
Todos los puntos del taller tienen el mismo peso

1. Suponga que se cuenta con una prueba para detectar la enfermedad A, que es positiva el 90% de las veces cuando se realiza en un paciente que tiene dicha enfermedad, y es negativa el 95% de las veces cuando se realiza en una persona que no tiene la enfermedad. También se sabe que la enfermedad afecta a un 1% de la población.
2. Construya una muestra aleatoria de tamaño 100000, que contenga “Sí” y “No”, con probabilidades de 1% y 99%, respectivamente.
3. Construya una muestra aleatoria a partir del vector de valores (“Negativo” y “Positivo”), que de cuenta de que la probabilidad de que el test salga “Negativo” dado que “No” tiene la enfermedad A es del 90%. Presente tablas de contingencia cruzadas condicionadas de acuerdo con si tiene o no tiene la enfermedad.
4. Construya una muestra aleatoria a partir del vector de valores (“Negativo” y “Positivo”), que de cuenta de que la probabilidad de que el test salga “Positivo” dado que “Sí” tiene la enfermedad A es del 90%. Presente tablas de contingencia cruzadas condicionadas de acuerdo con si tiene o no tiene la enfermedad.
5. Calcule la probabilidad de tener la enfermedad dado que el test salió positivo. Realice los cálculos utilizando las variables simuladas.
6. Realice los cálculos del punto anterior, utilizando la información del enunciado y el Teorema de Bayes. ¿Qué puede concluir?
7. Simule 1000 valores para cada una de las distribuciones de probabilidad uniforme discreta, binomial, Poisson, uniforme continua, normal y Exponencial. Especifique libremente los parámetros para cada una de ellas. Encuentre media y desviación estándar muestral para cada uno de los vectores simulados y compare dichos resultados con los obtenidos con las fórmulas de valor esperado y desviación estándar teóricos.
8. Realice secuencialmente la simulación del lanzamiento de un dado, de manera que en cada lanzamiento encuentre la proporción de veces que sale el número 5 (es decir, vamos a estimar de manera secuencial con el enfoque frecuentista del evento que al lanzar un dado se obtenga el número 5). Comente los resultados de cómo es la probabilidad cuando se hacen: 2 lanzamientos, 6 lanzamientos, 10 lanzamientos, 100 lanzamientos, 1000 lanzamientos. ¿La proporción de veces que sale el dado es equivalente a la teórica? Adicionalmente, construya un gráfico donde se evidencie la evolución de la proporción vs la cantidad de veces que se lanza el dado, interprete los resultados
9. Realice la simulación de 10000 conjuntos de datos diferentes provenientes de una distribución (desarrolle el ejercicio primero utilizando la distribución uniforme y posteriormente una exponencial, utilice los parámetros que desee de las distribuciones), obteniendo 1000 muestras de cada conjunto de datos. Luego, va a obtener el promedio en cada uno de los conjuntos de

datos y proceda a analizar la distribución de las medias obtenidas. ¿Qué evidencia en los histogramas? ¿A cuál de las distribuciones de la clase se le asemeja dicha distribución?

10. Suponga que una forma nueva de calcular un promedio (en una muestra aleatoria normal x_1, x_2, \dots, x_n) está dada por: $\tilde{Y} = \frac{100n}{n^2+1} + \sum_{i=1}^n \frac{x_i}{n}$, donde se sabe que este estimador de la media está sesgado. Simule 10, 100, 1000, 10000 y 100000 datos de una distribución normal con media 50 y varianza 10, y realice la estimación del promedio utilizando el nuevo estimador, y el estimador habitual. Comente los resultados, ¿qué comportamiento ve de esta nueva estimación?
11. Considere x_1, x_2, \dots, x_n una muestra aleatoria que sigue una distribución exponencial con parámetro λ , además se tienen los siguientes estimadores para el parámetro $\theta = \frac{1}{\lambda}$: $\hat{\theta}1 = x_1$, $\hat{\theta}2 = \sum_{i=1}^{n-1} \frac{x_i}{n-1}$, $\hat{\theta}3 = \bar{X}$ y $\hat{\theta}4 = \min \{x_1, x_2, \dots, x_n\}$ realice la simulación de 10, 100 y 1000 muestras aleatorias con el parámetro λ que desee, y realice la estimación del parámetro θ con los estimadores propuestos. Comente los resultados ¿Cuál de ellos se acerca más al valor real?
12. Considere el archivo SolicitudesDiarias.csv, en el cual se encuentran la cantidad de solicitudes diarias hechas en una institución de financiera por clientes. El equipo de mercadeo a partir del primero de febrero de 2022 implementó una campaña que buscaba aumentar la cantidad de solicitudes diarias realizadas por los clientes, además de que el primero de junio de 2022 lanzó una modificación a la campaña que tenía el mismo fin (aumentar la cantidad de solicitudes diarias). ¿Será que las campañas impartidas por el equipo de mercadeo tuvieron el efecto esperado? Obtenga los intervalos de confianza al 95% que considere para determinar si efectivamente el promedio diario de solicitudes aumentó con las campañas que lanzó el equipo de mercadeo. ¿Qué puede concluir al respecto? ¿Qué campaña fue más efectiva?
13. Sea X una variable aleatoria que tiene una distribución binomial con parámetros n y p . Dos estimadores propuestos para p son:
- $$\hat{P} = \frac{Y}{n} \text{ y } \tilde{P} = \frac{Y+1}{n+2}$$
- ☐ ¿Cuál de los estimadores es sesgado?
 - ☐ ¿Alguno es insesgado asintóticamente?
14. Realice los siguientes ejercicios del libro “An Introduction to Statistical Learning with Applications in Python” (Link para descargar los conjuntos de datos: <https://www.statlearning.com/resources-python>):
- ☐ Ejercicios Sección 2.4, página 64:
 - ☐ 1, 3, 6, 8
 - ☐ Ejercicios Sección 3.7, página 127:
 - ☐ 2, 4, 7, 10, 13 y 14

Entregable: Se recibirá un trabajo por equipo (equipos creados en EAFIT Interactiva), en el cual debe de contener (1) informe/documento escrito que contenga el enunciado de cada punto, además de su respectiva solución y análisis, adicionalmente (2) el código organizado y documentado

Estadística en Analítica
Taller 1, Semestre 2023-1
Profesores: Pablo Saldarriaga y Leandro Higueta

utilizado para dar solución a la implementación computacional realizada en el taller. Código que no ejecute o compile, se tomará como razón para disminuir la puntuación de los ejercicios computacionales.

Fecha de entrega: martes 7 de noviembre 2023.