



**Universidad de Valladolid**



**ESCUELA DE INGENIERÍAS  
INDUSTRIALES**

**UNIVERSIDAD DE VALLADOLID**

**ESCUELA DE INGENIERIAS INDUSTRIALES**

**Grado en Ingeniería de Organización Industrial**

**Caracterización de intersecciones y tramos curvos  
urbanos que deben ser protegidos ante posibles  
salidas de calzada de vehículos motorizados**

**Autor:**

**Álvarez Salamanqués, Patricia**

**Tutor:**

**Fernández Temprano, Miguel Alejandro**

**Gordaliza Ramos, Alfonso**

**Estadística e Investigación Operativa**

**Valladolid, Julio 2016**



*A toda mi familia, por su apoyo  
incondicional.*

*A Carlos y a todos aquellos amigos que me  
han acompañado durante esta etapa.*

*A mis tutores, por todo lo que me han  
enseñado a lo largo de estos meses.*



## **RESUMEN**

Las salidas de calzada son el tipo de accidente con una mayor letalidad en entorno urbano y suponen de media sesenta víctimas mortales anualmente en España (según datos de la DGT), siendo las intersecciones y las curvas los dos escenarios más habituales donde suceden.

En el presente Trabajo Fin de Grado se evaluará la influencia de distintos parámetros geométricos, considerados como posibles factores de riesgo, sobre la ocurrencia de las salidas de calzada. Para conseguir este objetivo se llevará a cabo un estudio caso-control, que será analizado mediante modelos estadísticos de regresión logística y que permitirán valorar la asociación entre las diferentes variables seleccionadas y el desenlace estudiado.

De esta forma, se pretende ayudar a las administraciones titulares de las vías para que determinen de manera fácil y rápida qué intersecciones o curvas deben estar protegidas mediante las adecuadas barreras de contención, y así evitar posibles salidas de calzada.

## **PALABRAS CLAVE**

Accidentes de tráfico urbanos, Factores de riesgo, Estudios caso-control, Regresión logística, Validación de modelos.



## **ABSTRACT**

Traffic accidents caused due to the car leaving the road are the type of accident with higher lethality in an urban environment and involve, on average, sixty fatal victims annually in Spain. It is important to take into account that intersections and curves are the most common scenarios where that kind of accident happens.

In this document, the influence of different geometric parameters, considered as possible risk factors on the occurrence of swerves off the road, will be evaluated. To achieve this objective, a case-control study will be performed, which will be analyzed by logistic regression. Logistic models will allow to assess the association between different variables selected and the outcome studied.

Thus, it is intended to help administrations, in terms of road safety, to determine quickly and easily what intersections or curved sections must be protected by the appropriate barriers, in order to avoid possible swerves off the road.

## **KEY WORDS**

Urban traffic accidents, Risk factors, Case-control studies, Logistic regression, Model validation.



# ÍNDICE

---

<b>INTRODUCCIÓN.....</b>	<b>1</b>
ANTECEDENTES .....	3
JUSTIFICACIÓN.....	5
OBJETIVOS .....	6
<b>CAPÍTULO 1. MÉTODOS ESTADÍSTICOS UTILIZADOS .....</b>	<b>9</b>
1.1. ESTUDIOS CASO-CONTROL .....	9
1.1.1. <i>Definición de estudio caso-control</i> .....	9
1.1.2. <i>Utilidad de los estudios caso-control</i> .....	13
1.1.3. <i>Selección de los controles y sesgo</i> .....	14
1.1.4. <i>Medida de la asociación: Odds-Ratio</i> .....	15
1.1.5. <i>Adecuación de la metodología caso-control al presente estudio</i> ...	18
1.2. REGRESIÓN LOGÍSTICA .....	19
1.2.1. <i>Conceptos básicos del modelo logístico</i> .....	19
1.2.2. <i>Regresión logística aplicada a estudios caso-control</i> .....	26
1.3. VALIDACIÓN LEAVE-ONE-OUT Y ANÁLISIS ROC.....	27
1.3.1. <i>Validación Leave-One-Out</i> .....	27
1.3.2. <i>Análisis ROC</i> .....	28
<b>PARTE I. OBTENCIÓN DE LA MUESTRA.....</b>	<b>33</b>
<b>CAPÍTULO 2. BASES DE DATOS DE PARTIDA.....</b>	<b>35</b>
2.1. DESCRIPCIÓN DE LA BASE DE DATOS “ARENA” .....	35
2.2. DESCRIPCIÓN DE LA BASE DE DATOS “PIMUSSVA” .....	40
<b>CAPÍTULO 3. OBTENCIÓN DE LOS CASOS Y CONTROLES .....</b>	<b>45</b>
3.1. OBTENCIÓN DE LOS CASOS.....	45
3.1.1. <i>Obtención de los accidentes de interés en la base de datos “ARENA”</i> .....	45
3.1.2. <i>Obtención de los accidentes de interés en la base de datos del</i> <i>“PIMUSSVA”</i> .....	47
3.1.3. <i>Unificación de los casos</i> .....	49
3.2. OBTENCIÓN DE LOS CONTROLES .....	51

<b>CAPÍTULO 4. VARIABLES DEL MODELO .....</b>	<b>55</b>
<b>PARTE II. ANÁLISIS DE LOS DATOS .....</b>	<b>65</b>
<b>CAPÍTULO 5. ANÁLISIS DESCRIPTIVO DE LAS VARIABLES Y ODDS-RATIO .....</b>	<b>67</b>
5.1. TIPO DE INTERSECCIÓN .....	68
5.2. RADIO DE CURVATURA.....	72
5.3. ANCHO DE CALZADA.....	76
5.4. ANCHO DE CARRIL.....	82
5.5. NÚMERO DE CARRILES PARA EL SENTIDO .....	87
5.6. SENTIDO DE CIRCULACIÓN.....	91
5.7. LONGITUD DEL TRAMO RECTO ANTERIOR.....	95
5.8. DISTANCIA AL SEMÁFORO ANTERIOR .....	101
5.9. PENDIENTE.....	105
5.10. PRIORIDAD.....	108
<b>CAPÍTULO 6. MODELOS DE REGRESIÓN LOGÍSTICA.....</b>	<b>115</b>
<b>CAPÍTULO 7. SELECCIÓN Y VALIDACIÓN DEL MODELO .....</b>	<b>121</b>
7.1. VALIDACIÓN <i>LEAVE-ONE-OUT</i> Y ANÁLISIS ROC EN CURVAS E INTERSECCIONES.....	122
7.2. VALIDACIÓN <i>LEAVE-ONE-OUT</i> Y ANÁLISIS ROC EN LAS INTERSECCIONES..	127
7.3. VALIDACIÓN <i>LEAVE-ONE-OUT</i> Y ANÁLISIS ROC PARA LAS CURVAS.....	133
7.4. MODELO FINAL.....	138
<b>CONCLUSIONES Y LÍNEAS FUTURAS DE TRABAJO.....</b>	<b>143</b>
CONCLUSIONES FINALES.....	145
<i>Conclusiones sobre las bases de datos.....</i>	145
<i>Conclusiones sobre los resultados finales .....</i>	146
LÍNEAS FUTURAS DE TRABAJO .....	149
<b>BIBLIOGRAFÍA .....</b>	<b>151</b>

<b>ANEXOS.....</b>	<b>157</b>
ANEXO 1. CUESTIONARIO DE ARENA .....	159
ANEXO 2. MAPA CON LOS MEDIDORES DE IMD EN VALLADOLID .....	161
ANEXO 3. ÍNDICE DE TABLAS .....	163
ANEXO 4. ÍNDICE DE IMÁGENES .....	166
ANEXO 5. ÍNDICE DE ECUACIONES .....	170





# INTRODUCCIÓN



Este trabajo se enmarca en el proyecto de investigación “Desarrollo de los criterios de aplicación de los elementos de contención de vehículos, ciclistas y peatones en intersecciones y tramos curvos de zona urbana”, que se lleva a cabo a través de la Universidad de Valladolid, junto con la spin-off Cidro de la Fundación Cidaut. El proyecto ha concurrido a la convocatoria de la Dirección General de Tráfico (en adelante DGT), realizada por Resolución de 1 de julio de 2015 («BOE» de 6 de julio), para la concesión de subvenciones destinadas al desarrollo de proyectos de investigación en el área de tráfico, movilidad y seguridad vial.

El proyecto se encuadra en dicha convocatoria dentro de la Línea 4A del Plan de Investigación de la DGT: *Análisis del comportamiento de los usuarios vulnerables y medidas de protección (Área programas de movilidad): Nuevos retos en la protección de usuarios vulnerables, en relación a variables sociodemográficas (envejecimiento, obesidad, discapacidad) y en relación a vehículos menos protectores (bicicletas, vehículos ecológicos, etc.)*, y ha sido seleccionado para su financiación con la referencia SPIP2015-01762, como se recoge en la Resolución de 24 de noviembre de 2015, de la DGT, por la que se conceden subvenciones para el desarrollo de proyectos de investigación en el área de tráfico, movilidad y seguridad vial, correspondientes al ejercicio 2015, publicado en «BOE» núm. 284, de 27 de noviembre de 2015, páginas 112154 a 112158.

## ANTECEDENTES

Las diferentes políticas a nivel nacional y europeo centran sus esfuerzos en conseguir vías urbanas más seguras (en busca de “cero lesionados”) a la vez que integradoras de los diferentes modos de transporte, con especial hincapié en los usuarios vulnerables (peatones, ciclistas y motociclistas).

Paralelamente, las cifras de accidentes muestran que se producen, anualmente, alrededor de 4000 salidas de calzada en entorno urbano dentro del territorio nacional [1], con un balance medio de 60 víctimas mortales al año. Por otro lado, si se analizan las localizaciones donde se producen estas salidas se obtiene que más del 40% se producen en intersecciones y el 20% en tramos curvos [1].

También es importante tener en cuenta que, en términos asociados a las consecuencias del accidente, los tipos más lesivos son la salida de la vía, con una letalidad (fallecidos por cada 100 víctimas) de 2,2, y después los atropellos con una letalidad de 1,8.

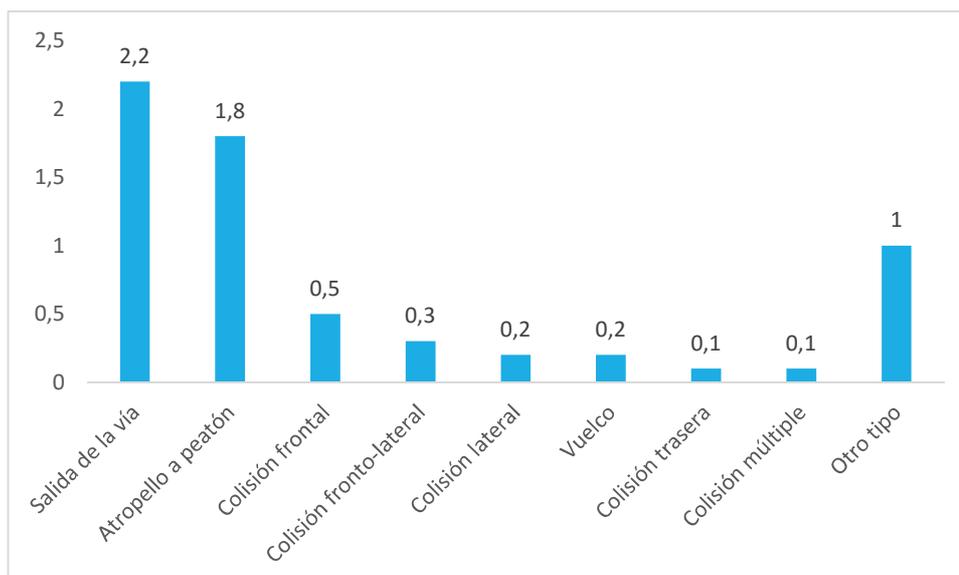


Ilustración 1. Letalidad en vías urbanas en función del tipo de accidente con víctimas. Año 2013 [1].

Por estos motivos han surgido, a nivel europeo, numerosos proyectos (SafetyNet [2], MAIDS [3], Aprosys [4], etc.) que han analizado las bases de datos nacionales existentes (CARE [5], MAIDS [3], GIDAS [6], CCIS [7], IRTAD [8], ARENA; etc.) con el fin de caracterizar los escenarios donde se producen las salidas de calzada.

A nivel nacional existen también estudios que definen estas características, como “Accidentalidad e infraestructura vial en carreteras de Diputaciones Provinciales” [9]; o incluso estudios que definen los parámetros de riesgo en lugares donde se pueden producir salidas de vía para determinados usuarios, como “Determinación de escenarios donde instalar Sistemas de Protección para Motociclistas en carreteras de la Red Regional de Castilla y León” [10].

Sin embargo, no existen estudios que hayan detallado las características geométricas (anchura de calzada, número de carriles, tipo de intersección, radio de curvatura; etc.), de tráfico (IMD – Intensidad Media Diaria) y de tipo y gravedad de los accidentes en los dos escenarios donde se producen mayor número de salidas de calzada en zona urbana, que son las intersecciones y tramos curvos.

Por otro lado, si se buscan estudios en los que se defina cómo se deberían proteger estos emplazamientos urbanos, no se encuentra documento alguno. Más aún, siendo la “Orden Circular OC 35/2014 sobre criterios de aplicación de los SCV (Sistemas de Contención de Vehículos)” [11] del Ministerio de Fomento un documento normativo que define los lugares que deben protegerse de posibles salidas de calzada de vehículos y, sobre todo, que especifica las características de contención que debe tener cualquier sistema que se utilice

para mitigar las consecuencias de estos accidentes, esta orden circular no es aplicable a zona urbana.

De igual manera, si se analiza la bibliografía existente sobre cómo deben ser los SCV en relación con la severidad hacia los usuarios vulnerables (ciclistas y peatones), se observa, por ejemplo, que el “Manual de recomendaciones para el diseño de carril bici” (DGT) [12] o documentos normativos como “Austroads Guide to Traffic – Bicycles” [13] o la “Specification for pedestrian restraint systems. BRITISH STANDARD 7818:1995” [14], no definen los requisitos que tienen que tener esos sistemas desde un punto de vista de contención del vehículo motorizado ni, sobre todo, de severidad de cara al ciclista en caso de colisionar contra ellos.

Cabe destacar que en el proyecto “Desarrollo del modelo de barrera de segregación de espacios seguros urbanos para usuarios vulnerables” (DGT 2015) [15], ya se están abarcando los requisitos en cuanto a especificaciones técnicas que deben tener los SCV, pero no los criterios de implantación de estos sistemas.

En la actualidad los diferentes sistemas, barandillas y demás elementos de infraestructura urbana se están instalando, en la mayoría de los casos, sin conocer su comportamiento ante impacto, ni las características necesarias de implantación para su buen funcionamiento: disposiciones longitudinales, condiciones del terreno, etc. Tampoco se implantan con criterios rigurosos en base al riesgo de accidente y tipo de usuarios, sino que los técnicos y empresas correspondientes determinan los criterios en base a su experiencia particular, debido a que no existe ninguna normativa al respecto.

Todo esto hace necesario desarrollar una investigación para establecer, en primer lugar, una caracterización de los tipos de intersecciones y zonas curvas que poseen un mayor riesgo a la hora de producirse salidas de la calzada urbana y, seguidamente, conocer qué criterios de aplicación deben cumplir los SCV a instalar en estos emplazamientos con el fin de proteger a todos los usuarios (tanto ocupantes de vehículos de cuatro ruedas, como usuarios vulnerables – motociclistas, peatones y ciclistas). El primero de los retos anteriores es de naturaleza eminentemente estadística, dentro del cual se enmarca este Trabajo Fin de Grado.

## JUSTIFICACIÓN

Los estudios que se desarrollan en este proyecto de investigación contribuirán a avanzar en la consecución del objetivo “cero lesionados”; así como a conseguir un crecimiento inteligente, sostenible e integrador de los distintos modos de

transporte en una ciudad, aportando una mayor seguridad en las infraestructuras viarias a la vez que mayor protección de los usuarios más vulnerables (Políticas sobre seguridad vial 2011 – 2020 de la Comisión Europea) [16].

Las conclusiones de este trabajo y del proyecto mencionado permitirán realizar una serie de recomendaciones destinadas a las administraciones que, de esta forma, tendrán una herramienta útil y rápida para poder distinguir los tipos de intersecciones o tramos curvos en vías urbanas que presentan un mayor riesgo de que se produzca una salida de vía. Igualmente, permitirán conocer las condiciones que deben cumplir los sistemas a instalar para conseguir una correcta protección de todos los usuarios, especialmente de los más vulnerables.

Por otro lado, el carácter innovador de este trabajo de investigación radica en el hecho de que hasta ahora sólo hay estudios que caracterizan los escenarios donde se han producido los accidentes de tráfico urbanos en general, pero no existe ningún estudio sobre parámetros de riesgo específicos que caractericen qué tipo de intersecciones o tramos curvos son los más susceptibles de sufrir una salida de calzada. De igual manera, este trabajo de investigación supone un nuevo reto para la protección de usuarios vulnerables (de acuerdo con la línea investigación “4A” de la Dirección General de Tráfico).

## OBJETIVOS

El proyecto de investigación en el que se enmarca este trabajo plantea como objetivo general el desarrollo de los criterios de aplicación de los elementos de contención de vehículos, ciclistas y peatones en intersecciones y tramos curvos de zona urbana, basándose en las características geométricas de intersecciones y curvas de tramo urbano de la ciudad de Valladolid.

El hecho de que el estudio se realice únicamente para salidas de calzada dentro de la ciudad de Valladolid se debe, principalmente, a dos motivos. El primero de ellos es que se trata de una ciudad de tamaño intermedio que podría ser identificada como ciudad tipo para, posteriormente y con las cautelas necesarias, tratar de extrapolar las conclusiones del estudio a otras ciudades. El segundo de los motivos radica en una imposibilidad de abordar un muestreo de accidentes de esta tipología, a nivel nacional, dentro de este Trabajo Fin de Grado.

Como ya se ha mencionado, otro objetivo importante del proyecto, derivado del anterior, es tratar de proporcionar herramientas a las administraciones titulares de las vías urbanas para que puedan determinar de manera fácil y rápida si una

determinada intersección o un determinado tramo curvo deben ser protegidos para evitar una posible salida de calzada. Igualmente, se detallarán posibles criterios de priorización dentro de los escenarios detectados en esta tarea (criterios de priorización basados en la presencia de áreas peatonales, presencia de carriles bici anexos, etc.).

A continuación, se detallan los objetivos específicos a abordar, en orden cronológico, a lo largo del proyecto de investigación en su conjunto:

- 1) Caracterizar las intersecciones y tramos curvos urbanos que deben ser protegidos ante posibles salidas de calzada de vehículos motorizados, basándose en características de la vía (parámetros de riesgo).
- 2) Detallar los requisitos de contención, severidad y accesibilidad que deben cumplir los sistemas de protección a instalar en los emplazamientos definidos en el primero de los objetivos específicos, tanto desde un punto de vista de impacto de vehículos motorizados como desde un punto de vista de impacto de usuarios vulnerables (peatones, ciclistas y motociclistas) contra esos sistemas.
- 3) Realizar una prueba piloto en dos escenarios seleccionados según los criterios desarrollados a lo largo del primer objetivo, y en base a la posterior protección con un sistema adecuado determinado con el segundo objetivo.

Como ya se ha mencionado, el propósito concreto de este trabajo se corresponde con el primero de los tres objetivos planteados en el proyecto de investigación realizado para la DGT: caracterización de intersecciones y tramos curvos urbanos en los que se producen las salidas de calzada, para así establecer los principales parámetros geométricos que influyen en que ocurra este tipo de accidente.

Para la caracterización de las intersecciones y tramos curvos urbanos se empleará la metodología estadística de estudio “caso – control”, diseñada para estudiar poblaciones en las que ciertos sujetos sufren el evento objeto de estudio (“casos”) y otros sujetos no (“controles”), pudiendo determinar así la importancia de los distintos factores de riesgo relacionados con la ocurrencia del evento. En el caso que nos ocupa, la aplicación de la mencionada metodología nos permitirá conocer qué características geométricas influyen, y de qué manera, en que en un determinado tipo de intersección o zona curva urbana haya mayor riesgo de producirse una salida de calzada urbana.



# CAPÍTULO 1. MÉTODOS ESTADÍSTICOS UTILIZADOS

---

## 1.1. ESTUDIOS CASO-CONTROL

### 1.1.1. Definición de estudio caso-control

Los estudios caso-control son ampliamente utilizados en el área de la Epidemiología, disciplina científica de extensa aplicación en el campo de la salud que se encarga de estudiar los factores que determinan la frecuencia y distribución de enfermedades en poblaciones humanas. Se apoya fundamentalmente en el método científico, utilizando de manera intensiva modelos y métodos del campo de la Estadística [17] [18] [19].

El carácter transversal de la metodología estadística hace que las herramientas de la Epidemiología se apliquen no sólo en el ámbito de las Ciencias de la Salud, sino también en otros muchos campos, con las consideraciones y adaptaciones pertinentes. Según puede verse en numerosas publicaciones (ver por ejemplo [20]), éste es también el caso del área de la Seguridad Vial, donde los elementos habituales de la Epidemiología pueden reconocerse sobre los accidentes de tráfico, considerados como sucesos que, aunque alejados del concepto habitual de enfermedad humana, suponen un desenlace no deseado provocado por un conjunto de causas o factores a estudiar [21].

A la hora de analizar la existencia de asociación (o incluso una relación ‘causa-efecto’) entre un grupo de variables (posibles factores de riesgo) y la ocurrencia de un desenlace o evento determinado, un aspecto importante es la exposición de la población a los posibles factores de riesgo. Es por ello que la Epidemiología no sólo analiza aquellos sujetos que han sufrido el desenlace objeto de estudio, sino también aquellos que no lo han sufrido. De esta forma, mediante la aplicación de los modelos estadísticos pertinentes, es posible analizar qué variables están relacionadas en mayor medida con la ocurrencia del evento estudiado.

Para materializar estos estudios epidemiológicos, y por motivos éticos claros, no podemos elegir qué sujetos estarán expuestos a unos determinados factores de riesgo ni, lo cual es más importante, cuáles sufrirán o no la “enfermedad” objeto

de estudio (Ej., efectividad de un nuevo medicamento en la lucha contra el cáncer de mama). Por esta razón, la Epidemiología ha desarrollado las metodologías necesarias para estudiar grupos de población en los que ciertos sujetos ya han sufrido el desenlace objeto de estudio (casos) y otros sujetos no (controles), pudiendo comparar así el efecto de las distintas exposiciones al riesgo [22]. Este es el caso de los estudios de casos y controles, donde los individuos se seleccionan basándose en la presencia (casos) o ausencia (controles) de una “enfermedad” cuya etiología queremos estudiar. A continuación, se valora la presencia o no de factores de riesgo (exposición) en el pasado, para estimar si existe una mayor prevalencia de exposición en “enfermos” en comparación con los sujetos “sanos”.

Otro motivo que avala el interés de los estudios caso-control es el hecho de que en algunos casos los estudios tradicionales de cohorte pueden ser problemáticos como, por ejemplo, cuando se investigan resultados que sólo se producen en una pequeña fracción de la población. También pueden ser problemáticos cuando es caro o muy difícil obtener información sobre la exposición de una población amplia. En estas situaciones, un diseño de casos y controles ofrece una alternativa que es mucho más eficiente. El objetivo de los estudios caso-control es el mismo que el de los estudios de cohorte; es decir, estimar la magnitud de la asociación entre una exposición y un resultado. Sin embargo, los estudios de casos y controles emplean una estrategia de muestreo diferente que les da una mayor eficiencia [23].

Para entender mejor las diferencias entre ambos diseños, dada una población de partida y aplicando un estudio de cohorte, se definirían los grupos de individuos expuestos y no expuestos y de ellos se obtendrían las tasas de incidencia o riesgo que se calcularían para cada cohorte. El objetivo del análisis sería identificar el número de casos que ocurren en cada grupo de expuestos o no expuestos y calcular la tasa de riesgo o la incidencia de cada uno de los factores de exposición [24].

En cambio, en un estudio de casos y controles se identifican los casos y se clasifican en expuestos o no expuestos para el factor estudiado; y en lugar de obtener las tasas de riesgo (en este tipo de estudios no se pueden calcular), se determina un grupo de controles como muestra de toda la población. Finalmente, los individuos en el grupo de controles son clasificados como expuestos o no expuestos.

El propósito del grupo de controles es determinar el tamaño relativo de individuos expuestos y no expuestos en la población de origen; es decir, se emplea para estimar la distribución de la exposición de un determinado factor en la población, de la que se obtuvieron los individuos seleccionados como casos. Por lo tanto, es de especial importancia que el grupo de controles sea una muestra independiente del estado de exposición.

Los estudios de casos y controles, al igual que los estudios de cohorte, pueden ser tanto prospectivos como retrospectivos. Los estudios de casos y controles tienen un carácter prospectivo cuando, después de iniciarse el estudio, se espera a que aparezcan casos nuevos de “enfermedad” en la población (casos incidentes) para ir incorporándolos al grupo de individuos “enfermos”. En este tipo de estudios los controles se suelen incorporar paulatinamente, aunque a veces se pueden escoger aleatoriamente de una lista de posibles controles, una vez conseguidos todos los casos necesarios.

Por el contrario, los estudios de casos y controles tienen un carácter retrospectivo cuando los “enfermos” ya están presentes al comienzo del estudio; es decir, los casos son prevalentes y los incorporamos de entrada al grupo de “enfermos”.

Para entender de una forma más clara la diferencia entre los tipos de diseños de estudios observacionales: caso-control y cohorte, se hace referencia al siguiente esquema:

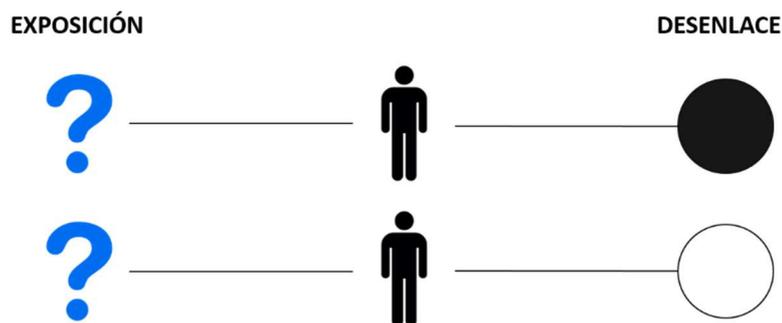


Ilustración 2. Procedimiento para un estudio caso-control prospectivo.

En esta primera imagen se muestra el procedimiento para un estudio caso-control prospectivo, en el que se trata de conocer los posibles factores de riesgo para un desenlace conocido, incorporando los individuos “enfermos” una vez comenzado el estudio.

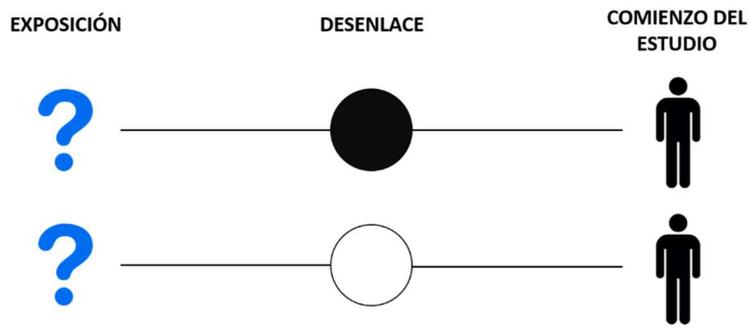


Ilustración 3. Procedimiento para un estudio caso-control retrospectivo.

En esta segunda imagen se muestra el procedimiento habitual para un estudio caso-control retrospectivo en el que, una vez conocida la “enfermedad” o desenlace, se trata de determinar qué factores de exposición en el pasado influyen para que se produzca el evento. Es decir, lo que se conoce es el resultado y se trata de determinar qué factores influyen para que éste ocurra.

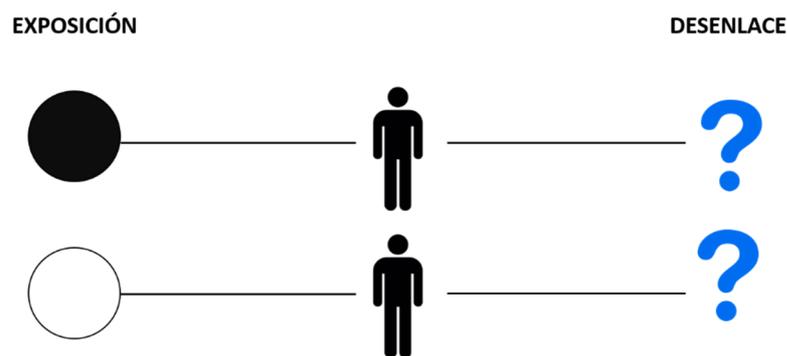


Ilustración 4. Procedimiento para un estudio de cohorte prospectivo.

En la imagen anterior se muestra el procedimiento para un estudio de cohorte prospectivo; es decir, aquel en el que se conocen los datos de medida de la exposición a los distintos factores y en el que se pretende determinar cuál es el resultado o desenlace.

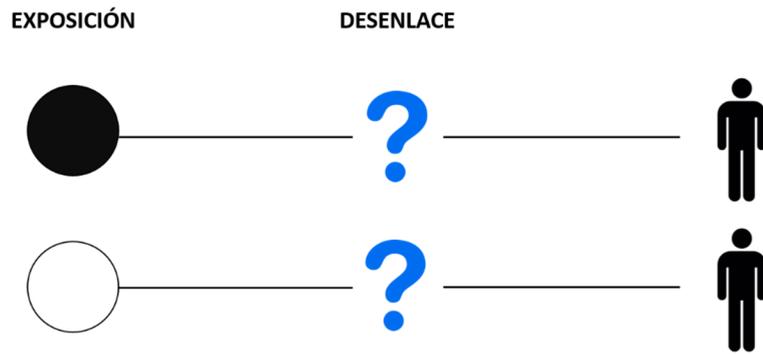


Ilustración 5. Procedimiento para un estudio de cohorte retrospectivo.

Por último, se muestra el procedimiento para un estudio de cohorte retrospectivo en el que se trata de determinar el desenlace o resultado (al igual que en el caso anterior), pero con la diferencia de que el estudio comienza posteriormente a la aparición de la “enfermedad”, mientras que en el prospectivo aparece después del inicio del análisis.



Ilustración 6. Leyenda para los esquemas de los procedimientos de los distintos diseños de estudios observacionales.

### 1.1.2. Utilidad de los estudios caso-control

Los estudios de caso-control son especialmente recomendables en las siguientes situaciones [17]:

- Cuando la “enfermedad” o resultado estudiado es poco común, ya que permiten, en lugar de medir la exposición a un factor en toda la

población, medirlo únicamente para una muestra de la población que será representada a través de los controles.

- Cuando el desenlace tiene un largo periodo de latencia (el tiempo entre la exposición y la manifestación de la “enfermedad”).
- Cuando la obtención de los datos de exposición es difícil o muy cara (de esta forma se miden los datos para una muestra de la población y no para la población entera).
- Cuando la información acerca de los factores de riesgo es escasa, los estudios caso-control son una forma fácil de probar la asociación con múltiples factores de riesgo.

A diferencia de los estudios de cohorte en los que se puede estimar el riesgo o la frecuencia de la “enfermedad” dentro de toda la población de estudio, en los estudios caso-control no es posible, ya que no se trabaja con la población completa sino con una muestra de tamaño suficiente.

Sin embargo, sí es posible obtener una medida de la probabilidad relativa de “enfermedad” en los individuos expuestos frente a los no expuestos. Esta medida generalmente se lleva a cabo a través de lo que se conoce como *Odds-Ratio* o razón de ventajas. Por lo tanto, cuando el evento estudiado no es muy habitual, el riesgo puede ser estimado de forma mucho más eficiente a través de un estudio caso-control.

### 1.1.3. Selección de los controles y sesgo

Para la selección de los controles es importante tener en cuenta los siguientes principios [25]:

- El grupo de comparación (controles) debería ser representativo de la población en la que se produjeron los casos.
- Deben ser muestreados de forma independiente de los factores de exposición; es decir, no deberían seleccionarse de forma que sea más probable su elección si tienen o no el factor de estudio.

Los controles deben ser recogidos de tal forma que sean similares en muchos aspectos a los casos. Además, el grupo de controles seleccionado debe tener un riesgo similar a la hora de desarrollar la variable resultado [26]. Si alguno de estos principios no se cumple adecuadamente pueden aparecer sesgos [27] en el resultado, con lo que el estudio no sería válido.

Además, debido a que la muestra seleccionada como grupo de control debe ser representativa de la población origen de los casos, el tamaño de muestreo de este grupo debe ser, como mínimo, igual al número de casos existentes. En aquellas situaciones en las que la toma de datos no consuma demasiado tiempo ni sea costosa, podría ser adecuado medir un mayor número de controles que de casos, para así conocer de forma más fiable el tamaño relativo de los individuos expuestos y no expuestos en la población origen.

Sin embargo, en aquellas situaciones en que sí sea costoso obtener toda la información necesaria, es suficiente con medir el mismo número de controles que de casos existentes en el estudio, ya que el incremento del número de controles no marca una diferencia sustancial en la eficiencia y fiabilidad del modelo.

