Laboratorio 3
Introducción a los Métodos Estadísticos

Condiciones:

- Fecha de asignación: 31 de Octubre.
- Fecha de entrega: 14 de Noviembre.
- 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 al campus.
- El informe se debe entregar impreso el 14 de octubre al inicio de la clase.

[1] *(0.5)* Genere 5000 muestras aleatorias de tamaño $n = 10$ de una población normal con media $\mu = 5$ y Varianza 1. Con cada una de ellas construya un intervalo de confianza del 95 % para la media y para la varianza. Cuente que porcentaje de los 5000 intervalos atrapan la verdadera media y la verdadera varianza poblacionales respectivamente, adional a esto calcule la longitud esperada de cada intervalo. Comente el resultado (*asuma un nivel de confianza del 95 %*).

a) Repita el ejercicio para los tamaños de muestra ($n = 30, 50, 100$). Represente gráficamente el % de cubrimiento observado y la longitud esperada de intervalo.

b) Repita el ejercicio realizando las simulaciones desde una distribución exponencial con media $\mu = 5$. Compare los resultados de forma adecuada. Se observan similaridades o diferencias? A que puede deberse?

[2] *(0.5)* En vista de que el tema sobre la fluoración del agua de consumo ha sido objeto de mucha polémica. En una población se decidió tomar una muestra aleatoria para consultar la opinión de la población sobre si estaban o no de acuerdo con que se agregara fluoruro de sodio de agua. Al tomar una muestra aleatoria de 200 personas 110 estaban a favor. Construya un intervalo del 99 % de confianza para la verdadera proporción de personas que están de acuerdo con la fluoración del agua. Con base en dicho intervalo, usted está confiado que en realidad la mayoría aprueba el proyecto de fluoración? (Explique muy bien su razonamiento)

[3] *(0.5)* Una cadena de negocios de electrodomésticos quiere estudiar la efectividad de una nueva campaña televisiva sobre la venta de una cierta marca de heladeras. Para ello se recoge el número de unidades vendidas durante un mes antes y un mes depués de la campaña, en 12 de los negocios que componen la cadena. Los resultados obtenidos están dados en la siguiente tabla.

Antes	12	10	15	8	19	14	12	21	16	11	8	15
Después	11	11	17	9	21	13	16	25	20	18	10	17

a) Mediante análisis exploratorio evalué la efectividad de la campaña.

b) Con un nivel de error del 5 % hallar un intervalo de confianza para la diferencia de medias de unidades vendidas durante un mes antes y un después.

c) ¿Se puede considerar efectiva la campaña publicitaria?.

[4] *(1.0)* Dada una muestra de tamaño n de una población con una proporción p (desconocida) y la proporción de muestra estimada $\hat{p}$ (conocida).

a) Construya una función en R para calcular el intervalo de confianza para p con un nivel de significancia α. La función debe devolver un vector de dos elementos: el primero es el límite inferior del IC, y el segundo es el límite superior del IC.

b) Genere 5000 muestras de tamaño 40 de una distribución binomial con una proporción de 0.85 y utilice la función construida en el literal a para obtener el intervalo de confianza del 95 % de la proporción de población p para cada muestra, basándose en la proporción de muestra estimada p (pretendiendo que p es desconocido). ¿Cuántas veces de las 5000 simulaciones los intervalos de confianza contienen la verdadera proporción de población p (es decir, 0.85)? ¿Cómo interpretaría el intervalo de confianza basado en este problema?

c) Repita el punto b para muestras de tamaño $n = 10, 20, 30, 50$ y 100. ¿Qué puede concluir acerca del porcentaje de cobertura?

[5] *0.5* Para verificar el estado de calibración de un báscula ($B1$) se ha realizado un experimento consistente en verificar su medición y confrontarla contra aquella realizada por una báscula certificada ($B2$), para ello se han tomado 10 especímenes, que son pesados en cada una de las basculas, obteniendo los siguientes resultados.

