

1. Objetivos

- Desarrollar habilidades para el manejo y análisis de datos a través del software R.
- Argumentar soluciones a problemas propuestos a partir del comportamiento de los datos.
- Ayudar a desarrollar un pensamiento que reconoce el elemento de la variabilidad.

2. Pregunta problematizadora:

Una base de datos aparentemente no contiene información alguna sobre el fenómeno que se está estudiando, sin embargo algunas técnicas básicas, acompañada de las herramientas adecuadas nos permite sacar conclusiones muy valiosas sobre nuestro objeto de estudio. ¿Cuales son las técnicas que nos ayudará a entender mejor el comportamiento de nuestra base de datos?

3. Resultados de aprendizaje:

- Diferenciar variables categóricas de variables cuantitativas (discretas y continuas).
- Analiza adecuadamente la información de variables cualitativas o categóricas.
- Representa de manera tabular y gráfica (diagrama de barras, diagrama circular) variables categóricas.
- Analiza adecuadamente la información de variables cuantitativas.
- Representa de forma tabular y gráfica (histogramas, polígonos, diagrama de cajas) variables cuantitativas.
- Calcula e interpreta medidas de tendencia central (media, mediana, moda), medidas o índices de posición (percentiles, cuartiles) y medidas o índices numéricos de dispersión o volatilidad (rango, varianza, desviación estándar, coeficiente de variación).

4. Recursos

Software R. RStudio o Rstudio Cloud (editores recomendados para trabajar con R).

5. Preguntas a responder por los estudiantes

Los datos en el archivo `Info_clientes_banco.xlsx` corresponden a la información contenida en el sistema de una entidad bancaria. Cada cliente es identificado con un código y en ella se encuentra información sobre los productos que tiene con el banco, como de las garantías que ha declarado para obtener créditos.

Los datos contienen:

- **ID_cliente**: código de identificación del cliente.
- **Asesor**: usuario en el sistema del asesor de servicios del cliente.
- **Sucursal**: sucursal a la que está adscrito el cliente.
- **Segmento**: clasificación del cliente según la información completa del cliente.
- **Tipo_Cliente**: clasificación del cliente según su comportamiento observado.
- **Cuenta_corriente**: 1 si tiene cuenta corriente con el banco, 0 en caso contrario.
- **Saldo_cuenta_corriente**: saldo en la cuenta corriente, si la tiene.
- **Cuenta_ahorros**: 1 si tiene cuenta de ahorros con el banco, 0 en caso contrario.
- **Saldo_cuenta_ahorros**: saldo en la cuenta de ahorros, si la tiene.
- **Credito_vivienda**: 1 si tiene crédito de vivienda con el banco, 0 en caso contrario.
- **Saldo_credito_vivienda**: saldo del crédito de vivienda, si lo tiene.
- **Vehículo_1**: 1 si tiene vehículo propio, 0 en caso contrario.
- **Valor_vehículo_1**: valor comercial del vehículo, si lo tiene.
- **Vehículo_2**: 1 si tiene un segundo vehículo propio, 0 en caso contrario.
- **Valor_vehículo_2**: valor comercial del segundo vehículo, si lo tiene.
- **Inmueble_1**: 1 si tiene inmueble propio, 0 en caso contrario.
- **Valor_inmuble_1**: valor comercial del inmueble, si lo tiene.
- **Inmueble_2**: 1 si tiene un segundo inmueble propio, 0 en caso contrario.
- **Valor_inmuble_2**: valor comercial del segundo inmueble, si lo tiene.

Bloque 0

Problema 1.

- Asigne a una variable llamada `x` el número 256.
- Imprima la variable `x` de modo que se muestre en el reporte.
- Asigne a una variable llamada `y` el número 559 y agregue `;y` al final de la línea de código. Explique qué efecto tuvo agregar `;y`.
- Genere una secuencia de números del 20 al 35 utilizando la función `seq` ¿se utilizó `print` o `;` para imprimir la secuencia? ¿por qué?
- Guarde en una variable llamada `secuencia.1` una secuencia de números del 122 al 137. Imprima.

- Guarde en una variable llamada `secuencia_2` una secuencia de números del 122 al 137 sin utilizar la función `seq`. Imprima. ¿Encuentra alguna diferencia con `secuencia_1`?
- Modifique la variable `secuencia_1` de modo que sólo contenga la secuencia de números de 3 en 3. Asigne esta secuencia a la variable `secuencia_3`. Imprima.
- Intente desarrollar el ejercicio anterior utilizando el código `122:137,by=3` ¿Funciona?

Problema 2.

