

# *Análisis Univariante: Accidentes de tránsito en Guatemala*

*Prof. Miguel Lafuente Blasco*

## 1. Contexto del estudio

En este ejercicio práctico utilizaremos un conjunto de datos real (`Accidentes Guatemala.csv`) que contiene información de víctimas de accidentes de tránsito recopilada por la Policía Nacional Civil (PNC) de Guatemala entre enero y junio de 2023.

Guatemala, al igual que muchos países de su entorno, sufre un serio problema de siniestralidad vial con fuertes implicaciones sociales. Los accidentes de tráfico tienen un impacto significativo en las comunidades: además de la pérdida de vidas, generan lesiones, discapacidades y afectación socioeconómica a las familias.

A lo largo de esta práctica se analizarán las características de las víctimas con el objetivo de trazar un perfil de la víctima típica. Como se verá más adelante, un dato especialmente relevante en el contexto guatemalteco es la alta implicación de las motocicletas en los accidentes de tránsito: cerca de la mitad de los casos registrados involucran este tipo de vehículo. Además, ciertos grupos —como los jóvenes conductores, en su mayoría hombres— podrían estar particularmente expuestos. De hecho, las estadísticas nacionales muestran de forma consistente que los hombres sufren casi el doble de accidentes que las mujeres, y que los jóvenes de entre 18 y 29 años son el grupo más afectado por muertes en el tránsito.

Analizar **quiénes** son las víctimas (edad, género, etc.) y **cómo** ocurren los accidentes (tipo de vehículo, hora del día, entre otros) es fundamental para identificar patrones de riesgo y orientar mejor las intervenciones en prevención y atención. Desde la perspectiva del Trabajo Social, además, estos patrones permiten reflexionar sobre las causas estructurales del problema: las condiciones de movilidad, los comportamientos de riesgo, las desigualdades en el acceso a medios de transporte seguros y la necesidad de políticas públicas centradas en la seguridad vial.

La descripción detallada de las variables incluidas en el estudio se encuentra en el Anexo A.

## 2. Cuestiones

### 2.1. Individuos del estudio y familiarización con los datos

- Describe si estos datos corresponden a una muestra o a un censo, y señala quién es el individuo estadístico (casos) de este estudio.

Además de los análisis cuantitativos, es útil a veces acercarse a los datos de manera cualitativa, *observando casos individuales*. Esto ayuda a humanizar las estadísticas y a verificar la consistencia de los datos.

- Localiza en la base de datos las filas (individuos) del número 9 al 20. Sin realizar cálculos estadísticos, lee la información de esas filas. ¿Notas que comparten algo? ¿Cómo explicarías esas similitudes?

- (c) Observa al segundo individuo de la base de datos: resultó fallecido pero no fue hospitalizado. ¿Cómo explicarías esto? Este aspecto debe ser tenido en cuenta a la hora de interpretar la variable Hospitalizado.

## 2.2. Edad de las víctimas

La edad de las personas involucradas en accidentes de tránsito es un factor crucial. Nos indica qué grupos etarios están más expuestos o vulnerables en la siniestralidad vial. En Guatemala, suele preocupar especialmente la incidencia en poblaciones jóvenes frente a adultas mayores, ya que esto puede relacionarse con estilos de vida, tipos de movilidad (por ejemplo, jóvenes que usan motocicleta para trabajar o estudiar) y otros factores de riesgo.

- (a) Resume mediante un gráfico adecuado y una tabla las edades de las víctimas. Realiza para ello una agrupación de las edades por décadas.
- (b) Examina las edades mínima y máxima, calcula la edad promedio, la mediana, el coeficiente de variación y el coeficiente de asimetría.
- (c) A la vista de los resultados de los apartados anteriores, da una breve explicación sobre las edades de las víctimas.
- (d) Realiza ahora una tabla resumen alternativa donde se clasifique a las personas según su grupo etario:
- **Menores de edad:** 0–17 años
  - **Jóvenes adultos:** 18–29 años
  - **Adultos:** 30–59 años
  - **Personas mayores**<sup>1</sup>: 60 años o más

