**Guías 02 y 03. Comparación de medias tradicional y con regresión**

La resolución de esta guía requiere realizar una serie de análisis estadísticos en R e interpretar brevemente los resultados obtenidos. Usted deberá entregar – mediante correo electrónico – dos archivos. Un primer documento en formato Word o PDF que incluya las respuestas de las preguntas presentes en la guía. Un segundo documento, será un archivo de R con el código que ejecutó para responder las preguntas. Recuerde que el código que elabore debe ser reproducible, es decir, el/las ayudantes podrán ejecutar las líneas de código y replicar exactamente los hallazgos reportados en el informe con las respuestas.

El plazo de entrega de las Guías 02 y 03 es el viernes 22 de abril a las 23:59. El informe de resultados (archivo Word o PDF) debe tener una extensión máxima de dos páginas y junto al archivo de R se deben enviar por correo electrónico: [docpsi4035@gmail.com](mailto:docpsi4035@gmail.com)

**Base de datos**

Para llevar a cabo la presente guía, recuerde descargar la base de datos *datos\_fl\_crime.csv*, disponible en la carpeta compartida de Google Drive (2022 PSI4035/Guías/Guías 02 y 03). En la última sección de este documento encontrará una breve descripción de la base de datos.

**Consideraciones generales**

Recuerde tener presente los siguientes aspectos:

1. Incorporar en el archivo de R los comandos desde cargar la base de datos en adelante.

2. Una vez calculadas los puntajes de las variables contempladas, presente una tabla para describirlas.

3. Considere, para efectos de examinación e interpretación, las tres preguntas que guías que hemos discutido en el texto de Vik:

- ¿Hay una relación entre dos variables?

- ¿Cuál es la dirección de la relación?

- ¿Qué tan fuerte es la relación?

**Análisis**

Utilizando la base de datos sobre criminalidad en condados de Florida:

1. Genere una matriz de diagramas de dispersión mostrando las relaciones entre las variables “tasa de criminalidad”, “renta mediana de los ingresos”, “porcentaje de finalización de la escuela secundaria” y “porcentaje urbano”.

1. Genere versiones dicotomizadas (divididas en dos grupos) de las variables “renta mediana de los ingresos”, “porcentaje de finalización de la escuela secundaria” y “porcentaje urbano”. Use el promedio de cada variable como punto de corte, creando variables indicadoras en donde el valor “1” indique valores sobre el promedio en cada una de estas tres variables.
2. Usando la variable indicadora que distingue entre los puntajes sobre la media en “renta mediana de los ingresos”, haga una prueba t entre los dos grupos e interprete el resultado.

1. Usando la variable indicadora que distingue entre los puntajes sobre la media en “porcentaje de finalización de la escuela secundaria”, haga una prueba ANOVA entre los dos grupos e interprete el resultado.
2. Transforme la variable indicadora usada en el punto (d) en una variable codificada como desviaciones (-1 y 1), donde -1 represente los condados con valores bajo la media en el “porcentaje de finalización de la escuela secundaria”.
3. Usando la variable de desviaciones creada en (e), haga un análisis de regresión entre los dos grupos e interprete el resultado indicando explícitamente la interpretación de cada parámetro.
4. Brevemente indique que información pudo obtener en el análisis realizado en (f) que no pudo obtener directamente del análisis realizado en (d).
5. Usando la variable indicadora que distingue entre los puntajes sobre la media en “porcentaje urbano”, haga un análisis de regresión entre los dos grupos e interprete el resultado indicando explícitamente la interpretación de cada parámetro.

**Descripción de la base de datos**

Datos sobre la delincuencia en los condados de Florida. Tabla 9.16 o 9.17 de la 4ª edición del libro “*Statistical Methods for the Social Sciences*” de Alan Agresti y Barbara Finlay.

El archivo tiene 67 observaciones con las siguientes 5 variables:

**County** — Nombre del condado

**C —** tasa de criminalidad

I **—** renta mediana de los ingresos

**HS —** porcentaje de finalización de la escuela secundaria

**U** — porcentaje urbano