#### 1.1.4. Medida de la asociación: *Odds-Ratio*

El *Odds-Ratio* es una medida epidemiológica utilizada habitualmente en los estudios de casos y controles. Se ha traducido como razón de probabilidades o razón de ventajas. El *Odds-Ratio* se define como el cociente entre dos *odds*, lo que en estadística se define como el cociente entre la probabilidad de que un evento suceda, frente a que no ocurra [23].

Lo más habitual en un estudio caso-control es que el posible factor de riesgo sea binario (expuesto frente a no expuesto). Bajo estas circunstancias, el *Odds-Ratio* es el cociente entre la *odds* de exposición observada en casos y la *odds* de exposición en el grupo de control. Se define a través de la siguiente fórmula:

$$\text{Odds Ratio (OR)} = \frac{\text{odds de exposición en casos}}{\text{odds de exposición en controles}}$$

	CASOS	CONTROLES
Expuestos	a	b
No Expuestos	c	d

Tabla 1. Tabla de contingencia de Caso/Control frente a Expuestos/No expuestos.

- *Odds* de exposición en casos:

$$\frac{a}{c} = \frac{\text{casos en expuestos}}{\text{casos en no expuestos}}$$

Ecuación 1. *Odds* de exposición en casos.

- *Odds* de exposición en controles:

$$\frac{b}{d} = \frac{\text{controles en expuestos}}{\text{controles en no expuestos}}$$

Ecuación 2. *Odds* de exposición en controles.

Por lo tanto, el *Odds-Ratio* se define como:

$$OR = \frac{a \cdot d}{c \cdot b}$$

Ecuación 3. *Odds-Ratio*.

Sin embargo, en aquellas variables que sean cualitativas de más de dos categorías, el *Odds-Ratio* (OR) se define como la razón de ventajas respecto a una categoría de referencia, que generalmente es aquella que se sospecha que menos influirá para que se produzca el desenlace objeto de estudio. De la misma forma, en el caso de variables continuas, para la obtención del OR se codifica la variable en el número de categorías necesarias y se toma una de ellas como referencia, para medir la “ventaja” de las demás respecto a ella.

El *Odds-Ratio* es una estimación de la asociación de un determinado factor con la aparición de un evento, por lo que resulta necesario calcular la medida de la variabilidad de esta estimación. El Intervalo de Confianza es el rango en el que se encuentra el verdadero valor del *Odds-Ratio* (OR). Sin embargo, permite obtener una buena estimación cuando el OR se aproxima a 1, pero se hace menos estable para valores mayores.

$$IC = OR^{(1 \pm \frac{Z}{xhm})}$$

Ecuación 4. Intervalo de confianza para el *Odds-Ratio*.

Z = constante dependiente del porcentaje (para un IC al 95%, Z=1,96).

$X_{hm}$  = Chi cuadrado de HM (Haenszel-Mantel).

$$X_{hm} = \sqrt{\frac{(n-1)(a \cdot d - b \cdot c)^2}{(a+b)(c+d)(a+c)(b+d)}}$$

Ecuación 5. Chi cuadrado de Haenszel-Mantel.

La interpretación del *Odds-Ratio* es la siguiente:

- Si el resultado es significativamente mayor que 1 la asociación es positiva; es decir, que la presencia del factor se asocia a la mayor ocurrencia del evento. Por lo tanto, se le considera como un Factor de Riesgo.
- Si el resultado es significativamente menor que 1 la asociación es negativa; es decir, que la presencia del factor no se asocia con la mayor ocurrencia del evento, sino al revés. Por lo tanto, se le considera como un Factor de Protección.
- Si el resultado es próximo a 1 no hay asociación entre las variables; es decir, que la probabilidad de que el evento ocurra va a ser igual con o sin la presencia del factor.

Para una mejor interpretación del *Odds-Ratio* puede ser traducido, en términos de probabilidad, mediante la siguiente fórmula:

$$Probabilidad = \frac{OR}{OR + 1}$$

Ecuación 6. Transformación de Odds-Ratio en probabilidad.

Por ejemplo, si el OR es 0,4 interpretaríamos como que es 0,4 veces más probable que un evento esté asociado a la presencia de un factor. Al traducirlo a probabilidad se obtiene 0,286, lo que es igual a 28,6%. En este caso la interpretación sería que existe un 28,6% de probabilidad de que un evento esté asociado a la presencia de un factor.

Por otro lado, los resultados del Intervalo de Confianza (IC) permiten establecer si una asociación es estadísticamente significativa. Cuando el IC incluye el 1 podemos decir que la asociación encontrada no es estadísticamente significativa, y si no incluye el 1 sí lo es.

La siguiente tabla recoge, de forma resumida, la interpretación de los resultados en función de los valores del OR y de su correspondiente Intervalo de Confianza:

VALOR "ODDS RATIO"	INTERVALO DE CONFIANZA		TIPO DE ASOCIACIÓN
	Inferior	Superior	
= 1			No hay asociación
> 1	> 1	> 1	Significativa (Factor de Riesgo)
	< 1	> 1	No significativa
< 1	< 1	< 1	Significativa (Factor de Protección)
	< 1	> 1	No significativa

Tabla 2. Interpretación del *Odds-Ratio*.

### 1.1.5. Adecuación de la metodología caso-control al presente estudio

En el trabajo que nos ocupa, la población objeto de estudio está constituida por las intersecciones y tramos curvos en vías urbanas de la ciudad de Valladolid en las que se produjo alguna salida de calzada entre los años 2009 y 2014, así como por aquellas intersecciones y curvas en las que no se produjo dicho desenlace a lo largo de ese periodo. Se ha considerado que el diseño más adecuado para este estudio es un diseño caso-control, ya que concurren las siguientes circunstancias:

- Lo que se conoce es el resultado (ocurrencia o no de un accidente con salida de calzada).
- Lo que se trata de conocer son los factores de tipo geométrico que incrementan las probabilidades de ocurrencia de este tipo de accidente (ilustración 3).
- El periodo de latencia es largo y muy variable, ya que podrían pasar muchos años hasta tener un número significativo de salidas de calzada en intersecciones y tramos curvos en la ciudad de Valladolid.

Por lo tanto, teniendo en cuenta las ideas recogidas en los apartados anteriores y dado que el evento cuyas causas se quieren analizar son las salidas de calzada en intersecciones y curvas de zona urbana, se designan como casos aquellas ubicaciones en intersecciones y tramos curvos urbanos de la ciudad de

Valladolid en las que se produjo alguna salida de calzada entre los años 2009 y 2014. Así mismo, los controles deberán ser también intersecciones y curvas de Valladolid en las que no se hayan producido salidas de vía a lo largo de dicho periodo, seleccionados de manera aleatoria y siempre cumpliendo los dos principios básicos para evitar la aparición de sesgos.

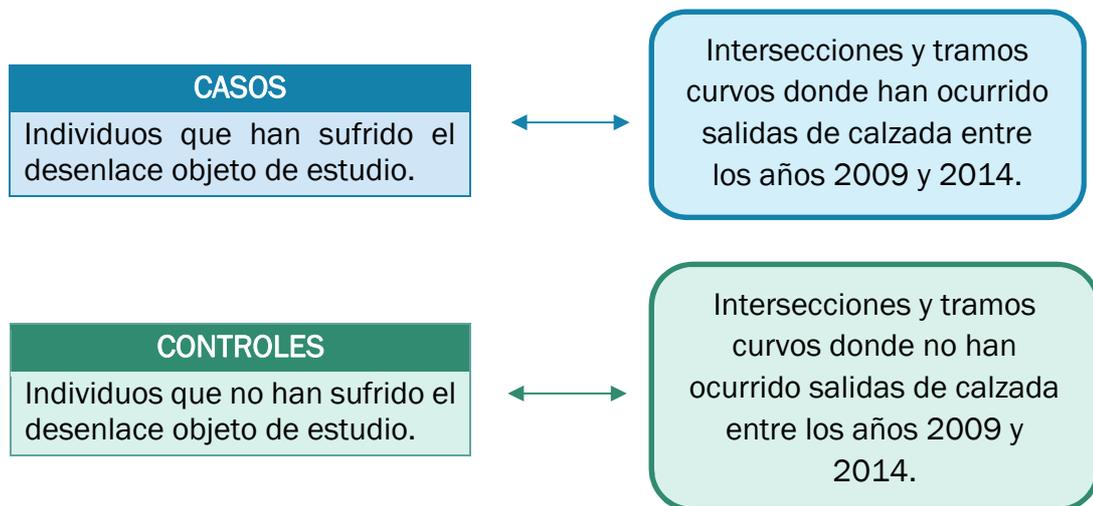


Ilustración 7. Esquema de casos y controles para el estudio.

Por último, hay que destacar que el objeto de este trabajo de investigación es un estudio caso-control retrospectivo, ya que se obtuvieron como casos aquellos lugares en los que se produjeron accidentes considerados como salida de calzada entre los años 2009 y 2014; es decir, en el momento en que se comenzó el análisis ya se conocían los individuos a incluir en la investigación (en un estudio prospectivo, como se explicó anteriormente, habría que esperar a que sucediera el desenlace).

## 1.2. REGRESIÓN LOGÍSTICA

### 1.2.1. Conceptos básicos del modelo logístico

Un modelo de regresión múltiple (no necesariamente de regresión lineal) permite explicar el comportamiento de una variable dependiente  $Y$  en función de una serie de variables independientes  $X_1, X_2, \dots, X_k$  y de un término de perturbación  $u$ :

$$Y = f(X_1, X_2, \dots, X_k, u)$$

Cuando la variable dependiente  $Y$  es continua, resulta frecuente utilizar un modelo de regresión lineal múltiple, para el que tendremos una expresión de la siguiente forma:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k + u$$

El objetivo de la regresión lineal será estimar los parámetros  $\beta_1, \beta_2, \dots, \beta_k$ , de forma que el modelo resultante se ajuste lo mejor posible a las observaciones. En tal caso, dicha estimación se lleva a cabo mediante los métodos de Mínimos Cuadrados (MCO o MCG).

Por otro lado, puede ocurrir que la variable dependiente  $Y$  sea cualitativa o categórica. En tales circunstancias, el modelo de regresión lineal presentará una serie de inconvenientes serios que se exponen a continuación, por lo que será necesario recurrir a la llamada regresión logística.

Si se considera el caso de una variable dependiente binaria,  $Y$ , explicada por un conjunto de predictores  $X_1, \dots, X_k$ , al tratarse de una variable dicotómica sólo podrá tomar valores entre 0 y 1, y siempre se cumplirá que:

$$E[Y] = 0 \cdot P(Y = 0) + 1 \cdot P(Y = 1) = P(Y = 1)$$

Por otra parte, se podría pensar en utilizar un modelo de regresión lineal múltiple para explicar el comportamiento de la variable  $Y$ . Bajo el supuesto habitual de que  $E[u] = 0$ , y suponiendo conocidos los valores que toman las variables explicativas (observaciones), tendremos que:

$$E[Y] = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

Igualadas las dos expresiones obtenidas para  $E[Y]$ , llegamos al resultado que da nombre al modelo de probabilidad lineal:

$$P(Y = 1) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k = Y - u$$

Ecuación 7. Modelo de probabilidad lineal.

Esta expresión nos viene a decir que podemos expresar la variable dependiente binaria  $Y$  como la probabilidad de “éxito” más un término de perturbación  $u$ :

$$Y = P(Y = 1) + u = E[Y] + u$$

Sin embargo, este modelo inicial no será válido para explicar el comportamiento de variables dependientes binarias [28], pues presenta varios problemas.

- El término de perturbación  $u = Y - (\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)$  ya no será una variable aleatoria continua (como ocurría en el modelo de regresión lineal múltiple o MRLM), sino que será una variable aleatoria discreta. Es decir, conocidos los valores de las variables explicativas,  $u$  sólo puede tomar dos valores determinados. Por tanto,  $u$  ya no se distribuirá de forma normal (uno de los supuestos básicos del MRLM). Si bien este supuesto no resulta estrictamente necesario para explicar MCO, sí es fundamental a la hora de realizar cualquier tipo de inferencia posterior sobre el modelo (intervalos de confianza para los parámetros estimados, contrastes de hipótesis, etc.).
- El término de perturbación  $u$  no cumple la hipótesis de homocedasticidad (la varianza de dicho término no es constante). Debido a este problema, los estimadores MCO no serán eficientes, por lo que resultará necesario recurrir a la estimación por MCG.
- Como la variable dependiente  $Y$  sólo puede tomar los valores 0 y 1, si representamos gráficamente la nube de puntos formada por los pares de observaciones de  $Y$  con una de las variables explicativas  $X_i$ , obtendremos puntos situados sobre las rectas  $Y=1$  e  $Y=0$ :

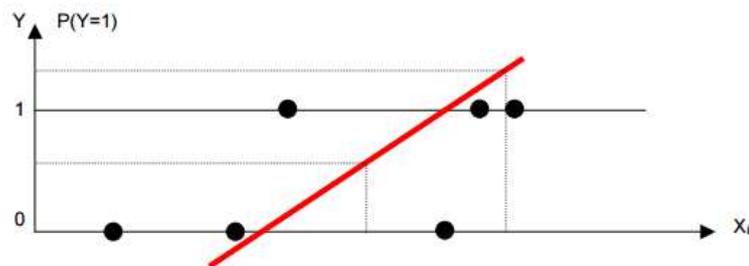


Ilustración 8. Ajuste de una variable dependiente binaria con un modelo de regresión lineal [28].

Al estimar los parámetros del modelo de probabilidad lineal, estaremos ajustando una recta a la nube de puntos anterior. El uso de dicha recta para predecir nuevos valores de  $Y$  a partir de valores dados de  $X_i$  puede proporcionar valores mayores que 1 o menores que 0 (lo cual está en contradicción con la definición de probabilidad).

- Finalmente, la expresión  $P(Y = 1) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$  nos dice que la probabilidad del suceso “éxito” viene determinada por una combinación lineal de variables explicativas. De ello se deduce que:

$$\frac{\delta P(Y=1)}{\delta X_i} = \beta_i \quad \forall i = 1, 2, \dots, k$$

En otras palabras: la variación en  $P(Y=1)$  causada por cambios en alguna de las variables explicativas es constante (y, por tanto, independiente del valor actual de dicha variable explicativa), lo cual es una hipótesis muy poco realista.

### Modelos Logit y Probit

Como hemos visto, dada una variable dependiente binaria  $Y$ , el modelo de regresión lineal presenta serias inconsistencias. Para evitarlas se han desarrollado modelos no lineales que tratan de resolver los dos últimos problemas anteriores. La idea consiste en realizar un modelo de la forma:

$$Y = f(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k) + u$$

En dicho modelo,  $f$  es una función real que depende de la expresión lineal  $\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$ , y que deberá ser distinta de la función identidad (si así fuese, se recuperaría el modelo lineal). Además, es necesaria una función que esté acotada por los valores 0 y 1 (puesto que su valor coincidirá con el de la probabilidad); es decir, se requiere una función de la siguiente forma:

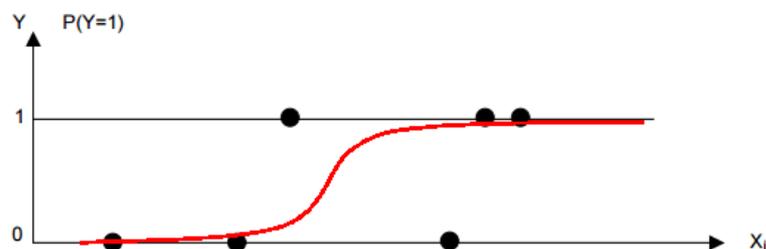


Ilustración 9. Tipo de función necesaria para la regresión con una variable dependiente binaria [28].

Pues bien, de entre las funciones  $f$  que presentan una forma similar a la de la gráfica hay dos que son las que se utilizan con mayor frecuencia: la función logística (que da lugar a los modelos Logit), y la función de distribución de una normal estándar (asociada a los modelos Probit).

Como acabamos de ver, una posible solución a las inconsistencias que presentaba el modelo de probabilidad lineal para explicar el comportamiento de una variable dependiente binaria es usar un modelo Logit, donde  $f$  es la función logística definida a continuación:

$$f(z) = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}}$$

Ecuación 8. Función logística.

Tal y como se mencionó anteriormente es necesario que el rango de  $f(z)$  esté comprendido entre 0 y 1, independientemente del valor de  $z$ . Es decir, el modelo debe estar diseñado para describir una probabilidad que será siempre un valor entre 0 y 1. El hecho de que con el modelo logístico nunca se pueda superar un valor de probabilidad de 1 ni obtener valores inferiores a 0 no ocurre siempre para otros posibles modelos, lo que hace que el modelo logístico resulte especialmente interesante.

Finalmente, sustituyendo el valor  $z$  por la correspondiente expresión lineal  $\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$ , tendremos el modelo logístico siguiente:

$$E[Y] = P(Y = 1) = \frac{e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}}{1 + e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}} = \frac{1}{1 + e^{-(\alpha + \beta_1 X_1 + \dots + \beta_k X_k)}}$$

Ecuación 9. Modelo logístico.

En conclusión, se puede definir la regresión logística como una técnica de análisis de datos multivariantes en la que la variable dependiente o respuesta es dicotómica (con dos valores a los que habitualmente nos referimos nominalmente, pero que suelen codificarse como “1” y “0” para un tratamiento matemático más sencillo) y las variables independientes pueden ser tanto cualitativas como cuantitativas.

Con modelos de regresión logística podemos investigar factores relacionados con una determinada característica de la población, y también estudiar cuáles modifican la probabilidad de un suceso determinado. Además, la regresión logística resulta útil para aquellas situaciones en las que se desea predecir la presencia o ausencia de una característica o resultado según los valores de un conjunto de variables predictoras. Es similar a un modelo de regresión lineal, pero está adaptado para situaciones en las que la variable dependiente es categórica (tal y como se ha explicado con anterioridad), siendo la situación más habitual aquella en la que la variable dependiente es dicotómica, en cuyo caso se denomina regresión logística binaria [29].

Como hemos comentado al principio de esta sección, la estimación en modelos Logit y Probit se realiza mediante el método de Máxima Verosimilitud (MV). Además, en este tipo de modelos no resulta posible interpretar directamente las estimaciones de los parámetros  $\beta$ , ya que son modelos no lineales. Lo que haremos en la práctica es fijarnos en el signo de los estimadores. Si el estimador es positivo significará que incrementos en la variable asociada causan aumentos en  $P(Y=1)$ . Por el contrario, si el estimador muestra un signo negativo supondrá que incrementos en la variable asociada causarán disminuciones en  $P(Y=1)$ .

En el modelo Logit se suelen usar otros dos conceptos para profundizar más en la interpretación de los estimadores:

- Se llama *odds* al siguiente cociente de probabilidades:

$$Odds = \frac{P(Y = 1)}{1 - P(Y = 1)} = e^{(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}$$

Ecuación 10. Definición de *odds*.

Tomando logaritmos neperianos en la expresión anterior obtenemos una expresión lineal para el modelo:

$$Logit[P(Y = 1)] = \ln\left(\frac{P(Y = 1)}{1 - P(Y = 1)}\right) = \beta_0 + \beta_1 X_1 + \dots + \beta_k X_k$$

Ecuación 11. Definición del modelo Logit.

Aquí se aprecia claramente que el estimador del parámetro  $\beta$  se podrá interpretar como la variación en el término Logit (el logaritmo neperiano del cociente de probabilidades) causada por una variación unitaria en la variable  $X_i$  (suponiendo constantes el resto de variables explicativas).

- Cuando se hace referencia al incremento unitario en una de las variables explicativas del modelo aparece el concepto del *Odds-Ratio* como el cociente entre los dos *odds* asociados: el obtenido tras realizar el incremento y el anterior al mismo. Así, si definimos  $P_1$  como la probabilidad de que suceda el evento cuando se incrementa en una unidad la variable  $X_i$  -mientras las demás permanecen constantes-, se obtiene el *Odds-Ratio* de la siguiente forma:

$$Odds\ Ratio = \frac{P_1/(1 - P_1)}{P_0/(1 - P_0)} = \frac{e^{\beta_0} \cdot e^{\beta_1 X_1} \cdot e^{\beta_i(X_i+1)} \cdot \dots \cdot e^{\beta_k X_k}}{e^{\beta_0} \cdot e^{\beta_1 X_1} \cdot e^{\beta_i X_i} \cdot \dots \cdot e^{\beta_k X_k}} = e^{\beta_i}$$

Ecuación 12. Definición de Odds-Ratio en regresión logística.

De la expresión anterior se deduce que un coeficiente  $\beta_i$  cercano a cero, o equivalentemente un *Odds-Ratio* cercano a uno, significará que cambios en la variable explicativa  $X_i$  asociada no tendrán efecto alguno sobre la variable dependiente  $Y$ .

Es frecuente que en el ajuste de estos modelos no se tenga la seguridad de que todas las variables consideradas sean significativas a la hora de predecir la respuesta, o incluso, como en nuestro caso, que la determinación de qué variables son más relevantes a la hora de predecir dicha respuesta sea uno de los objetivos fundamentales del estudio. La selección de variables es, por tanto, un punto esencial en estos casos en los que el uso de métodos de selección de variables por pasos es muy frecuente [31]. Entre los métodos más habituales empleados en los modelos de regresión se encuentran los siguientes: *forward*, *backward* y *stepwise*.

- En el método *forward* se comienza con un modelo que no contiene ninguna variable explicativa, y en cada paso se incorpora aquella variable que aporta una mayor capacidad explicativa al modelo. El procedimiento se detiene cuando el incremento en el coeficiente de determinación debido a la inclusión de una nueva variable explicativa en el modelo ya no es importante.
- En el método *backward* se comienza por considerar incluidas en el modelo teórico a todas las variables disponibles. En cada paso se van eliminando de una en una según su capacidad explicativa. Este procedimiento se lleva a cabo hasta llegar a una situación en la que la eliminación de una variable más suponga un descenso demasiado acusado en el coeficiente de determinación.
- Por último, el método *stepwise* consiste en una combinación de los dos anteriores. En el primer paso se procede como en el método *forward*, pero a diferencia de éste en el que cuando una variable entra en el modelo ya no vuelve a salir, en el procedimiento *stepwise* es posible que la inclusión de una nueva variable haga que otra que ya estaba en el modelo resulte redundante y sea “expulsada” de él.

Una vez obtenido el modelo ajustado es conveniente valorar la capacidad del mismo para producir predicciones correctas en futuras observaciones. La clasificación de observaciones en una de las dos clases del modelo se hace a

través de la estimación de la probabilidad de pertenencia a cada grupo, obtenida a partir de los coeficientes estimados del modelo y de la ecuación 9.

Mediante esta ecuación obtenemos una estimación de la probabilidad de pertenencia a una clase. Si decidimos establecer como punto de corte entre las dos clases un valor de la probabilidad entre 0 y 1, podemos utilizar las observaciones para obtener una tabla de clasificación a partir de los resultados, así como el porcentaje de observaciones correctamente clasificadas, con el objetivo de evaluar la capacidad de clasificación aparente de la regla.

Sin embargo, esto puede conducir a una sobreestimación de la capacidad de clasificación, ya que se toma como muestra de prueba la misma que se utilizó para obtener el modelo. Por este motivo, para corregir este sesgo y evaluar de una manera más adecuada la capacidad de clasificación de la regla establecida, se utilizan diferentes métodos de estimación de la capacidad de clasificación, como el método de validación cruzada *Leave-One-Out*, que se explicará más detenidamente en el apartado 1.3 de este mismo capítulo.

Hay que destacar que si hacemos variar el punto de corte entre las dos clases antes mencionadas, recorriendo el intervalo (0,1), obtendremos una familia de reglas de clasificación asociadas al modelo que estamos valorando. La calidad de las clasificaciones obtenidas con estas reglas permitirá valorar globalmente el modelo y comparar las diferentes posibilidades mediante las denominadas curvas ROC, que también se explican con más detalle en el apartado 1.3 de este capítulo.

### 1.2.2. Regresión logística aplicada a estudios caso-control

La orientación con la que se lleva a cabo el modelo logístico es la misma con la que se desarrollan los estudios de cohorte; es decir, el modelo describe la probabilidad de desarrollar una “enfermedad” de interés expresada en función de una serie de variables independientes, presuntamente medidas al principio del periodo de desarrollo. Por esta razón, es conveniente analizar si el modelo es adecuado para ser aplicado a estudios de tipo caso-control.

Tanto Breslow y Day, en 1981, como Prentice y Pike, en 1979, identificaron ciertas condiciones bajo las cuales el modelo logístico puede ser utilizado en estudios de tipo caso-control. En términos de este tipo de estudios se ha demostrado que, aunque los casos y controles se seleccionan en primer lugar y después los factores de exposición previos, el procedimiento del análisis se lleva a cabo como si fuese al revés; es decir, como si se tratase de un estudio de cohorte.

En conclusión, de la misma forma que para estudios de cohorte, se puede realizar el modelo logístico con la variable referente a la “enfermedad” como variable respuesta y los factores de exposición como variables independientes. Es decir, el tratamiento de los datos es igual que si se tratase de un estudio de cohorte, obteniendo resultados igualmente válidos.

Sin embargo, aunque el modelo logístico es igualmente aplicable a los estudios caso-control, existe una limitación importante en dicho análisis. Tal y como se ha mencionado anteriormente, mientras que para un estudio de cohorte se puede obtener una estimación del riesgo de un individuo con las variables independientes especificadas, este modelo no puede ser empleado para predecir el riesgo de un individuo para los estudios de caso-control. De hecho, únicamente se puede realizar una estimación del riesgo a través de las *Odds-Ratio* (OR) [30]. De este modo, aunque para un estudio caso-control no se puede obtener una estimación del riesgo directamente, es posible estimar este riesgo indirectamente, para lo que se obtiene el denominado *Odds-Ratio* (OR) o razón de ventajas. De hecho, el OR es la única medida de la asociación directamente estimada de un modelo logístico, independientemente de si se trata de un estudio de cohorte, caso-control o estudio transversal.

Una de las principales razones por las que la regresión logística ha sido tan popular, especialmente en los estudios epidemiológicos, es la facilidad para obtener una estimación de los OR de los coeficientes de las variables independientes ( $\beta_i$ ), así como contrastar su significación estadística cuando la muestra se lleva a cabo condicionada por la variable resultado, tal y como ocurre en los estudios caso-control. De la misma forma, se puede obtener el Intervalo de Confianza para dicha estimación, intervalo en el que se encontrará el valor real del *Odds-Ratio*. La forma en que el modelo logístico estima el OR se explicó con anterioridad en la ecuación 12.

## 1.3. VALIDACIÓN *LEAVE-ONE-OUT* Y ANÁLISIS ROC

### 1.3.1. Validación *Leave-One-Out*

La validación cruzada o *cross-validation* es una técnica utilizada para evaluar los resultados de un análisis estadístico y garantizar que éstos son independientes de la partición entre datos de entrenamiento y prueba. Generalmente, el conjunto de datos se divide en dos subconjuntos: el primero de ellos, denominado de entrenamiento, se emplea para la obtención del modelo; por el contrario, el subconjunto de prueba sirve para comprobar que el modelo clasifica

correctamente a dichos individuos con la regla obtenida a partir del subconjunto de entrenamiento.

En general, la validación cruzada consiste en repetir y calcular la media aritmética obtenida de las medidas de evaluación sobre diferentes particiones en ambos subconjuntos de datos, para así obtener una mejor estimación del error real del modelo, frente al error que se obtiene cuando el subconjunto de entrenamiento y prueba coinciden (tal y como ocurre en la regresión logística obtenida con el software SPSS). En resumen, este tipo de técnica se utiliza en entornos donde el objetivo principal es la predicción y en los que se quiere estimar cómo de preciso es un modelo que se llevará a cabo en la práctica.

Las técnicas de validación están motivadas esencialmente por los dos principales problemas que surgen a la hora de realizar un análisis estadístico: la selección del modelo y la estimación de su error o del rendimiento del mismo. En el primero de los casos se trata de seleccionar el modelo óptimo cuando éste depende de una serie de parámetros. En el segundo de los casos, una vez seleccionado un modelo, se trata de valorar su rendimiento real, lo que generalmente se mide a través de la tasa de clasificación correcta.

La validación cruzada dejando uno fuera o *Leave-one-out cross-validation* (LOOCV) implica separar los datos de forma que para cada iteración tengamos un solo individuo dentro del subconjunto de datos de prueba, y el resto dentro del subconjunto de entrenamiento. De forma más detallada, este método de validación cruzada deja fuera un individuo en cada iteración, y comprueba si lo ha clasificado correctamente en el modelo obtenido a partir del resto de datos. Por lo tanto, habrá un número de iteraciones igual al número de datos existentes. Finalmente, una vez que se ha calculado cada una de las iteraciones, se calcula la media aritmética del error, con lo que se obtiene una estimación más realista de éste [32].

La evaluación viene dada por el error que, en este tipo de validación cruzada, es muy bajo; en cambio, a nivel computacional es muy costoso puesto que se tienen que realizar un elevado número de iteraciones, tantas como N muestras tengamos, y para cada una analizar los datos tanto de entrenamiento como de prueba.

### 1.3.2. Análisis ROC

#### ***Sensibilidad y Especificidad***

En el caso de variables respuesta dicotómicas, la exactitud de un modelo estadístico puede medirse en términos de Sensibilidad y Especificidad. La

Sensibilidad es la probabilidad de obtener un resultado positivo cuando el individuo tiene la “enfermedad” objeto de estudio [33].

En cambio, la Especificidad indica la probabilidad de obtener un resultado negativo cuando el individuo no sufre el desenlace, lo que se traduce en términos de su capacidad para descartar la “enfermedad” cuando ésta no está presente. No existe ninguna manera teórica para la medición de estas dos características en un modelo; el único procedimiento es el experimental. Éste consiste en someter a un grupo de individuos al modelo, con y sin el evento objeto de estudio o “enfermedad”, para así obtener los datos necesarios. La Sensibilidad se obtiene con el subgrupo de “enfermos”, y la Especificidad con el de “sanos”.

En el caso concreto del estudio objeto de este proyecto, se trabaja con la técnica estadística caso-control anteriormente definida. Por lo tanto, los individuos “enfermos” corresponden a lugares donde se han producido salidas de calzada (el evento de estudio), mientras que los individuos “sanos” equivalen a lugares donde no se han producido salidas de calzada. En conclusión, la Sensibilidad será la probabilidad de obtener un resultado igual a “salida de calzada” cuando se trata de un lugar donde se ha producido este hecho. Por el contrario, la Especificidad indica la probabilidad de obtener un resultado equivalente a “no salida de calzada” cuando el individuo es un lugar en el que no se ha producido esa eventualidad.

RESULTADO DE LA PRUEBA	RESULTADO REAL	
	Salida de calzada	No salida de calzada
Salida de calzada	Verdadero positivo (VP)	Falso positivo (FP)
No salida de calzada	Falso negativo (FN)	Verdadero negativo (VN)

Tabla 3. Comparación de los resultados de la prueba a evaluar frete al resultado real.

Por lo tanto, a continuación se definen tanto la Sensibilidad como la Especificidad:

$$Sensibilidad = \frac{\text{"Enfermos" positivos}}{\text{Total "enfermos"}} = \frac{VP}{VP + FN}$$

Ecuación 13. Sensibilidad.

$$\text{Especificidad} = \frac{\text{"Sanos" negativos}}{\text{Total "sanos"}} = \frac{VN}{VN + FP}$$

Ecuación 14. Especificidad.

A la hora de interpretar los resultados de estos parámetros, el modelo ideal debería tener una Sensibilidad y una Especificidad lo más próximas posibles al 100%. A partir de la tabla de contingencia anterior (tabla 1.3.), se puede definir el valor predictivo del resultado positivo como la proporción de resultados válidos entre los resultados positivos de la prueba:

$$VPP = \frac{VP}{VP + FP}$$

Ecuación 15. Valor predictivo del resultado positivo.

El valor predictivo del resultado negativo será la proporción de resultados válidos entre los negativos:

$$VPN = \frac{VN}{VN + FN}$$

Ecuación 16. Valor predictivo del resultado negativo.

Por último, el valor global de la proporción de resultados válidos entre la totalidad de pruebas efectuadas se define como:

$$VG = \frac{VP + VN}{VP + FP + VN + FN}$$

Ecuación 17. Valor global de resultados válidos.

### **Curva ROC**

La curva ROC es un gráfico en el que se observan todos los pares Sensibilidad/Especificidad resultantes de la variación continua de los puntos de corte en todo el rango de resultados observados. En el eje Y de coordenadas se sitúa la Sensibilidad o fracción de verdaderos positivos, definida como se expuso anteriormente y calculada para el grupo de "enfermos". En el eje X de

coordenadas se sitúa la fracción de falsos positivos; es decir,  $1 - \text{Especificidad}$ , definida como  $\text{FP}/(\text{VN} + \text{FP})$  y calculada a partir del subgrupo no afectado por el evento de estudio [33].

