

Taller de práctica 4

Gráficos de dispersión, Espacios muestrales y Eventos aleatorios
(secciones 1.5.5, 2.1.1 y 2.1.2)

1. En un estudio para explicar la tasa máxima de flujo de seis sectores de drenaje después de una tormenta, se recolectaron los siguientes datos en un base (link):

- **area**: área del sector de drenaje.
- **area2**: área impermeabilizada.
- **pendiente**: pendiente media del sector.
- **largo**: largo del flujo.
- **absorbencia**: índice de absorbencia de la superficie (0: absorbencia total, 100: no absorbencia).
- **capacidad**: capacidad estimada de almacenamiento del suelo.
- **infiltracion**: tasa de infiltración del agua en el suelo.
- **lluvia**: pulgadas de lluvia caída.
- **tiempo**: tiempo en el cual la lluvia excedió 1/4 pulgadas/hora.
- **logq**: logaritmo de la tasa de flujo máxima de seis sectores de drenaje después de una tormenta.

Nota: para cargar la base de datos, utilice el comando `read.table(, header = TRUE)`.

- a) Estudie las relaciones entre todas las variables cuantitativas de la base de datos. Utilice el comando `pairs()`. Mencione los casos en los que observe un clara tendencia lineal, interpretando dichos casos.
2. La siguiente base (link) contiene datos relacionados a características de ingresos de estudiantes a un determinada universidad. Las columnas de la base de datos son las siguientes.

- **Sexo**: Hombre o Mujer.
- **Ingreso**: indica la vía de ingreso del estudiante a la universidad, se clasifica en PTU u Otra.
- **Logro**: corresponde a la proporción de logro (número entre 0 y 1, un logro de 0.4 indica que el estudiante respondió correctamente un 40 % de la prueba) del estudiante en el diagnóstico de “Comunicación escrita” aplicado por la universidad a la cual ingresa.
- **LEN**: Puntaje PTU - Lenguaje.
- **NEM**: Puntaje NEM del estudiante.

- a) Realice un gráfico de dispersión entre la variable Logro y la variable NEM.
 - b) Considerando el gráfico generado en a), segmente el gráfico por sexo. Utilice el comando `facet_wrap()`. Interprete
 - c) Considerando el gráfico generado en a), diferencia los puntos por color según la vía de ingreso de los estudiantes. Interprete.
3. Considerando la base de datos del segundo enunciado defina los evento aleatorios, según los experimentos señalados.

- a) Experimento: Seleccionar un estudiante. Evento: seleccionar un estudiante mujer.
- b) Segmente las proporciones de logro obtenidas en el diagnóstico de “Comunicación escrita” como “Insuficiente” (menor a 0.4), “Suficiente” (de 0.4 a 0.7) y “Bueno” (sobre 0.7). Experimento: Seleccionar un estudiante hombre. Evento: seleccionar un estudiante hombre que no esté en la categoría Insuficiente de la variable Logro.
- c) Segmente las proporciones de logro obtenidas en la PTU de Lenguaje como “Insuficiente” (menor a 480 puntos), “Suficiente” (de 480 a 630 puntos), “Bueno” (de 630 a 700 puntos) y “Excelente” (sobre 700 puntos). Experimento: Seleccionar un estudiante. Evento: seleccionar un estudiante categorizado como Suficiente en el diagnóstico de “Comunicación escrita” y “Bueno” en la PTU de Lenguaje.