Báscula 1	11.23	14.36	8.33	10.5	23.42	9.15	13.47	6.47	12.4	19.38
Báscula 2	11.27	14.41	8.35	10.52	23.41	9.17	13.52	6.46	12.45	19.35

Con base en estos resultados usted debe emitir un diagnóstico respecto al estado de la calibración de la báscula a través de intervalo de confianza al 98 % de confiabilidad.

[6] *(0.5)* El deterioro de muchas redes de tubería municipales en todo el país es una preocupación creciente. Una tecnología, propuesta para la rehabilitación de tubería, utiliza un revestimiento flexible insertado en la tubería existente. En el artículo *Effect of welding on a Hige-density Poliyethylene liner* (Journal of Materials in Civil Engr) se proporcionan los siguientes datos acerca de la resistencia a la tensión $lb/pulg^2$ de revestimientos cuando se utilizó y cuando no se utilizó el revestimiento. *Asuma un nivel de confianza del 98 %*

Sin revestimiento: 2748, 2700, 2655, 2822, 2511, 3149, 3257, 3213, 3220, 2753
Con revestimiento: 3027, 3356, 3359, 3297, 3125, 2910, 2889, 2902

a) Considera usted que las varianzas poblacionales pueden ser iguales?

b) Recomienda el uso del revestimiento como mecanismo complementario?

[7] *(1.5)* El cociente de variación ($cv = \mu/\sigma$ con $\mu \neq 0$) es una medida adimensional (sin unidad de medida) de dispersión de una distribución de probabilidad. Su utilidad esta en que es una medida para comparar variables poblaciones que tiene diferente unidad de medida. Dada la

dificultad de encontrar un intervalo de confianza para este cociente, se han propuesto varias alternativas, algunas de estas son (*Asuma un nivel de confianza del 95 %*):

Estimador 1:

$$IC\,(cv) = \left(\hat{cv} + z_{\alpha/2}\sqrt{(n-1)^{-1}\hat{cv}^2\left(0.5+\hat{cv}^2\right)}\,;\hat{cv} + z_{1-\alpha/2}\sqrt{(n-1)^{-1}\hat{cv}^2\left(0.5+\hat{cv}^2\right)}\right)$$

Estimador 2:

$$IC\,(cv) = \left(\frac{\sqrt{(n-1)cv}}{\sqrt{\chi^2_{n-1;1-\alpha/2}}};\frac{\sqrt{(n-1)\hat{cv}}}{\sqrt{\chi^2_{n-1;\alpha/2}}}\right)$$

Estimador 3:

$$IC\,(cv) = \left(\hat{cv}(x) + z_{\alpha/2}\sqrt{\frac{1}{n}\left(\hat{cv}^4+0.5\hat{cv}^2\right)};\hat{cv}(x) + z_{1-\alpha/2}\sqrt{\frac{1}{n}\left(\hat{cv}^4+0.5\hat{cv}^2\right)}\right)$$

Método de simulación Para resolver las dudas anteriores es necesario realizar simulaciones, para ello se plantea el siguiente algoritmo:

- Se selecciona una distribución de probabilidad y determine el valor de σ o de σ/mu.
- Genere una muestra aleatoria de tamaño n y calcule los intervalos de confianza propuestos (desviación estándar o cociente de variación).
- Calcule la longitud de cada intervalo de confianza.
- Determine si el intervalo contiene el valor del parámetro σ o de σ/μ.
- Realice 5000 simulaciones y compare los intervalos (porcentaje de cobertura y longitud promedio), haga este paso para $n = 5, 10, 20, 30, \ldots, 100$.

Distribuciones a evaluar: Se proponen tres distribuciones para las simulaciones. El primer caso una distribución normal, el segundo una distribución asimétrica y otro caso con una distribución simétrica no normal.

Caso	**Distribución**
I	$Normal\,(\mu,\sigma)$
II	$Gamma\,(\alpha,\beta)$
III	$Uniforme\,(a,b)$

Para las comparaciones deben escoger los parámetros de tal forma que los parámetros a comparar sean los mismos en cada distribución. Para el caso de la comparación del CV la media de la distribución debe ser diferente de cero (preferiblemente positiva). Compare el desempeño de los tres estimadores a partir del porcentaje de cobertura y longitud promedio.

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