

Prueba de Evaluación Continua de la asignatura «Estadística (Ing. Informática / Ing. T. I.)» (código 7190105-)

Curso 2024-2025

1. Descripción general de la prueba

La prueba de evaluación continua (PEC) de la asignatura «Estadística (Ing. Informática / Ing. T. I.)» (código 7190105-) consiste en resolver uno de los dos ejercicios propuestos en este documento. Ambos ejercicios tienen como objetivo revisar de una manera práctica algunos de los contenidos que se estudian en las unidades 1, 2, 3 y 4 del programa de la asignatura mediante el uso de procedimientos sencillos de simulación estocástica.

Los ejercicios no requieren emplear funciones complejas para la simulación, con lo cual se pueden resolver utilizando las funciones estadísticas elementales de una hoja de cálculo como Excel. Aunque este equipo docente no impone la herramienta a utilizar, sí recomienda el uso de la programación en R. Este lenguaje de programación es un estándar para el análisis científico de datos, tanto en el ámbito académico como en el sector privado; junto con Python, R se ha convertido en uno de los lenguajes de programación más empleados para la computación, la visualización y el análisis de datos en el campo de Estadística y Ciencia de Datos.

R es un software de distribución libre que puede descargarse de la página web

<https://www.r-project.org/>

Para una mejor experiencia de desarrollo al programar en R se recomienda utilizar algún editor avanzado como por ejemplo RStudio. Si utiliza programación con código R para resolver algún ejercicio, se recomienda leer algún manual introductorio de R antes de comenzar a programar. La ayuda sobre la sintaxis de cualquier función se obtiene tecleando `?nombre de función` en la ventana de comandos.

Si no se opta por esta alternativa (programar en R) se puede emplear cualquier otro lenguaje o herramienta; en la resolución de la PEC hay que hacer constar la herramienta utilizada.

Recordamos que, según figura en el documento de organización del curso, la evaluación continua supone un 10 % de la nota final y que el peso del examen en la calificación final es del 90 %. También recordamos que, según aparece en la información específica del Grado sobre la asignatura y de acuerdo con la

decisión tomada en la reunión de las dos Comisiones de Grado de la Escuela de Informática, celebrada el día 31 de Mayo de 2010, la nota final no puede ser 10 si no se ha realizado esta prueba de evaluación continua.

2. Entrega de la prueba

La prueba se entregará en el plazo establecido en el curso virtual, se considerará el horario peninsular para determinar la hora límite a efectos de entrega. Dado que el sistema bloquea la entrega una vez que ha pasado la hora límite establecida, es recomendable no esperar al último momento para remitir la prueba porque no habrá tiempo de reacción ante cualquier eventualidad que pueda surgir; no se admitirán entregas fuera del plazo y hora establecidos. A fin de facilitar la labor de los tutores en la corrección de la prueba, ésta se debe entregar en un único fichero con formato *pdf*.

Se ruega encarecidamente disciplina en observar las indicaciones anteriores.

3. Ejercicios propuestos

El alumno debe elegir uno de los dos siguientes ejercicios.

3.1. Ejercicio 1

En este ejercicio se propone simular $B = 50000$ valores de las distribuciones de los estimadores cuasivarianza muestral y varianza muestral:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{y} \quad s^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

a partir de muestras de tamaño n tomadas de una población $\mathcal{N}(0, \sigma)$.

El objetivo del ejercicio es mostrar la manera en que las simulaciones ilustran los resultados teóricos conocidos relativos al sesgo de ambos estimadores (capítulo 4 del texto base). Para ello el alumno deberá obtener y analizar la siguiente información a partir de las simulaciones.

1. Calcular, a partir de las simulaciones, los valores aproximados de $E\{S^2\}$ y $E\{s^2\}$ para los tamaños muestrales $n = 25$, $n = 50$ y $n = 100$; tómese $\sigma = 1$. Comente los resultados obtenidos (*máximo 1 página*).
2. Utilizar los valores anteriores para aproximar el sesgo de ambos estimadores. ¿En qué manera están reflejando los resultados de las simulaciones los que proporciona la teoría? (*máximo 1 página*).

Recomendación. Si utiliza programación con código R para resolver este ejercicio, las siguientes funciones de R le serán de utilidad:

- `numeric` para crear un vector donde guardar los valores de S^2 y s^2 .
- `for` para programar el bucle con las simulaciones.
- `rnorm` para generar muestras de la distribución normal.

- Para cada muestra, la función `var`, con argumento un vector numérico, permite hallar la cuasivarianza muestral.
- `mean` para calcular la media de un vector numérico.

3.2. Ejercicio 2

En este ejercicio se propone simular 50000 observaciones de una variable aleatoria Z con distribución $\mathcal{N}(0, 1)$, a partir de los valores simulados de una variable uniforme en $(0, 1)$, utilizando el método de BOX-MULLER (capítulo 3 del texto base).

1. Describir los pasos del procedimiento que le permitiría obtener los valores simulados de la $\mathcal{N}(0, 1)$ e implementar el procedimiento en un algoritmo para la simulación.
2. Explicar la manera en que un histograma de frecuencias nos ayuda a representar la función de densidad de una variable continua; para ello puede consultar cualquier manual introductorio de Estadística. Representar el histograma de los valores simulados. ¿Qué se puede decir de la forma del histograma obtenido? (*máximo 1 página*).
3. Utilizar los resultados de la simulación para aproximar la probabilidad $P(Z > 1,645)$.

Recomendación. Si utiliza programación con código R para resolver este ejercicio, las siguientes funciones de R le serán de utilidad:

- `runif` para generar valores de una variable uniforme en $(0, 1)$.
- `hist` para representar los valores simulados de la $\mathcal{N}(0, 1)$ mediante un histograma de frecuencias.