Estadística (M)
PRIMER TRABAJO PRÁCTICO
Un análisis mediante Bootstrap de datos

Miembros del grupo:
Cantidad Total de Hojas:

La resolución del TP incluirá

1. *El script que permite efectuar los cálculos que deberá enviarse por mail para verificar el programa.*
2. *Un documento con las respuestas a las preguntas planteadas que puede entregarse por alguna de las siguientes vías*
   - *Personalmente, la copia impresa que incluya todos los gráficos necesarios.*
   - *Por mail, enviando un doc, LATEX o pdf que contenga los gráficos necesarios.*

- Justifique todas sus respuestas -

Sean $X_1, \ldots, X_n$ v.a. i.i.d. con distribución $N(\mu, \sigma^2)$. Se desea estimar el coeficiente de variación $\theta = \mu/\sigma$, es decir, el cociente entre la media y el desvío .

1. Hallar el estimador $\widehat{\theta}_{\mathrm{MV}}$ de máxima verosimilitud de θ.
2. Sea $\widehat{\sigma}^2 = (1/n)\sum_{i=1}^n (X_i - \overline{X})^2$. Muestre que
$$\sqrt{n}\left(\left(\begin{array}{c} \overline{X} \\ \widehat{\sigma}^2 \end{array}\right) - \left(\begin{array}{c} \mu \\ \sigma^2 \end{array}\right)\right) \xrightarrow{D} N\left(\mathbf{0}, \left(\begin{array}{cc} \sigma^2 & 0 \\ 0 & 2\sigma^4 \end{array}\right)\right)$$
   Sugerencia: Defina $\widehat{\sigma}_0^2 = (1/n)\sum_{i=1}^n (X_i - \mu)^2$ y muestre que $\widehat{\sigma}^2 = \widehat{\sigma}_0^2 + (\overline{X} - \mu)^2$. Deduzca de allí que $\sqrt{n}\left(\widehat{\sigma}^2 - \sigma^2\right) = \sqrt{n}\left(\widehat{\sigma}_0^2 - \sigma^2\right) + o_{\mathbb{P}}(1)$ lo que le permite reducir el problema.
3. Deducir de 2) que
$$\sqrt{n}(\widehat{\theta}_{\mathrm{MV}} - \theta) \xrightarrow{D} N\left(0, 1 + \frac{\theta^2}{2}\right)$$
   Sugerencia: Use la siguiente generalización del resultado visto en el caso unidimensional.

   Sean $\mathbf{W}_1, \ldots, \mathbf{W}_n, \ldots$ *vectores aleatorios* $\mathbf{W}_i \in \mathbb{R}^2$, $G : \mathbb{R}^2 \to \mathbb{R}$ *continuamente diferenciable en* $\boldsymbol{\nu} \in \mathbb{R}^2$. *Indiquemos por* $\nabla G(\mathbf{w})$ *el vector de derivadas parciales de* G *en* $\mathbf{w}$.
   Si $\sqrt{n}(\mathbf{W}_n - \boldsymbol{\nu}) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{\Sigma})$ *y* $\upsilon = \nabla G(\boldsymbol{\nu})^{\mathrm{T}} \boldsymbol{\Sigma} \nabla G(\boldsymbol{\nu}) \neq 0$, *entonces,*
$$\sqrt{n}(G(\mathbf{W}_n) - G(\boldsymbol{\nu})) \xrightarrow{D} N(0, \upsilon)$$

4. Si X_i no tiene distribución normal pero $\mathbb{E}(X_1^4) < \infty$, se sabe que

$$\sqrt{n}\left(\widehat{\theta} - \theta\right) \xrightarrow{D} N(0, \upsilon)$$

donde

$$\widehat{\theta} = \frac{\overline{X}}{\widehat{\sigma}} \qquad \upsilon = 1 + \theta\,\mathbb{E}(Z^3) + \frac{\theta^2}{4}\,\mathrm{VAR}(Z^2) \qquad \text{con} \qquad Z = \frac{X-\mu}{\sigma}$$

Muestre que de este resultado se deduce lo obtenido en 3) en el caso normal.

5. Los siguientes datos

62.90	56.50	43.30	61.50	45.90	58.60	56.60	57.00	63.80	63.20	63.70	40.00
57.00	51.00	61.00	52.90	60.00	63.00	50.50	50.50	53.80	62.80	58.80	58.10

corresponden la máxima cantidad de oxígeno consumida en ml/kg/min por 24 atletas profesionales. Se desea estimar el coeficiente de variación $\theta = \mu/\sigma$. Suponiendo que las mediciones corresponden realizaciones independientes de una variable aleatoria con distribución $F = F_{\mu,\sigma}$

(a) Calcular el valor aproximado del desvío estándar de $\widehat{\theta}$ dado por la teoría asintótica cuando se supone que los datos son normales, o sea, $F_{\mu,\sigma} = N(\mu, \sigma^2)$ y cuando no se hace este supuesto.

(b) En este item usaremos bootstrap para aproximar la distribución de $\widehat{\theta}$. Más precisamente, si $\widehat{\theta}$ tiene distribución $G = G(F)$ que depende de la distribución F de las observaciones X_i, el método bootstrap permite estimar G estimando F por $\widehat{F}$ y G por $\widehat{G} = G(\widehat{F})$.

Este ejercicio propone evaluar el comportamiento de la aproximación $\widehat{G}$ cuando se usa el bootstrap noparamétrico que toma como estimador $\widehat{F} = F_n$ la empírica de los datos y cuando utilizamos el bootstrap paramétrico que estima F por $F_{\widehat{\mu},\widehat{\sigma}}$.

i. Calcule un estimador del desvío de $\widehat{\theta}$, utilizando el bootstrap paramétrico suponiendo que $F = F_{\mu,\sigma} = N(\mu, \sigma^2)$y el bootstrap noparamétrico con $B = 5000$ réplicas. (Fije la semilla inicial en 123 en ambos casos para comparar resultados). Compárelos con los valores obtenidos en a). ¿Qué observa?

ii. Grafique el estimador de la densidad de las réplicas $\widehat{\theta}^{\star}$ obtenidas mediante ambos métodos bootstrap, en un mismo gráfico. ¿Qué observa?

(c) ¿Qué sugieren los resultados obtenidos sobre la aproximación normal para el estimador de θ? ¿Usaría dicha aproximación para hacer inferencia sobre θ, es decir, para aproximar la distribución de $\widehat{\theta}$?

¿Los resultados obtenidos permiten decir algo sobre el supuesto de normalidad hecho para estos datos o usaría otro método para ver si dicho supuesto es razonable?