Para ello, tienes dos alternativas en Excel:

- (1) Usa la opción de las tablas dinámicas en Excel para calcular cuántas personas hay en cada intervalo de edad. Para ello, deberás manipular convenientemente la amplitud y el inicio del primer intervalo para conocer, de uno en uno, la cantidad de víctimas en cada intervalo. A partir de estas cifras, construye la tabla de frecuencias.
- (2) Crea una nueva variable (columna) que agrupe las edades en los intervalos de edad mediante la función SI().
- 1) Supón que la edad de cada víctima está en la columna A; puedes copiar y pegar la información en una nueva hoja, empezando por la celda A2.
  - 2) En la nueva columna (por ejemplo, B2), escribe la siguiente fórmula:  

```
=SI(A2<18, "Menores de edad",  
SI(A2<30, "Jóvenes adultos",  
SI(A2<60, "Adultos",  
"Personas mayores")))
```

<sup>1</sup>En países con una esperanza de vida no tan alta, suele considerarse como personas mayores las mayores de 60 años en lugar de 65.

3) Arrastra esta fórmula hacia abajo en toda la columna para clasificar todas las filas del archivo.

- (e) A partir de la tabla anterior, habiendo calculado cuántas personas hay en cada grupo, calcula además qué porcentaje representan sobre el total. Reflexiona: ¿Qué grupo etario resulta más afectado? ¿Cuál crees que es la razón de que esto sea así?

### 2.3. Sexo de las víctimas

El análisis por sexo de las víctimas permite visibilizar posibles desigualdades en la siniestralidad vial. En muchos países, incluido Guatemala, es común que los hombres estén más involucrados en accidentes de tránsito que las mujeres. Esto puede deberse a diversos factores: mayor cantidad de hombres conduciendo vehículos (especialmente motocicletas y transporte de carga), diferentes patrones de movilidad o conductas de riesgo asociadas a los roles de género, entre otros. Es importante, por tanto, verificar si nuestra base de datos refleja esta tendencia.

- (a) ¿Qué sexo predomina entre las víctimas de estos accidentes registrados entre enero y junio de 2023? Obtén la frecuencia absoluta de cada categoría de la variable **Sexo**, su porcentaje relativo, y representa los datos con un gráfico adecuado para su mejor comprensión.
- (b) Comparación proporcional: Si un sexo aparece con mucha mayor frecuencia que el otro, cuantifica esa diferencia. Por ejemplo, ¿los hombres duplican, triplican, etc., a las mujeres en número de casos? Contrasta tus hallazgos con datos conocidos: a nivel nacional, los hombres suelen representar alrededor de dos tercios o más de las víctimas de tránsito, según informes del gobierno de Guatemala. ¿Coincide tu análisis con esa proporción aproximada?

**Reflexión sociológica:** Consideremos brevemente por qué podría existir la diferencia observada. Si los hombres son mayoritarios, ¿qué factores sociales podrían influir? Pensemos en la distribución de roles: es más frecuente que los hombres conduzcan vehículos de dos ruedas, trabajen como pilotos de transporte público o de carga, o adopten conductas de mayor riesgo vial. Por otro lado, las mujeres representan una minoría entre las víctimas. Una rápida reflexión sobre los posibles motivos puede apuntar a que muchas mujeres viajan más como pasajeras o en transporte público en lugar de conducir, tienden a ser más precavidas al volante o están menos presentes en modalidades de mayor riesgo, como las motocicletas. Como se puede vislumbrar, los roles de género y las oportunidades de movilidad influyen directamente en la exposición al riesgo.

### 2.4. Tipo de vehículo involucrado y cómo suceden los accidentes

El tipo de vehículo en el que viajaba la víctima (o con el que estuvo involucrada) es otro eje de análisis fundamental. Diferentes vehículos implican distintos niveles de exposición y protección: por ejemplo, un motociclista está mucho más expuesto físicamente que alguien que viaja dentro de un automóvil; un peatón o ciclista lo está aún más. Dado que en Guatemala las motocicletas son un medio de transporte muy extendido, especialmente entre la población joven y trabajadora, no es sorprendente que tengan una participación elevada en los accidentes. Veamos qué revela la base de datos al respecto.

