



## Trabajo Práctico Estadística Descriptiva

### Objetivo

Lograr que el estudiante aplique las herramientas adquiridas durante el curso sobre estadística descriptiva en un conjunto de datos reales, haciendo uso de un software apropiado.

### Presentación

El trabajo se realizará en grupos de dos o tres alumnos. Dado el carácter virtual que posee el cursado de la materia actualmente, el informe será entregado a través de la plataforma utilizada para la materia.

Para lograr la regularización, el trabajo práctico debe estar aprobado. De acuerdo a las correcciones que se realicen, puede requerirse la defensa en forma oral por uno o más de los integrantes de los grupos en el caso de que se considere necesario. Tenga en cuenta que la evaluación es individual.

### Descripción del problema

EcoBici es un sistema de bicicletas compartidas que funciona en la Ciudad de Buenos Aires, cuenta con más de 200 estaciones y 1.200 rodados. Actualmente el sistema está presente en 30 de los 48 barrios de la Ciudad Autónoma de Buenos Aires, funcionando los 365 días del año 24 horas. Existen 2 formas de usar el sistema: mediante una app o mediante una tarjeta. Mediante la app, el usuario debe registrarse y, mediante su nombre de usuario y password, ingresa al sistema, elige la estación en la que se encuentra y luego elige la bici que quiere. Mediante tarjeta, el usuario debe apoyarla en un lector y el sistema le asigna una bicicleta. El usuario se moviliza por la ciudad durante un tiempo limitado (30 minutos) y luego debe devolver la bicicleta en alguna de las estaciones habilitadas. El sistema le asigna al usuario un id y registra datos tales como la edad, el género, estación de origen, estación de destino. . . etc.

### Variables registradas

En los datos provistos por la cátedra <sup>1</sup> para la realización del informe existen dos unidades de análisis diferentes. Por un lado, existe información relativa a los usuarios que utilizaron el servicio durante el año 2020. Por el otro, se presenta información sobre los recorridos que realizó cada uno de los usuarios.

Las variables registradas referidas a los *usuarios* del sistema EcoBici son las que se presentan a continuación.

---

<sup>1</sup>Los datos fueron descargados desde <https://data.buenosaires.gob.ar/dataset/estaciones-bicicletas-publicas>. Las tablas que recibió cada grupo corresponden a una elección aleatoria sobre las observaciones totales. A su vez, los datos fueron procesados por la cátedra previamente.

Nombre	Descripción
ID de usuario	Identificación del usuario.
Género	Género del usuario. La variable puede tomar los siguientes valores: F (femenino), M (masculino), OTRO (otro género distinto a femenino y masculino, o dato faltante).
Edad	Edad del usuario (en años).

Se considera como *recorrido* hecho por un usuario al trayecto que este realiza a partir de que retira el rodado desde la estación de origen hasta su devolución en la estación destino. Las variables registradas referidas a los recorridos son las que se presentan a continuación.

Nombre	Descripción
Duración	Duración del recorrido realizado en segundos (s).
Distancia	Distancia aproximada del recorrido realizado en metros (m).
Día	Día de la semana en que se realizó el recorrido. La variable puede tomar 7 valores distintos, correspondientes al día de la semana.
Dirección de origen	Dirección de la estación de EcoBici desde donde inició el recorrido.
Dirección de destino	Dirección de la estación de EcoBici donde finalizó el recorrido.

## Consigna

Elabore un informe que refleje las características más notables acerca de los datos brindados. Para llevar esto a cabo deberá utilizar un software estadístico que le permita efectuar un estudio descriptivo que incluya: tablas de distribución de frecuencias, gráficos y medidas descriptivas.

Se debe incluir un análisis univariado de cada una de las variables presentadas. A su vez, se solicita incluir un gráfico donde realice un análisis comparativo de una variable según los niveles de otra, con sus adecuada interpretación bivariada.

Por último, redacte un breve informe, a modo de conclusión, acerca de los resultados obtenidos en el análisis.

Trabaje teniendo en mente que el informe lo leerá quien está interesado en analizar los datos presentados y no los docentes. Es interesante ejercitar una primera aproximación a la forma en la cual se resumen los datos en la práctica profesional.

