

# **Análisis estadístico descriptivo de los accidentes de circulación en Madrid en 2024**

PL1 – Estadística – Curso 2024/25  
Grado en Ingeniería de Computadores

Roberto Pais Bustamante  
Pablo García Hernández

# ÍNDICE

---

<b>1. Introducción .....</b>	<b>Pág. 3</b>
<b>2. Metodología .....</b>	<b>Pág. 4</b>
2.1. Origen de los datos .....	
2.2. Herramientas .....	
2.3. Diseño .....	
<b>3. Resultados .....</b>	<b>Pág. 5</b>
<b>4. Análisis de resultados .....</b>	<b>Pág. 8</b>
<b>5. Conclusiones .....</b>	<b>Pág. 10</b>
<b>6. Referencias .....</b>	<b>Pág. 10</b>

# 1. Introducción

España es un lugar en el que se circula muy a menudo debido a diferentes razones, ya sea para ir al trabajo, visitar a la familia o incluso para ir a un concierto. Madrid, la capital de España, cuenta con uno de los mayores números de ciudadanos del país, por lo que resulta frecuente que ocurran accidentes a menudo por la zona, siendo en España uno de los lugares con mayor número de accidentes anualmente.

El objetivo de esta investigación llevada a cabo ha sido realizar un análisis detallado sobre los accidentes de circulación que han ocurrido en las carreteras de los diferentes distritos de Madrid este año (2024), así pudiendo comprobar si existen diferencias, similitudes, o cualquier otro dato interesante entre los diferentes distritos de Madrid.

La investigación se limita a los accidentes de circulación ocurridos en Madrid en el año 2024, desde el primer día de enero hasta la fecha en la que se ha realizado este trabajo (14/10/2024).

## 2. Metodología

### 2.1. Origen de los datos

Este trabajo se ha realizado basándose del archivo “2024\_Accidentalidad.csv”, un archivo del portal de datos del Ayuntamiento de Madrid (sitio web al que puedes acceder pinchando en el primer enlace en el apartado de referencias<sup>6</sup>), en el que se han recopilado todos los accidentes ocurridos en Madrid este año.

En dicho archivo se encuentra diferentes tipos de datos sobre cada accidente, como la fecha, el género del conductor, o el tipo de vehículo que conducían.

Para este trabajo nos hemos centrado en estudiar los datos en base a los distritos donde ocurrieron los accidentes, y, adicionalmente, en el sexo del conductor involucrado.

### 2.2. Herramientas

Estos datos han sido procesados mediante el lenguaje de programación R, utilizando como entorno de desarrollo el programa RStudio para Windows, versión 2022.12.0.

### 2.3. Diseño

El resumen del diseño del trabajo es el siguiente:

- **Variable estadística:** Las variables estadísticas que hemos utilizado para este trabajo son el “distrito” y el “sexo”, ambas de tipo cualitativo, observando la frecuencia en la que ocurren los accidentes en cada distrito y el sexo del conductor involucrado.
- **Población:** Para este trabajo se ha utilizado una población de 31357 accidentes ocurridos en diferentes carreteras de los distritos de Madrid.
- **Organización de datos:** Se han eliminado del archivo “2024\_Accidentalidad.csv” cualquier dato erróneo que tuviera.

### 3. Resultados

A continuación, se presentan las siguientes tablas y graficas sobre los resultados obtenidos al analizar los datos propuestos para este trabajo:

