

11.67 ESTADÍSTICA APLICADA - TRABAJO PRÁCTICO

1 Búsqueda del set de datos

- Buscar un set de datos que contenga entre 3 y 6 variables, de las cuales 1(una) debe ser categórica y el resto deben ser numéricas. La cantidad de observaciones debe estar entre 100 y 500. Si los datos originales tienen más de 500 observaciones, se deberá tomar una muestra aleatoria.
- Identificar si el set de datos consiste en una serie de tiempo o en una muestra de tipo transversal.
- Enviar la propuesta de set de datos al docente asignado antes de comenzar el trabajo.

2 Análisis descriptivo

- Para cada una de las variables numéricas se deberá hacer un análisis descriptivo que incluya todos los aspectos vistos en el ejercicio T2.1 realizado en clase. No olvidar incluir una definición clara de las variables.
- El grupo deberá elegir una de las variables numéricas y hacer un análisis gráfico de cómo cambia la distribución de los datos en los distintos niveles de la variable categórica. Se sugiere hacer un análisis similar al que se pide en los incisos a. y b. del ejercicio T2.4.
- Producir un programa en R que permita ejecutar los cálculos y gráficos que el grupo considere necesarios, partiendo de la importación de datos. **Importante:** Usar únicamente las funciones y librerías vistas en clase, a menos que el grupo proponga armar alguna función propia.
- **Importante:** Se deberá trabajar creando un *proyecto* en R. Investigar la opción `File > New Project...` `> New Directory`, que permite armar una carpeta local que contendrá un archivo `.Rproj`. Allí se podrán incluir todos los scripts del trabajo y los archivos de datos. Sobre esta carpeta se trabajará a lo largo del curso.

3 Producción de informe

- Armar la base del informe. Debe tener carátula e índice al principio y una sección de referencias bibliográficas al final. En el cuerpo de este documento se añadirán sucesivamente todas las etapas del trabajo, a medida que avance el curso.
- A esta altura deberá incluirse solamente lo hecho en los puntos 1 y 2. Resumir los resultados en tablas y gráficos. **Importante:** Incluir todos los aspectos de análisis que el grupo considere importantes. Tener en cuenta que no se aceptarán resultados sueltos.
- Se deberá conservar un formato claro y unificado. Se evaluará la calidad de presentación de los contenidos. No olvidar citar claramente la fuente de los datos y otras referencias necesarias.

4 Primera entrega

El grupo debe entregar una carpeta titulada `Proyecto Grupo ...` que tenga lo siguiente: 1) El archivo `.Rproj` que debe ejecutar la sesión de R Studio, 2) El archivo `Punto2.R` con el programa realizado en el punto 2, 3) El archivo del informe completo, en formato **.pdf**, 4) El o los archivos con los datos originales, en formato `.csv`, `.xlsx` o `.txt`.

5 Estimadores puntuales de la mediana

Para la variable de estudio elegida, se busca estimar la mediana como medida de posición central:

- En forma paramétrica, usando dos modelos de distribución propuestos por el grupo.
- En forma no paramétrica.

Indicar las expresiones de cálculo de los estimadores en función de la muestra aleatoria.

Nota: En caso de que la distribución de la variable cambie notoriamente en los distintos niveles de la variable categórica, elegir uno de los niveles para realizar el análisis (es decir, trabajar con una de las variables aleatorias condicionadas).

6 Implementación de estimadores

El grupo debe programar en R los tres estimadores propuestos, mediante funciones que tengan como argumento el vector de datos `x` y devuelvan la estimación correspondiente. Aplicar las funciones a los datos del trabajo, para obtener las estimaciones correspondientes de la mediana y resumir los resultados en una única salida.

7 Análisis del error

Estimar el error estándar de cada uno de los tres estimadores mediante bootstrap y resumir los resultados. Resumir brevemente el procedimiento realizado. ¿Es posible cuantificar el sesgo de los estimadores mediante este procedimiento? ¿Por qué?

8 Ajuste de las distribuciones

Realizar un gráfico en el que se superponga el histograma con los dos modelos de distribución ajustados por máxima verosimilitud. Para cada caso, calcular las log-verosimilitudes máximas obtenidas. ¿Cuál ajusta mejor? ¿Son útiles los modelos paramétricos para describir los datos? En función del trabajo realizado, ¿Qué estimador es el más recomendable? Realizar conclusiones de acuerdo a lo visto en clase.

9 Producción de informe

Resumir el trabajo realizado en una nueva sección del informe. Incluir todas las observaciones cualitativas que el grupo considere apropiadas. Actualizar el informe en la carpeta del proyecto, e incluir en un nuevo archivo `.R` los programas necesarios para esta entrega.