Por lo tanto, cada punto de la curva representa un par Sensibilidad/( $1 - \text{Especificidad}$ ) correspondiente a un nivel de decisión determinado. Una prueba con discriminación perfecta, sin solapamiento de resultados en las dos poblaciones, tiene una curva ROC que pasa por la esquina superior izquierda, donde la Sensibilidad y la Especificidad toman valores máximos ( $S$  y  $E = 1$ ). Por el contrario, una prueba sin discriminación, con igual distribución de resultados en los dos subgrupos, da lugar a una línea diagonal de  $45^\circ$  desde la esquina inferior izquierda hasta la superior derecha. Sin embargo, la mayoría de las curvas ROC caen entre estos dos extremos.

En resumen, cuanto más próxima es una curva ROC a la esquina superior izquierda más alta es la exactitud global de la prueba. De la misma forma, si se dibujan en un mismo gráfico las curvas obtenidas con distintos modelos, aquel que esté situado más hacia arriba y hacia la izquierda tiene mayor exactitud: por simple observación se puede obtener una comparación cualitativa [34].

Pero, además, las curvas ROC no sólo aportan información cualitativa, sino que también permiten realizar los análisis estadísticos para la evaluación cuantitativa de los distintos modelos. En concreto, hay que tener en cuenta que el área bajo la curva ROC es una medida global de la calidad de un modelo. A través de dicha área referente a varios modelos es posible realizar un contraste ROC, para así seleccionar el modelo más adecuado, tal y como se realizará en el capítulo 7.

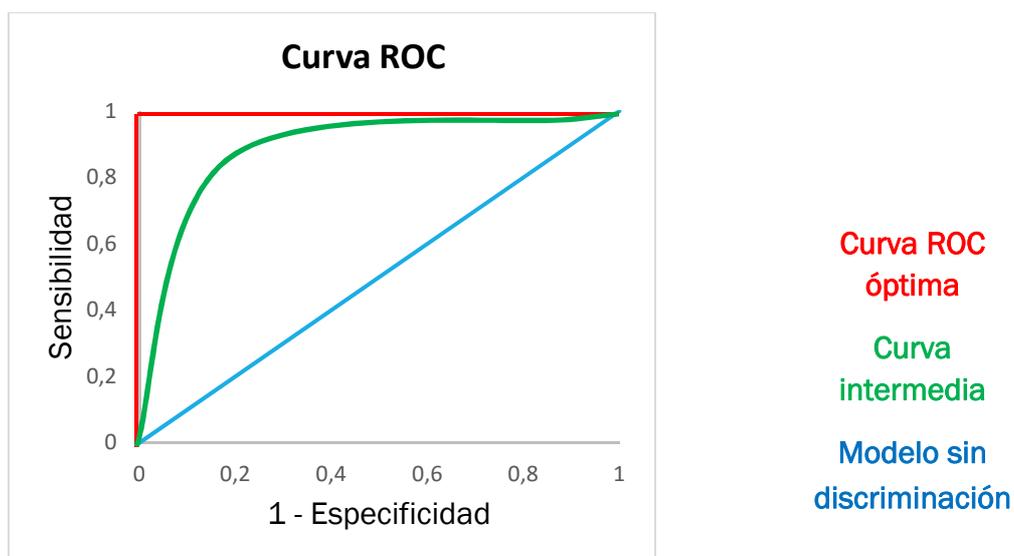


Ilustración 10. Descripción Curva ROC.





# PARTE I. OBTENCIÓN DE LA MUESTRA



## CAPÍTULO 2. BASES DE DATOS DE PARTIDA

---

Para la obtención de los datos correspondientes a las salidas de calzada en intersecciones y tramos curvos urbanos, se han tomado dos bases de datos de partida. Por un lado, se ha trabajado con la base de datos de la DGT denominada ARENA, en la que se recogen todos aquellos accidentes con víctimas; es decir, aquellos en los que alguna persona ha resultado lesionada.

Por otro lado, se parte de la base de datos realizada por la empresa Exeleria para el PIMUSSVA (Plan Integral de Movilidad Urbana Sostenible y Segura de la ciudad de Valladolid), herramienta de planificación estratégica desarrollada por el Ayuntamiento con el objetivo de satisfacer las necesidades de movilidad de las personas y empresas en la ciudad y sus alrededores, en busca de una mejor calidad de vida.

Los motivos de recurrir a esta segunda base de datos son, principalmente, los siguientes:

- Obtener un mayor número de casos, con el objetivo de ampliar los datos obtenidos a partir de la base de datos ARENA, y así poder analizar un mayor rango de accidentes.
- Contar con datos más actuales, debido a que en la base de datos ARENA, por publicarse con cierto desfase, solamente se pudo acceder a los accidentes ocurridos entre los años 2009 – 2013. En cambio, en la base de datos elaborada para el PIMUSSVA, la información almacenada es referente a accidentes sucedidos hasta el año 2014.
- Valorar también la posibilidad de incluir los accidentes sin víctimas, ya que en ARENA únicamente se incluyen aquellos en los que alguna persona ha resultado lesionada. Este otro tipo de accidentes, sin lesionados, también pueden ser interesantes a la hora de ser analizados a lo largo de este trabajo de investigación.

### 2.1. DESCRIPCIÓN DE LA BASE DE DATOS “ARENA”

ARENA es un sistema de captura, almacenamiento y gestión de la información sobre accidentes de la DGT, cuya principal finalidad es proporcionar un interfaz único de inserción de datos por parte de los usuarios autorizados. Así mismo, el

sistema permite el tratamiento de la información recogida para la elaboración de estadísticas e informes sobre accidentalidad [35]. En el Anexo 1 se muestra la plantilla a rellenar por los agentes de tráfico en el momento de la toma de información referente a un accidente.

En este sistema de almacenamiento de datos se recogen todos aquellos accidentes en los que ha habido víctimas, es decir, personas con cualquier tipo de lesión.

Dicha base de datos engloba gran cantidad de variables que se agrupan en tres categorías:

- Datos generales.
- Datos de vehículos, tantos grupos de datos como de vehículos implicados.
- Datos de personas, tantos grupos de datos como ocupantes o peatones.

Para este estudio, únicamente se han empleado variables correspondientes al grupo de datos generales, debido a que en él se encuentra la información de interés, relacionada con las características geométricas de la ubicación.

Las variables recogidas en el grupo de datos generales son las que se recogen en la siguiente tabla.

CODIFICACIÓN	ETIQUETA	VALORES
<b>provin</b>	Provincia	Adquiere un valor de dos dígitos en función de la provincia.
<b>isla</b>	Isla	Tendrá valor "0" debido a que únicamente toma valores positivos para las distintas islas de España.
<b>año</b>	Año	Es el año en que ha ocurrido el accidente.
<b>mes</b>	Mes	Toma valores de 1 a 12, en función del mes en que ha ocurrido el accidente.
<b>diasem</b>	Día de la semana	Toma valores de 1 a 7, en función del día de la semana en que ha ocurrido el accidente, empezando por "Lunes".
<b>día</b>	Día del mes	Es el día del mes en que ha ocurrido el accidente.
<b>tipodia</b>	Tipo de día	"L" "Laborable" "F" "Festivo" "P" "Posterior a festivo" "A" "Anterior a festivo"
<b>zona</b>	Zona	1 "Carretera" 2 "Zona Urbana"

		3 "Travesía" 4 "Variante"
<b>red</b>	Titularidad de la vía	1 "Estatal" 2 "Autonómica" 3 "Provincial" 4 "Municipal" 5 "Otras"
<b>sent</b>	Sentido	"A" "Ascendente" "D" "Descendente" "N" "No especificado"
<b>Codmun</b>	Código de municipio	
<b>habit</b>	Habitantes	1 "Más de cien mil" 2 "De 50 a 100 mil" 3 "De 5 a 50 mil" 4 "De mil a cinco mil" 5 "Menos de mil"
<b>tipovia</b>	Tipo de vía	1 "Autopista" 2 "Autovia" 3 "Vía rápida" 4 "Vía convencional con carril lento" 5 "Vía convencional" 6 "Camino vecinal" 7 "Vía de Servicio" 8 "Ramal de enlace" 9 "Otro tipo"
<b>numcarr</b>	Número de carriles	
<b>anchcalz</b>	Anchura de calzada	1 "Menos de 5,99 metros" 2 "Entre 6 y 6,99 metros" 3 "De 7 metros o más"
<b>anchcarr</b>	Anchura de carril	1 "De más de 3,75 metros" 2 "De 3,25 a 3,75 metros" 3 "Menos de 3,25 metros"
<b>marcvial</b>	Marcas viales	1 "Inexistentes o borradas" 2 "Solo separación de carriles" 3 "Separación carriles y bordes" 4 "Solo separación de bordes"
<b>arcen</b>	Arcén	1 "Inexistente o impracticable" 2 "Menor de 1,5 metros" 3 "De 1,5 a 2,49 metros" 4 "De 2,5 metros en adelante"
<b>arcenpav</b>	Arcén pavimentado	"S" "Sí" "N" "No" " " "arcen inexistente o impracticable"
<b>mediana</b>	Mediana entre calzadas	"S" "Sí" "N" "No"
<b>barrera</b>	Barrera de seguridad	"S" "Sí" "N" "No"

<b>panel</b>	Paneles direccionales	“S” “Sí” “N” “No”
<b>hitos</b>	Hitos de arista	“S” “Sí” “N” “No”
<b>captafar</b>	Captafaros	“S” “Sí” “N” “No”
<b>fueraint</b>	Fuera de intersección	1 “Recta” 2 “Curva suave” 3 “Curva fuerte sin señalizar” 4 “Curva fuerte con señal y sin velocidad señalizada” 5 “Curva fuerte con señal y velocidad señalizada”
<b>tipointe</b>	Tipo de intersección	1 “En T o Y” 2 “En X o +” 3 “Enlace de entrada” 4 “Enlace de salida” 5 “Giratoria” 6 “Otros”
<b>aconint</b>	Acondicionamiento de intersección	1 “Nada especial” 2 “Sólo isletas o paso para peatones en vía secundaria” 3 “Paso para peatones o isleta en centro de la vía principal” 4 “Carril central de espera” 5 “Raqueta de giro a izquierda” 6 “Otro tipo”
<b>priori</b>	Prioridad regulada por	1 “Agente” 2 “Semáforo” 3 “Señal de stop” 4 “Señal de ceda el paso” 5 “Sólo marcas viales” 6 “Paso para peatones” 7 “Otra señal” 8 “Ninguna (sólo norma)”
<b>superf</b>	Superficie	1 “Seca y limpia” 2 “Umbría” 3 “Mojada” 4 “Helada” 5 “Nevada” 6 “Barrillo” 7 “Gravilla suelta” 8 “Aceite” 9 “Otra”
<b>lumin</b>	Luminosidad	1 “Pleno día” 2 “Crepúsculo” 3 “Iluminación suficiente (noche)” 4 “Iluminación insuficiente (noche)”

		5 "Sin iluminación (noche)"
<b>facatm</b>	Factores atmosféricos	1 "Buen tiempo" 2 "Niebla intensa" 3 "Niebla ligera" 4 "Lloviznando" 5 "Lluvia fuerte" 6 "Granizando" 7 "Nevando" 8 "Viento fuerte" 9 "Otro"
<b>visibili</b>	Visibilidad restringida por	1 "Edificios" 2 "Configuración del terreno" 3 "Vegetación" 4 "Factores atmosféricos" 5 "Deslumbramiento" 6 "Polvo o humo" 7 "Otra causa" 8 "Sin restricción"
<b>se_alpel</b>	Señalización de peligro	1 "Señalización existente" 2 "Señalización inexistente" 3 "Señalización innecesaria (no hay peligro)"
<b>acera</b>	Acera	"S" "Si" "N" "No"
<b>se_alver</b>	Visibilidad de la señalización vertical	1 "Buena" 2 "Deficiente" 3 "Nula"
<b>tipoacc</b>	Tipo de accidente	11 "Colisión frontal" 12 "Colisión fronto-lateral" 13 "Colisión lateral" 14 "Colisión por alcance" 15 "Colisión múltiple o en caravana" 21 "Colisión con vehículo estacionado o averiado" 22 "Colisión con valla de defensa" 23 "Colisión con barrera de paso a nivel" 24 "Colisión con otro objeto o material" 31 "Atropello a peatón sosteniendo bicicleta" 32 "Atropello a peatón reparando el vehículo" 33 "Atropello a peatón aislado o en grupo" 34 "Atropello a conductor de animales" 35 "Colisión con animal conducido o rebaño" 36 "Colisión con animales sueltos" 41 "Vuelco en la calzada"

		51 "Salida de la vía por la izda. y choque con árbol o poste"
		52 "Salida de la vía por la izda. y choque con muro o edificio"
		53 "Salida de la vía por la izda. y choque con cuneta o bordillo"
		54 "Salida de la vía por la izda. y otro choque"
		55 "Salida de la vía por la izda. con despeñamiento"
		56 "Salida de la vía por la izda. con vuelco"
		57 "Salida de la vía por la izda. en llano"
		58 "Salida de la vía por la izda. otra"
		61 "Salida de la vía por la dcha. y choque con árbol o poste"
		62 "Salida de la vía por la dcha. y choque con muro o edificio"
		63 "Salida de la vía por la dcha. y choque con cuneta o bordillo"
		64 "Salida de la vía por la dcha. y otro choque"
		65 "Salida de la vía por la dcha. con despeñamiento"
		66 "Salida de la vía por la dcha. con vuelco"
		67 "Salida de la vía por la dcha. en llano"
		68 "Salida de la vía por la dcha. otra"
		71 "Otro"
<b>denscir</b>	Densidad de la circulación	1 "Fluida" 2 "Densa" 3 "Congestionada"

Tabla 4. Descripción de las variables generales de la base de datos ARENA.

## 2.2. DESCRIPCIÓN DE LA BASE DE DATOS “PIMUSSVA”

A partir de los datos proporcionados por el Ayuntamiento de Valladolid, la empresa Exeleria desarrolló una base de datos con la información correspondiente a todos los accidentes en la ciudad de Valladolid entre los años

2009 y 2014 para la elaboración del PIMUSSVA (Plan Integral de Movilidad Urbana Sostenible y Segura).

Esta base de datos contiene una cifra mayor de accidentes respecto a ARENA, ya que, además de los datos correspondientes a “atestados”, también contiene los correspondientes a “partes de accidentes”; es decir, contiene la información tanto de los casos con víctimas (personas con algún tipo de lesión) como de los casos sin víctimas.

Las variables que conforman dicha base de datos están menos centradas en parámetros geométricos, a diferencia de las del apartado “datos generales” de ARENA. Además, al igual que la base de datos ARENA, estas variables pueden clasificarse en tres grupos: datos generales, datos de vehículos y datos de personas. Por último, cabe destacar que el número de variables es más reducido en esta segunda base de datos, comprendiendo la siguiente información:

VARIABLE	VALOR
<b>Fecha</b>	
<b>Día semana</b>	1 "Lunes" 2 "Martes" 3 "Miércoles" 4 "Jueves" 5 "Viernes" 6 "Sábado" 7 "Domingo"
<b>Día mes</b>	
<b>Mes</b>	1 "Enero" 2 "Febrero" 3 "Marzo" 4 "Abril" 5 "Mayo" 6 "Junio" 7 "Julio" 8 "Agosto" 9 "Septiembre" 10 "Octubre" 11 "Noviembre" 12 "Diciembre"
<b>Año</b>	
<b>Hora</b>	Hh:mm
<b>Hora entrada</b>	Hh
<b>Tipo de documento</b>	“Parte de accidente” “Atestado”
<b>Lugar</b>	
<b>Coordenada X</b>	Es la coordenada X de localización en metros.

<b>Coordenada Y</b>	Es la coordenada Y de localización en metros.
<b>Tipo de ubicación</b>	“Aparcamiento” “Intersección” “Tramo de vía”
<b>Daños</b>	“Sin daños” “Escasos” “Menor cuantía” “Consideración” “Gran consideración” “Siniestro total”
<b>Lesionado</b>	Registra las personas lesionadas y las características de las lesiones.
<b>Nº personas lesionadas</b>	Registra el número de personas que han resultado lesionadas a causa del accidente.
<b>Nº fallecidos</b>	
<b>Nº graves</b>	
<b>Nº leves</b>	
<b>Suma lesionados pronóstico conocido</b>	
<b>Nº heridos sin información</b>	
<b>Nº conductores</b>	0 1 2
<b>Nº conductores fallecidos</b>	
<b>Nº conductores graves</b>	
<b>Nº conductores leves</b>	
<b>Nº conductores lesionados sin información</b>	
<b>Nº pasajeros</b>	
<b>Nº pasajeros fallecidos</b>	
<b>Nº pasajeros graves</b>	
<b>Nº pasajeros leves</b>	
<b>Nº pasajeros lesionados sin información</b>	
<b>Nº peatones</b>	
<b>Nº peatones fallecidos</b>	
<b>Nº peatones leves</b>	
<b>Nº peatones lesionados sin información</b>	
<b>Nº hombres</b>	
<b>Nº mujeres</b>	
<b>Nº niños</b>	Considerando niño de 0 a 14 años.
<b>Nº jóvenes</b>	Considerando joven de 15 a 34 años.
<b>Nº adultos</b>	Considerando adulto de 35 a 59 años.

<b>Nº personas mayores</b>	Considerando persona mayor de más de 60 años.
<b>Sexo lesionado</b>	“Hombre” “Mujer”
<b>Edad lesionado</b>	
<b>Categoría lesionado</b>	“Conductor” “Pasajero” “Peatón” “Sin información”
<b>Gravedad lesionado</b>	“Leve” “Grave” “Fallecido” “Sin información”
<b>Implicados</b>	Datos de los vehículos implicados en el accidente (modelo de coche y matrícula).
<b>Nº de vehículos implicados</b>	
<b>Nº turismos</b>	
<b>Nº autobús</b>	
<b>Nº bicicletas</b>	
<b>Nº ciclomotores</b>	
<b>Nº motocicletas</b>	
<b>Nº furgonetas</b>	
<b>Nº camiones</b>	
<b>Nº otros</b>	
<b>Categoría vehículo</b>	“Turismo” “Autobús” “Bicicleta” “Camión” “Ciclomotor” “Furgoneta” “Motocicleta” “Otros” “Sin información”
<b>Accidente por</b>	“Alcance” “Alcance – atropello” “Atropello” “Atropello de perro” “Caída” “Caída de motocicleta” “Caída de vehículo” “Choque” “Choque y colisión” “Choque – vuelco” “Colisión” “Colisión frontal” “Colisión frontolateral”

	“Colisión frontolateral y choque” “Colisión lateral” “Colisión marcha atrás” “Colisión y choque” “Colisión y vuelco” “Fricción” “Salida de vía” “Vuelco” “Otros”
<b>Presencia alcohol</b>	En mg/l de aire espirado.
<b>Posible causa</b>	

Tabla 5. Descripción de las variables de la base de datos del PIMUSSVA.

Para el estudio de esta base de datos, únicamente tendremos en cuenta las variables generales sobre lugar, tipo de ubicación, coordenada X, coordenada Y, fecha y hora del accidente; así como el tipo de accidente y el número de personas lesionadas.

# CAPÍTULO 3. OBTENCIÓN DE LOS CASOS Y LOS CONTROLES

---

## 3.1. OBTENCIÓN DE LOS CASOS

### 3.1.1. Obtención de los accidentes de interés en la base de datos "ARENA"

Partiendo de todas las variables comentadas en la tabla anterior se realizó la selección de los accidentes de interés a través del programa SPSS, en base a los siguientes criterios:

#### ***Zona***

Teniendo en cuenta que el estudio se centra en estudiar aquellos casos de tramo urbano, se seleccionan los accidentes en los que la variable "zona" toma los valores:

- 2 "Zona Urbana"
- 3 "Travesía"

#### ***Tipo de accidente***

Como las salidas de calzada son el desenlace a estudiar se seleccionan los tipos de accidentes siguientes, relacionados con la salida de la vía del vehículo:

- 22 "Colisión con valla de defensa"
- 41 "Vuelco en la calzada"
- 51 "Salida de la vía por la izda. y choque con árbol o poste"
- 52 "Salida de la vía por la izda. y choque con muro o edificio"
- 53 "Salida de la vía por la izda. y choque con cuneta o bordillo"
- 54 "Salida de la vía por la izda. y otro choque"
- 55 "Salida de la vía por la izda. con despeñamiento"
- 56 "Salida de la vía por la izda. con vuelco"
- 57 "Salida de la vía por la izda. en llano"

- 58 "Salida de la vía por la izda. otra"
- 61 "Salida de la vía por la dcha. y choque con árbol o poste"
- 62 "Salida de la vía por la dcha. y choque con muro o edificio"
- 63 "Salida de la vía por la dcha. y choque con cuneta o bordillo"
- 64 "Salida de la vía por la dcha. y otro choque"
- 65 "Salida de la vía por la dcha. con despeñamiento"
- 66 "Salida de la vía por la dcha. con vuelco"
- 67 "Salida de la vía por la dcha. en llano"
- 68 "Salida de la vía por la dcha. otra"

### ***Provincia***

Por otro lado, como el estudio se localiza en la ciudad de Valladolid el valor de la variable "provin" tiene que ser:

- 47 "Valladolid"

### ***Código de municipio***

Así mismo, hay que establecer el código de municipio de la ciudad de Valladolid, correspondiente a:

- 1 "Valladolid" para los años 2009, 2010 y 2011.
- 186 "Valladolid" para el año 2013.

### ***Tipo de vía***

Para asegurar que sólo se toman casos de tramo urbano, se eligen aquellos accidentes en los que la variable "tipovia" toma valores distintos de:

- 1 "Autopista"
- 2 "Autovía"

### ***Fuera de intersección***

Debido a que el estudio se focaliza en intersecciones y tramos curvos se eliminan los accidentes en recta, descartando aquellos casos en los que la variable "fueraint" tome el valor:

- 1 "Recta"

### ***Tipo de intersección***

Por último, debido a que los casos objeto del estudio son las intersecciones y curvas donde ocurrió alguna salida de calzada entre los años 2009 y 2013, se seleccionan los accidentes producidos en intersecciones; es decir, en los que la variable “tipointe” tome cualquier valor distinto de 0.

Con todos estos criterios se realiza la búsqueda en SPSS que se detalla a continuación. Con ella se obtiene un total de 46 salidas de vía en intersecciones y curvas de tramo urbano para la ciudad de Valladolid, correspondientes a la base de datos ARENA:

```
((a_o = 2009 | a_o = 2010 | a_o = 2011) & ((zona = 2 | zona = 3) & (tipoacc = 22 | tipoacc = 41 | tipoacc = 51 | tipoacc = 52 | tipoacc = 53 | tipoacc = 54 | tipoacc = 55 | tipoacc = 56 | tipoacc = 57 | tipoacc = 58 | tipoacc = 61 | tipoacc = 62 | tipoacc = 63 | tipoacc = 64 | tipoacc = 65 | tipoacc = 66 | tipoacc = 67 | tipoacc = 68) & provin = 47 & codmun = 001 & (tipovia ~= 2 & tipovia ~= 1) & (fueraint ~= 1 | tipointe ~= 0))) | ((a_o = 2012 | a_o = 2013) & ((zona = 2 | zona = 3) & (tipoacc = 22 | tipoacc = 41 | tipoacc = 51 | tipoacc = 52 | tipoacc = 53 | tipoacc = 54 | tipoacc = 55 | tipoacc = 56 | tipoacc = 57 | tipoacc = 58 | tipoacc = 61 | tipoacc = 62 | tipoacc = 63 | tipoacc = 64 | tipoacc = 65 | tipoacc = 66 | tipoacc = 67 | tipoacc = 68) & provin = 47 & codmun = 186 & (tipovia ~= 2 & tipovia ~= 1) & (fueraint ~= 1 | tipointe ~= 0)))
```

### **3.1.2. Obtención de los accidentes de interés en la base de datos del “PIMUSSVA”**

En primer lugar, se hizo una búsqueda comparativa entre los accidentes de ARENA y los de la base de datos elaborada por Exeleria, para tratar de verificar la correspondencia entre ambas. Se pretendía observar los accidentes coincidentes en una y otra, y comprobar cuál era la codificación empleada en Exeleria, respecto a la variable “Tipo de accidente”, en comparación con la codificación existente en ARENA. Esta información es esencial para identificar cómo se codifican las salidas de calzada en la base de datos Exeleria, con el

propósito de determinar cuáles son los tipos de accidentes que se deberán seleccionar de este conjunto de datos.

De esa comparativa se observó que en la base de datos para la realización del PIMUSSVA había muy pocos casos clasificados como “Salida de vía”. Además, la mayoría de los accidentes codificados dentro de la categoría de “Salida de la calzada” en ARENA (valores 51-58 y 61-68 para la variable “tipo de accidente”) se encontraban clasificados principalmente como “Choques” en la base de datos del PIMUSSVA. Esta falta de coordinación entre las dos bases de datos supuso, en muchos casos, una gran dificultad a la hora de localizar los accidentes coincidentes en ambas.

Tras realizar esta comparativa de las dos bases de datos, se decidieron los tipos de accidentes a seleccionar como casos de interés para nuestro estudio; es decir, aquellos que podrían estar relacionados con la aparición de una salida de la calzada:

### ***Tipo de accidente***

- “Caída”
- “Caída de motocicleta”
- “Caída del vehículo”
- “Choque”
- “Choque y colisión”
- “Choque - vuelco”
- “Colisión y choque”
- “Colisión y vuelco”
- “Salida de vía”
- “Vuelco”

Aplicando dicho filtro a la variable “tipo de accidente” se determinó que había 652 casos. Teniendo en cuenta que lo que se trata de analizar son las intersecciones y curvas de zona urbana en las que ha ocurrido alguna salida de calzada, a este número de accidentes habría que restar los producidos en tramos rectos.

A pesar de ello, se trataba de un número muy elevado de casos de cara a la posterior medición de variables, por lo que se trató de reducir dicha cantidad aplicando un nuevo filtro: que la variable “Número de lesionados” tuviera un valor distinto de 0. Es decir, se trató de recoger aquellas salidas de calzada en las que hubo lesionados, teniendo en cuenta que el principal objetivo del presente estudio es determinar las características geométricas que favorecen

una salida de calzada, en vistas a implantar barreras de contención adecuadas para una mayor protección de los peatones (es decir, para tratar de reducir el número de víctimas y llegar al objetivo “cero lesionados”). En un principio podría parecer que este aspecto contradice el tercero de los motivos por los que se decidió trabajar también con esta base de datos, además de con ARENA. Sin embargo, se decidió emplear este filtro de selección de datos para tratar de reducir el número de accidentes a considerar, a pesar de que inicialmente se valoró introducir también aquellos sin víctimas.

Finalmente se obtuvieron 116 casos con la tipología de accidente seleccionada y en los que hubo algún tipo de víctima. Sin embargo, no todos ellos se produjeron en intersecciones y curvas, sino que algunos fueron en tramo recto.

Por este motivo se aplicó un nuevo filtro sobre la variable “Tipo de ubicación” que toma dos posibles valores: “intersección” y “tramo de vía”. Si se selecciona el tipo de ubicación “intersección” se obtienen 48 casos, a los que habría que sumar los correspondientes a tramos curvos.

Para detectar las ubicaciones curvas se aplicó el filtro “tramo de vía”, obteniendo 68 accidentes. A través del software ArcGIS y la cartografía de la ciudad de Valladolid se detectó visualmente si cada uno de dichos casos se correspondía con un tramo recto o curvo, gracias a las coordenadas X e Y contenidas en la base de datos. Con ello se llegó a la conclusión de que 20 de esos 68 casos se produjeron en curvas.

Gracias a todo este procedimiento, a partir de la base de datos para la elaboración del PIMUSSVA se obtuvieron 68 salidas de vía en intersecciones y curvas entre los años 2009 y 2014. Sin embargo, algunos de estos accidentes ya estaban recogidos en ARENA, tal y como se observó en la comparativa inicial entre ambas bases de datos.

Por consiguiente, eliminando los accidentes ya contenidos en ARENA se obtuvieron 45 casos útiles procedentes de la base de datos del PIMUSSVA.

### 3.1.3. Unificación de los casos

A partir de los accidentes seleccionados en las dos bases de datos de partida, se obtiene un total conjunto de 91 casos. Tras tener localizados todos los datos de interés se encontraron problemas a la hora de ubicar correctamente cada uno de ellos, principalmente en los provenientes de la base de datos ARENA debido a varios motivos:

- Únicamente se encuentra recogida la calle en la que se ha producido el accidente, lo que dificulta saber en cuál de todas las intersecciones existentes en dicha calle se ha producido.
- En algunas situaciones se introduce la calle principal y, además, dos calles cercanas de referencia entre las que se produjo el accidente (a través de las variables “calle1” y “calle2” de la base de datos ARENA). Sin embargo, esa referencia resulta poco válida porque, en muchos casos, entre esas dos calles existen varias intersecciones o curvas, con lo que no termina de precisarse con exactitud la ubicación del accidente.

Por este motivo se recurrió a la ayuda del Subinspector del Departamento de Atestados de la Policía Municipal de Valladolid, quien a través del registro interno de la Policía Municipal, con información más detallada, colaboró para conseguir localizar de forma exacta cada uno de ellos. Así mismo, nos sugirió otros cuatro accidentes que fueron salidas de calzada y cuya tipología se encontraba codificada de forma distinta en la base de datos (en tres de los casos como colisión frontolateral y en uno de ellos como atropello a peatón aislado o en grupo).

Por lo tanto, como punto de partida existían 95 posibles casos a estudiar, que finalmente se redujeron a 82 puesto que, al estudiar cada uno de ellos detenidamente junto con el Subinspector, se llegó a la conclusión de que no eran accidentes de interés para este estudio. Es decir, no eran salidas de calzada sino otra tipología de accidente. En conclusión: existían errores en su codificación.

Para resumir todo el proceso de la recogida de casos de interés se realizó el siguiente procedimiento:

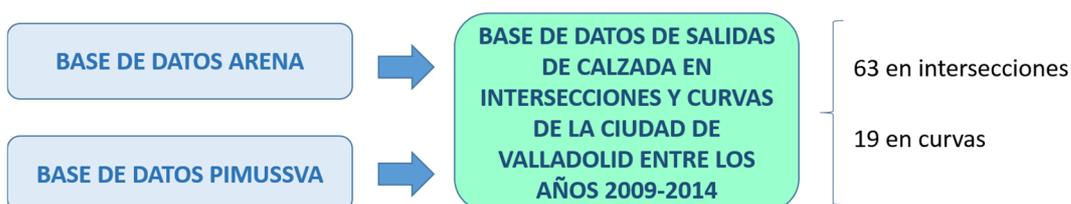


Ilustración 11. Esquema de la obtención de los casos.

Por último, se localizaron todos los casos en el mapa de la ciudad de Valladolid a través del software ArcGIS, tal y como se muestra en la siguiente imagen:

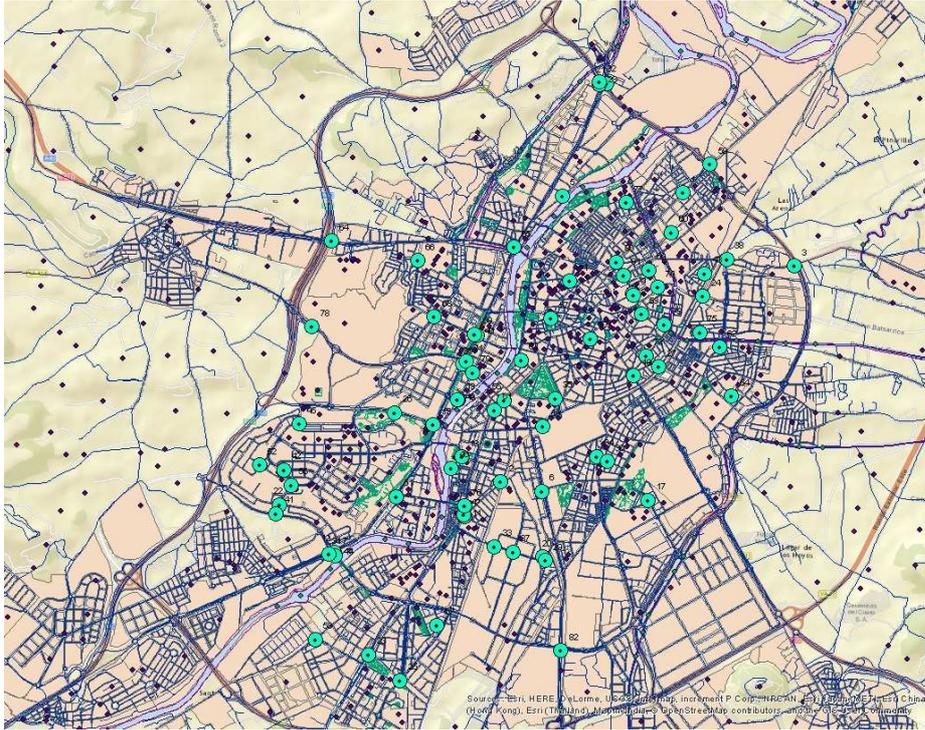


Ilustración 12. Localización de los casos en la cartografía de Valladolid.

## 3.2. OBTENCIÓN DE LOS CONTROLES

La selección de los controles es la parte más crítica de un estudio caso-control retrospectivo. Como está bien documentado en la literatura, no es sencillo realizar una selección exenta de crear algún tipo de sesgo inesperado en las conclusiones.