- **Tablas de resultados**

**Tabla 1. Medidas estadísticas descriptivas**

Medida	Todos	Mujer	Hombre	Desconocido
Tamaño muestra	31357	9053	18935	3369
<b>Medidas de centralización</b>				
Media	1493	431	902	160
Mediana	1522	419	905	150
<b>Medidas de dispersión</b>				
Mínimo	531	153	307	61
Máximo	2547	712	1485	350
Rango	2016	559	1178	289
Desviación típica	539.16	161.61	337.96	71.59
Coefficiente de variación	0.3611	0.3750	0.3747	0.4474
<b>Medidas de posición</b>				
Primer cuartil (Q1)	1183	372	634	114
Tercer cuartil (Q3)	1796	574	1110	187
Rango intercuartílico (IQR)	613	202	476	73
<b>Medidas de forma</b>				
Coefficiente asimetría	-0.0149	0.0302	-0.1029	1.1009
Coefficiente apuntamiento (curtosis)	-0.7493	-0.9626	-1.0458	0.8813

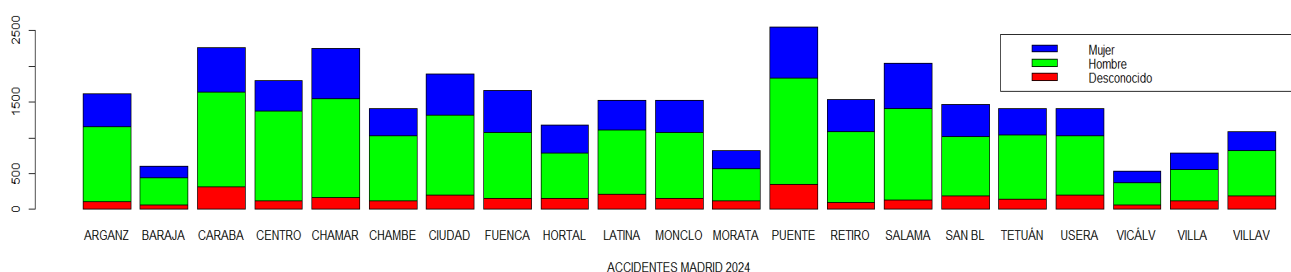
**Tabla 2. Distribuciones de frecuencia absoluta de los accidentes en cada distrito según el sexo del conductor**

DISTRITO \ SEXO	Desconocido	Hombre	Mujer	Sum
ARGANZUELA	103	1048	469	1620
BARAJAS	66	379	153	598
CARABANCHEL	318	1321	617	2256
CENTRO	124	1253	419	1796
CHAMARTÍN	164	1387	699	2250
CHAMBERÍ	122	905	377	1404
CIUDAD LINEAL	205	1110	574	1889
FUENCARRAL-EL PARDO	157	923	578	1658
HORTALEZA	157	630	396	1183
LATINA	215	898	417	1530
MONCLOA-ARAVACA	150	928	444	1522
MORATALAZ	114	458	253	825
PUENTE DE VALLECAS	350	1485	712	2547
RETIRO	98	988	452	1538
SALAMANCA	136	1279	625	2040
SAN BLAS-CANILLEJAS	187	835	444	1466
TETUÁN	146	897	372	1415
USERA	199	832	383	1414
VICÁLVARO	61	307	163	531
VILLA DE VALLECAS	114	438	235	787
VILLAVERDE	183	634	271	1088
Sum	3369	18935	9053	31357

