X_1 + \frac{2}{3}X_2\end{aligned}$$

1. Probar que los tres estimadores son insesgados.
2. ¿Cuál de los tres estimadores es más eficiente?
3. Hallar la eficiencia relativa de $\overline{X}$ con respecto a los otros estimadores.

1.1 Solución: Estimación puntual

1. Para probar que los tres estimadores son insesgados calculemos sus esperanzas:

- $E\left(\overline{X}\right) = E\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) = \frac{1}{2}\mu + \frac{1}{2}\mu = \mu.$
- $E\left(\hat{\mu}^{(1)}\right) = E\left(\frac{1}{4}X_1 + \frac{3}{4}X_2\right) = \frac{1}{4}\mu + \frac{3}{4}\mu = \mu.$
- $E\left(\hat{\mu}^{(2)}\right) = E\left(\frac{1}{3}X_1 + \frac{2}{3}X_2\right) = \frac{1}{3}\mu + \frac{2}{3}\mu = \mu.$
  Por lo tanto los tres estimadores son insesgados para el parámetro μ.

2. ¿Cuál de los tres estimadores es más eficiente? Calculemos las varianzas de cada estimador. Como las variables son independientes la varianza de la suma es la suma de las varianzas

- $Var\left(\overline{X}\right) = Var\left(\frac{1}{2}X_1 + \frac{1}{2}X_2\right) = \frac{1}{4}\sigma^2 + \frac{1}{4}\sigma^2\right) = \frac{1}{2}\cdot\sigma^2.$
- $Var\left(\hat{\mu}^{(1)}\right) = Var\left(\frac{1}{4}X_1 + \frac{3}{4}X_2\right) = \frac{1}{16}\sigma^2 + \frac{9}{16}\sigma^2 = \frac{5}{8}\cdot\sigma^2.$
- $Var\left(\hat{\mu}^{(2)}\right) = Var\left(\frac{1}{3}X_1 + \frac{2}{3}X_2\right) = \frac{1}{9}\sigma^2 + \frac{4}{9}\sigma^2\right) = \frac{5}{9}\cdot\sigma^2.$
  Es evidente que $\frac{1}{2} < \frac{5}{9} < \frac{5}{8}$, por lo que el estimador mas eficiente, como ya sabíamos por teoría, es $\overline{X}$.

3. Hallar la eficiencia relativa de $\overline{X}$ con respecto a los otros estimadores son:

- $\frac{Var\left(\overline{X}\right)}{Var\left(\hat{\mu}^{(1)}\right)} = \frac{\frac{\sigma^2}{2}}{\frac{5}{8}\sigma^2} = \frac{8}{10} = 0.8$
- $\frac{Var\left(\overline{X}\right)}{Var\left(\hat{\mu}^{(2)}\right)} = \frac{\frac{\sigma^2}{2}}{\frac{5}{9}\sigma^2} = \frac{9}{10} = 0.9$
  Lo que se interpreta como que, respecto a su varianza la media aritmética tiene un 80% y un 90% de la varianza de los otros dos estimadores.

2 Intervalos de confianza

En una clase de estadística asisten estudiantes del Grado de Matemáticas y de Ingeniería Telemática. En una muestra de 18 estudiantes se observaron las siguientes calificaciones en el examen final

$$62, 57, 85, 59, 64, 63, 71, 58, 77, 72, 73, 79, 85, 73, 62, 51, 60, 57$$

1. Utilizar un método de estimación insesgado para obtener una estimación puntual de las calificación media de los estudiantes.
2. Utilizad un método de estimación insesgado para obtener una estimación puntual de la proporción poblacional de estudiantes que obtuvieron una calificación mayor que 70.
3. Calculad manualmente y con `R` un intervalo de confianza al 95% para la nota media (suponiendo población normal). Interpretad el resultado
4. Calculad con R un intervalo de confianza del 95% para la proporción poblacional de notas mayores de 70.

Indicación: utilizad si queréis el código siguiente comentando que es es lo que hace cada linea. Interpretad el resultado.

```
data=c(62,57,85,59,64,63,71,58,77,72,73,79,85,73,62,51,60,57)
length(data)
```

```
## [1] 18
```

```
library(epitools)
binom.approx(sum(data>70),length(data),0.95)
```

```
##   x  n proportion     lower     upper conf.level
## 1 8 18  0.4444444 0.2148907 0.6739982       0.95
```

```
binom.exact(sum(data>70),length(data),0.95)
```

```
##   x  n proportion     lower     upper conf.level
## 1 8 18  0.4444444 0.2153015 0.6924283       0.95
```

```
binom.wilson(sum(data>70),length(data),0.95)
```

```
##   x  n proportion     lower     upper conf.level
## 1 8 18  0.4444444 0.2455952 0.6628358       0.95
```

2.1 Solución: Intervalos de confianza

Explicamos primero el código de `R`

```
data=c(62,57,85,59,64,63,71,58,77,72,73,79,85,73,62,51,60,57)
n=length(data)
n
```

```
## [1] 18
```

De estos resultados obtenemos que el tamaño de la muestra es $n = 18$.

```
library(epitools)
```

