Reglas del TP:

- Este trabajo debe hacerse de forma individual o de a dos.
- Deben enviarse el script de R y un informe que contenga las respuestas a todas las preguntas y los gráficos pedidos. No hace falta explicar en el informe que es lo que hace cada una de las funciones del script.
- El script debe estar prolijo. Esto en particular implica que las variables tienen que tener un nombre descriptivo (es decir, no llamar `a`, `b`, `c` a las variables).

Intervalos Bootstrap y asintóticos para la varianza

Supongamos que tenemos variables aleatorias i.i.d. $X_1, \ldots, X_n$ tal que $\mathbb{E}[X_1] = \mu$ y $\text{VAR}(X_1) = \sigma^2$. Nos interesa hacer un intervalo de confianza de nivel 0.95 para σ^2 a partir de su estimador

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Dada una muestra, vamos a considerar las siguientes cuatro estrategias para hacer intervalos de confianza para σ^2:

- $IC_1(\sigma^2)$: asumiendo que la muestra tiene distribución normal y usando como pivote $(n-1)s^2/\sigma^2$ (y que este pivote se distribuye como una χ^2_{n-1}).
- $IC_2(\sigma^2)$: googleando cuál es la distribución asintótica de s^2 y armando el intervalo a partir de dicha distribución (esimando los parámetros que sean necesarios).
- $IC_3(\sigma^2)$: siguiendo el siguiente procedimiento:
  - Obtener $N_{boot} = 1000$ muestras Bootstrap a partir de la muestra original.
  - Obtener los valores $\widetilde{s}^2_1, \ldots, \widetilde{s}^2_{N_{boot}}$ correspondientes al aplicar el estimador s^2 a cada una de las muestras Bootstrap.
  - Obtener $IC_3(\sigma^2)$ a partir de los cuantiles 0.025 y 0.975 del vector $\widetilde{s}^2_1, \ldots, \widetilde{s}^2_{N_{boot}}$

- $IC_4(\sigma^2)$: estimando la varianza de s^2 a partir de los valores $\widetilde{s}^2_1, \ldots, \widetilde{s}^2_{N_{boot}}$ definidos en el método anterior. Si llamamos v^2 a dicha estimación, definir el intervalo $IC_4(\sigma^2) = [s^2 - v\, z_{0.025}, s^2 + v\, z_{0.025}]$, donde $z_{0.025}$ es el percentil 0.975 de una distribución normal estándar.

1. Hacer cuatro funciones (una para cada uno de los métodos descriptos) que, dada una muestra, calculen el intervalo de confianza correspondiente. Aplicar cada uno de estos métodos a la siguiente muestra:

   1.688 0.819 0.053 0.068 0.767 0.251 0.316 0.500 0.149 0.282

2. Realizar el siguiente estudio de simulación:

   - Generar $N_{rep} = 500$ muestras independientes, cada una de ellas de tamaño $n = 500$ y donde cada uno de sus elementos tiene distribución $N(\mu = 0, \sigma^2 = 9)$ (y son todos independientes).
   - Para cada una de estas muestras, aplicar los cuatro métodos (usando las funciones que ya fueron programadas), obteniendo así 4 intervalos para cada muestra.
   - Para cada uno de los cuatro métodos, obtener la proporción empírica de intervalos que cubren al verdadero valor de σ^2. Obtener además, para cada método, el promedio de las longitudes de los N_{rep} intervalos generados.

   Finalmente, comparar los resultados entre los 4 métodos. ¿Cuál resulta mejor en este caso?

3. Repetir el item anterior, reemplazando $n = 50$ por $n = 100$ y la distribución $N(0, 9)$ por una distribución exponencial con parámetro $\lambda = 1/3$ (notar que en este caso $\mu = 3$ y $\sigma^2 = 9$). ¿Cómo compararía los cuatro métodos en este caso?

4. A partir de los resultados obtenidos, ¿qué ventajas tiene Bootstrap respecto de los otros métodos?

5. (Optativo) Extender este trabajo en la dirección que le parezca natural (por ejemplo, variando el valor de n, el nivel de los intervalos, la distribución de los datos, pensando algún nuevo método para hacer un intervalo, etc.)