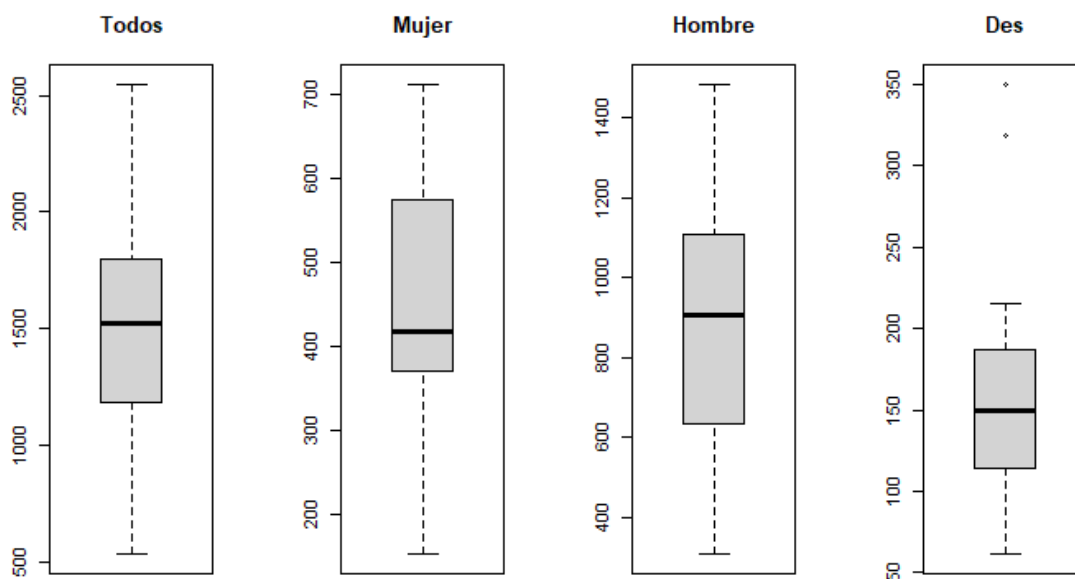
**Tabla 3. Distribuciones de frecuencia relativa de los accidentes en cada distrito según el sexo del conductor**

DISTRITO \ SEXO	Desconocido	Hombre	Mujer	Sum
ARGANZUELA	0.003285	0.033422	0.014957	0.051663
BARAJAS	0.002105	0.012087	0.004879	0.019071
CARABANCHEL	0.010141	0.042128	0.019677	0.071946
CENTRO	0.003954	0.039959	0.013362	0.057276
CHAMARTÍN	0.00523	0.044233	0.022292	0.071754
CHAMBERÍ	0.003891	0.028861	0.012023	0.044775
CIUDAD LINEAL	0.006538	0.035399	0.018305	0.060242
FUENCARRAL-EL PARDO	0.005007	0.029435	0.018433	0.052875
HORTALEZA	0.005007	0.020091	0.012629	0.037727
LATINA	0.006857	0.028638	0.013298	0.048793
MONCLOA-ARAVACA	0.004784	0.029595	0.01416	0.048538
MORATALAZ	0.003636	0.014606	0.008068	0.02631
PUENTE DE VALLECAS	0.011162	0.047358	0.022706	0.081226
RETIRO	0.003125	0.031508	0.014415	0.049048
SALAMANCA	0.004337	0.040788	0.019932	0.065057
SAN BLAS-CANILLEJAS	0.005964	0.026629	0.01416	0.046752
TETUÁN	0.004656	0.028606	0.011863	0.045125
USERA	0.006346	0.026533	0.012214	0.045094
VICÁLVARO	0.001945	0.00979	0.005198	0.016934
VILLA DE VALLECAS	0.003636	0.013968	0.007494	0.025098
VILLAVERDE	0.005836	0.020219	0.008642	0.034697
Sum	0.10744	0.603852	0.288707	1

## - Gráficas de resultados



*Figura 1. Diagrama de barras de accidentes por distrito de Madrid y según el sexo*



*Figura 2. Diagrama de cajas de accidentes en Madrid según el sexo*

## 4. Análisis de los resultados

De los resultados obtenidos (tabla 1) y observando los diagramas obtenidos (figura 2), podemos observar que tanto la media como la mediana de los hombres es mucho más mayor que el de las mujeres o de sexo desconocido, llegando a ser más que el doble la media y mediana de los hombres (902 y 905 respectivamente) que de las mujeres (431 y 419).

Viendo la tabla de distribuciones de frecuencias absolutas (tabla 2), podemos observar que el distrito en el que ocurren el mayor número de accidentes es Puente de Vallecas, con un total de 2547 accidentes, mientras que el distrito con menos accidentes es Vicálvaro, con un total de 531 accidentes (en Puente de Vallecas han ocurrido aproximadamente 4,8 veces más accidentes que en Vicálvaro este año, por lo que es mucho más peligroso conducir en Puente de Vallecas que en Vicálvaro).