- (a) Examina la variable **Tipo Vehículo**. ¿Cuáles son las categorías presentes (por ejemplo: *motocicleta*, *automóvil*, *pickup*, etc.) y cuántos casos hay en cada categoría? Elabora una tabla de frecuencias para responder.
- (b) Elabora un gráfico de barras que permita visualizar claramente qué tipo de vehículo está más frecuentemente involucrado en los accidentes con víctimas. Asegúrate de que los nombres de las categorías sean legibles en el gráfico.  
*Nota:* En Excel puedes girar las etiquetas del eje o cambiar la orientación del gráfico para facilitar la lectura.
- (c) ¿Es la motocicleta el tipo de vehículo con mayor número de víctimas en estos datos? Si es así, determina aproximadamente qué porcentaje del total de víctimas viajaba en motocicleta. Si se observa que cerca de la mitad o más de los casos corresponden a motociclistas, estaríamos ante un hallazgo coherente con los datos nacionales.
- (d) Identifica el segundo y el tercer tipo de vehículo más frecuentes después de la motocicleta.
- (e) Redacta unas breves conclusiones sobre la variable **Tipo Vehículo**, considerando qué implicaciones sociológicas o de movilidad podría tener la distribución observada.

Por otro lado, también resulta interesante analizar qué tipo de eventos viales están detrás de los accidentes más comunes. Para ello, observa la variable **Tipo Evento**, que indica cómo ocurrió el accidente (por ejemplo, *colisión*, *choque*, *derrape*, entre otros).

- (f) Realiza una tabla de frecuencias con las distintas categorías de **Tipo Evento**. ¿Cuál es el evento más frecuente?
- (g) Representa visualmente los datos con un gráfico de barras que muestre claramente la frecuencia de cada tipo de evento.
- (h) Redacta una conclusión clara en dos o tres frases sobre qué tipos de accidentes predominan en los datos y cómo se podrían interpretar esos resultados desde la realidad guatemalteca.

## 2.5. Momento del accidente

Comprender cuándo ocurren más accidentes permite identificar patrones temporales relevantes: por ejemplo, puede haber picos en horas de tráfico intenso, en altas horas de la noche —potencialmente asociados a fatiga o consumo de alcohol— o durante los fines de semana. Conocer estos momentos de mayor riesgo facilita la planificación de acciones preventivas, el refuerzo de la vigilancia vial y una mejor preparación de los servicios de emergencia.

- (a) Utiliza la variable **Hora Ocurrencia** para representar, sin agrupar, un gráfico de barras o un polígono de frecuencias que muestre cuántos accidentes ocurrieron en cada hora del día. Interpreta el gráfico: ¿hay horas especialmente conflictivas?
- (b) Elabora una tabla de frecuencias agrupando las horas en tramos de 6 horas (por ejemplo, 0–5, 6–11, etc.). ¿Qué tramo concentra más siniestros?

- (c) Ahora trabaja con la variable **Franja Horaria**. Elabora su tabla de frecuencias y compárala con los resultados del apartado anterior. ¿Coinciden? ¿Por qué crees que se da esa correspondencia (o diferencia)?
- (d) Representa las frecuencias porcentuales de los accidentes por franja horaria en un gráfico adecuado. Antes de ordenar el gráfico, reflexiona: ¿cuál debe ser la primera franja del día desde una mirada social? ¿Empieza el día a las 0:00 (madrugada)? ¿O sería más adecuado el orden empezando por la mañana? En base a esa decisión, ordena la tabla y el gráfico.
- (e) A la luz de los datos, identifica las horas o franjas más peligrosas. ¿Por qué podrían concentrarse allí más accidentes? Aporta alguna interpretación sociológica, por ejemplo:
- Si los accidentes se concentran en *madrugada o noche*, considera factores como menor visibilidad, fatiga, consumo de alcohol, menor presencia policial o peores condiciones de iluminación.
  - Si se dan más en la *tarde*, piensa en el tráfico de salida del trabajo, el cansancio acumulado o la congestión urbana.
- (f) Usando la variable **Día semana**, construye una tabla de frecuencias y un gráfico que represente visualmente la distribución de los accidentes a lo largo de la semana.
- (g) ¿Qué día presenta la mayor cantidad de accidentes? ¿Y cuál la menor?
- (h) Calcula el porcentaje de accidentes ocurridos durante el fin de semana (sábado y domingo). ¿Es una proporción elevada o reducida? ¿Qué proporción de accidentes debería haber si los accidentes tuvieran la misma probabilidad que entre semana?
- (i) A la luz de los datos, ¿podemos decir que los accidentes se concentran más entre semana o durante el fin de semana? ¿Por qué crees que ocurre esta distribución?
- (j) Dado que tenemos los datos de enero a junio y estos están recogidos en la variable **Mes**, ¿observas diferencias entre meses? Cuidado con sobreinterpretar, pues medio año es poca ventana para analizar la estacionalidad, pero si hubiera un pico notable en un mes (tal vez relacionado a vacaciones, festividades, clima, etc.), vale la pena mencionarlo.