Cargamos la librería `epitools`.

```
binom.approx(sum(data>70),length(data),0.95)
```

```
##   x  n proportion     lower     upper conf.level
## 1 8 18  0.4444444 0.2148907 0.6739982       0.95
```

Esta función calcula el intervalo de confianza aproximado por la fórmula de Laplace

```r
binom.exact(sum(data>70),length(data),0.95)
```

```
##   x  n proportion     lower     upper conf.level
## 1 8 18  0.4444444 0.2153015 0.6924283       0.95
```

Esta función calcula el intervalo de confianza exacto siguiendo una distribución binomial por la fórmula de Laplace

```r
binom.wilson(sum(data>70),length(data),0.95)
```

```
##   x  n proportion     lower     upper conf.level
## 1 8 18  0.4444444 0.2455952 0.6628358       0.95
```

Esta función calcula el intervalo de confianza aproximado de Wilson.

1. Esta estimación no está en el código. El estimador insesgado de la media poblacional de notas es la media aritmética. Así que la calculamos manualmente

$$\begin{aligned}\overline{x} &= \frac{62+57+85+59+64+63+71+58+77+72+73+79+85+73+62+51+60+57}{18}\\ &= \frac{1208}{18} = 67.11111\end{aligned}$$

2. Como se ve en cada salida de los resultados del código se estima la proporción muestral, que es el estimador insesgado de la proporción poblacional. Notemos que la entrada es `sum(data>70)` que nos da el número de observaciones mayores que 70 en la muestra. Por lo tanto la respuesta es que el estimador insesgado es

$$\hat{p} = \frac{8}{18} = 0.4444$$

.

3. Ahora necesitamos calcular la desviación típica muestral. Calculemos primero la media de los cuadrados de la muestra (de forma manual)

$$\begin{aligned}\frac{\sum_i^n x_i^2}{n} &= \frac{62^2+57^2+85^2+59^2+64^2+63^2+71^2+58^2+77^2+72^2+73^2+79^2+85^2+73^2+62^2+51^2+60^2+57}{18}\\ &= \frac{82800}{18} = 4600\end{aligned}$$

por lo tanto la varianza muestral es

$$S^2 = 4600 - 67.1111^2 = 96.0989.$$

Ahora calculamos la cuasivarianza $\tilde{S}^2$

$$\tilde{S}^2 = \frac{18}{18-1}\cdot S^2 = 90.7600722 \text{ y } \tilde{S} = +\sqrt{101.7517765} = 10.08721$$

Sabemos, que bajo las condición de normalidad y σ desconocida el intervalo de confianza del $(1-\frac{\alpha}{2})\cdot 100\%$ es

$$\left]\overline{x} - t_{n-1,1-\frac{\alpha}{2}}\cdot\frac{\tilde{S}}{\sqrt{n}}, \overline{x} + t_{n-1,1-\frac{\alpha}{2}}\cdot\frac{\tilde{S}}{\sqrt{n}}\right[.$$

Donde $t_{n-1}1-\frac{\alpha}{2}$ es el cuantil $1-\frac{\alpha}{2}$ de una distribución t de Student con $n-1$ grados de libertad. En nuestro caso $n = 18$ y $1-\alpha = 0.95$ por lo tanto, consultando la tabla de cuantiles de la t de Student

$$t_{n-1,1-\frac{\alpha}{2}} = t_{17,0.975} \approx 2.109816$$

Sustituyendo en la fórmula del intervalo

$$\left]67.11111 - 2.109816 \cdot \frac{10.08721}{\sqrt{18}}, 67.11111 + 2.109816 \cdot \frac{10.08721}{\sqrt{18}}\right[= \left]62.0949, 72.1274\right[.$$

Por lo tanto y con un nivel de confianza del 95% la media poblacional de las notas se encontrará (redondeando) entre 62.01 y 72.13 puntos.

Con calculos más precisos con `R`:

```r
mean(data)
```

```
## [1] 67.11111
```

```r
sd(data)
```

```
## [1] 10.0872
```

```r
qt(0.975,17)
```

```
## [1] 2.109816
```

```r
c(mean(data)-qt(0.975,17)*sd(data)/sqrt(18),mean(data)+qt(0.975,17)*sd(data)/sqrt(18))
```

```
## [1] 62.09486 72.12736
```

4. Para el intervalo de las proporciones sí utilizamos el código. Como la muestra es pequeña no es mayor que 100 no podemos utilizar Laplace y al no ser mayor que 40 no podemos utilizar Wilson. Así que el resultado debe de ser el intervalo exacto o de Clopper-Pearson que se calcula con la siguiente función

```r
sum(data>70)
```

```
## [1] 8
```

```r
binom.exact(sum(data>70),length(data),0.95)
```

```
##   x  n proportion     lower     upper conf.level
## 1 8 18  0.4444444 0.2153015 0.6924283       0.95
```

Por lo tanto la proporción poblacional de notas mayores de 70 se encontrará (redondeando) entre el 21.53% y el 69.24% con un nivel de confianza del 95%.