## Recomendaciones generales para la elaboración del informe de resultados

La elaboración de un informe escrito es fundamental para comunicar los resultados obtenidos en una investigación. Este informe, como elemento de comunicación, debe poseer una serie de características para que cumpla con su cometido primordial. Estas características van desde su presentación visual (ordenado y legible), pasando por los elementos estructurales fundamentales como la lógica de presentación, la exposición de ideas, la calidad de las fuentes mencionadas y la interpretación de los datos. En particular, debe poseer al menos dos características esenciales: a) que las personas a las que va dirigido lo lean porque es bueno y consistente; b) que otras personas, que no necesariamente sean especializadas en la materia de la que trata el documento, lo puedan comprender sin mayores dificultades.

Las características globales mencionadas abarcan una serie de recomendaciones especiales. A continuación, se mencionan algunas de ellas:

- La redacción de los informes, en general, se hace de forma impersonal, sin acudir a las primeras personas (en lugar de “podemos ver”, decir “se puede ver”; en lugar de “observamos”, “se observa”; etc.). Utilizar un lenguaje técnico dota al escrito de claridad y precisión. La tercera persona es un componente clave de la escritura académica, ya que es una voz que pretende transmitir un punto de vista objetivo.
- En cuanto a la forma de presentación, es necesario mantener un mismo formato a lo largo de todo el trabajo. Esto incluye desde el tipo o tamaño de letra durante la redacción, el uso de sangría, ajuste de párrafo y el interlineado, entre otras cuestiones. Lo mismo debe ocurrir con los gráficos y con las tablas. Por ejemplo, si se decide indicar con “Cantidad de personas” a la frecuencia absoluta, se debe usar la misma denominación en todo el trabajo para que sea uniforme. Lo mismo sucede al trabajar con números decimales, donde es necesario determinar pautas sobre la cantidad de decimales que se decide usar. El objetivo es que exista consistencia a lo largo del informe.

Particularmente al hablar de la descripción de datos a través de tablas y gráficos, los títulos, las fuentes y los nombres de los ejes juegan un rol fundamental, ya que depende de ellos la claridad con la que se transmite la información. Se recomienda que a lo largo del trabajo sea uniforme la forma de redacción y el formato, teniendo en cuenta las siguientes características:

- Los títulos deben responder a las preguntas qué, cómo, cuándo y dónde fueron recopilados los datos (por ejemplo: Distribución de los encuestados según la franja horaria en la que realizaron su viaje. Rosario, año 2008).
- Las fuentes deben indicar de dónde provienen los datos que se presentan (por ejemplo: “Fuente: elaboración propia a partir de datos suministrados por la Municipalidad de Rosario”). Como alternativa, la fuente de obtención de los datos puede ser indicada al comienzo del trabajo y entonces no hace falta que figure al pie de cada tabla o figura.
- En los gráficos, cada eje debe contar con su nombre y la unidad de medida correspondiente.

Por su parte, las tablas deben llevar una fila con el encabezado de las columnas y una fila de totales, y en la primera columna deben aparecer los valores que toma la variable o bien los niveles que la constituyen. Es fundamental que estos valores se expresen en términos del problema, y que no se utilicen las etiquetas o códigos con los que fueron ingresados en la base de datos (“Femenino”, “Masculino” en lugar de “1”, “2”, por ejemplo). Las cantidades dentro de las tablas deben alinearse a la derecha. Debe analizarse siempre de antemano la naturaleza de cada variable, el

tipo de valores que puede tomar, qué representa la misma y usar la lógica y los conocimientos adquiridos durante el cursado para determinar el tipo de análisis que se debe aplicar en cada caso.

Las variables cuantitativas en general son susceptibles de un análisis más amplio que las variables cualitativas. Permiten gran variedad de medidas resumen, pero ¡ajojo!: deben emplearse medidas que se correspondan con la forma de la distribución de la variable. A su vez, permiten los cálculos de distribuciones acumuladas, que pueden aportar información enriquecedora.

No es necesaria la interpretación de una fila de la tabla, sino más bien se prefiere una interpretación global de la tabla presentada (por ejemplo, se puede hablar en general de la forma de la distribución si la variable lo permite, o de las medidas descriptivas correspondientes para cada caso, o resaltar algunas frecuencias que resulten llamativas, si las hay, etc).

Para las variables cuantitativas continuas, debe haber correspondencia entre la tabla de distribución de frecuencias, el histograma y el gráfico acumulativo en lo que respecta a la amplitud de los intervalos elegidos.

Como paso final, es importante la lectura del informe, intentando realizar una revisión crítica sobre lo ya hecho. Deben revisarse todas las pautas que fueron propuestas como consignas y recomendaciones, y también aquellas que hayan surgido con el criterio propio o el consenso del grupo de autores.