- Genere un vector con números del 122 al 137 y guárdelo en una variable llamada `vector_1`. Utilice la función `matrix`. Imprima. ¿Cuál es la diferencia con `secuencia_1`?
- Genere una matriz cuadrada con los números del 122 al 137 y guárdela en una variable llamada `matriz_1`.
- Genere una variable llamada `z` con una secuencia de números del 0 al 10. Genere un vector con la función z^3 y asígnele una variable llamada `vector_2`. Imprima.
- Haga una gráfica de puntos de la variable `z`.
- Haga una gráfica de línea del `vector_2`.
- Modifique la `matriz_1` cambiando la posición `[3,2]` de la matriz por un 135. Imprima nuevamente la `matriz_1`.
- Haga un gráfico de puntos de la columna 2 de la `matriz_1`.

Cierre: Los problemas de este Bloque están diseñados para lograr familiaridad con el lenguaje de programación R.

Bloque 1

Problema 3.

- Instale el paquete “readxl” y actívelo.
- Cargue el archivo `Info_clientes_banco.xlsx`.

Problema 4. Clasifique las variables de la base de datos. Es decir, para cada variable de la base defina si es cualitativa o numérica. Si es numérica, a su vez defina si es discreta o continua, y si es cualitativa, defina si es nominal u ordinal.

Problema 5. Realice e interprete un análisis de frecuencias absolutas y relativas para las variables `Asesor`, `Sucursal`, `Segmento` y `Tipo_Cliente`.

Problema 6. Cree una lista que contenga la información del asesor con el que se abrió cada una de las cuentas corrientes.

Problema 7. Obtenga una tabla de frecuencias donde se muestre la cantidad de cuentas corrientes que abrió cada asesor. Realice e interprete un diagrama de barras para esta información.

Problema 8. Desarrolle los dos ejercicios anteriores para cuentas de ahorros y créditos de vivienda.

Problema 9. Obtenga un listado que contenga los saldos de cada una de las cuentas corrientes abiertas en el banco.

Problema 10. ¿Cuál es el saldo total que tiene el banco en cuentas corrientes?

Problema 11. ¿El banco tiene más dinero en cuentas corrientes o de ahorros? ¿Cuál es la distribución porcentual de estas cantidades?

Problema 12. Obtenga e interprete tablas de frecuencias absolutas y relativas para el segmento de clientes y tipos de clientes.

Problema 13. Obtenga e interprete diagramas de barras para la información anterior.

Cierre: La clasificación de variables nos permite utilizar las herramientas de estadística descriptiva correctamente para tener resultados cuya interpretación sea clara, por otro lado, el análisis de frecuencia es el primer paso para conocer el comportamiento de nuestras variables.

Bloque 2

Problema 14. Cree una variable en el conjunto de datos que indique la cantidad de vehículos que tiene cada cliente y otra para el valor total de sus vehículos. Denote estas nuevas variables como `Vehículos` y `Valor_vehículos`. Al nuevo conjunto de datos llámelo `Info_clientes_banco_ampliada`.

Problema 15. Obtenga un listado que contenga el valor total de los vehículos de los clientes que tienen por lo menos un vehículo.

Problema 16. Genere e interprete histogramas con la información anterior, bajos dos escenarios: uno en que se muestre la cantidad de clientes y otro en que se muestre el porcentaje de clientes. En cada caso modifique el título del gráfico y las etiquetas de los ejes.

Problema 17. Obtenga un listado de las identificaciones de los clientes que tienen cuenta de ahorros o cuenta corriente.

Problema 18. Obtenga un listado de las identificaciones de los clientes que siendo propietarios de dos inmuebles cuyo valor conjunto supere 600 millones, solicitaron crédito de vivienda.

Problema 19. Realice análisis descriptivos para los saldos en cuentas corrientes y para los créditos de vivienda. Compare por sucursales, segmentos y tipos de cliente. Interprete los resultados comparándolos. ¿Cuál de los asesores tiene mejor desempeño? ¿Cuál sucursal? ¿Existen diferencias significativas entre los tipos de cliente? ¿Tiene sentido la segmentación que hace el banco? Argumente sus respuestas con base a las medidas descriptivas y en gráficos apropiados.

Cierre: Los problemas de este bloque apuntan a entender la creación de nuevas variables y los filtros para profundizar en el análisis exploratorio de una base de datos. Además, nos permite entender la importancia de las medidas descriptivas en el análisis de una variable.

6. Condiciones de entrega:

1. Se deben incluir los nombres de los integrantes en el archivo entregable.
2. Solo se debe entregar UN ARCHIVO y este debe ser con extensión `.rmd`
3. Se debe marcar el archivo con nombre del profesor tallerista y el nombre y apellidos (ambos apellidos) del autor de correspondencia (el estudiante encargado de subir el archivo a Interactiva Virtual).