## 2.6. Resultado del Accidente y gravedad de las consecuencias

El desenlace de un accidente puede ser muy diferente: algunas personas resultan **lesionadas** y otras, lamentablemente, **fallecen**. Además, en ciertos casos la persona es **internada** (hospitalizada), y en otros no. Estas variables (**Resultado Accidente** y **Hospitalizado**) permiten hacernos una idea de la gravedad de los siniestros, aunque no contemos con el detalle del diagnóstico clínico.

**Nota:** En este apartado no se estudian relaciones entre variables —eso se hará en una actividad posterior dedicada al análisis bivariante—, sino que se describe cada variable de forma independiente.

- (a) Comienza con la variable **Resultado Accidente**. Elabora una tabla de frecuencias: ¿cuántos casos resultaron **lesionados** y cuántos **fallecidos**? Calcula también los porcentajes correspondientes.

- (b) Representa esta variable mediante un gráfico adecuado. Usa los colores de forma clara y coherente. ¿Cuál es el desenlace más común en el conjunto de víctimas?
- (c) Repite el procedimiento con la variable **Hospitalizado**: ¿cuántas personas fueron **internadas** y cuántas no? ¿Qué porcentaje representa cada grupo? Reflexiona: ¿cómo de frecuente es que una víctima deba ser hospitalizada?

### 3. Conclusiones

A partir de los apartados anteriores, redacta un párrafo que sintetice el perfil de la víctima más común en estos datos: tanto desde el punto de vista sociodemográfico (sexo, edad), como por las características del accidente (tipo de vehículo, momento del día o de la semana) y su desenlace (lesión, fallecimiento, hospitalización).

## A. Descripción de las variables

A continuación se describen las variables presentes en el conjunto de datos:

- **Hora Ocurrencia:** Indica la hora en la que ocurrió el accidente. Ejemplo: *15*.
- **Franja Horaria:** Franja del día en la que ocurrió el accidente. Ejemplo: *Tarde*.
- **Mes:** Mes en el que ocurrió el accidente. Ejemplo: *Enero*.
- **Día semana:** Día de la semana en el que ocurrió el accidente. Ejemplo: *Domingo*.
- **Departamento:** Departamento en el que ocurrió el accidente. Ejemplo: *Guatemala*.
- **Sexo Víctima:** Sexo de la víctima del accidente. Ejemplo: *Hombre*.
- **Edad Víctima:** Edad de la víctima al momento del accidente. Ejemplo: *17*.
- **Mayoría Edad Víctima:** Indica si la víctima es mayor o menor de edad. Ejemplo: *Menor*.
- **Tipo Vehículo:** Tipo de vehículo involucrado en el accidente. Ejemplo: *Motocicleta*.
- **Tipo Evento:** Tipo de evento de tránsito que resultó en el accidente. Ejemplo: *Colisión*.
- **Resultado Accidente:** Desenlace del accidente para la víctima. Ejemplo: *Fallecido*.
- **Hospitalizado:** Indica si la víctima fue hospitalizada o no. Ejemplo: *Internado*.