Mientras que los conductores de sexo desconocido y las mujeres tienen una media más grande que su mediana, se puede notar que para los hombres es lo contrario, aunque la diferencia no sea muy grande, indicando que de promedio las mujeres y de sexo desconocido tienen menos accidentes en distritos que lo que representaría la media, mientras que para los hombres resultaría al revés.

Se puede observar que el rango de los hombres (1178) es mucho más grande que el de las mujeres (559) y los de sexo desconocido (289), y esto también resulta similar cuando vemos la desviación típica. Pero al observar los coeficientes de variación, podemos ver que el de las mujeres (0.3750) y hombres (0.3747) son similares, mientras que el de los desconocidos es mayor (0.4474), por lo que se puede ver que los datos de mujeres y hombres son homogéneos y la media representa mejor los datos, mientras que en los conductores de sexo desconocido los datos son más heterogéneos y la media no es tan buen representante.

En cuanto al diagrama de cajas (figura 2), podemos observar que el 50% de los datos de los accidentes de mujeres rondan entre 372 y 574, mientras que en el de los hombres son 634 y 1110, y en sexo desconocido es 114 y 187, esto indica que hay mucha dispersión entre estos datos en los hombres, mientras que en los de sexo desconocido la dispersión es muy poca. Eso sí, cabe mencionar que el diagrama de los de sexo desconocido cuenta con dos datos atípicos, mientras que el resto no cuenta con ninguno, por lo que existen dos distritos muy extremos y distintos al promedio de accidentes de género desconocido.

Finalmente, en el diagrama de caja podemos observar que los bigotes del diagrama de los hombres son muy similares, por lo que los accidentes son más repartidos y homogéneos durante el año, mientras que tanto el de las mujeres como el de los de sexo desconocido tienen el bigote inferior más largo que el superior (sobre todo el de sexo desconocido), por lo que la dispersión de accidentes en ambos diagramas es mayor que en el diagrama de los hombres.

Por último, podemos observar que, entre las 3, solo el coeficiente asimétrico de los hombres es negativo, mientras que el de las mujeres y desconocidos es positivo, indicando que los números de accidentes más altos de los hombres ocurren después de la media, ocurriendo lo contrario para el resto. Eso sí, cabe mencionar que aun así el coeficiente de asimetría total es negativo, por lo que, aunque los hombres solo tengan el coeficiente de asimetría negativo, el número

de accidentes de hombres total es tan alto que supera a los otros dos, dejando un resultado negativo en el coeficiente de asimetría total.

Por otro lado, el coeficiente de apuntamiento de los conductores desconocidos es positivo, siendo la distribución de frecuencias de estos “leptocúrtica”, mientras que el resto (incluido el total), es negativo, siendo la distribución “platicúrtica”.

## 5. Conclusiones

A partir del análisis de los resultados obtenidos, se puede concluir que, entre otras cosas, la zona más peligrosa para conducir en Madrid es Puente de Vallecas, con mayor número de accidentes que el resto de los distritos de otras zonas. Los datos totales, aunque cuenten con un rango grande, son homogéneos, la curva de la asimetría es “platicúrtica” y, específicamente para los conductores de sexo desconocido, existen dos datos atípicos relevantes, destacando dos distritos con números de accidentes mucho mayores que el resto.

## 6. Referencias

A continuación, aquí tienes los enlaces que utilizamos y que sirven de referencia para entender mejor el trabajo realizado:

<https://datos.madrid.es/sites/v/index.jsp?vgnextoid=7c2843010d9c3610VgnVCM2000001f4a900aRCRD&vgnnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD>

[https://uah.blackboard.com/ultra/courses/\\_35576\\_1/outline/file/\\_2812841\\_1](https://uah.blackboard.com/ultra/courses/_35576_1/outline/file/_2812841_1)