

Actividad: estadística descriptiva con R

Contexto

Vimos que uno de los usos de los modelos estadísticos es describir o caracterizar un conjunto de datos, lo que aprendimos en la enseñanza secundaria y en cursos anteriores con el nombre de estadística descriptiva. Son muchas las herramientas, pero es importante saber usar la que es adecuada para lo que se quiere describir.

Objetivos de aprendizaje

1. Tener conciencia de la existencia de numerosas alternativas para describir datos.
2. Seleccionar una herramienta adecuada considerando los datos que se tienen y las características que se quiere describir.
3. Describir datos usando herramientas disponibles en el entorno R.

Éxito de la actividad

1. El equipo revisa los datos entregados y reconoce la(s) variable(s) involucrada(s) en la pregunta.
2. El equipo elige una herramienta de visualización adecuada para los datos y pregunta(s) planteada(s).
3. El equipo consigue caracterizar los datos entregados, usando el entorno R, y responder la(s) pregunta(s) planteada(s) en consecuencia.

Actividades

1. Obtener el archivo con una selección de datos públicos recopilados por la encuesta Casen 2017 en la Región Metropolitana. La siguiente es una breve descripción de las columnas seleccionadas:
 - folio: identificador del hogar
 - o: número de orden de la persona dentro del hogar
 - id.vivienda: identificador de la vivienda
 - hogar: identificación del hogar en la vivienda
 - region: region
 - provincia: provincia
 - comuna: comuna
 - ing.comuna: posición en el ranking histórico del ingreso de la comuna (ascendente)
 - zona: área geográfica (Urbano, Rural)
 - sexo: sexo de la persona registrada
 - edad: edad de la persona registrada
 - ecivil: estado civil de la persona registrada
 - ch1: situación ocupacional de la persona registrada
 - ytot: ingreso total
2. Considerar las siguientes preguntas:

Sala 1

¿Se encuestaron más o menos la misma cantidad de gente en cada provincia de la RM?

Sala 2

¿Cómo diría que es el ingreso de los hombres de la RM (simétrico/asimétrico, concentrado/disperso, unimodal/multimodal, etc.)?

Sala 3

¿Tienen hombres y mujeres ingresos similares en la RM?

Sala 4

¿Se distribuye de igual manera la situación ocupacional de los hombres que viven en áreas rurales y quienes viven en áreas urbanas de la RM?

Sala 5

¿Son similares los ingresos registrados en las diferentes provincias de la RM?

Sala 6

¿Van los ingresos de las mujeres de la RM incrementándose con la edad?

Sala 7

¿Tiene relación el ingreso de las mujeres de la RM con la riqueza del municipio donde habita?

Sala 8

¿Cómo diría que es el ingreso de las mujeres de la RM (simétrico/asimétrico, concentrado/disperso, unimodal/multimodal, etc.)?

Sala 9

El nivel de ingreso de las mujeres de la RM ¿varía con el estado civil?

Sala 10

Los ingresos de los habitantes de la RM, ¿van incrementándose con la edad de forma similar en hombres y mujeres?

3. Discutir y consensuar qué medidas estadísticas (media, mediana, moda, etc.) y qué forma gráfica ayudaría a responder la pregunta asignadas.
4. Construir un script en R que produzca los estadísticos y el gráfico seleccionado.
5. Documentar en el script el objetivo de cada estadístico, una descripción de lo que se ve en el gráfico y la respuesta a la pregunta planteada.
6. Subir la respuesta con el nombre "EP02-respuesta-sala-i", donde i es el número de la sala asignada, a la carpeta compartida para este fin.
7. Calcular $j = (i + 4) \bmod n + 1$, donde n es la cantidad de salas usadas; descargar la respuesta de la sala j y aplicar la rúbrica publicada para este EP.
8. Subir la rúbrica completada, en formato PDF, con el nombre "EP02-resultado-sala-j" a la carpeta compartida.
9. Las respuestas deben subirse antes del mediodía del sábado y las revisiones, antes del mediodía del lunes.