Para la obtención de los controles se realizaron dos enfoques distintos que se detallan a continuación.

### *Enfoque I*

Como un primer enfoque, al igual que para otros estudios epidemiológicos de este mismo tipo en otras áreas como la medicina, se decidió obtener los controles a partir de otros sujetos almacenados en las bases de datos y que no sufrieran el evento de estudio, con el objetivo de facilitar la recogida de datos. En el caso concreto de este estudio se determinaron como posibles controles todos aquellos accidentes en intersecciones y tramos curvos que no fueran salida de vía; y de entre todos ellos se extrajo una muestra aleatoria de controles del mismo tamaño (82 accidentes) que la muestra de casos disponibles.

Sin embargo, tras analizar ese conjunto de datos (casos y controles) con las variables escogidas como posibles factores de riesgo, no se obtuvieron resultados que determinaran una diferencia sustancial entre salida de calzada y otro tipo de accidente, con lo que se podría concluir que podrían existir factores ocultos que oscurecen los resultados.

### ***Enfoque II***

Por este motivo, se decidió realizar una nueva medición de controles a través de un sorteo aleatorio de todas las intersecciones y curvas de Valladolid, independientemente de si en ellas hubo algún otro tipo de accidente distinto de salida de calzada o no. Para ello se recurrió a la cartografía de la ciudad en AutoCAD, sobre la que se trazaron cuadrículas de 150 metros de ancho para la realización del sorteo.

Para la selección aleatoria de las casillas se realizó un sorteo tanto en las filas como en las columnas, partiendo de una distribución uniforme cuyo rango se limitó en el número máximo de filas o de columnas, respectivamente. Una vez seleccionada la fila y la columna, se podía conocer la cuadrícula del mapa donde habría que buscar una intersección o una curva. En el caso de no encontrar ninguna intersección o curva dentro de la casilla, se ignoró dicha cuadrícula y se pasó a la siguiente obtenida con el sorteo aleatorio. Además, ante la posibilidad de que en una misma celda hubiese más de una intersección o curva, también se sorteó el vértice de la cuadrícula, desde 1 a 4, tal y como muestra la siguiente imagen:

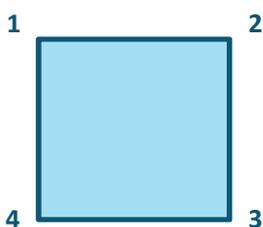


Ilustración 13. Esquema de la cuadrícula para la obtención de los controles.

De esta forma, se seleccionaron aquellas curvas o intersecciones más cercanas al vértice resultante del sorteo para cada una de las celdas. Finalmente, en cada una de las ubicaciones se realizaron las mediciones de las variables elegidas como posibles factores de riesgo, cuya relación con el desenlace estudiado (salida de la calzada) se estudiará en los siguientes capítulos.

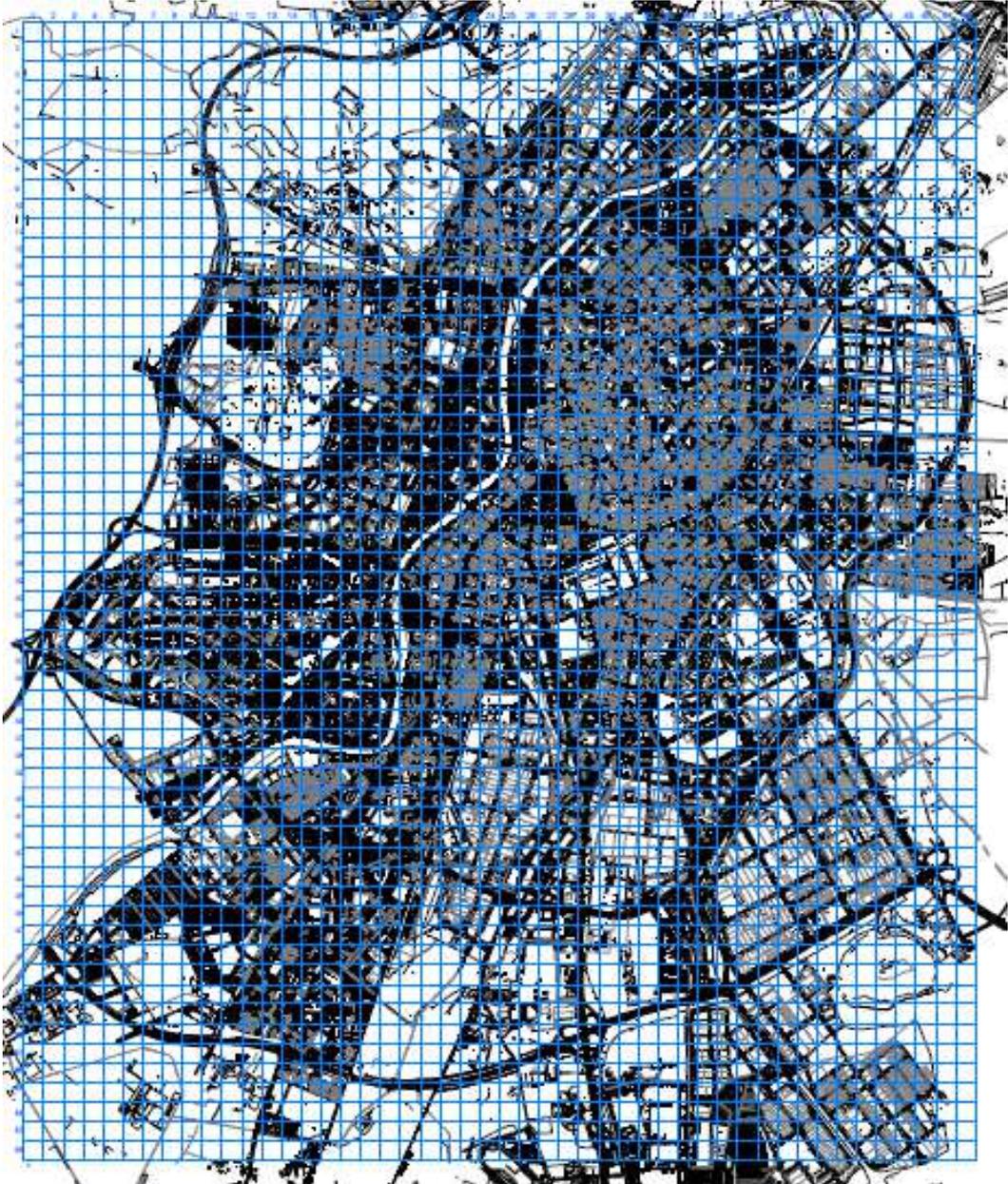


Ilustración 14. Formato para la selección de los controles en la cartografía en AutoCAD.

Es importante tener en cuenta que una ubicación que ya se ha considerado como caso no puede almacenarse también como control, por lo que si alguna de las intersecciones o curvas seleccionadas coincidiese con cualquiera de los casos (lugar donde se ha producido una salida de calzada) no se tendrá en cuenta como posible control.

Finalmente, al igual que para los casos, se localizaron todos aquellos controles seleccionados aleatoriamente en el mapa de la ciudad de Valladolid a través del software ArcGIS. Seguidamente, se muestran las imágenes con la localización de los controles, así como de los casos y controles en conjunto.



## CAPÍTULO 4. VARIABLES DEL MODELO

---

Las variables recogidas tanto para los casos como para los controles se pueden clasificar en dos grupos: unas almacenadas en las bases de datos convencionales de accidentes, y otras específicas que habrá que recoger para complementar a las anteriores.

### ***Grupo 1: Variables recogidas en las bases de datos convencionales***

Las variables incluidas en este grupo son las que forman parte de la base de datos de accidentes ARENA, perteneciente a la Dirección General de Tráfico (ofertada al público por parte del portal Estadístico de la DGT desde el 22 de Julio de 2015).

Estas variables tienen como objetivo aglutinar información sobre las intersecciones y zonas curvas de España donde han ocurrido salidas de calzada urbanas. Algunas de estas variables son las que se recogen a continuación: Tipo de accidente, Visibilidad, Anchura de la calzada y el carril, Tipo de curva, Tipo de intersección, Prioridad regulada por, etc.

### ***Grupo 2: Variables complementarias***

Las variables de este grupo recogen otro tipo de información adicional a complementar a las del grupo 1: Radio de curvatura, Pendiente, IMD (Intensidad Media Diaria), Longitud del tramo recto anterior, Distancia al semáforo anterior, Número de carriles para el sentido, Sentido de la circulación, entre otras.

Los valores o mediciones de todas las variables seleccionadas, tanto del Grupo 1 como del Grupo 2, tendrán que ser recogidos tanto en intersecciones y tramos curvos donde han ocurrido salidas de calzada en el periodo 2009-2014 (lugares que se denominarán como “casos”), como en intersecciones y tramos curvos donde no ha ocurrido este tipo de accidente en dicho periodo (que se denominarán “controles”).

De esta manera, se dispondrá del conjunto de datos de trabajo o “muestra” donde poder aplicar la técnica estadística elegida. Dicha técnica es la regresión logística, escogida con el objetivo de estudiar la asociación entre el grupo de

variables seleccionadas y un desenlace determinado (ocurrencia de salida calzada). La selección de dicha técnica se debe a que la variable respuesta (caso o control, en otras palabras, salida de calzada o no) es categórica y no continua, a diferencia de lo que ocurre en otros modelos de regresión, como por ejemplo en la regresión lineal.

A continuación, se detallan las variables que inicialmente se determinaron como relevantes; es decir, aquellas que se consideró que podrían actuar como posibles factores de riesgo de cara a la ocurrencia de salidas de calzada.

### 1. Tipo de intersección

Es una variable categórica, ya incluida en la base de datos ARENA con las siguientes categorías:

- 1 - “T o Y”
- 2 - “X o +”
- 3 - “Enlace de entrada”
- 4 - “Enlace de salida”
- 5 - “Giratoria”
- 6 - “Otros”

Sin embargo, se decidió modificar dichas categorías ligeramente en vistas a las necesidades observadas al analizar cada caso individualmente. Por ejemplo, aunque las denominadas comúnmente “glorietas partidas” técnicamente se engloban dentro de la categoría “X o +”, se decidió clasificarlas en un grupo independiente porque su complejidad es mucho mayor, lo que podría influir en mayor medida a la hora de producirse una salida de calzada frente a una intersección en “X o +”, más simple.

Así mismo, se eliminaron las categorías referentes a enlaces de entrada y de salida ya que en tramo urbano no se observó ninguno de estos tipos de intersecciones. Finalmente, las categorías establecidas fueron las siguientes:

- 1 - “T o Y”

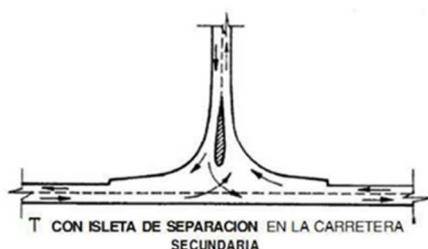


Ilustración 18. Ejemplo de intersección en T.

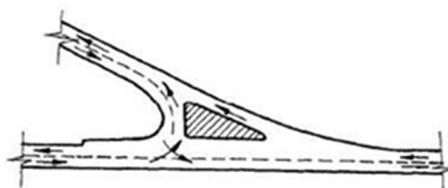


Ilustración 17. Ejemplo de intersección en Y.

- 2 - "X o +"

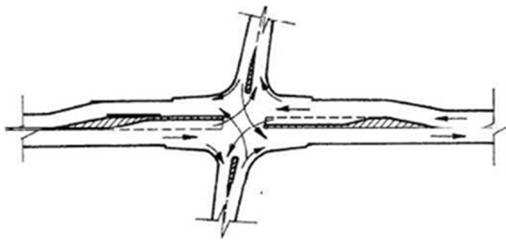


Ilustración 19. Ejemplo de intersección en X.

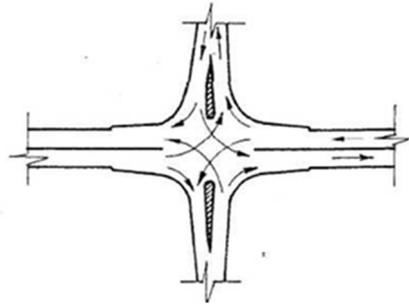


Ilustración 20. Ejemplo de intersección en cruz.

- 3 - "Giratoria" (refiriéndose a glorietas o rotondas)
- 4 - "Glorieta partida"

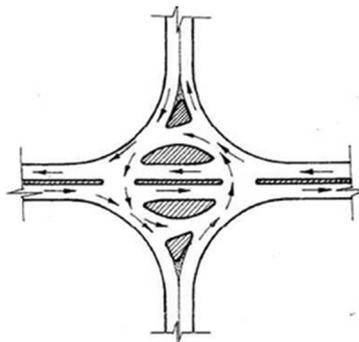


Ilustración 21. Ejemplo de glorieta partida.

- 5 - "Curva"

Hay que destacar que la última de las categorías se añade para identificar aquellos lugares que no son intersecciones, sino que son tramos curvos (el estudio se centra en estudiar tanto intersecciones como tramos curvos de zona urbana).

## 2. Radio de curvatura

El radio de curvatura se medirá ubicando cada accidente en el mapa de la ciudad de Valladolid mediante el software AutoCAD. Para su medida se ha utilizado la siguiente metodología [36]:

1. Medir en línea recta la cuerda “L”, desde el punto PC al punto PT.
2. Hallar el punto medio de la cuerda, “c”.
3. Trazar una perpendicular desde “c” a la huella, y obtener la longitud “M” de este segmento.

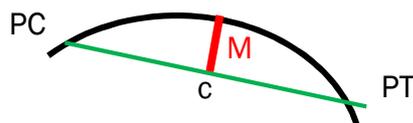


Ilustración 22. Esquema para la medida del radio de curvatura.

Finalmente, el radio de curvatura se obtendrá a través de la siguiente ecuación:

$$R = \frac{L^2}{8 \cdot M} + \frac{M}{2}$$

Ecuación 18. Medida del radio de curvatura [36].

### 3. Pendiente

Debido a que Valladolid no es una ciudad que presente importantes pendientes, salvo en zonas localizadas como el barrio de Parquesol, se decidió medir la variable “Pendiente” como una variable categórica dicotómica que toma dos posibles valores:

- 1 - “Sí”
- 2 - “No”

### 4. Longitud del tramo recto anterior

Se trata de una variable continua, que trata de recoger la distancia que tiene el usuario para acelerar el vehículo a velocidades superiores a las permitidas, con lo que aumentaría el riesgo de sufrir un accidente.

Con la información recogida en las bases de datos de partida únicamente se puede saber la localización del accidente y no el recorrido que hace el vehículo hasta llegar a ese punto; es decir, se desconoce la procedencia del vehículo. Por

lo tanto, se decidió medir esta variable como la máxima distancia posible con el fin de considerar la peor de todas las situaciones hipotéticas.

### **5. Distancia al semáforo anterior**

Al igual que la longitud del tramo recto se trata de una variable continua que pretende medir la distancia que tiene el usuario para acelerar el vehículo.

De la misma forma que para el caso anterior, se mide la máxima distancia posible para tener como dato la peor de las situaciones factibles.

### **6. Anchura de calzada**

Se considera como calzada el ancho de vía que permite al vehículo moverse libremente; es decir, en el caso de existir medianas con isletas divisorias, únicamente se ha tenido en cuenta el ancho para todos los carriles de un lado de la mediana, excluyendo los del otro lado.

A pesar de que se trata de una variable que se encuentra dentro del Grupo 1 (variables almacenadas en las bases de datos convencionales), en ARENA es una variable categórica, por lo que se decidió medir como una variable continua, con el objetivo de conseguir una mayor exactitud a la hora de incluirla en el modelo logístico posterior.

### **7. Anchura de carril**

Se mide el ancho del carril mediante el software AutoCAD y con la cartografía de la ciudad de Valladolid, al igual que para el ancho de calzada. De la misma forma, aunque esta información ya se encontraba almacenada en ARENA, se decidió medirla como una variable continua y no categórica.

### **8. Número de carriles en el mismo sentido**

Esta variable mide el número de carriles existentes para el mismo sentido de la circulación.

### **9. Prioridad**

Como punto de partida se tomó la variable recogida en ARENA que contenía las siguientes categorías:

- 1 - “Agente”
- 2 - “Semáforo”
- 3 - “Señal de STOP”
- 4 - “Señal de CEDA EL PASO”
- 5 - “Sólo marcas viales”
- 6 - “Paso para peatones”
- 7 - “Otra señal”
- 8 - “Ninguna (solo norma)”

Finalmente, se decidió modificar en cierta medida dichas categorías, resultando las siguientes:

- 1 - “Semáforo”
- 2 - “Señal de STOP”
- 3 - “Señal de CEDA EL PASO”
- 4 - “Ninguna (solo norma)”

## **10. Sentido**

Se trata de una variable cualitativa que define si se trata de una calzada de sentido único, doble o de doble sentido con isleta divisoria. Por lo tanto, está formada por las tres categorías siguientes:

- 1 - “Sentido único”
- 2 - “Doble sentido”
- 3 - “Doble sentido con mediana divisoria”

El motivo de la anterior categorización se debe a que las consecuencias para un vehículo o conductor no son las mismas si el impacto se produce con otros vehículos que circulan en el mismo sentido o en el contrario.

## **11. IMD (Intensidad Media Diaria)**

La IMD (Intensidad Media Diaria) es útil con el objetivo de determinar el flujo de vehículos en la localización a estudiar; es decir, la demanda de tráfico en el lugar del accidente.

Es un parámetro que se obtiene a través de una serie de aforos distribuidos a lo largo de la ciudad (el plano de la ciudad con los aforos existentes se encuentra en el Anexo 2) que mide el cociente entre el número de vehículos que atraviesan

una sección y el tiempo de medida. Dichos aforos cuantifican la intensidad de cada uno de los carriles y se codifican de la siguiente forma:

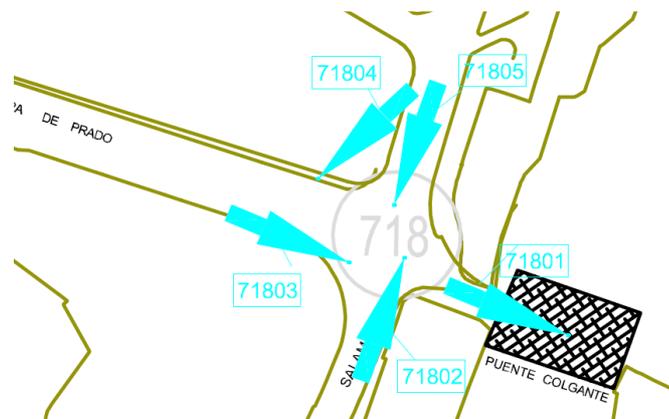


Ilustración 23. Ejemplo de medidor de IMD en una intersección.

El número de intersección se codifica con tres dígitos y cada punto de medida con cinco: los tres primeros son los correspondientes al código de la intersección y los dos últimos indican el número de medidor al que se refiere dentro de esa intersección.

Generalmente la IMD se refiere a la media anual de vehículos que circulan por esa sección a través de la siguiente ecuación:

$$IMD = \frac{\text{Número de vehículos que pasan en 1 año}}{365 \text{ días}}$$

Ecuación 19. Intensidad Media Diaria (IMD).

Sin embargo, para nuestro estudio no se trata de una medida suficientemente representativa, por lo que se trabajaría con la intensidad media para la hora en que se ha producido el accidente (no será lo mismo el tráfico a las 03:00 que a las 14:00, por ejemplo).

Tras observar el mapa con las ubicaciones de todos los aforos de la ciudad de Valladolid, se concluyó que en muchas de las ubicaciones objeto de medida no existía un aforo cercano para la obtención de este parámetro (muchos de esos medidores se ubican en puntos de elevada demanda y calles principales de la ciudad, no en las calles de poca circulación dentro de los barrios). Por consiguiente, a falta de una información suficiente sobre los valores de esta

variable y ante la imposibilidad de recoger dicha información en el marco de este trabajo, se decidió no incluirla en el estudio final.

## **12. Velocidad máxima**

En un principio se pensó que podía tratarse de una variable de interés, ya que la limitación de velocidad es un factor importante en la accidentalidad, sobre todo de cara a las lesiones producidas en los peatones (principal objeto de este estudio): cuanta mayor es la velocidad mayor es el riesgo de que se produzcan lesiones de más gravedad.

Sin embargo, debido a que en zona urbana generalmente el límite es de 50 km/h, salvo en ciertas zonas residenciales en las que la limitación es de 30 km/h, se decidió no incluir dicho parámetro dentro del grupo de variables independientes.

## **13. Tipo de pavimento**

Inicialmente se valoró incluir esta variable debido a que el rozamiento que sufren las ruedas del vehículo en función del tipo de pavimento podría ser de relevancia a la hora de que el vehículo salga de la calzada. Por ejemplo, un pavimento con adoquines hace que el vehículo sufra una menor adherencia comparada con el asfalto, lo que podría provocar que el conductor controlase peor el vehículo.

Sin embargo, debido a que existen pocos puntos en la ciudad que tengan un tipo de pavimento diferente al asfalto, finalmente se decidió no medir esta variable.

## **14. Visibilidad**

Aunque es un aspecto que puede ser significativo a la hora de que se produzca cualquier tipo de accidente, finalmente no se tuvo en cuenta debido a la dificultad para su medición; ya que habría que observar el entorno concreto de la ubicación para detectar la existencia de vegetación, edificios y otros elementos que habrían podido impedir la fácil visualización de otros vehículos o peatones en la calzada. Así mismo, para conocer mejor este parámetro habría que saber exactamente cómo se produjo el accidente, lo que resulta bastante difícil a partir de los datos existentes.

Finalmente, las variables que se decidieron recoger como posibles factores determinantes a la hora de que se produzca una salida de calzada fueron las siguientes:

<b>TIPO DE INTERSECCIÓN</b>	1 - En T o Y 2 - En X o + 3 - Giratoria 4 - Glorieta partida 5 - Curva
<b>RADIO DE CURVATURA</b>	Variable continua
<b>ANCHO DE CALZADA</b>	Variable continua
<b>ANCHO DE CARRIL</b>	Variable continua
<b>NÚMERO DE CARRILES PARA EL SENTIDO DE LA CIRCULACIÓN</b>	Variable discreta
<b>SENTIDO</b>	1 - Sentido único 2 - Doble sentido 3 - Doble sentido con mediana divisoria
<b>LONGITUD DEL TRAMO RECTO ANTERIOR</b>	Variable continua
<b>DISTANCIA AL SEMÁFORO ANTERIOR</b>	Variable continua
<b>PENDIENTE</b>	1 - Sí 2 - No
<b>PRIORIDAD</b>	1 - Semáforo 2 - Señal de STOP 3 - Señal de CEDA EL PASO 4 - Ninguna (sólo norma)

Tabla 6. Resumen de variables finales.





# PARTE II. ANÁLISIS DE LOS DATOS



## CAPÍTULO 5. ANÁLISIS DESCRIPTIVO DE LAS VARIABLES Y *ODDS-RATIO*

---

Una vez medidas todas las variables, tanto para los casos como para los controles, se realizó un análisis descriptivo con el objetivo de ver el comportamiento de las variables elegidas como posibles factores de riesgo. En el caso de las variables categóricas la descripción se hace mediante tablas de frecuencias y diagramas de barras múltiples. Por otro lado, en el caso de las variables cuantitativas utilizaremos histogramas y diagramas de cajas múltiples.

El propósito de este análisis es observar cuáles son las diferencias principales entre las distribuciones de cada una de las variables para los casos y los controles. Es decir, se trata de detectar la presencia de una diferencia notable que se relacione con una mayor o menor probabilidad de cara a la producción de una salida de vía, y así poder anticipar el posible papel de las distintas variables en el modelo final.

Posteriormente, para estudiar mejor la asociación entre cada variable y el desenlace (caso o control) se obtuvieron las tablas cruzadas de cada uno de los factores frente a la variable respuesta “Caso o control”. En estadística, las tablas de contingencia [37] se emplean para analizar la asociación entre dos o más variables, habitualmente de naturaleza cualitativa (nominales u ordinales).

En este caso, vamos a comparar las variables independientes, tanto categóricas como continuas, frente a la variable dependiente “caso o control”, para así ver la distribución de frecuencias de dichas variables en función de que sea caso o control, y poder estimar las interrelaciones entre ambas, así como su posible influencia como factor de riesgo o de protección. Para poder obtener la tabla cruzada de las variables independientes continuas se categorizarán en tantos grupos como sea oportuno.

Por último, se calculó la estimación del *Odds-Ratio* como medida de la asociación estadística entre la variable de interés y la variable “Caso o control”, con objeto de determinar si alguna de las características geométricas establecidas supone un factor de riesgo, incrementando la probabilidad de que ocurra una salida de calzada.

## 5.1. TIPO DE INTERSECCIÓN

Como se trata de una variable categórica, a continuación se recoge la tabla de frecuencias tanto para los casos (intersecciones y curvas urbanas donde han ocurrido accidentes considerados como salidas de vía) como para los controles. Así mismo, para ver mejor las diferencias entre las frecuencias de esta variable, en función de si se trata de caso o control, se ha obtenido la tabla cruzada frente a la variable “Caso o control”.

**Tabla de frecuencias de “Tipo de Intersección” para los casos**

	Frecuencia	Porcentaje
En T o Y	21	25,6
En X o +	12	14,6
Giratoria	9	11,0
Glorieta partida	21	25,6
Curva	19	23,2
Total	82	100,0

Tabla 7. Tabla de frecuencias de "Tipo de intersección" para los casos.

**Tabla de frecuencias de “Tipo de Intersección” para los controles**

	Frecuencia	Porcentaje
En T o Y	37	45,1
En X o +	19	23,2
Giratoria	4	4,9
Glorieta partida	3	3,7
Curva	19	23,2
Total	82	100,0

Tabla 8. Tabla de frecuencias de "Tipo de intersección para los controles.

Las tablas anteriores permiten observar, de forma comparativa, las frecuencias de la variable “Tipo de intersección” tanto para los casos (salida de calzada) como para los controles. Cabe destacar que para las intersecciones en T o Y existe una mayor preponderancia en los controles (hay una frecuencia de aproximadamente el doble) y que las glorietas partidas son mucho más

abundantes en los casos (un 25,6% de los casos se produjeron en glorieta partida mientras que únicamente un 3,7% lo fueron en los controles).

Así mismo, las intersecciones giratorias también predominan más en los casos, aunque la diferencia de frecuencias sea mucho menor que para las glorietas partidas. Por otro lado, hay que decir que la categoría “curva” tiene la misma frecuencia en ambos grupos de individuos, ya que se midieron como controles el mismo número de curvas que las existentes en los casos. Estas conclusiones se pueden ver de forma más clara en la siguiente tabla cruzada.

**Tabla de contingencia para la variable “Tipo de intersección”**

		Caso o control		Total
		Control	Caso	
Tipo de intersección	En T o Y	37 45,1%	21 25,6%	58 35,4%
	En X o +	19 23,2%	12 14,6%	31 18,9%
	Giratoria	4 4,9%	9 11,0%	13 7,9%
	Glorieta partida	3 3,7%	21 25,6%	24 14,6%
	Curva	19 23,2%	19 23,2%	38 23,2%
Total		82 100,0%	82 100,0%	164 100,0%

Tabla 9. Tabla cruzada de "Tipo de intersección" frente a "Caso o control".

Con estas observaciones se podría deducir que las intersecciones más simples en T o Y son las que menos influyen en la ocurrencia de una salida de la vía, ya que existe una mayor preponderancia para los controles. Sin embargo, otras intersecciones más complejas, como las glorietas o glorietas partidas, suponen a priori un mayor riesgo de cara a que se produzca este tipo de accidente. Por lo tanto, debido a que la primera de las categorías parece influir menos en la ocurrencia del evento de estudio, es la que se tomará como referencia en los análisis posteriores.

En el siguiente gráfico de barras conjunto se muestra de forma más visual la asociación entre ambas variables: “Tipo de intersección” y “Caso o control”.

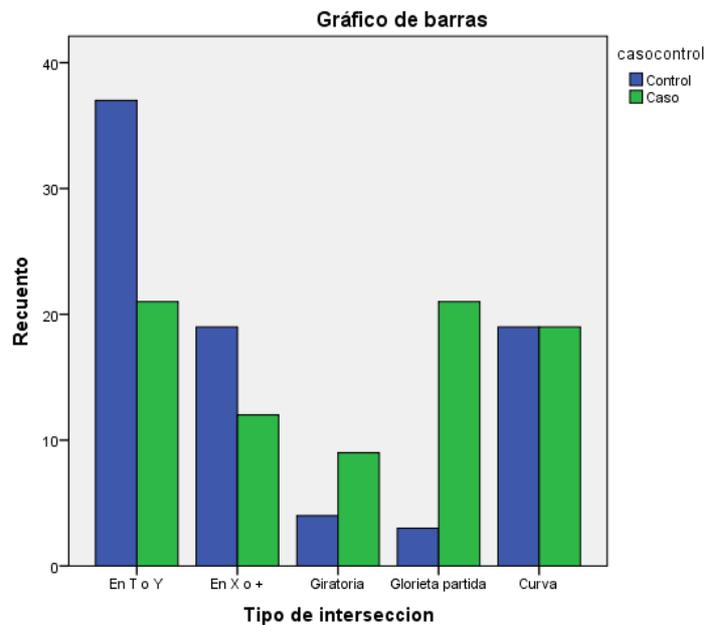


Ilustración 24. Gráfico de barras agrupadas de "Tipo de intersección" respecto a "Caso o control".

### ***Odds-Ratio de la variable "Tipo de intersección"***

Para valorar la asociación estadística entre ambas variables se recurre a la obtención del *Odds-Ratio* (OR), junto con su significación estadística (p-valor) y el Intervalo de Confianza al 95%. Primeramente, hay que tener en cuenta que "Tipo de intersección" no es una variable dicotómica, por lo que para el cálculo del OR se tomará como categoría de referencia la que se sospecha que influirá menos en una salida de calzada, según el análisis descriptivo anterior.

Para la obtención de estos valores hay que realizar una regresión logística binaria con la variable "Caso o Control" como dependiente y "Tipo de intersección" como independiente, cuya categoría de referencia se establece en la primera de todas: "intersección en T o Y", tal y como se determinó en el análisis descriptivo previo.

El OR se calcula como  $\exp(\beta)$  en la regresión logística, como se puede ver en la siguiente tabla.

	$\beta$	Error estándar	Wald	gl	Sig.	Exp(B)	95% C.I. para EXP( $\beta$ )	
							Inferior	Superior
En T o Y			17,24	4	,002			
En X o +	,107	,459	,054	1	,816	1,113	,453	2,736
Giratoria	1,377	,660	4,353	1	,037	3,964	1,087	14,456
Glorieta partida	2,512	,675	13,85	1	,000	12,333	3,285	46,305
Curva	,566	,424	1,783	1	,182	1,762	,767	4,046
Constante	-,566	,273	4,298	1	,038	,568		

Tabla 10. Odds-Ratio de la variable "Tipo de intersección".

En cuanto a los resultados observados se puede llegar a las siguientes conclusiones:

- El OR, en el caso de las intersecciones en "X o +", es 1,113. Sin embargo, el Intervalo de Confianza (IC) al 95% es muy amplio y contiene al 1, por lo que la asociación no es significativa. Es decir, no existen diferencias significativas entre las intersecciones en "T o Y" y las intersecciones en "X o +".
- Las intersecciones giratorias tienen un OR de 3,964, es decir, mayor que 1, lo que indicaría que el riesgo de que se produzca una salida de calzada se multiplica prácticamente por 4 cuando hay una glorieta frente a cuando hay una intersección de tipo "T o Y". Además, se trata de una asociación significativa porque tanto el límite superior como inferior del Intervalo de Confianza son mayores que 1. Por lo tanto, se puede concluir que las intersecciones giratorias son un factor de riesgo, ya que aumentan las probabilidades de una salida de la vía.
- Por último, las glorietas partidas tienen un OR de 12,333, lo que indica que es doce veces más probable que se produzca una salida de calzada en una glorieta partida que en una intersección en "T o Y". El p-valor es prácticamente 0, por lo que se trata de un valor significativo. Se concluye entonces que, al igual que las intersecciones giratorias, las glorietas partidas actúan como un factor de riesgo.

## 5.2. RADIO DE CURVATURA

A continuación se estudia el comportamiento de la variable continua “Radio de curvatura” a través de dos histogramas, uno para los casos y otro para los controles, seguidos de un box-plot múltiple para comparar mejor el comportamiento de dicha variable en función de si se trata de caso o control.

### *Histograma de “Radio de curvatura” para los casos*

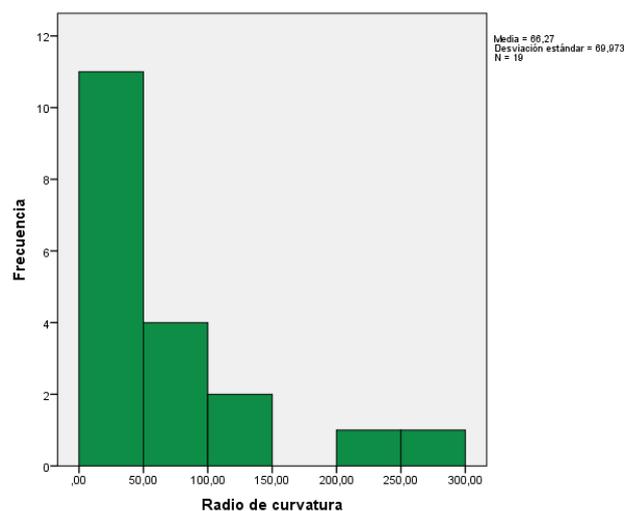


Ilustración 25. Histograma de "Radio de curvatura" para los casos.

