Laboratorio 1
Introducción a los Métodos Estadísticos

Condiciones:

- Fecha de asignación: 18 de septiempre.
- Fecha de entrega: 2 de octubre.
- Realizar en grupos de 3 persones.
- El informe se debe realizar en *LaTeX* utilizando la plantilla que se encuentran en el campus virtual.
- Se debe adjuntar el código en R.

1) En una población hay un número θ de carros que trabajan con la aplicación UBER, este número es desconocido. Supongamos que dichos carros, están numerados visiblemente en forma consecutiva: $1, 2, \ldots, \theta$. Con el propósito de estimar θ, usted registra una muestra aleatoria de n carros y anotando cada vez el número X correspondiente. Así dispone de una muestra aleatoria: $X_1, X_2, \ldots, X_n$.

   Para cumplir con el objetivo se han propuesto los siguientes estimadores:

$$\hat{\theta}_1 = 2 \cdot \bar{X} - 1$$

$$\hat{\theta}_2 = X_{Max} + X_{Min} - 1$$

$$\hat{\theta}_3 = X_{Max} \text{ (el máximo de la muestra)}$$

$$\hat{\theta}_4 = Me(X) \text{ (la mediana de la muestra)}$$

   Su labor consiste en analizar, mediante la implementación de escenarios de simulación, la eficiencia de los estimadores planteados, valorando desde los resultados el sesgo del estimador, su varianza y el error cuadrático medio. Asuma que el valor del parámetro $\theta = 1200$. Trabaje bajo 1000 simulaciones iníciales a partir de las cuales se evalúan cada una de las condiciones solicitadas para los estimadores, a partir de muestras de tamaño $n = 20$, $n = 30$, $n = 40$, $n = 50$, $n = 100$ y $n = 250$. Evalué todas las propiedades de los estimadores y el efecto del tamaño de muestra sobre ellas.

2) Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados de un parámetro θ, tales que: $V(\hat{\theta}_1) = \sigma_1^2$ y $V(\hat{\theta}_2) = \sigma_2^2$

   a) Demuestre que si a es un número real, entonces: $\hat{\theta}_3 = a \cdot \hat{\theta}_1 + (1 - a) \cdot \hat{\theta}_2$ es otro estimador insesgado para θ.

   b) ¿Cuál es el coeficiente de variación de $\hat{\theta}_3$?

c) Si los estimadores $\hat{\theta}_1$ y $\hat{\theta}_2$ son independientes, ¿cómo debe escogerse a para que la varianza de $\hat{\theta}_3$ sea mínima?

3) El número de fallas por semana de cierto mini computador es una variable aleatoria X con distribución Poisson de parámetro λ. Se dispone de una muestra aleatoria de 30 semanas del año.

a) Sugiera un estimador insesgado para λ.

b) Suponga que los costos de reparación en que incurre la empresa de forma semanal se encuentra en función del número de fallas, y que su relación se puede aproximar mediante la siguiente expresión

$$C = 3 \cdot X + X^2$$

Demuestre que $E(C) = 4 \cdot \lambda + \lambda^2$ y sugiera un estimador insesgado para $E(C)$.

4) Una población es definida por una variable aleatoria X con distribución exponencial:

$$f(x) = 5 \exp\{-5x\}, \; x > 0$$

a) Genere 1000 muestras aleatorias de tamaño n de dicha población y calcule para cada una su media muestral. Construya un histograma con los 1000 valores de la media, sobreponiendo al histograma la curva normal con parámetros μ y $\sigma/\sqrt{n}$. Haga esto para tamaños de muestra de 5, 10, 20, 30, 50, 100. Coloque los 7 gráficos incluyendo el de la población, el cual encabezará la sucesión de gráficos que usted colocará en orden en una misma página para poder visualizar el impacto del tamaño de muestra en la aproximación a la distribución normal. Observe, reflexione, comente y concluya.

b) Haga el mismo ejercicio partiendo de una distribución t-student con 7 grados de libertad. Compare y comente los resultados obtenidos en (a) y (b).

5) Un fabricante de varillas metálicas para construcción, cuando el proceso está funcionando adecuadamente, produce varillas con resistencia a la tracción variable, que puede modelarse con la distribución Normal. Cuando el proceso está bajo control, la resistencia media (μ) es de 3800 psi y su desviación estándar $\sigma = 400$ psi. Cierto comprador, con el ánimo de monitorear la media de los lotes que este proveedor surte, pues su varianza es bastante estable, ha establecido la siguiente regla de decisión. Si al tomar una muestra de tamaño $n = 4$, encuentra que su media es menor que 3334 psi entonces se preocupa y retorna el lote a su proveedor, pues le parece suficiente evidencia de que el proceso de este proveedor esta fuera control.

a) ¿Cuál es el riesgo que corre este proveedor bajo este criterio?.

b) La media del proceso ha cambiado hasta un nivel $\mu = 3700$, que probabilidad existe de que un lote conformado bajo estas condiciones sea detectado?.

c) Represente gráficamente la probabilidad planteada en el punto (b) para valores de $\mu =$ 3700, 3600, 3500, 3400, 3300, 3200......2500. *Pista: Encuentre una expresión general y realice los cálculos y gráficos con la ayuda de alguna herramienta computacional.*

d) Que sucede con la gráfica observada en (c), cuando el tamaño de muestra se incrementa de 4 a 5, 6, Represente todas las gráficas obtenidas en una sola en la que pueda apreciar las diferencias, analice y concluya acerca del efecto de incrementar el tamaño de muestra sobre la sensibilidad de la prueba.

"Si quieres conseguir lo que nunca has tenido, debes estar dispuesto a hacer lo que nunca has hecho"