### *Histograma de “Radio de curvatura” para los controles*

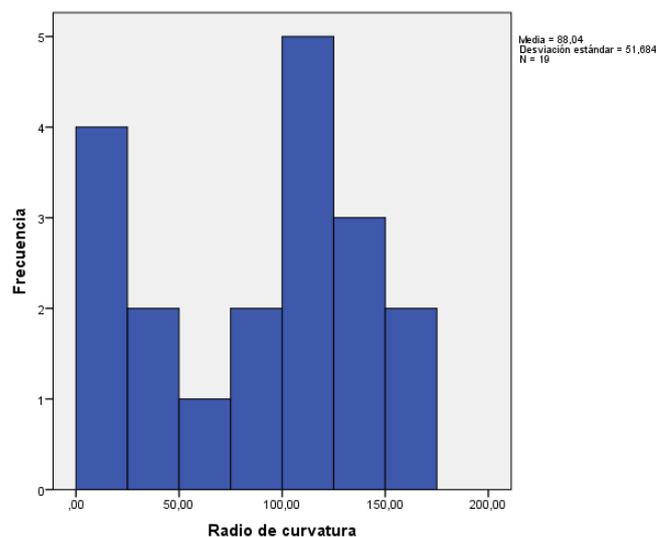


Ilustración 26. Histograma de "Radio de curvatura" para los controles.

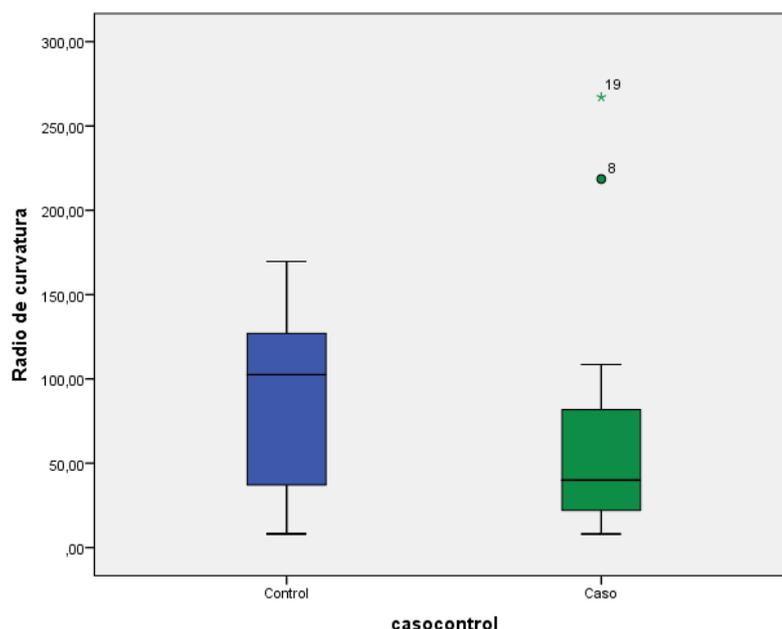
**Box-plot múltiple de “Radio de curvatura”**

Ilustración 27. Box-plot múltiple de "Radio de curvatura" frente a "Caso o control".

A través de la visualización de los histogramas anteriores se puede comprobar que la variable “Radio de curvatura” para los casos presenta una asimetría positiva, mientras que para los controles muestra una distribución plana con algunos dientes de sierra, aspecto frecuente en aquellas situaciones en las que existen pocos datos, tal y como ocurre en este estudio en el que sólo se dispone de 19 datos medidos en curvas.

Por otro lado, los datos se encuentran especialmente agrupados para los casos, ya que el grueso de la distribución se concentra en valores menores. Así mismo, la distribución de éstos presenta una menor dispersión que para los controles. Aunque existen dos datos fuera del diagrama de caja para los casos, son datos que entran dentro de la asimetría positiva, ya que pertenecen a la cola de esta distribución.

**Tabla de contingencia y Odds-Ratio de “Radio de curvatura”**

Al tratarse de una variable continua, lo primero que debemos hacer es categorizarla para poder obtener la tabla de contingencia y los OR; aunque posteriormente, al realizar la regresión logística con todas las variables, se incluirá como cuantitativa y no como categórica.

Las categorías seleccionadas para esta variable han sido las siguientes:

- Menor o igual que 25 metros.
- Entre 25,01 y 50 metros.
- Entre 50,01 y 100 metros.
- Entre 100,01 y 200 metros.
- Mayor o igual que 200,01 metros.
- Intersección (equivale a un radio de curvatura de 1000 metros, que es el valor que se decidió incluir en el caso de las intersecciones).

A continuación, se muestra el gráfico de barras para la variable codificada en el que se observa que el porcentaje de curvas en cada tramo de la variable es prácticamente el mismo.

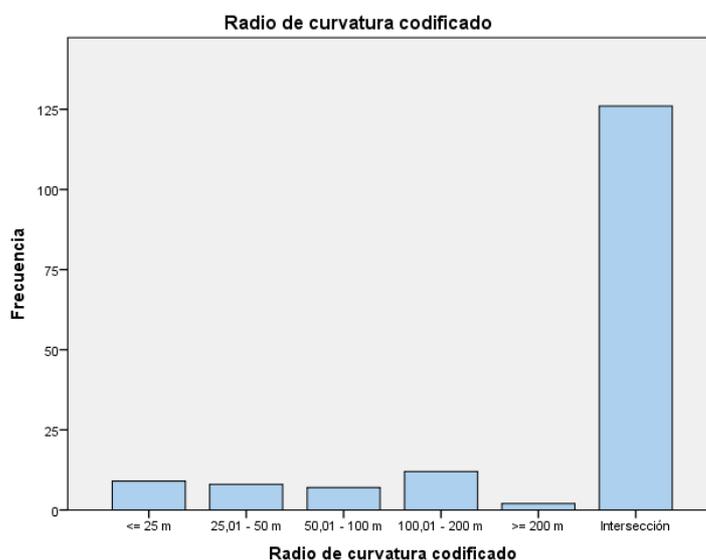


Ilustración 28. Gráfico de barras de la variable "Radio de curvatura codificado".

Seguidamente, se muestra la tabla cruzada para la variable "Radio de curvatura codificado" frente a "Caso o control". Observando las frecuencias de cada categoría del radio de curvatura en casos o controles se comprueba que los porcentajes son aproximadamente similares para todas ellas, salvo para la de curvas de radio entre 100,01 y 200 metros, en la que hay una mayor cantidad de controles que de casos. Esto concuerda con las conclusiones extraídas del análisis descriptivo anterior, ya que la mediana de dicha variable en los controles se encontraba en 100 metros, mientras que los casos toman valores más bajos por lo general.

		Caso o control		Total
		Control	Caso	
Radio de curvatura codificado	<= 25 m	4 4,9%	5 6,1%	9 5,5%
	25,01 - 50 m	2 2,4%	6 7,3%	8 4,9%
	50,01 - 100 m	3 3,7%	4 4,9%	7 4,3%
	100,01 - 200 m	10 12,2%	2 2,4%	12 7,3%
	>= 200 m	0 0,0%	2 2,4%	2 1,2%
	Intersección	63 76,8%	63 76,8%	126 76,8%
Total		82 100,0%	82 100,0%	164 100,0%

Tabla 11. Tabla cruzada de "Radio de curvatura codificado" frente a "Caso o control".

Para analizar de forma estadística la asociación entre ambas variables, se calcula el *Odds-Ratio* a través de una regresión logística con la variable "Radio de curvatura codificado" como única variable independiente, y con la primera categoría como referencia (radio de curvatura menor de 25 metros). De este análisis se obtuvieron los siguientes resultados:

	$\beta$	Error estándar	Wald	gl	Sig.	Exp(B)	95% C.I. para EXP( $\beta$ )	
							Inferior	Superior
<= 25 m			6,38	5	,271			
25,01 - 50 m	,875	1,057	,686	1	,407	2,400	,303	19,041
50,01 - 100 m	,065	1,017	,004	1	,949	1,067	,145	7,822
100,01 - 200 m	-1,83	1,025	3,19	1	,074	,160	,021	1,192
>= 200 m	20,98	28420,7	,000	1	,999	12923798	,000	.
Intersección	-,223	,694	,103	1	,748	,800	,205	3,118
Constante	,223	,671	,111	1	,739	1,250		

Tabla 12. *Odds-Ratio* de la variable "Radio de curvatura codificado".

Tras observar los valores de los límites inferiores y superiores de los Intervalos de Confianza, o equivalentemente los p-valores de cada categoría, se puede concluir que ninguna de las asociaciones es significativa, ya que todos los IC contienen el valor 1. Por lo tanto, no hay diferencias significativas entre cualquiera de las categorías del radio de curvatura frente a los radios pequeños. Esto se traduce en que, independientemente del tamaño del radio de curvatura (pequeño o grande), éste no influye significativamente para incrementar el riesgo de que ocurra una salida de calzada.

### 5.3. ANCHO DE CALZADA

Las siguientes gráficas recogen los histogramas para los casos y los controles en cuanto al ancho de calzada. Es importante destacar que, tal y como se mencionó anteriormente, para la medición de esta variable ante la existencia de medianas con isletas divisorias, únicamente se ha tenido en cuenta el ancho para todos los carriles de un lado de la mediana, excluyendo los del otro lado de ésta.

#### *Histograma de "Ancho de calzada" para los casos*

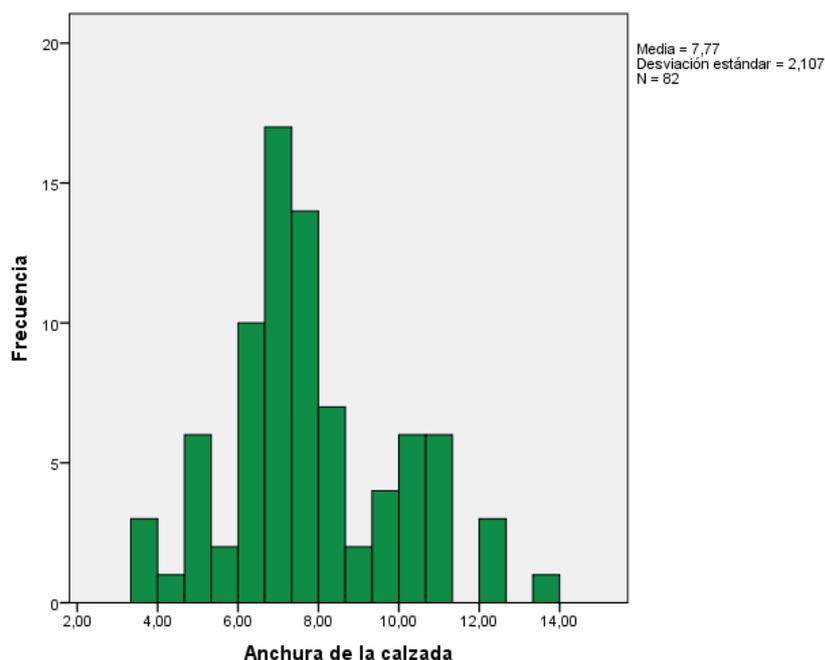


Ilustración 29. Histograma de "Ancho de calzada" para los casos.

### Histograma de "Ancho de calzada" para los controles

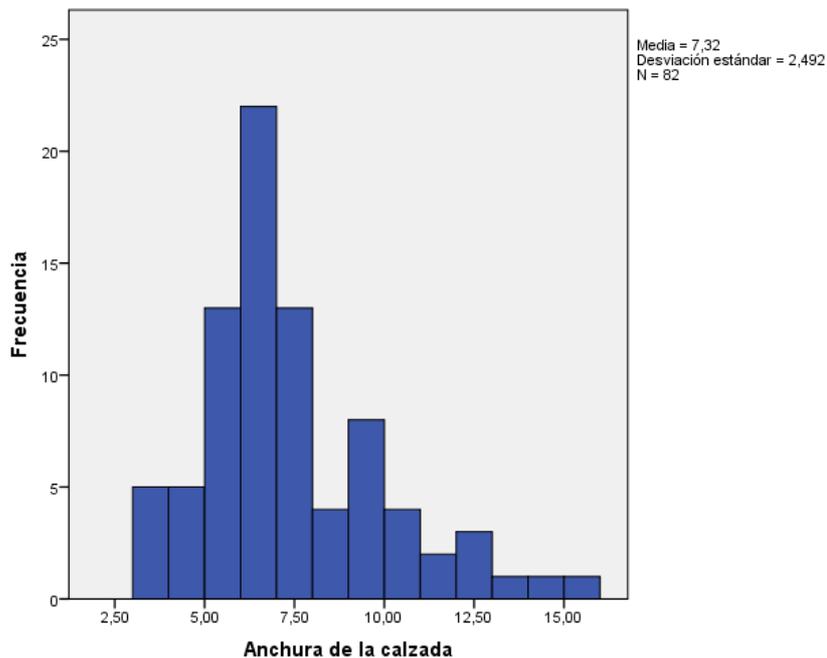


Ilustración 30. Histograma de "Ancho de calzada" para los controles.

### Box-plot múltiple de "Ancho de calzada"

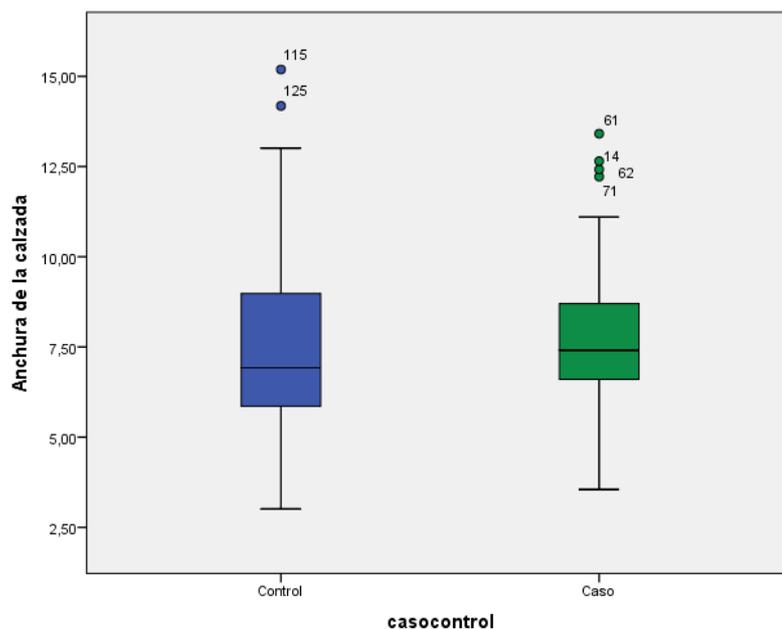


Ilustración 31. Box-plot múltiple de "Ancho de calzada" frente a "Caso o control".

Tras observar tanto los histogramas de forma conjunta como el box-plot múltiple, se puede determinar que el ancho de calzada se comporta de forma similar independientemente de si es caso o control, con una distribución bastante simétrica y una dispersión ligeramente mayor en los controles.

### **Tabla de contingencia y Odds-Ratio de “Ancho de calzada”**

Al igual que el radio de curvatura, el ancho de la calzada es una variable continua, por lo que para poder realizar el análisis actual (tabla de contingencia y *Odds-Ratio*) se requiere su categorización, aunque en la regresión logística posterior se incluya como una variable independiente cuantitativa. Se decidió dividir dicha variable en las categorías que se muestran a continuación, escogidas en base a los distintos rangos de anchura de calzada habituales en las calles de zona urbana.

- Menor de 6 metros.
- Entre 6,01 y 8 metros.
- Entre 8,01 y 10 metros.
- Mayor de 10,01 metros.

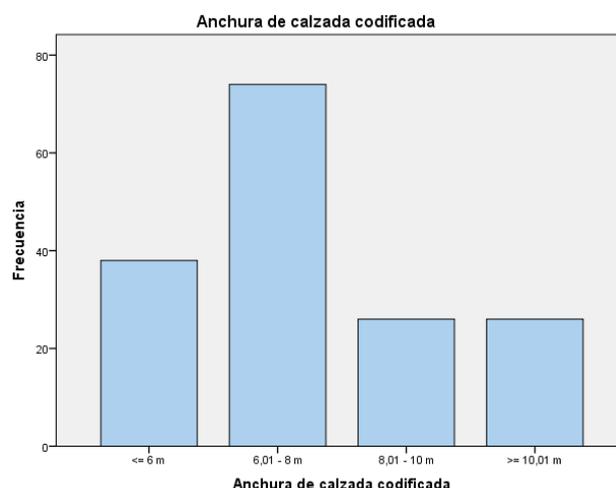


Ilustración 32. Gráfico de barras de la variable "Anchura de calzada codificada".

Una vez categorizada la variable se procedió a obtener su tabla de contingencia, así como los OR con el mismo procedimiento que para las variables anteriores.

		Caso o control		Total
		Control	Caso	
Anchura de calzada codificada	<= 6 m	25 30,5%	13 15,9%	38 23,2%
	6,01 - 8 m	34 41,5%	40 48,8%	74 45,1%
	8,01 - 10 m	12 14,6%	14 17,1%	26 15,9%
	>= 10,01 m	11 13,4%	15 18,3%	26 15,9%
Total		82 100,0%	82 100,0%	164 100,0%

Tabla 13. Tabla cruzada de "Anchura de calzada codificada" frente a "Caso o control".

A través de la tabla cruzada se puede observar que los porcentajes de cada categoría de ancho de calzada no difieren mucho entre casos y controles, siendo la diferencia más notable en las calzadas de menor ancho (menos de 6 metros), para las que hay aproximadamente el doble de controles que de casos. Por lo tanto, esta es la categoría que se tomará como referencia para calcular el OR, ya que es la que se sospecha que influirá en menor medida a la hora de que se produzca una salida de calzada.

Para la obtención de los *Odds-Ratio* o razón de ventajas se ha realizado el análisis de tres formas distintas:

- Empleando todo el conjunto de datos, tanto curvas como intersecciones.
- Únicamente el subconjunto de intersecciones.
- Únicamente el subconjunto de curvas.

El objetivo de realizar estos tres análisis es determinar si es mejor estudiar intersecciones y curvas en un mismo grupo, o si se obtienen mejores resultados tratando ambos subconjuntos por separado. Este mismo procedimiento se empleará para todas las demás variables medidas.

Con objeto de obtener el valor estimado del OR así como su Intervalo de Confianza al 95% se realiza una regresión logística binaria con la variable "Ancho de calzada codificado", estableciendo como categoría de referencia la primera de todas (ancho de calzada menor o igual que 6 metros).

---

**INTERSECCIONES Y CURVAS**

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
<= 6 m			4,903	3	,179			
6,01 - 8 m	,816	,414	3,891	1	,049	2,262	1,005	5,092
8,01 - 10 m	,808	,521	2,403	1	,121	2,244	,808	6,232
>= 10,01 m	,964	,524	3,386	1	,066	2,622	,939	7,323
Constante	-,654	,342	3,657	1	,056	,520		

Tabla 14. Odds-Ratio de la variable "Ancho de calzada codificado" para curvas e intersecciones.

Con este primer modelo logístico se pueden extraer las siguientes conclusiones:

- Un ancho de calzada entre 6,01 y 8 metros frente a una calzada estrecha (ancho menor de 6 metros) es un factor de riesgo, lo que se traduce en que los lugares con estas características son más propensos a que ocurra una salida de la vía. En concreto, si se dan estas circunstancias el riesgo se multiplica aproximadamente por dos.
- Anchos de calzada entre 8,01 y 10 metros no actúan ni como factor de riesgo ni de protección, ya que el límite inferior del Intervalo de Confianza es menor que 1 y el límite superior mayor que 1.
- Lo mismo ocurre con anchos mayores de 10 metros; por lo que no existen diferencias significativas frente a anchos inferiores a 6 metros.

---

**INTERSECCIONES**

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
<= 6 m			5,289	3	,152			
6,01 - 8 m	1,040	,473	4,847	1	,028	2,830	1,121	7,147
8,01 - 10 m	1,004	,570	3,108	1	,078	2,730	,894	8,338
>= 10,01 m	,742	,608	1,488	1	,222	2,100	,638	6,916
Constante	-,742	,384	3,729	1	,053	,476		

Tabla 15. Odds-Ratio de la variable "Ancho de calzada codificado" para las intersecciones.

Al tratar por separado el subconjunto de intersecciones, las conclusiones son similares a cuando se analizan conjuntamente todos los datos:

- Si el ancho de la calzada está comprendido entre 6,01 y 8 metros actúa como factor de riesgo frente a un ancho inferior a 6 metros. A diferencia del caso anterior el OR es algo mayor, 2,83 frente a 2,262, con lo que se advierte que al tratar por separado las intersecciones existe un mayor riesgo de que ocurra salida de calzada en lugares con estas características.
- Aunque el ancho de calzada entre 8,01 y 10 metros no es significativo a un nivel del 5%, sí lo es al 10% (el p-valor es 0,078). Por consiguiente, bajo estas consideraciones, actuaría como un factor de riesgo.
- De la misma forma que en el caso anterior los anchos de calzada superiores no actúan ni como factor de riesgo ni de protección frente a las calzadas más estrechas, ya que los p-valores son elevados.

CURVAS

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
<= 6 m			2,462	3	,482			
6,01 - 8 m	,087	,886	,010	1	,922	1,091	,192	6,196
8,01 - 10 m	-,405	1,443	,079	1	,779	,667	,039	11,285
>= 10,01 m	1,386	1,118	1,537	1	,215	4,000	,447	35,788
Constante	-,288	,764	,142	1	,706	,750		

Tabla 16. Odds-Ratio de la variable "Ancho de calzada codificado" para las curvas.

Por último, al analizar el subconjunto de curvas se comprueba que ninguna de las asociaciones es significativa, con lo que no existe ningún resultado determinante que influya en la ocurrencia de una salida de la vía.

## 5.4. ANCHO DE CARRIL

### Histograma de "Ancho de carril" para los casos

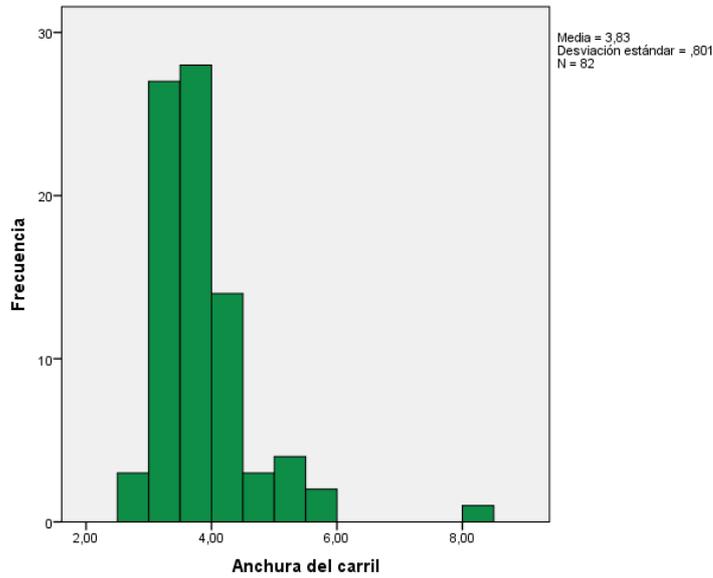


Ilustración 33. Histograma de "Ancho de carril" para los casos.

### Histograma de "Ancho de carril" para los controles

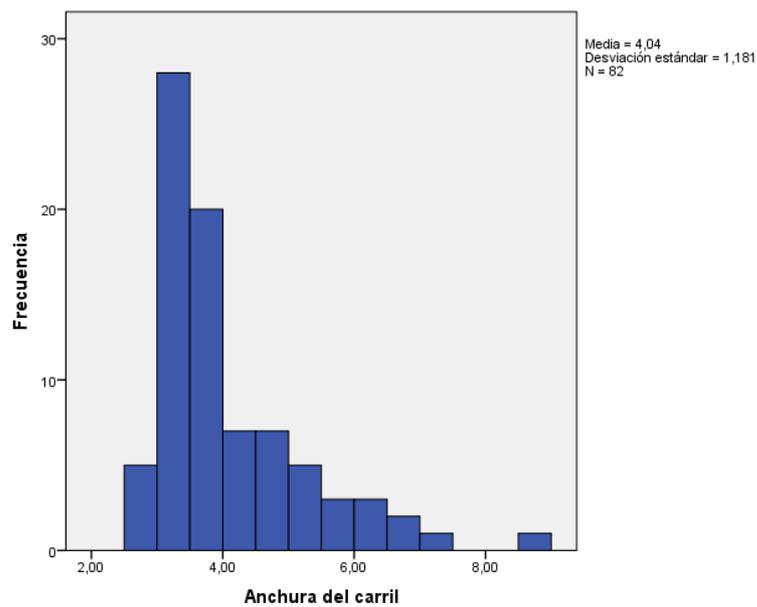


Ilustración 34. Histograma de "Ancho de carril" para los controles.

### Box-plot múltiple de "Ancho de carril"

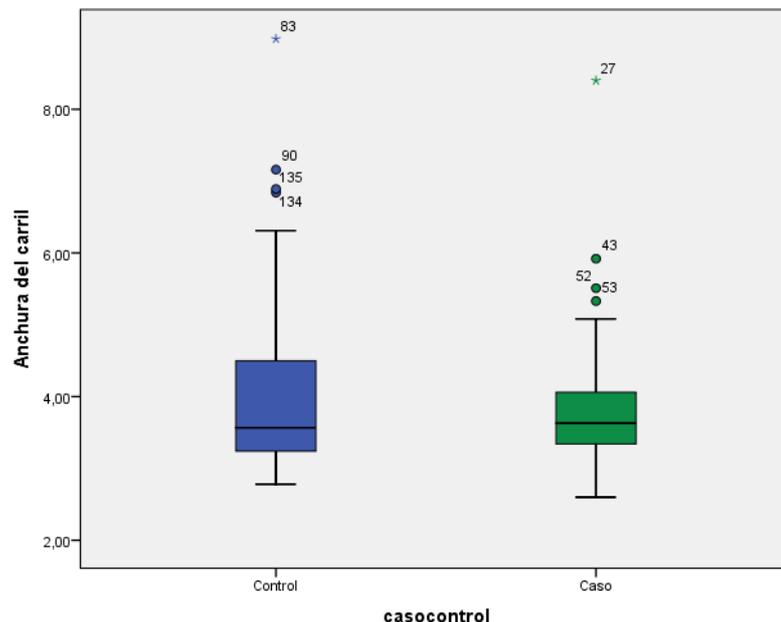


Ilustración 35. Box-plot múltiple de "Ancho de carril" frente a "Caso o control".

Tras observar los dos histogramas se puede comprobar que ambos tienen formas similares, con una ligera asimetría positiva y el grueso de la distribución centrado en valores parecidos. Además, a través del *box-plot* múltiple se puede concluir que, aunque con comportamientos análogos, el ancho de carril para los controles tiene una mayor dispersión y llega a tomar valores algo superiores.

### Tabla de contingencia y Odds-Ratio de "Ancho de carril"

La categorización de esta variable se ha realizado teniendo en cuenta cuáles suelen ser los valores estándar para el ancho de carril en zona urbana (categorías similares a las incluidas en el cuestionario de ARENA para la variable "anchcarr"). Se plantearon las siguientes categorías:

- Menor que 3,25 metros.
- Entre 3,26 y 3,7 metros.
- Entre 3,71 y 4 metros.
- Mayor que 4,01 metros.

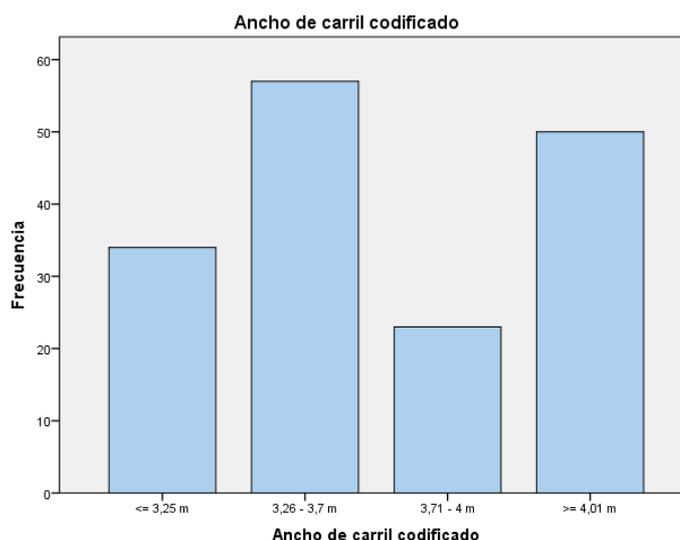


Ilustración 36. Gráfico de barras de la variable "Ancho de carril codificado".

Para observar las interrelaciones entre dicha variable, ya codificada, y el desenlace estudiado, se realizó la tabla cruzada siguiente:

		Caso o control		Total
		Control	Caso	
Ancho de carril codificado	<= 3,25 m	22 26,8%	12 14,6%	34 20,7%
	3,26 - 3,7 m	26 31,7%	31 37,8%	57 34,8%
	3,71 - 4 m	6 7,3%	17 20,7%	23 14,0%
	>= 4,01 m	28 34,1%	22 26,8%	50 30,5%
Total		82 100,0%	82 100,0%	164 100,0%

Tabla 17. Tabla cruzada de "Ancho de carril codificado" frente a "Caso o control".

Al comparar los datos porcentuales de cada categoría para casos y controles se comprueba que, en la mayoría de ellas, los porcentajes son prácticamente equiparables, siendo las diferencias más notables en los carriles de menos de 3,25 metros en los que predominan los controles. Así mismo, en el caso de carriles entre 3,71 y 4 metros de ancho destaca el número de casos frente a controles, ya que es más del doble.

Por lo tanto, se puede deducir que los carriles más estrechos y los más anchos son más frecuentes en los controles, mientras que las salidas de calzada se presentan en ubicaciones con anchos de carriles medios.

Con el propósito de conseguir una mejor interpretación de la influencia del ancho de carril en las salidas de vía, se realizó una regresión logística binaria con la variable "Ancho de carril codificado" como variable independiente y con la primera de las categorías como referencia. Al igual que para la variable anterior, "Anchura de calzada codificada", se realizaron tres análisis diferentes con los distintos subconjuntos de datos, con el fin de determinar la forma en que se obtienen los mejores resultados.

### INTERSECCIONES Y CURVAS

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
<= 3,25 m			8,808	3	,032			
3,26 - 3,7 m	,782	,447	3,065	1	,080	2,186	,911	5,246
3,71 - 4 m	1,648	,595	7,662	1	,006	5,194	1,618	16,680
>= 4,01 m	,365	,458	,634	1	,426	1,440	,587	3,536
Constante	-,606	,359	2,853	1	,091	,545		

Tabla 18. Odds-Ratio de la variable "Ancho de carril codificado" para intersecciones y curvas.

De este primer análisis se pueden deducir las siguientes conclusiones:

- Un ancho de carril entre 3,26 y 3,7 metros frente a un carril más estrecho no es una asociación estadísticamente significativa, debido a que el Intervalo de Confianza contiene el valor 1.
- Sin embargo, un ancho de carril entre 3,71 y 4 metros actúa como factor de riesgo con un OR significativamente mayor que 1: 5,194. Esto quiere decir que en curvas o intersecciones donde el carril tiene un ancho entre 3,71 y 4 metros el riesgo es cinco veces mayor que si tuviera un ancho de carril menor que 3,25 metros.
- Los carriles más amplios, de más de 4,01 metros, tampoco son determinantes a la hora de que se produzca el evento estudiado.

## INTERSECCIONES

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
<= 3,25 m			7,875	3	,049			
3,26 - 3,7 m	,742	,502	2,181	1	,140	2,100	,785	5,621
3,71 - 4 m	1,417	,712	3,956	1	,047	4,125	1,021	16,667
>= 4,01 m	-,134	,529	,064	1	,801	,875	,310	2,467
Constante	-,405	,408	,986	1	,321	,667		

Tabla 19. Odds-Ratio de la variable "Ancho de carril codificado" para las intersecciones.

Al analizar las intersecciones por separado los resultados son similares al caso anterior.

- Una anchura de carril entre 3,26 y 3,7 metros, al igual que las mayores de 4,01 metros, no son estadísticamente significativos.
- De la misma forma que al analizar curvas e intersecciones conjuntamente, los carriles cuyo ancho está entre 3,71 y 4 metros actúan como factor de riesgo para que se produzca una salida de calzada, ya que su Odds-Ratio toma un valor de 4,125, algo inferior que en el modelo logístico anterior.

## CURVAS

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
<= 3,25 m			6,490	3	,090			
3,26 - 3,7 m	,560	1,069	,274	1	,601	1,750	,215	14,224
3,71 - 4 m	2,351	1,144	4,222	1	,040	10,500	1,115	98,914
>= 4,01 m	1,946	1,009	3,720	1	,054	7,000	,969	50,567
Constante	-1,25	,802	2,441	1	,118	,286		

Tabla 20. Odds-Ratio de la variable "Ancho de carril codificado" para las curvas.

Por último, al estudiar únicamente las curvas, se puede ver que las conclusiones son las mismas:

- Las anchuras de carril entre 3,26 y 3,7 metros y las mayores de 4,01 metros no presentan diferencias significativas frente a las menores de 3,25 metros.
- Así mismo, si el carril tiene un ancho entre 3,71 y 4 metros entraña un riesgo mayor para que suceda una salida de la vía. Además, cabe destacar que, en este caso, el OR es muy superior ya que toma un valor de 10,5, con lo que se puede concluir que el riesgo, en estas circunstancias, es más elevado para las curvas que para las intersecciones.

## 5.5. NÚMERO DE CARRILES PARA EL SENTIDO

Aunque no sea una variable categórica sino discreta, a continuación se recogen las tablas de frecuencias para los casos y los controles por separado, así como la tabla cruzada correspondiente.

**Tabla de frecuencias de "Número de carriles" para los casos**

	Frecuencia	Porcentaje
1	33	40,2
2	35	42,7
3	13	15,9
4	1	1,2
Total	82	100,0

Tabla 21. Tabla de frecuencias de "Número de carriles" para los casos.

**Tabla de frecuencias de "Número de carriles" para los controles**

	Frecuencia	Porcentaje
1	63	76,8
2	16	19,5
3	3	3,7
Total	82	100,0

Tabla 22. Tabla de frecuencias de "Número de carriles" para los controles.

**Tabla de contingencia para la variable "Número de carriles"**

		Caso o control		Total
		Control	Caso	
Número de carriles	1	63 76,8%	33 40,2%	96 58,5%
	2	16 19,5%	35 42,7%	51 31,1%
	3 o más	3 3,7%	14 17,1%	16 9,8%
Total		82 100,0%	82 100,0%	82 100,0%

Tabla 23. Tabla cruzada de "Número de carriles" frente a "Caso o control".

Como se puede observar en las tablas de frecuencias y la tabla cruzada anteriores, la mayor parte de las salidas de calzada se produjeron en lugares con uno o dos carriles para el mismo sentido de la circulación de forma prácticamente equitativa. Sin embargo, aparecen algunas ubicaciones con más de dos carriles para el mismo sentido (principalmente tres carriles).

A través del análisis descriptivo de esta variable para los controles se puede observar que la mayor parte de ellos se sitúan en vías con un único carril para el sentido de la circulación, seguidos de vías con dos carriles. A diferencia de los casos, hay un porcentaje muy pequeño de ubicaciones con tres o más carriles (únicamente un 3,7% de los controles se ubican en calzadas con estas características, mientras que un 17,1% corresponde a los casos).

En conclusión, el riesgo de que ocurra una salida de calzada parece crecer cuando aumenta el número de carriles. Por consiguiente, se puede deducir que hay una preponderancia de vías con pocos carriles para el mismo sentido en los controles (generalmente un carril), mientras que en las salidas de vía hay una mayor tendencia a calzadas de más carriles. Todo ello se puede ver de forma más visual en el siguiente gráfico de barras conjunto.

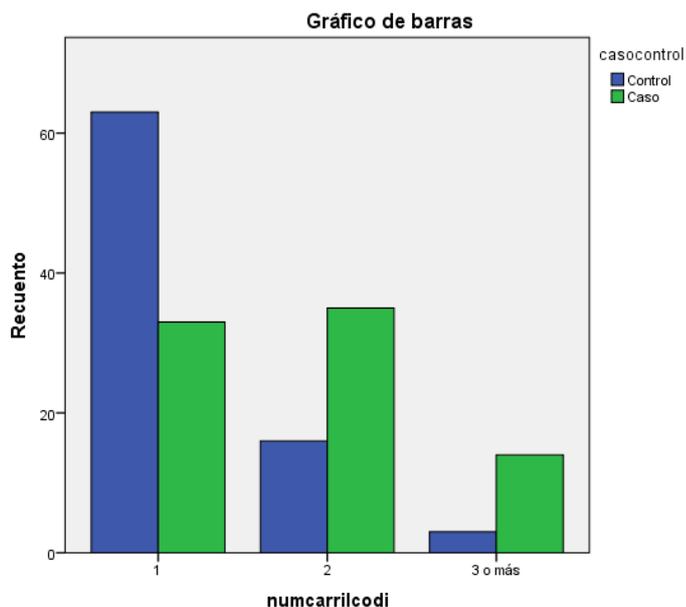


Ilustración 37. Gráfico de barras agrupadas de "Número de carriles" frente a "Caso o control".

### Odds-Ratio de la variable "Número de carriles"

Con el objetivo de conocer mejor la asociación entre la variable "Número de carriles para el sentido de la circulación" y "Caso o control", se realiza una regresión logística binaria en la que la variable independiente es el número de carriles, interpretada como una variable cualitativa cuyo valor de referencia es la primera de las categorías (un carril).

#### INTERSECCIONES Y CURVAS

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
1			21,572	2	,000			
2	1,429	,370	14,887	1	,000	4,176	2,020	8,632
3 o más	2,187	,672	10,607	1	,001	8,909	2,389	33,222
Constante	-,647	,215	9,055	1	,003	,524		

Tabla 24. Odds-Ratio de la variable "Número de carriles" para intersecciones y curvas.

Las conclusiones para este análisis en el que se trabaja con los datos de curvas e intersecciones son las siguientes:

- La presencia de dos carriles para el mismo sentido frente a un carril actúa como factor de riesgo, ya que el OR toma un valor de 4,176 y el IC no comprende el valor 1 entre sus límites. Esto quiere decir que es aproximadamente cuatro veces más probable que ocurra una salida de calzada cuando hay dos carriles que cuando hay uno.
- En intersecciones y curvas con tres o más carriles para el sentido de la circulación se incrementa el riesgo de una salida de calzada. En este caso el valor del OR es aún mayor (el doble), con lo que se confirman las hipótesis iniciales del análisis descriptivo anterior: a medida que aumenta el número de carriles crece el riesgo de que se produzca una salida de la vía.

### INTERSECCIONES

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
1			26,302	2	,000			
2	2,148	,473	20,583	1	,000	8,568	3,387	21,672
3 o más	2,246	,694	10,485	1	,001	9,455	2,427	36,830
Constante	-,860	,254	11,439	1	,001	,423		

Tabla 25. Odds-Ratio de la variable "Número de carriles" para las intersecciones.

Al tratar únicamente con el subconjunto de intersecciones las conclusiones son similares a cuando se trabaja con todos los datos.

- La presencia de dos carriles para el sentido de la circulación también actúa como factor de riesgo. Sin embargo, en este análisis se ve una mayor influencia sobre la casuística de salida de calzada, ya que el valor del OR es prácticamente el doble del anterior. Por lo tanto, al separar las intersecciones se aprecia un incremento del riesgo de producirse una salida de la vía cuando hay dos carriles para el sentido de la circulación.
- Las calles con tres o más carriles para el sentido también actúan como factor de riesgo, con un valor prácticamente igual que en el modelo anterior.

## CURVAS

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
1			,175	2	,916			
2	-,288	,688	,175	1	,676	,750	,195	2,889
3 o más	21,203	28420,72	,000	1	,999	1615474842	,000	.
Constante	,000	,426	,000	1	1,000	1,000		

Tabla 26. Odds-Ratio de la variable "Número de carriles" para las curvas.

Al analizar por separado el subconjunto de curvas recogidas en la base de datos no se obtiene ninguna asociación significativa, debido a que todos los Intervalos de Confianza al 95% contienen el valor 1, con lo que no se llega a ningún resultado concluyente.

## 5.6. SENTIDO DE CIRCULACIÓN

A continuación, se recogen las tablas de frecuencia de esta variable para casos y controles por separado, con el objetivo de ver la frecuencia con la que ocurre cada una de las tres posibles categorías en función de si se trata de caso o control. Así mismo, para ver mejor la asociación entre ambas variables se ha realizado la tabla cruzada correspondiente y la regresión logística necesaria para la obtención de la estimación del Odds-Ratio y su Intervalo de Confianza.

### Tabla de frecuencias de "Sentido" para los casos

	Frecuencia	Porcentaje
Sentido único	22	26,8
Doble sentido	22	26,8
Doble sentido con mediana	38	46,3
Total	82	100,0

Tabla 27. Tabla de frecuencias de "Sentido" para los casos.

**Tabla de frecuencias de “Sentido” para los controles**

	Frecuencia	Porcentaje
Sentido único	34	41,5
Doble sentido	40	48,8
Doble sentido con mediana	8	9,8
Total	82	100,0

Tabla 28. Tabla de frecuencias de "Sentido" para los controles.

**Tabla de contingencia para la variable “Sentido”**

		Caso o control		Total
		Control	Caso	
Sentido	Sentido único	34 41,5%	22 26,8%	56 34,1%
	Doble sentido	40 48,8%	22 26,8%	62 37,8%
	Doble sentido con mediana	8 9,8%	38 46,3%	46 28,0%
Total		82 100,0%	82 100,0%	164 100,0%

Tabla 29. Tabla cruzada de "Sentido" frente a "Caso o control".

La tabla cruzada de las variables “Sentido” y “Caso o control” muestra las frecuencias de cada categoría: “sentido único”, “doble sentido” y “doble sentido con mediana”, frente a si se trata de caso o control. Con ello se puede ver una comparativa de la frecuencia de esta variable en ambas situaciones.

Gracias a esta tabla se puede apreciar que hay más controles que casos en los que la configuración de la calzada sea de sentido único o doble (hay aproximadamente el doble). Por el contrario, se producen más casos en ubicaciones en las que hay doble sentido con mediana divisoria, siendo esta diferencia de frecuencias mucho más notoria ya que es un porcentaje aproximadamente cuatro veces mayor.

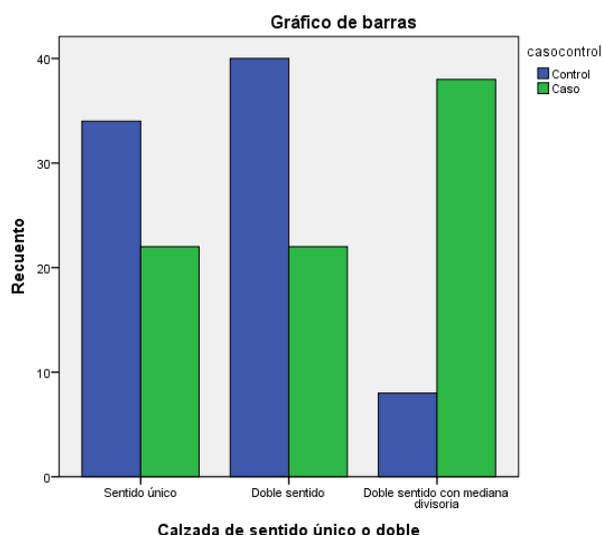


Ilustración 38. Gráfico de barras agrupadas de "Sentido" frente a "Caso o control".

### Odds-Ratio de la variable "Sentido"

Seguidamente, se obtuvo el OR ( $\exp(\beta)$ ) mediante la regresión logística binaria con la variable "Caso o Control" como dependiente y con "Sentido" como única variable independiente, cuya categoría de referencia se estableció en la primera de todas: sentido único.

De la misma forma que para el resto de las variables, se realizó dicho análisis para curvas e intersecciones de forma conjunta, al igual que para intersecciones y curvas de forma separada; todo ello con el objetivo de comparar los resultados conjuntos de todos los datos con los específicos para las intersecciones y las curvas.

### CURVAS E INTERSECCIONES

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
Sentido único			23,180	2	,000			
Doble sentido	-,163	,381	,182	1	,670	,850	,403	1,794
Doble sentido con mediana	1,99	,476	17,569	1	,000	7,341	2,89	18,64
Constante	-,435	,274	2,531	1	,112	,647		

Tabla 30. Odds-Ratio de la variable "Sentido" para curvas e intersecciones.

Las conclusiones extraídas de este análisis son las siguientes:

- La presencia de doble sentido en una ubicación, frente a sentido único, no es una asociación significativa, dado que el Intervalo de Confianza para el *Odds-Ratio* contiene el 1 entre sus extremos.
- Que la calzada sea de doble sentido con mediana divisoria, frente a que sea de sentido único, incrementa el riesgo de que ocurra una salida de la vía. Esto se debe a que su OR toma un valor de 7,341, con lo que se concluye que es siete veces más probable que ocurra el evento de estudio cuando la calle tiene doble sentido con mediana divisoria.

### INTERSECCIONES

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
Sentido único			23,602	2	,000			
Doble sentido	-,167	,449	,139	1	,709	,846	,351	2,040
Doble sentido con mediana	2,705	,614	19,393	1	,000	14,95	4,486	49,83
Constante	-,595	,311	3,647	1	,056	,552		

Tabla 31. *Odds-Ratio* de la variable "Sentido" para las intersecciones.

Al trabajar únicamente con las intersecciones los resultados son similares a cuando se analizan de forma conjunta con las curvas, tal y como se explica a continuación:

- De la misma forma, no existen diferencias significativas entre que la calzada sea de doble sentido y de sentido único.
- Al igual que en el análisis anterior, la presencia de doble sentido con mediana divisoria actúa como factor de riesgo. Sin embargo, el valor del OR en este caso es mucho mayor (aproximadamente el doble), por lo que si se analizan por separado las intersecciones el riesgo de que ocurra una salida de vía ante estas circunstancias es mayor que al analizar conjuntamente curvas e intersecciones.

CURVAS

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
Sentido único			,423	2	,810			
Doble sentido	-,405	,769	,278	1	,598	,667	,148	3,011
Doble sentido con mediana	,041	,904	,002	1	,964	1,042	,177	6,123
Constante	,182	,606	,091	1	,763	1,200		

Tabla 32. Odds-Ratio de la variable "Sentido" para las curvas.

Por último, al analizar únicamente el subconjunto de curvas se llega a la conclusión de que ninguna de las asociaciones es significativa, dado que todos los p-valores son elevados, es decir, bastante superiores a 0,05.

## 5.7. LONGITUD DEL TRAMO RECTO ANTERIOR

### Histograma de "Longitud del tramo recto anterior" para los casos

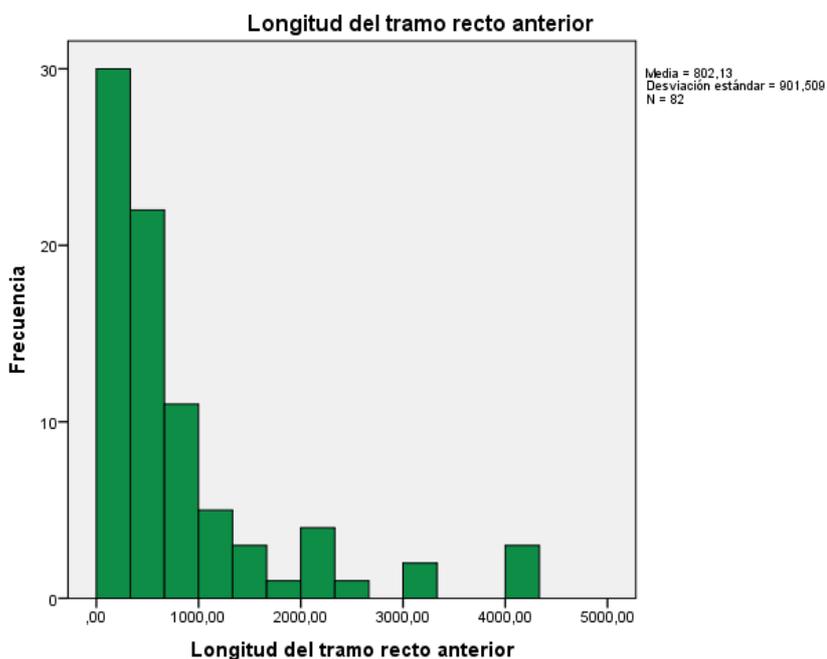


Ilustración 39. Histograma de "Longitud del tramo recto anterior" para los casos.

### Histograma de "Longitud del tramo recto anterior" para controles

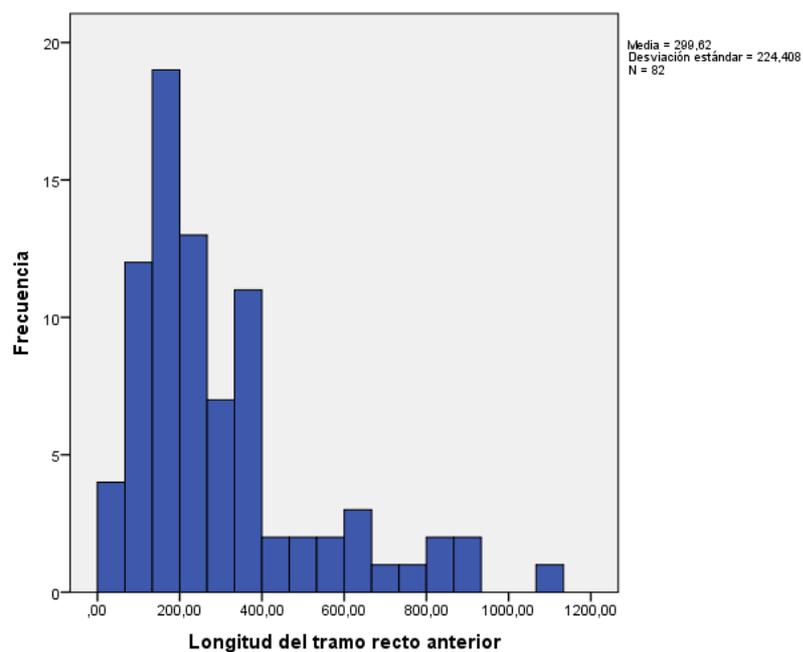


Ilustración 40. Histograma de "Longitud del tramo recto anterior" para los controles.

### Box-plot múltiple de "Longitud del tramo recto anterior"

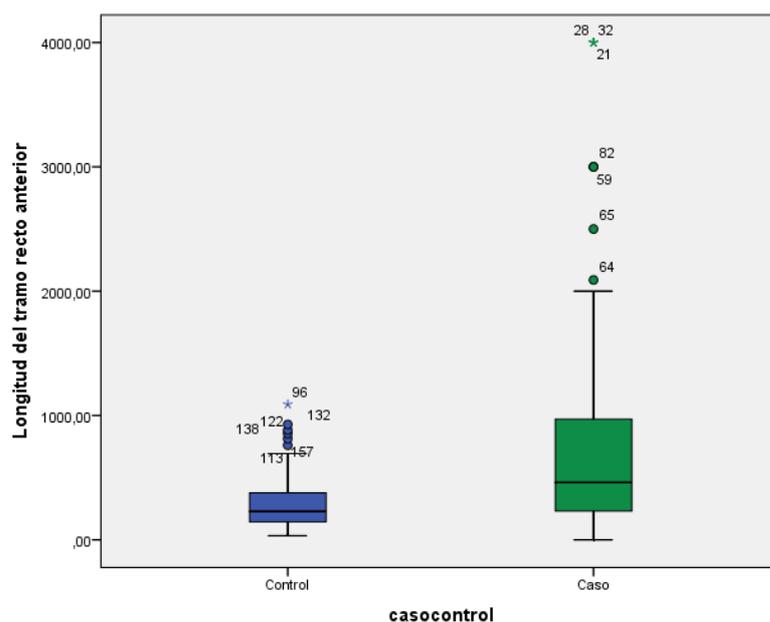


Ilustración 41. Box-plot múltiple de "Longitud del tramo recto anterior" frente a "Caso o control".

Tal y como se puede ver, ambos histogramas presentan una fuerte asimetría positiva. Además, los controles toman un rango de valores inferior a los casos, ya que alguno de estos datos toma una longitud del tramo recto anterior muy elevada (alrededor de 4000 metros).

Por otro lado, los valores de esta variable para los controles están mucho menos dispersos y son inferiores que para los casos, con una mediana de alrededor de 200 metros. Con todo ello se puede ver, a priori, una importante tendencia: cuanto mayor es la longitud del tramo recto anterior, mayor es la probabilidad de que se produzca una salida de calzada.

Teniendo en cuenta el objetivo por el que se decidió medir esta variable, conocer el tramo recto que podría inducir al conductor a acelerar el vehículo a velocidades más altas de las permitidas, se llega a la conclusión de que para la mayoría de los casos (salidas de calzada) existe una amplia posibilidad de que se produzca este fenómeno (con valores por encima de los 500 o 1000 metros el conductor ya tiene espacio suficiente para ello). En cambio, para los controles, debido a que esta variable toma valores más bajos, la posibilidad de que el conductor acelere en exceso el vehículo es menor.

### ***Tabla de contingencia y Odds-Ratio de “Longitud del tramo recto”***

Al tratarse de una variable continua, para la obtención de la tabla de contingencias y el cálculo de los OR hay que codificar la variable. Las categorías que se eligieron pretendían determinar qué distancias permitirían a un conductor acelerar el vehículo a velocidades por encima del límite, y son las siguientes:

- Menor o igual que 150 metros.
- Entre 150,01 y 300 metros.
- Entre 300,01 y 500 metros.
- Mayor o igual que 500,01 metros.

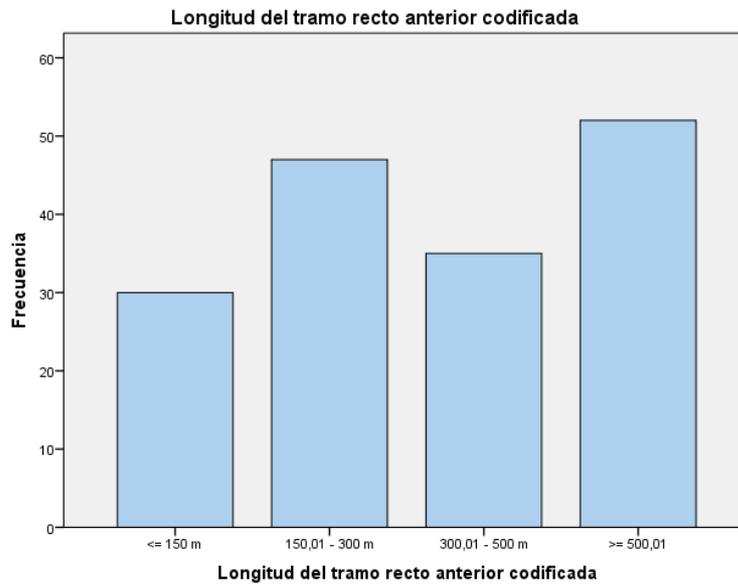


Ilustración 42. Gráfico de barras de la variable "Longitud del tramo recto anterior codificada".

Además, a través de la tabla cruzada siguiente, se pueden ver comparativamente las frecuencias de cada una de estas categorías en función de si se trata de caso o control.

		Caso o control		Total
		Control	Caso	
Longitud del tramo recto anterior codificada	<= 150 m	23 28,0%	7 8,5%	30 18,3%
	150 - 300 m	28 34,1%	19 23,2%	47 28,7%
	300,01 - 500 m	17 20,7%	18 22,0%	35 21,3%
	>= 500,01 m	14 17,1%	38 46,3%	52 31,7%
Total		82 100,0%	82 100,0%	164 100,0%

Tabla 33. Tabla cruzada de "Longitud del tramo recto anterior codificada" frente a "Caso o control".

Tras observar dichos porcentajes se comprueba que la mayor parte de los controles se producen en ubicaciones con longitudes pequeñas, principalmente en las dos primeras categorías (menor de 150 metros y entre 150,01 y 300 metros). Por el contrario, la mayoría de los casos se concentran en las longitudes mayores, de más de 500,01 metros. Es decir, se puede ver una tendencia clara:

en los controles, a medida que aumenta la longitud, menor es la frecuencia; mientras que en los casos ocurre al revés.

Las diferencias más notables se muestran, por lo tanto, en las categorías extremas, siendo en la más baja mucho más predominantes los controles, mientras que en la más alta los casos. Por consiguiente, se establece la primera (longitud del tramo recto anterior inferior a 150 metros) como categoría de referencia para el análisis posterior, ya que es la situación que se estima que menos favorece para que se produzca una salida de calzada.

Para estudiar de forma más concreta la asociación entre ambas variables se recurre a la regresión logística binaria con “Longitud del tramo recto anterior codificada” como variable independiente, tomando como categoría de referencia la primera de todas ellas (longitud menor de 150 metros).

### CURVAS E INTERSECCIONES

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
<= 150 m			19,526	3	,000			
150,01 - 300 m	,802	,524	2,341	1	,126	2,230	,798	6,228
300,01 - 500 m	1,247	,548	5,169	1	,023	3,479	1,188	10,19
>= 500,01 m	2,188	,533	16,854	1	,000	8,918	3,138	25,34
Constante	-1,190	,432	7,594	1	,006	,304		

Tabla 34. Odds-Ratio de la variable "Longitud del tramo recto anterior" para curvas e intersecciones.

De este primer modelo logístico se pueden extraer las siguientes conclusiones:

- Que la distancia que puede recorrer el vehículo esté comprendida entre 150,01 y 300 metros no actúa ni como factor de riesgo ni de protección debido a que el Intervalo de Confianza contiene el valor 1 entre sus extremos.
- Cuando la longitud del tramo recto anterior está comprendida entre 300,01 y 500 metros se incrementa el riesgo de que ocurra una salida de calzada, ya que el Odds-Ratio toma un valor significativamente mayor que 1: 3,479.
- Cuando aumenta la longitud por encima de 500,01 metros, el riesgo se incrementa aún más debido a que el OR toma un valor de 8,918, lo que

indica que es casi nueve veces más probable que suceda una salida de la vía bajo estas condiciones, frente a cuando la longitud es menor de 150 metros.

### INTERSECCIONES

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
<= 150 m			16,857	3	,001			
150,01 – 300 m	,959	,650	2,175	1	,140	2,609	,730	9,328
300,01 – 500 m	1,299	,698	3,468	1	,063	3,667	,934	14,39
>= 500,01 m	2,398	,653	13,493	1	,000	11,00	3,060	39,54
Constante	-1,386	,559	6,150	1	,013	,250		

Tabla 35. Odds-Ratio de la variable "Longitud del tramo recto anterior" para las intersecciones.

Al trabajar con las intersecciones por separado se pueden apreciar ciertas diferencias:

- En este modelo la tercera categoría ya no supone un factor de riesgo a un nivel de significación del 5%, puesto que el límite inferior del Intervalo de Confianza es menor que 1 y el límite superior es mayor. Sin embargo, a un nivel del 10% sí lo sería, tal y como muestra el p-valor asociado.
- Las longitudes mayores, por encima de 500,01 metros, continúan siendo un factor de riesgo, pero en este caso la estimación del Odds-Ratio es un valor superior: 11. Esto se traduce en que las intersecciones en las que las distancias máximas a recorrer por los vehículos son mayores, el riesgo es más elevado que para curvas e intersecciones conjuntamente.

CURVAS

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
<= 150 m			3,148	3	,369			
150,01 - 300 m	,624	,962	,421	1	,517	1,867	,283	12,31
300,01 - 500 m	1,184	,905	1,711	1	,191	3,267	,554	19,25
>= 500,01 m	1,764	1,08	2,644	1	,104	5,833	,696	48,87
Constante	-,847	,690	1,508	1	,220	,429		

Tabla 36. Odds-Ratio de la variable "Longitud del tramo recto anterior" para las curvas.

En este último análisis en el que sólo se trabaja con el subconjunto de curvas, ninguna de las asociaciones es significativa, dado que todos los Intervalos de Confianza para el Odds-Ratio contienen al 1.

## 5.8. DISTANCIA AL SEMÁFORO ANTERIOR

### Histograma de "Distancia al semáforo anterior" para los casos

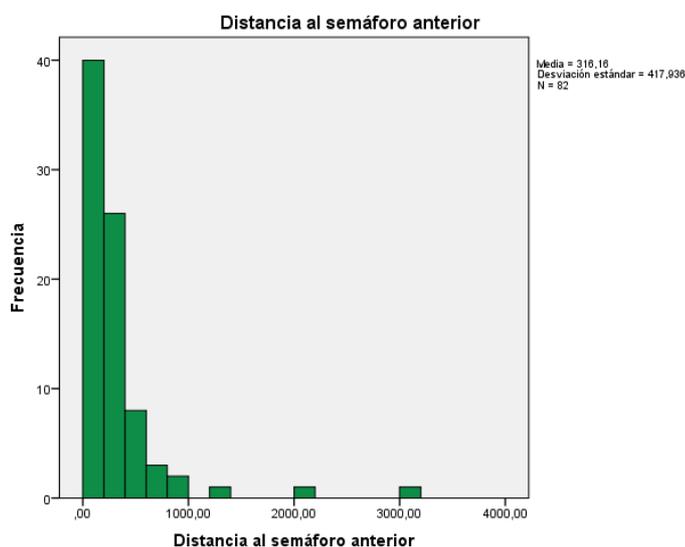


Ilustración 43. Histograma de "Distancia al semáforo anterior" para los casos.

**Histograma de "Distancia al semáforo anterior" para los controles**

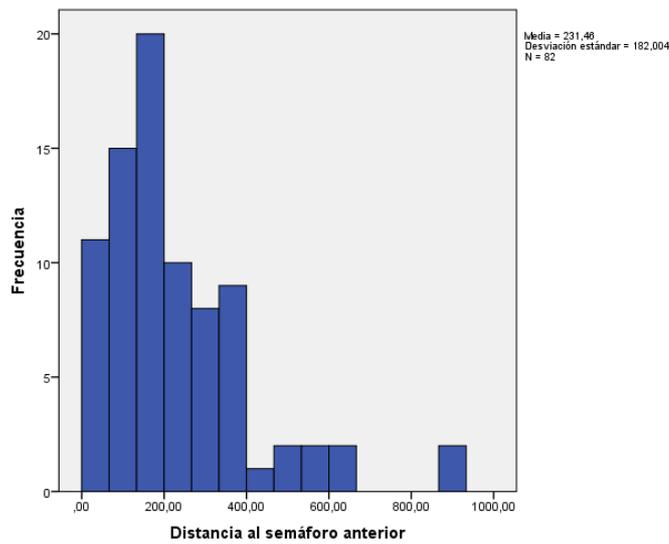


Ilustración 44. Histograma de "Distancia al semáforo anterior" para los controles.

**Box-plot múltiple de la variable "Distancia al semáforo anterior"**

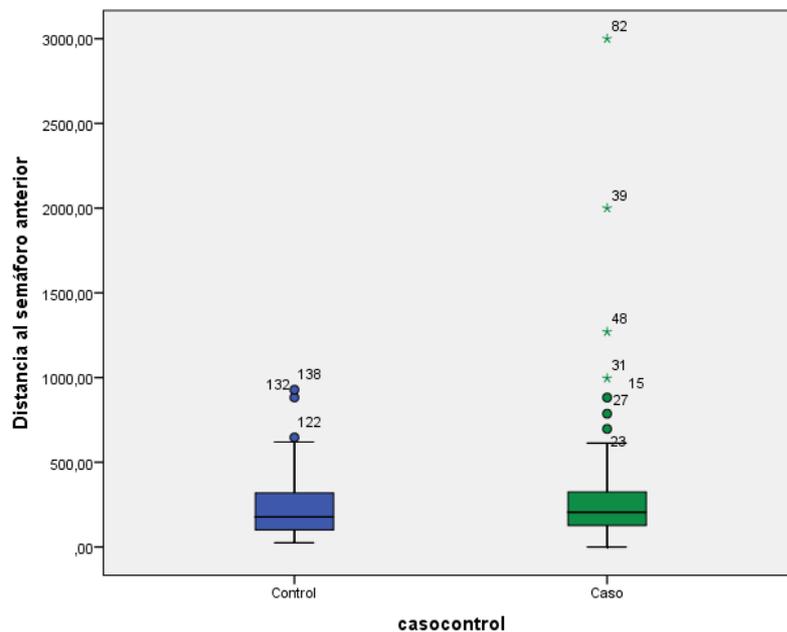


Ilustración 45. Box-plot múltiple de "Distancia al semáforo anterior" frente a "Caso o control".

Observando los histogramas y el diagrama de cajas múltiple anteriores, se puede ver que el rango de valores de la distancia al semáforo anterior para los casos llega hasta aproximadamente los 3000 metros, mientras que para los controles hasta alrededor de 1000 metros.

Es decir, los datos toman una forma más asimétrica para los casos, pero a pesar de ello, tal y como se ve en el *box-plot* múltiple, la posición y dispersión son parecidas para el grueso de la distribución.

### ***Tabla de contingencia y Odds-Ratio de “Distancia al semáforo”***

Como también se trata de una variable continua hay que codificarla, para así poder realizar la tabla de contingencia y calcular los OR de las diferentes categorías respecto a una de referencia. Las que se seleccionaron son las mismas que para la variable “Longitud del tramo recto anterior”, tal y como se puede ver en el siguiente gráfico de barras.

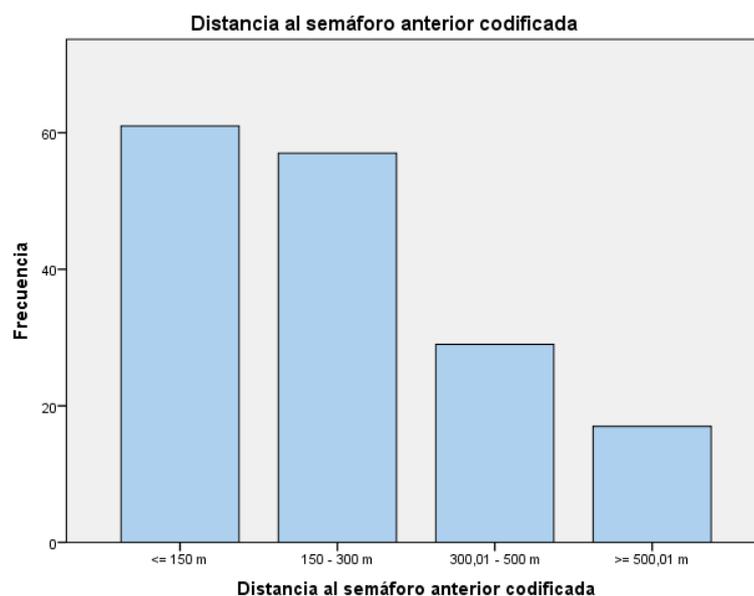


Ilustración 46. Gráfico de barras de la variable "Distancia al semáforo anterior codificada".

En este caso, las frecuencias decrecen de forma más o menos uniforme según aumenta la distancia al semáforo anterior, al contrario que para la longitud del tramo recto. Así mismo, para observar el comportamiento de esta variable frente a “Caso o control” se obtuvo la siguiente tabla cruzada:

		Caso o control		Total
		Control	Caso	
Distancia al semáforo anterior codificada	<= 150 m	34 41,5%	27 32,9%	61 37,2%
	150 - 300 m	25 30,5%	32 39,0%	57 34,8%
	300,01 - 500 m	15 18,3%	14 17,1%	29 17,7%
	>= 500,01 m	8 9,8%	9 11,0%	17 10,4%
Total		82 100,0%	82 100,0%	164 100,0%

Tabla 37. Tabla cruzada de "Distancia al semáforo anterior codificada" frente a "Caso o control".

Tras observar las frecuencias de cada categoría, se puede ver que apenas hay diferencia entre los porcentajes para casos o controles, con lo que se podría deducir, previo a un estudio más intensivo, que no hay asociación entre ambas variables. Para comprobarlo más detalladamente se calcula el OR de cada clase respecto a una establecida como referencia: distancias menores de 150 metros.

A continuación, se recogen los resultados de la regresión logística binaria con "Distancia al semáforo anterior codificada" como única variable independiente.

### CURVAS E INTERSECCIONES

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
<= 150 m			1,748	3	,626			
150,01 - 300 m	,477	,371	1,655	1	,198	1,612	,779	3,336
300,01 - 500 m	,162	,452	,128	1	,721	1,175	,484	2,852
>= 500,01 m	,348	,550	,401	1	,527	1,417	,482	4,164
Constante	-,231	,258	,800	1	,371	,794		

Tabla 38. Odds-Ratio de la variable "Distancia al semáforo anterior codificada" para curvas e intersecciones.

INTERSECCIONES

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
<= 150 m			1,209	3	,751			
150,01 - 300 m	,322	,430	,560	1	,454	1,380	,594	3,207
300,01 - 500 m	-,123	,520	,056	1	,814	,885	,319	2,451
>= 500,01 m	,391	,589	,440	1	,507	1,479	,466	4,694
Constante	-,140	,306	,209	1	,648	,870		

Tabla 39. Odds-Ratio de la variable "Distancia al semáforo anterior codificada" para las intersecciones.

CURVAS

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
<= 150 m			2,172	3	,537			
150,01 - 300 m	,922	,748	1,521	1	,217	2,514	,581	10,88
300,01 - 500 m	1,145	,992	1,333	1	,248	3,143	,450	21,95
>= 500,01 m	-20,75	40192	,000	1	1,000	,000	,000	.
Constante	-,452	,483	,874	1	,350	,636		

Tabla 40. Odds-Ratio de la variable "Distancia al semáforo anterior codificada" para las curvas.

En los tres estudios se llega a la misma conclusión: que la distancia al semáforo anterior no es significativa. Es decir, ninguna asociación resulta notable respecto a la categoría de referencia, debido a que todos los IC al 95% contienen al 1. De esta forma se confirman las hipótesis anteriores que se dedujeron de la tabla de contingencia y del *box-plot* múltiple: no se trata de una variable determinante a la hora de que ocurra una salida de calzada.

## 5.9. PENDIENTE

Al tratarse de una variable categórica se realizó la tabla de frecuencias tanto para los casos como para los controles, con el propósito de ver la posible influencia de esta variable sobre el desenlace a estudiar: salida de calzada o no.

**Tabla de frecuencias de "Pendiente" para los casos**

	Frecuencia	Porcentaje
Sí	10	12,2
No	72	87,8
Total	82	100,0

Tabla 41. Tabla de frecuencias de "Pendiente" para los casos.

**Tabla de frecuencias de "Pendiente" para los controles**

	Frecuencia	Porcentaje
Sí	5	6,1
No	77	93,9
Total	82	100,0

Tabla 42. Tabla de frecuencias de "Pendiente" para los controles.

**Tabla de contingencia para la variable "Pendiente"**

		Caso o control		Total
		Control	Caso	
Pendiente	Sí	5 6,1%	10 12,2%	15 9,1%
	No	77 93,9%	72 87,8%	149 90,9%
Total		82 100,0%	82 100,0%	164 100,0%

Tabla 43. Tabla cruzada de "Pendiente" frente a "Caso o control".

Observando ambas tablas de frecuencias y la tabla cruzada anteriores, se comprueba que los porcentajes de la presencia o ausencia de pendiente son prácticamente equivalentes, ya que existe una gran preponderancia de lugares donde no hay pendiente, independientemente de si se trata de caso o control. Aun así, el porcentaje de lugares en los que sí hay pendiente en los casos es aproximadamente el doble que en los controles (12,6% frente a 6,1%).

En conclusión, se puede determinar que la ausencia de pendiente es la categoría que parece menos determinante a la hora de que ocurra una salida de calzada.

Por lo tanto, es la que se tomará como referencia para realizar los estudios posteriores. Estas observaciones se pueden comprobar de una forma más visual en el siguiente gráfico de barras agrupadas.

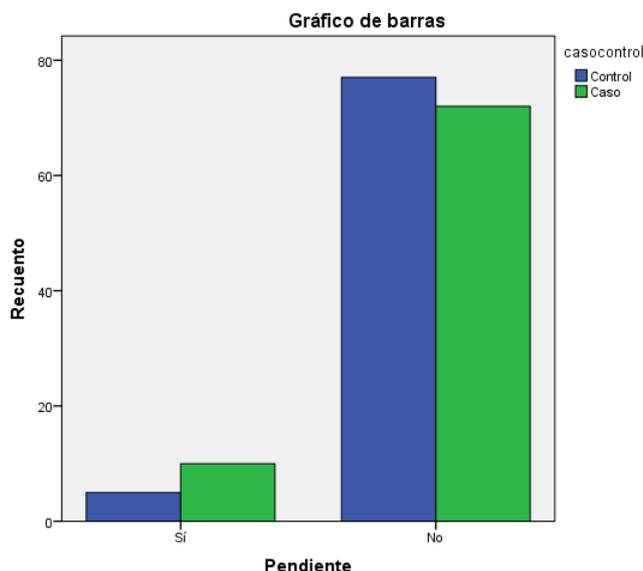


Ilustración 47. Gráfico de barras agrupadas de "Pendiente" frente a "Caso o control".

### Odds-Ratio de la variable "Pendiente"

Para estudiar mejor la asociación entre la variable "Pendiente" y "Caso o control" se realiza la regresión logística correspondiente, con el objetivo de obtener los valores del Odds-Ratio.

#### CURVAS E INTERSECCIONES

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
Pendiente	,760	,572	1,768	1	,184	2,139	,697	6,559
Constante	-,067	,164	,168	1	,682	,935		

Tabla 44. Odds-Ratio de la variable "Pendiente" para curvas e intersecciones.

## INTERSECCIONES

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
Pendiente	,612	,654	,875	1	,350	1,844	,512	6,643
Constante	-,052	,187	,078	1	,780	,949		

Tabla 45. Odds-Ratio de la variable "Pendiente" para las intersecciones.

## CURVAS

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
Pendiente	1,216	1,205	1,019	1	,313	3,375	,318	35,78
Constante	-,118	,344	,118	1	,732	,889		

Tabla 46. Odds-Ratio de la variable "Pendiente" para las curvas.

De cada uno de los tres análisis se obtiene la misma conclusión: ninguna de las asociaciones es significativa y la pendiente no es determinante para que ocurra una salida de calzada.

## 5.10. PRIORIDAD

*Tabla de frecuencias de "Prioridad" para los casos*

	Frecuencia	Porcentaje
Semáforo	35	42,7
Señal de STOP	3	3,7
Señal de CEDA EL PASO	9	11,0
Ninguna (sólo norma)	35	42,7
Total	82	100,0

Tabla 47. Tabla de frecuencias de "Prioridad" para los casos.

**Tabla de frecuencias de "Prioridad" para los controles**

	Frecuencia	Porcentaje
Semáforo	12	14,6
Señal de STOP	11	13,4
Señal de CEDA EL PASO	17	20,7
Ninguna (sólo norma)	42	51,2
Total	82	100,0

Tabla 48. Tabla de frecuencias de "Prioridad" para los controles.

**Tabla de contingencia para la variable "Prioridad"**

		Caso o control		Total
		Control	Caso	
Prioridad regulada por:	Semáforo	12 14,6%	35 42,7%	47 28,7%
	Señal de STOP	11 13,4%	3 3,7%	14 8,5%
	Señal de CEDA EL PASO	17 20,7%	9 11,0%	26 15,9%
	Ninguna (sólo norma)	42 51,2%	35 42,7%	77 47,0%
Total		82 100,0%	82 100,0%	164 100%

Tabla 49. Tabla cruzada de "Prioridad" frente a "Caso o control".

Tras observar las tablas de frecuencias y la tabla cruzada, se puede determinar que la mayor parte de los casos y controles almacenados en la base de datos son lugares en los que no existe ninguna señalización para la prioridad, siendo algo superior este porcentaje para los controles. En cuanto a las diferencias existentes entre los distintos tipos de regulación de la prioridad, destaca que en la mayor parte de los casos (aproximadamente la mitad) la regulación es a través de semáforo, mientras que en los controles predominan más las regulaciones del tipo: señal de STOP o de CEDA EL PASO.

Con todo ello se puede concluir que la forma de regulación de la prioridad que parece que es menos determinante para que se produzca una salida de la vía es la señal de STOP, por lo que esa es la categoría que se utilizará como referencia para estudios posteriores. Todas estas conclusiones se pueden ver en el gráfico de barras posterior.

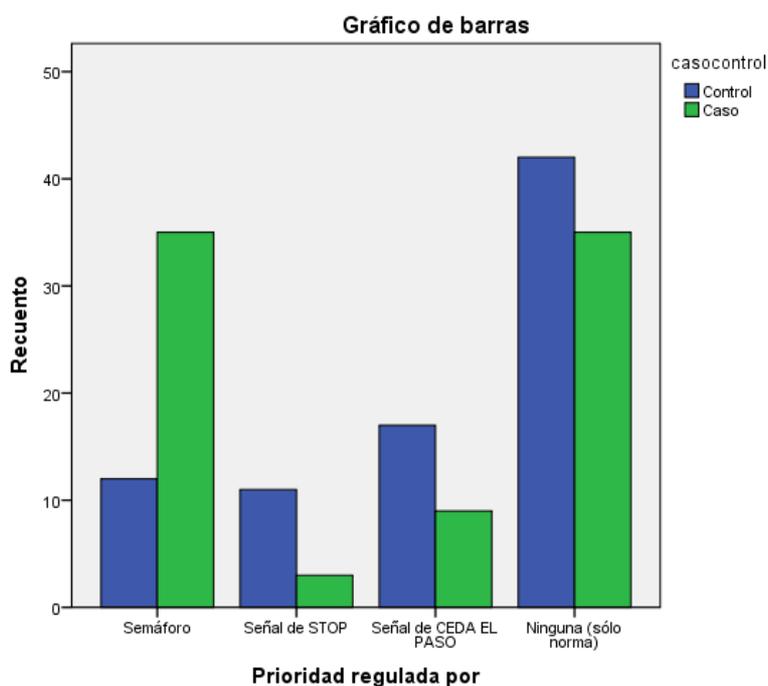


Ilustración 48. Gráfico de barras agrupadas de "Prioridad" frente a "Caso o control".

### Odds-Ratio de la variable "Prioridad"

Una vez observada la asociación entre las variables "Prioridad" y "Caso o control" en la tabla de contingencias, se realiza una regresión logística binaria con la primera de las variables como independiente (cuya categoría de referencia será la señal de STOP) y como dependiente la variable "Caso o control".

De la misma forma que para el resto de las variables, se realizó dicho análisis tanto para el conjunto completo de datos como para los subconjuntos de intersecciones y curvas por separado.

#### CURVAS E INTERSECCIONES

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
Señal de STOP			17,21	3	,001			
Semáforo	2,370	,732	10,47	1	,001	10,69	2,546	44,91
Señal de CEDA EL PASO	,663	,771	,740	1	,390	1,941	,428	8,794
Ninguna (sólo norma)	1,117	,690	2,618	1	,106	3,056	,790	11,82
Constante	-1,299	,651	3,979	1	,046	,273		

Tabla 50. Odds-Ratio de la variable "Prioridad" para curvas e intersecciones.

Tras observar los resultados de la regresión logística recogidos en la tabla anterior, se pueden determinar las siguientes conclusiones:

- Respecto a aquellas curvas o intersecciones en las que la regulación de la prioridad se realiza mediante semáforo, se ha podido comprobar que el riesgo de ocurrir una salida de calzada es mayor que si hubiera una señal de STOP. Esto no significa que el problema sea el semáforo, sino que podría deberse a que los usuarios de la vía no lo respetan adecuadamente. Otra posibilidad podría ser que el semáforo se hubiera colocado en esos lugares precisamente por ser considerados como puntos de riesgo. En ese caso, este hecho implicaría que la regulación con semáforo no es suficiente, por lo que su presencia no sería incompatible con la instalación de las barreras de contención adecuadas. Que dicha asociación sea significativa es debido a que el p-valor es inferior a 0,05 (existe significación a un nivel del 5%). El valor estimado del OR es 10,694, lo que indica que el riesgo de que ocurra una salida de calzada en lugares regulados por semáforo se multiplica aproximadamente por 10, frente a si lo estuvieran por una señal de STOP.
- La señalización de CEDA EL PASO, en cambio, no supone diferencias significativas frente a la señal de STOP de cara a que haya un mayor riesgo de que se produzca una salida de vía.
- Lo mismo ocurre cuando se contrastan la última categoría (ninguna regulación de la prioridad) con la señal de STOP. La asociación en este caso tampoco es significativa debido a que el IC al 95% contiene el valor 1 dentro de su rango.

### INTERSECCIONES

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
Señal de STOP			12,98	3	,005			
Semáforo	2,182	,736	8,781	1	,003	8,861	2,093	37,51
Señal de CEDA EL PASO	,663	,771	,740	1	,390	1,941	,428	8,794
Ninguna (sólo norma)	1,255	,716	3,068	1	,080	3,507	,861	14,28
Constante	-1,299	,651	3,979	1	,046	,273		

Tabla 51. Odds-Ratio de la variable "Prioridad" para las intersecciones.

Al tratar por separado el subconjunto de intersecciones, los resultados obtenidos de los valores del OR son similares que cuando se tratan en conjunto con las curvas.

- La presencia de semáforo frente a la señal de STOP supone un riesgo estadísticamente significativo de que ocurra una salida de la vía (aunque, atendiendo a lo explicado anteriormente, este hecho no significa que el problema sea el semáforo). El incremento del riesgo se debe a que el *Odds-Ratio* toma un valor de 8,861, algo menor que cuando se analizaron las intersecciones y curvas en conjunto.
- La regulación de la prioridad a través de la señal de CEDA EL PASO continúa sin ser factor de riesgo ni de protección.
- Lo mismo ocurre cuando no hay ningún tipo de regulación de la prioridad frente a la señalización de STOP.

CURVAS

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
Prioridad	21,58	16408,7	,000	1	,999	236107	,000	.
Constante	-,379	,360	1,112	1	,292	,684		

Tabla 52. *Odds-Ratio* de la variable "Prioridad" para las curvas.

Si se trabaja únicamente con las curvas los únicos valores que toma la variable "Prioridad" son la primera y la última de las categorías: "Semáforo" y "Ninguna (sólo norma)", y debido a que existe una mayor preponderancia de la última de estas categorías para los controles frente a los casos, es la que se ha tomado como referencia para dicha variable en el modelo logístico anterior.

Tal y como se puede ver en la tabla anterior, los estimadores para la variable "Prioridad" son muy inestables debido a que en la tabla cruzada para las curvas hay una celda sin ningún dato:

		Caso o control		Total
		Control	Caso	
Prioridad regulada por:	Semáforo	0 0,0%	6 31,6%	6 15,8%
	Ninguna (sólo norma)	19 100,0%	13 68,4%	32 84,2%
Total		19 100,0%	19 100,0%	38 100,0%

Tabla 53. Tabla cruzada de "Prioridad" frente a "Caso o control" para las curvas.

Sin embargo, aunque el ajuste del modelo sea inestable por cuestiones técnicas, parece un aspecto significativo y que debería tenerse en cuenta ya que, eligiendo al azar el grupo de curvas definidas como controles, no se encontró ninguna con semáforo. Por consiguiente, aquellas curvas en las que la regulación de la prioridad se realice mediante semáforo se considerarán como un factor de riesgo, aunque no sea posible evaluarlo bien cuantitativamente al tener una celda vacía en la tabla cruzada.



## CAPÍTULO 6. MODELOS DE REGRESIÓN LOGÍSTICA

---

En el capítulo anterior se analizó individualmente el efecto de cada una de las variables explicativas sobre la ocurrencia de salidas de calzada, con objeto de tener una idea de las que podrían ser más relevantes. Sin embargo, siempre es conveniente considerar modelos multivariantes para ver el efecto conjunto de todas las variables sobre la respuesta. A continuación, consideraremos los modelos descritos en la sección 2 del capítulo 1, realizados para los tres conjuntos de datos ya especificados en el capítulo anterior: curvas e intersecciones simultáneamente, e intersecciones y curvas por separado. En cada uno de estos conjuntos de datos se considerará primeramente el modelo con todas las variables, para posteriormente obtener los modelos con los métodos habituales de selección de variables.

Con el propósito de no redactar un capítulo excesivamente extenso y repetitivo, únicamente se presentan los resúmenes de los modelos en las tablas que vienen a continuación. En ellas, además, se registra el porcentaje de clasificación correcto aparente. A través de esta tabla, y comparando los porcentajes de los distintos modelos, se seleccionarán los más adecuados, es decir, aquellos que mejor explican el desenlace estudiado y que incluyen variables cuya presencia es razonablemente significativa.

En el capítulo siguiente se validarán y compararán dichos modelos, para lo que se obtendrá tanto el porcentaje de clasificación correcto con el método de validación cruzada *Leave-One-Out*, como los contrastes ROC para cada conjunto de datos. El objetivo que se pretende no es otro que seleccionar el modelo más apropiado en cada caso.

Hay que destacar que, aunque en el capítulo anterior se procedió a categorizar las variables continuas para el análisis univariante, a la hora de ajustar los modelos con todas las variables se incluyeron como cuantitativas, tal y como se mencionó con anterioridad. En cambio, para las variables cualitativas, el grupo de referencia es el que se indicó en el capítulo anterior y que se recoge de forma resumida en la siguiente tabla.

VARIABLE	CATEGORÍA DE REFERENCIA
Tipo de intersección	En T o Y
Sentido	Sentido único
Pendiente	No
Prioridad	Señal de STOP

Tabla 54. Resumen de las categorías de referencia para las variables cualitativas del modelo.

Seguidamente, se recogen las tablas comparativas de los resultados de los distintos métodos de introducción o selección de variables, además del porcentaje de clasificación correcto aparente, las variables dentro de la ecuación y las que son significativas (para las variables cualitativas puede ocurrir que no todas las categorías sean significativas, por lo que entre paréntesis se especifican las que sí lo son). Esta comparación se ha realizado tanto para las curvas e intersecciones simultáneamente como para los subconjuntos de intersecciones y curvas por separado, tal y como se mencionó anteriormente.

### CURVAS E INTERSECCIONES

MÉTODO	% DE CLASIFICACIÓN CORRECTO APARENTE	VARIABLES EN LA ECUACIÓN	VARIABLES SIGNIFICATIVAS
<b>Introducción de todas las variables</b>	77,4 %	Todas	Longitud del tramo recto anterior Radio de curvatura Tipo de intersección (curva)
<b>Hacia delante: condicional, LR y Wald</b>	71,3 %	Longitud del tramo recto anterior Sentido	Longitud del tramo recto anterior Sentido (sentido único y doble sentido con mediana divisoria)
<b>Hacia atrás: condicional, LR y Wald</b>	71,3 %	Longitud del tramo recto anterior Sentido	Longitud del tramo recto anterior Sentido (sentido único y doble sentido con mediana divisoria)

73,8 %	Número de carriles	Longitud del tramo recto anterior
	Sentido	Sentido (doble sentido y doble sentido con mediana divisoria)
	Longitud del tramo recto anterior	Radio de curvatura
	Radio de curvatura	Tipo de intersección (curva)

Tabla 55. Tabla comparativa de posibles modelos logísticos para curvas e intersecciones.

### INTERSECCIONES

MÉTODO	% DE CLASIFICACIÓN CORRECTO APARENTE	VARIABLES EN LA ECUACIÓN	VARIABLES SIGNIFICATIVAS
<b>Introducción de todas las variables</b>	77,8 %	Todas menos "Radio de curvatura"	Longitud del tramo recto anterior
<b>Hacia delante: condicional, LR y Wald</b>	73,8 %	Longitud del tramo recto anterior Sentido	Longitud del tramo recto anterior Sentido (sentido único y doble sentido con mediana divisoria)
<b>Hacia atrás: condicional, LR y Wald</b>	73,8 %	Ancho de calzada	Ancho de calzada
		Número de carriles	Número de carriles
		Longitud del tramo recto anterior	Longitud del tramo recto anterior
	77,8 %	Ancho de calzada	Ancho de calzada
		Número de carriles	Número de carriles
		Longitud del tramo recto anterior	Longitud del tramo recto anterior
		Pendiente	Tipo de intersección (giratoria)
		Tipo de intersección	

Tabla 56. Tabla comparativa de posibles modelos logísticos para las intersecciones.

## CURVAS

MÉTODO	% DE CLASIFICACIÓN CORRECTO APARENTE	VARIABLES EN LA ECUACIÓN	VARIABLES SIGNIFICATIVAS
<b>Introducción de todas las variables</b>	81,6 %	Todas menos "Tipo de intersección"	Ancho de carril
<b>Hacia delante: condicional y LR</b>	76,3 %	Ancho de carril Prioridad	Ancho de carril
<b>Hacia delante: Wald</b>	65,8 %	Prioridad	
<b>Hacia atrás: condicional y LR</b>	76,3 %	Ancho de carril Prioridad	Ancho de carril
	86,8 %	Ancho de carril Sentido	Ancho de carril
<b>Hacia atrás: Wald</b>	61,6 %	Prioridad	
		Ancho de carril Longitud del tramo recto anterior	Ancho de carril
		Prioridad	

Tabla 57. Tabla comparativa de posibles modelos logísticos para las curvas.

Tras ver los porcentajes de clasificación correctos en cada caso y el número de variables incluidas, se escogen los modelos completos con todas las variables, así como los modelos hacia atrás (*backward*) con el método condicional. En los modelos *backward*, además del último paso con el menor número de variables, se ha recogido un paso intermedio en el que hay un porcentaje de clasificación notablemente mayor, que podría ser un modelo interesante a considerar. Por consiguiente, en la siguiente tabla resumen se recogen aquellos modelos que van a ser estudiados en el capítulo siguiente:

CONJUNTO DE DATOS	MÉTODO	VARIABLES EN LA ECUACIÓN
<b>Curvas e intersecciones</b>	Modelo completo	Todas las variables Número de carriles Sentido
	Modelo intermedio hacia atrás: condicional	Longitud del tramo recto anterior Radio de curvatura Tipo de intersección
	Modelo hacia atrás: condicional	Longitud del tramo recto anterior Sentido
	Modelo completo	Todas las variables menos "Radio de curvatura" Ancho de calzada Número de carriles
<b>Intersecciones</b>	Modelo intermedio hacia atrás: condicional	Longitud del tramo recto anterior Pendiente Tipo de intersección
	Modelo hacia atrás: condicional	Ancho de calzada Número de carriles Longitud del tramo recto anterior
	Modelo completo	Todas las variables menos "Tipo de intersección" Ancho de carril
	Modelo intermedio hacia atrás: condicional	Sentido Prioridad
<b>Curvas</b>	Modelo hacia atrás: condicional	Ancho de carril Prioridad

Tabla 58. Tabla resumen con los modelos a comparar con el análisis ROC.



# CAPÍTULO 7. SELECCIÓN Y VALIDACIÓN DEL MODELO

---

Tras observar las tablas del capítulo anterior se determinan los tres modelos más interesantes teniendo en cuenta su porcentaje de clasificación correcto aparente y las variables incorporadas en su ajuste:

- Modelo con todas las variables, que denominaremos como “Modelo completo” o “modelo”.
- Modelo hacia atrás condicional, que denominaremos como “Modelo hacia atrás” o “*backward*”.
- Modelo hacia atrás condicional intermedio; es decir, aquel de un paso anterior al último y que presenta un elevado porcentaje de clasificación. Se denominará como “Modelo intermedio” o “intermedio”.

A continuación, se van a analizar los tres modelos correspondientes, tanto para todo el conjunto de datos (curvas e intersecciones), como para cada uno de los subconjuntos por separado (intersecciones por un lado y curvas por otro). Este procedimiento se realiza con el objetivo de seleccionar qué modelo es el que presenta un mejor comportamiento.

Para poder lograr este objetivo, en primer lugar, se recurre al método de validación cruzada *Leave-One-Out* explicado con anterioridad y que consiste en realizar tantas iteraciones como datos disponibles, para calcular la media aritmética del porcentaje de clasificación correcto en cada paso. En cada iteración se deja un individuo de la muestra fuera del subconjunto de entrenamiento a través del que se ajusta el modelo. Posteriormente, se obtiene la predicción de este individuo a través del modelo ajustado, con lo que se puede conocer si la clasificación ha sido correcta.

En segundo lugar, para seleccionar el modelo más satisfactorio dentro de todos los posibles, se recurre a un análisis ROC, a través del cual se obtiene de una manera gráfica la curva ROC así como la comparativa de las curvas para cada modelo. De esta forma, se puede analizar gráficamente cuál parece que es el modelo mejor (aquel que se acerque más a la esquina superior izquierda o el de mayor área bajo la curva ROC). Así mismo, se realiza un contraste ROC para conocer si hay diferencias significativas entre cada uno de los modelos en cuanto al área bajo la curva RO

## 7.1. VALIDACIÓN *LEAVE-ONE-OUT* Y ANÁLISIS ROC EN CURVAS E INTERSECCIONES

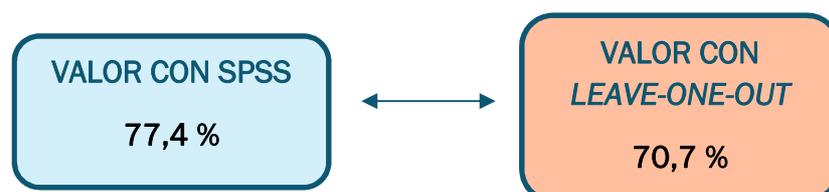
En primer lugar, se obtuvieron tanto la tabla de clasificación con el método de validación *Leave-One-Out* como las curvas ROC de cada uno de los modelos seleccionados para todo el conjunto de datos. A continuación, se muestra el código empleado en el software SAS para obtener todos los resultados necesarios:

```
libname acc 'd:/SAS';

proc logistic data=acc.cyctodo;
class tipointe sentido pend priori;
model casocontrol=tipointe radcurv anchocalz anchocarr numcarril sentido
longrec distsem pend priori/ctable;
*model casocontrol=numcarril sentido longrec radcurv tipointe/ctable;
*model casocontrol=longrec sentido/ctable;
roc 'intermedio' numcarril sentido longrec radcurv tipointe;
roc 'backward' longrec sentido;
roccompare reference('intermedio') / estimate e;
run;
```

### ***Modelo completo para curvas e intersecciones***

Observando la tabla de clasificación obtenida con *Leave-One-Out*, y teniendo en cuenta que el punto de corte que se había elegido para obtener los modelos de regresión logística con SPSS era 0,5, se pueden comparar los porcentajes de clasificación correctos tanto para el modelo obtenido con SPSS (porcentaje de clasificación correcto aparente) como para el valor obtenido en ese punto de corte con el método *Leave-One-Out*. Así mismo, se puede concluir que dicho valor coincide con el valor máximo dentro de la tabla; por lo tanto, se trata del punto de corte con un menor error de clasificación.



Seguidamente, se obtiene la curva ROC que representa cada uno de los valores de Sensibilidad y  $1 - \text{Especificidad}$  de la tabla anteriormente mencionada; es decir, los valores de estos parámetros para cada punto de corte. Además, se obtiene el área bajo la curva, que da una idea de la exactitud global del modelo.

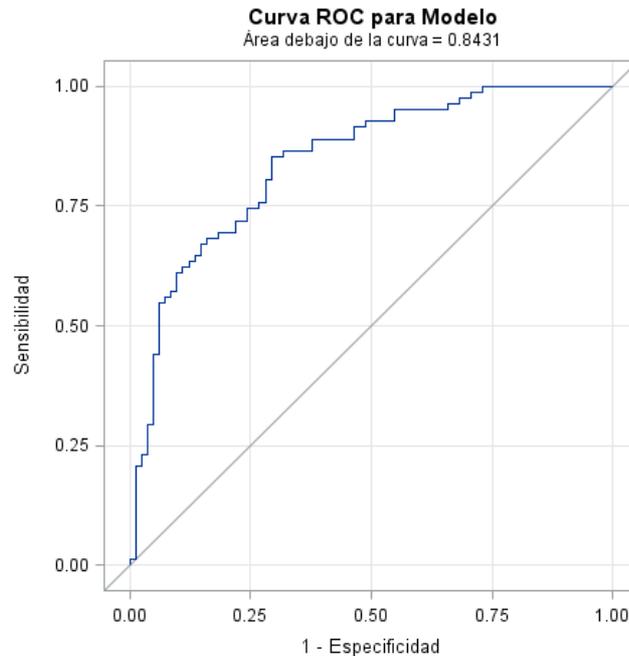


Ilustración 49. Curva ROC del modelo completo para curvas e intersecciones.

### ***Modelo intermedio para curvas e intersecciones***

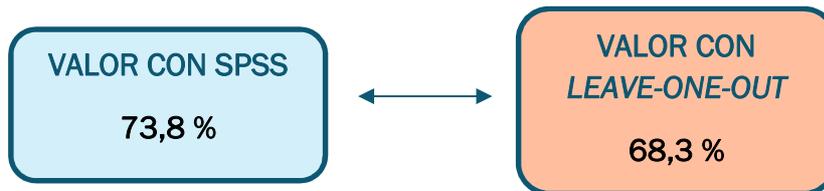
En primer lugar, se obtuvo la tabla de clasificación para el modelo intermedio, en la que se recogen tanto el porcentaje de clasificación correcto con el método *Leave-One-Out* como los porcentajes de la Sensibilidad y Especificidad, que indican, por un lado, la capacidad del modelo para clasificar correctamente aquellos individuos que tienen la “enfermedad” y, por otro, la capacidad del modelo para identificar como “sanos” a aquellos que no la tienen.

El modelo intermedio obtenido con el método de selección de variables hacia atrás condicional contenía las siguientes variables:

- Número de carriles
- Sentido
- Longitud del tramo recto anterior
- Radio de curvatura
- Tipo de intersección

Por lo tanto, éstas son las variables que se incluyen en el modelo para la obtención de la tabla de clasificación y la curva ROC.

De esta forma se comparan los porcentajes de clasificación correctos: tanto el inicial obtenido de la regresión logística con SPSS como el resultante tras la validación del modelo con el método *Leave-One-Out*, donde se obtienen los siguientes resultados:



Tal y como se puede ver, el porcentaje de clasificación correcto real es inferior al 70%, con lo que se deduce que el modelo completo no aporta buenos resultados.

Partiendo de la tabla de clasificación se obtiene la curva ROC para el modelo intermedio. Como puede apreciarse el área bajo la curva es algo menor que en el modelo completo, ya que está más alejada de la curva óptima (la más cercana a la esquina superior izquierda).

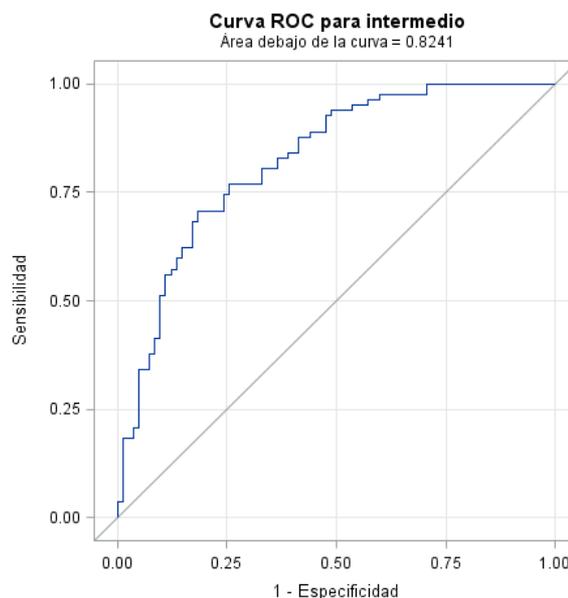


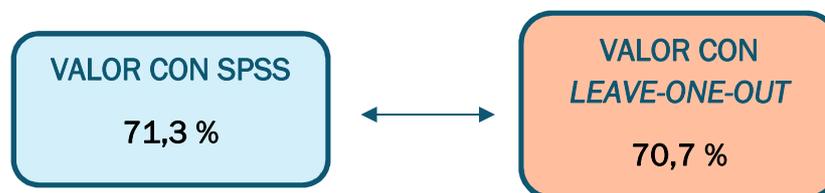
Ilustración 50. Curva ROC del modelo intermedio para curvas e intersecciones

### Modelo hacia atrás para curvas e intersecciones

A continuación, se obtienen la tabla de clasificación y la curva ROC para el último de los tres modelos: el modelo hacia atrás condicional, que incluía las siguientes variables:

- Longitud del tramo recto anterior
- Sentido

La comparación de los porcentajes de clasificación correctos, tanto para el modelo inicial con SPSS como para el método de validación *Leave-One-Out*, da lugar a los siguientes resultados, que difieren en menor medida que en el caso de los modelos anteriores.



La curva ROC obtenida es la que se muestra en la siguiente imagen, apreciándose que el área bajo la curva es aún más bajo que para el modelo intermedio.

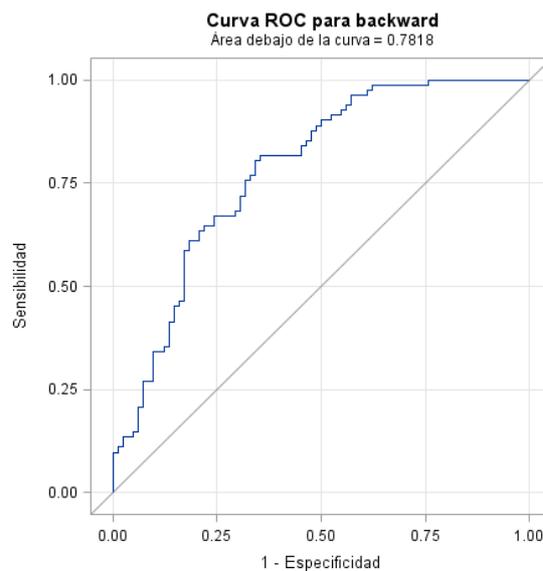


Ilustración 51. Curva ROC del modelo hacia delante para curvas e intersecciones.

### Comparación de los tres modelos para curvas e intersecciones

Finalmente, se comparan los tres modelos estudiados con el objetivo de comprobar si existen diferencias estadísticamente significativas entre ellos y así poder seleccionar el modelo más adecuado. Se calcula un contraste ROC en cuanto a las áreas bajo la curva, tomando como referencia el modelo intermedio. Los resultados obtenidos fueron los siguientes:

Estimación de contraste ROC y resultados de prueba por fila						
Contraste	Estimador	Error estándar	Límites de confianza al 95% de Wald		Chi-cuadrado	Pr > Chicua
Modelo - intermedio	0.0190	0.0139	-0.00821	0.0463	1.8748	0.1709
backward - intermedio	-0.0422	0.0220	-0.0854	0.000887	3.6850	0.0549

Tabla 59. Estimación de contraste ROC para curvas e intersecciones.

Debido a que el p-valor obtenido del contraste entre los modelos “backward” e “intermedio” roza un nivel de significación del 5%, se podría pensar que el modelo más adecuado es el “intermedio”. Sin embargo, si se observan los porcentajes de clasificación correctos obtenidos en ambos casos, se comprueba que el del modelo “intermedio” supone una peor capacidad predictiva ante posibles casos futuros, ya que toma un valor de 68,3% frente a un valor de 70,7% obtenido con el “backward”. Por consiguiente, el modelo seleccionado para este conjunto de datos es el obtenido con el método de selección de variables hacia atrás condicional.

Estas mismas conclusiones se pueden observar gráficamente a través de la comparativa de las tres curvas ROC (ilustración 52), en las que las áreas bajo la curva de cada modelo son indicadores de la exactitud global de cada uno de ellos. Tal y como se puede ver en la tabla posterior las áreas bajo las curvas de cada uno de ellos son las siguientes:

MODELO	ÁREA BAJO LA CURVA ROC
Completo	0,8431
Intermedio	0,8241
Backward	0,7818

Tabla 60. Tabla de las áreas bajo la curva ROC de cada modelo para curvas e intersecciones.

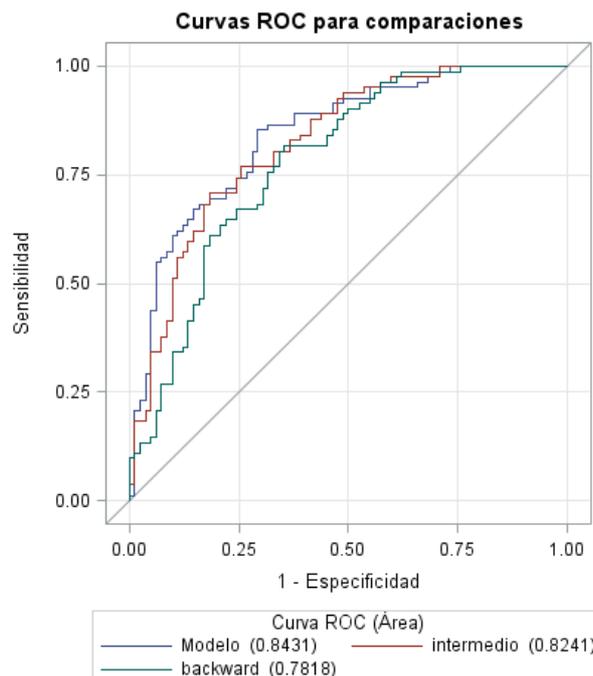


Ilustración 52. Comparación de curvas ROC para curvas e intersecciones.

## 7.2. VALIDACIÓN *LEAVE-ONE-OUT* Y ANÁLISIS ROC EN LAS INTERSECCIONES

En esta sección se realizarán los mismos procedimientos que para la anterior, pero únicamente con el subconjunto de datos de intersecciones. A continuación, se muestra el código empleado en el software SAS para obtener los resultados deseados:

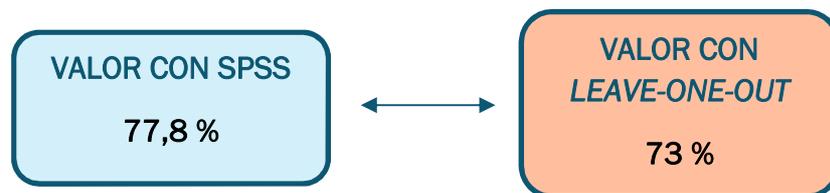
```
libname acc 'd:/SAS';

proc logistic data=acc.cycINTERSECCIONES;
class tipointe sentido pend priori;
model casocontrol=tipointe anchocalz anchocarr numcarril sentido longrec
distsem pend priori/ctable;
*model casocontrol=tipointe anchocalz numcarril longrec pend/ctable;
*model casocontrol=anchocalz numcarril longrec/ctable;
roc 'intermedio' tipointe anchocalz numcarril longrec pend;
roc 'backward' anchocalz numcarril longrec;
rocontrast reference('intermedio') / estimate e;
run;
```

### Modelo completo para las intersecciones

En primer lugar, se obtuvieron los resultados para el modelo completo, en el que se incluyen todas las variables excepto el radio de curvatura, ya que al tratarse únicamente de intersecciones no es una variable de interés.

Seguidamente, se comparan los porcentajes de clasificación correctos entre el modelo completo obtenido con SPSS y el obtenido tras realizar la validación cruzada *Leave-One-Out*, teniendo en cuenta que el punto de corte escogido era 0,5.



Tal y como se puede ver, el porcentaje con el método *Leave-One-Out* siempre es menor, ya que se ajusta más a la realidad. Esto se debe a que el individuo que se emplea como prueba no está incluido en el entrenamiento del modelo, a diferencia de lo que ocurre con el valor obtenido de SPSS.

En la curva ROC siguiente se han representado todos los pares Sensibilidad/(1-Especificidad) para cada punto de corte del modelo completo, ajustado únicamente con las intersecciones.

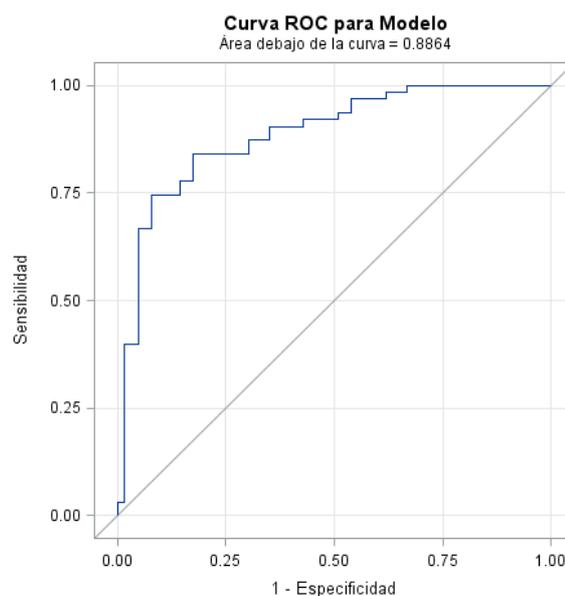


Ilustración 53. Curva ROC del modelo completo para las intersecciones.

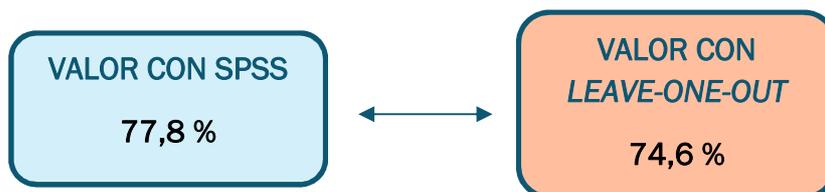
### **Modelo intermedio para las intersecciones**

En este caso el modelo intermedio obtenido con el método de selección de variables hacia atrás condicional estaba formado por las siguientes:

- Ancho de calzada
- Número de carriles
- Longitud del tramo recto anterior
- Pendiente
- Tipo de intersección

Por lo tanto, estas son las variables incluidas en el modelo intermedio en SAS para la obtención de la tabla de clasificación y la curva ROC.

A continuación se comparan los porcentajes de clasificación, tanto para el modelo inicial como para el obtenido con la validación *Leave-One-Out*. En este caso se comprueba que la diferencia entre los porcentajes es mucho menor que en el caso del modelo completo, en el que el porcentaje inicial con SPSS coincide con el del modelo intermedio estudiado actualmente.



A través de los pares Sensibilidad/(1 – Especificidad) para cada punto de corte contenido en la tabla de clasificación se obtiene la curva ROC. Además, se aprecia que el área bajo la curva (indicador de la exactitud global del modelo) es 0,8793, un valor similar al del modelo completo.

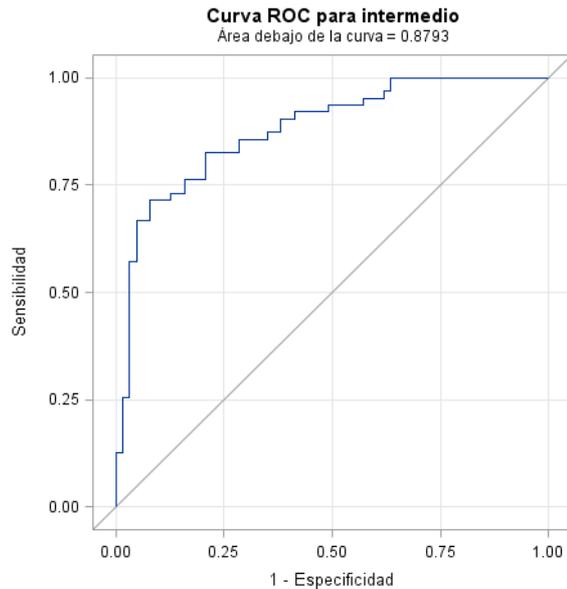


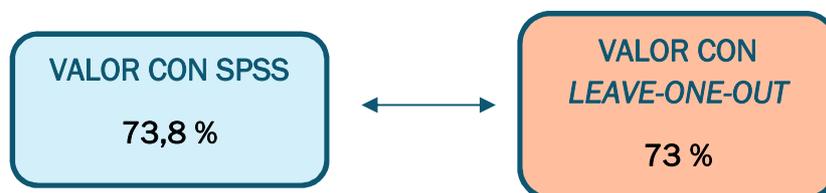
Ilustración 54. Curva ROC del modelo intermedio para las intersecciones.

### ***Modelo hacia atrás para intersecciones***

El modelo hacia atrás condicional seleccionado estaba formado por las siguientes variables:

- Ancho de calzada
- Número de carriles
- Longitud del tramo recto anterior

Por lo tanto, estas son las que se incluyen en el modelo intermedio en SAS para la obtención de la tabla de clasificación y curva ROC. Si se observa dicha tabla y se comprueba el porcentaje de clasificación correcto para el punto de corte 0,5, se obtiene un valor de 73%, prácticamente igual que el obtenido inicialmente con SPSS sin la validación del modelo. Por este motivo, se deduce que el porcentaje de clasificación correcto inicial se acercaba mucho a la realidad.



A partir de esta tabla de clasificación se obtiene la curva ROC que recoge todos los pares Sensibilidad/(1 - Especificidad) de cada punto de corte del modelo intermedio para las intersecciones.

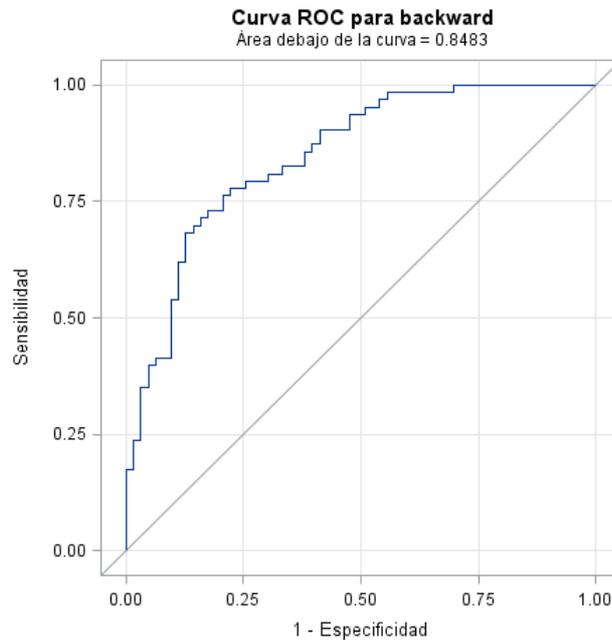


Ilustración 55. Curva ROC del modelo hacia atrás para las intersecciones.

### Comparación de los tres modelos para las intersecciones

Finalmente, se comparan los tres modelos estudiados para comprobar si existen diferencias estadísticamente significativas entre ellos, con el objetivo de seleccionar el modelo más adecuado. Se toma como categoría de referencia el modelo intermedio y se calcula el contraste ROC. Los resultados obtenidos fueron los siguientes:

Estimación de contraste ROC y resultados de prueba por fila						
Contraste	Estimador	Error estándar	Límites de confianza al 95% de Wald		Chi-cuadrado	Pr > Chicua
Modelo - intermedio	0.00705	0.00797	-0.00857	0.0227	0.7832	0.3762
backward - intermedio	-0.0310	0.0191	-0.0684	0.00645	2.6325	0.1047

Tabla 61. Estimación de contraste ROC para las intersecciones.

Debido a que los p-valores para ambos contrastes toman valores superiores a 0,05, se puede concluir que no existen diferencias significativas entre ninguno de los modelos completo y “backward” frente al “intermedio”. Por lo tanto, el modelo más satisfactorio para el subconjunto de intersecciones sería el modelo

final obtenido con el método de selección de variables hacia atrás condicional, que se ha denominado como “backward”.

Estas mismas conclusiones se pueden extraer de forma cualitativa del gráfico de curvas ROC comparativo (ilustración 56) para los tres modelos analizados. Si se observa el área bajo la curva en cada caso, medida de la exactitud global del modelo, se puede ver que no existen diferencias importantes entre las correspondientes a cada uno. Así mismo, también se aprecia que estas áreas son superiores respecto al análisis de las curvas e intersecciones conjuntamente. Por lo tanto, parece que los modelos son más adecuados cuando se analizan únicamente las intersecciones.

MODELO	ÁREA BAJO LA CURVA ROC
Completo	0,8864
Intermedio	0,8793
Backward	0,8483

Tabla 62. Tabla de las áreas bajo la curva ROC de cada modelo para las intersecciones.

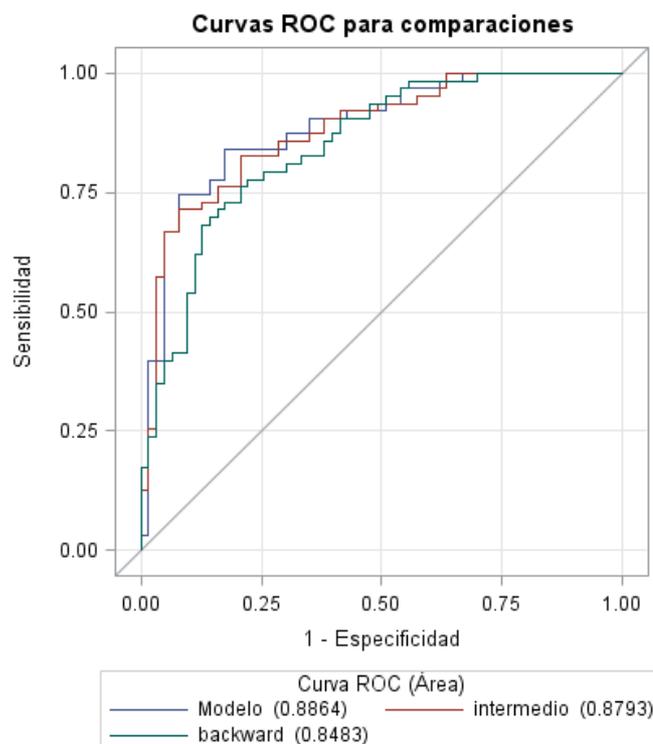


Ilustración 56. Comparación de los tres modelos para las intersecciones.

### 7.3. VALIDACIÓN *LEAVE-ONE-OUT* Y ANÁLISIS ROC PARA LAS CURVAS

En último lugar, se realiza el mismo procedimiento pero únicamente para el subconjunto de curvas. En el siguiente cuadro se recoge el código en SAS para obtener los resultados deseados (tabla de clasificación con *Leave-One-Out* y análisis ROC).

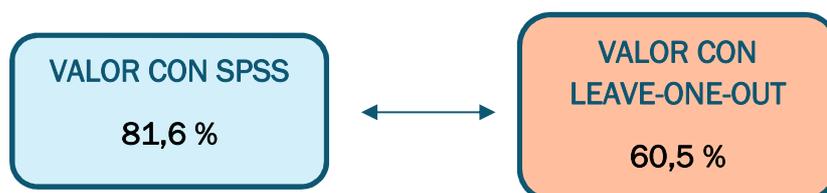
```
libname acc 'd:/SAS';

proc logistic data=acc.cycCURVAS;
class sentido pend priori;
model casocontrol=radcurv anchocalz anchocarr numcarril sentido longrec
distsem pend priori/ctable;
*model casocontrol=anchocarr sentido priori/ctable;
*model casocontrol=anchocarr priori/ctable;
roc 'intermedio' anchocarr sentido priori;
roc 'backward' anchocarr anchocarr priori;
roccompare reference('intermedio') / estimate e;
run;
```

#### *Modelo completo para las curvas*

En el caso del subconjunto de curvas el modelo completo incluye todas las variables salvo el tipo de intersección, ya que no se trata de una variable de interés.

A continuación, se realiza una comparativa de los porcentajes de clasificación correctos del modelo inicial, obtenido con la regresión logística binaria en SPSS, y el extraído de la tabla de clasificación con el método de validación cruzada *Leave-One-Out* para el punto de corte 0,5, ya que fue el seleccionado anteriormente en SPSS.



Tal y como se puede ver los porcentajes difieren aproximadamente en un 20%, lo que significa que el obtenido con SPSS sin la validación se aleja mucho de la clasificación real que se obtiene con el método *Leave-One-Out*. Es decir, el error real es mucho mayor de lo que se pensó inicialmente al observar los datos porcentuales de clasificación supuestos o aparentes.

A continuación, se muestra la curva ROC con la que se puede apreciar de forma gráfica como de óptimo es el modelo, además de ver el área bajo la curva que es un indicador de su exactitud global.

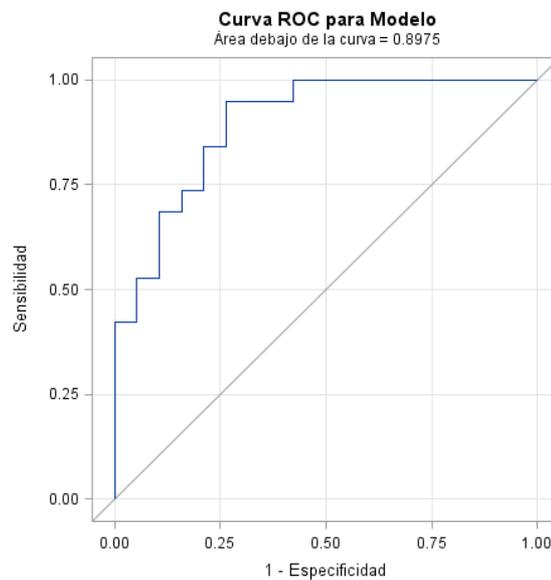


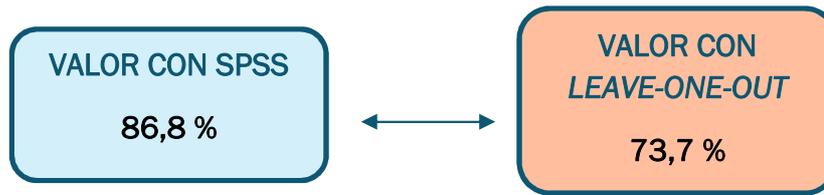
Ilustración 57. Curva ROC del modelo completo para las curvas.

### ***Modelo intermedio para las curvas***

Tal y como se mencionó anteriormente, el segundo de los modelos seleccionados para ser comparado es el intermedio, obtenido con el método de selección de variables hacia atrás condicional para los datos referentes a curvas y formado por las siguientes variables:

- Ancho de carril
- Sentido
- Prioridad

A continuación, se comparan los porcentajes de clasificación correctos para el modelo obtenido con SPSS frente al correspondiente a un punto de corte de 0,5 en la tabla de clasificación.



Se puede ver que el porcentaje de clasificación correcto es superior al del modelo completo, tanto para el caso sin la validación *Leave-One-Out* como con ella. Además, la diferencia de porcentajes, aunque aún bastante elevada, es menor que en el modelo anterior.

A pesar de ello, se puede apreciar que la curva ROC es similar, aunque el área bajo la curva es algo menor en este caso.

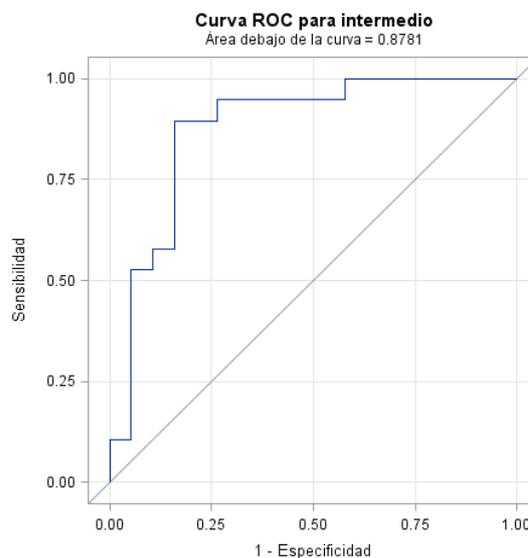


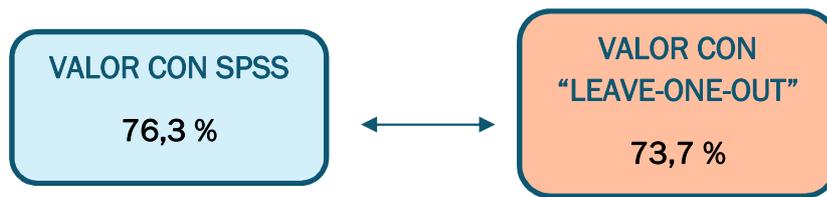
Ilustración 58. Curva ROC del modelo intermedio para las curvas.

### ***Modelo hacia atrás para curvas***

El modelo hacia atrás condicional para el subconjunto de datos de curvas estaba formado por las siguientes variables:

- Ancho de carril
- Prioridad

Para ver el efecto de la validación *Leave-One-Out* se comparan los porcentajes de clasificación correctos:



Cabe destacar que el porcentaje inicial difiere menos del obtenido con la validación, mucho más realista. Esto se traduce en que el error supuesto inicialmente se acercaba más a la realidad que en los casos anteriores.

Por último, a través de los pares de Sensibilidad/(1 - Especificidad) para los distintos puntos de decisión, se obtiene la curva ROC y se concluye que el área bajo la curva es algo menor que en los dos modelos anteriores, con lo que la exactitud global del modelo es inferior.

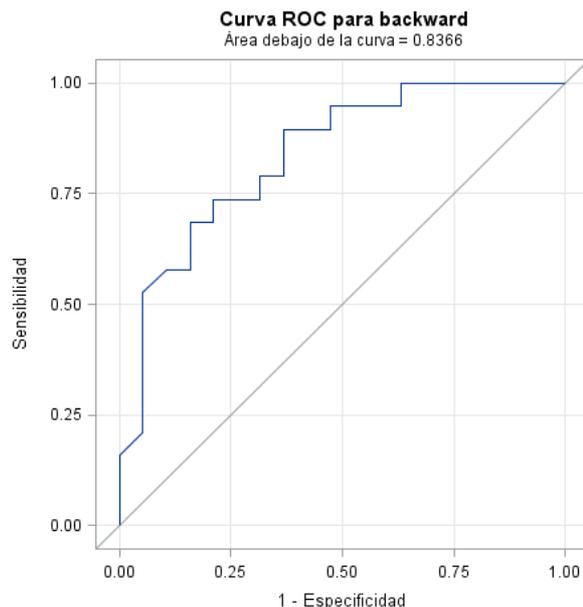


Ilustración 59. Curva ROC del modelo hacia delante para las curvas.

### **Comparación de los tres modelos para las curvas**

Para la comparación de los tres modelos posibles en las curvas y, con la finalidad de escoger el modelo más adecuado, se realizan dos contrastes ROC con el modelo intermedio como referencia, obteniéndose los siguientes resultados:

Estimación de contraste ROC y resultados de prueba por fila						
Contraste	Estimador	Error estándar	Límites de confianza al 95% de Wald		Chi-cuadrado	Pr > Chicua
Modelo - intermedio	0.0194	0.0343	-0.0477	0.0865	0.3205	0.5713
backward - intermedio	-0.0416	0.0388	-0.1177	0.0346	1.1441	0.2848

Tabla 63. Estimación de contraste ROC para las curvas.

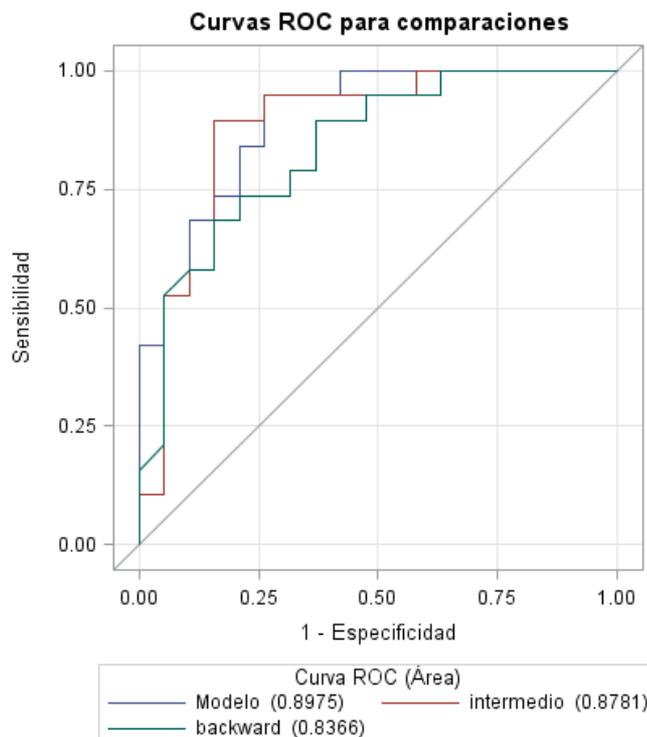


Ilustración 60. Comparación de los tres modelos para las curvas.

Tal y como demuestran los p-valores de cada contraste no existen diferencias significativas entre los modelos completo y hacia atrás, frente al intermedio. Por lo tanto, el modelo *backward* es el más adecuado para el subconjunto de curvas.

Estas conclusiones, además, se pueden ver cualitativamente porque las tres curvas ROC son muy similares (ilustración 60), al igual que las áreas bajo la curva de cada modelo recogidas en la siguiente tabla:

MODELO	ÁREA BAJO LA CURVA ROC
Completo	0,8975
Intermedio	0,8781
Backward	0,8366

Tabla 64. Tabla de las áreas bajo la curva ROC de cada modelo para las curvas.

## 7.4. MODELO FINAL

Finalmente, gracias a las conclusiones de los contrastes ROC realizados para los tres posibles conjuntos de datos, el modelo seleccionado en cada caso es el mismo: hacia atrás condicional. En la tabla siguiente se muestran las variables finales para cada conjunto de datos, además del porcentaje de clasificación correcto que mide la capacidad predictiva del modelo en cada caso.

CONJUNTO DE DATOS	VARIABLES EN EL MODELO	% DE CLASIFICACIÓN CORRECTO
Curvas e intersecciones	Longitud del tramo recto anterior	70,7 %
	Sentido	
Intersecciones	Ancho de calzada	73 %
	Número de carriles	
Curvas	Longitud del tramo recto anterior	73,7 %
	Ancho de carril	
	Prioridad	

Tabla 65. Resumen de variables para el modelo final.

Por lo tanto, se puede apreciar que en el modelo para curvas e intersecciones existe una variable específica que únicamente aparece si se trabaja con todos los datos: “Sentido”. Por otro lado, la otra variable incluida en el modelo, “Longitud del tramo recto anterior”, pasa a formar parte del ajustado para las intersecciones.

Además, en los modelos de las curvas e intersecciones por separado existen otras variables adicionales con respecto a las resultantes cuando se tratan todos los datos conjuntamente. Éstas son factores específicos que influyen en la

ocurrencia de una salida de calzada en esos dos escenarios concretos: intersección o curva.

En las intersecciones, además, aparecen las variables específicas: “Ancho de calzada” y “Número de carriles”. En cambio, en las curvas todas las variables incluidas en el modelo son específicas de este tipo de escenario: “Ancho de carril” y “Prioridad”.

Por otro lado, cabe destacar que los modelos obtenidos son más adecuados para las intersecciones y las curvas por separado, ya que los porcentajes de clasificación correctos obtenidos tras la validación *Leave-One-Out* son mayores que cuando se analizan ambos escenarios conjuntamente.

A continuación, se recogen las tablas correspondientes al modelo logístico final para cada uno de los conjuntos de datos:

### Modelo logístico final para curvas e intersecciones

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
Longitud del tramo recto	,002	,001	12,53	1	,000	1,003	1,001	1,004
Sentido único			14,90	2	,001			
Doble sentido	-,565	,418	1,821	1	,177	,569	,250	1,291
Doble sentido con mediana	1,402	,509	7,577	1	,006	4,065	1,498	11,03
Constante	-1,164	,349	11,11	1	,001	,312		

Tabla 66. Modelo logístico final para curvas e intersecciones.

Como se puede ver, las únicas variables significativas son la “Longitud del tramo recto anterior”, “Sentido único” y “Doble sentido con mediana”, ya que sus p-valores son inferiores a 0,05. En cuanto a la interpretación de este modelo hay que destacar que:

- Cuanto mayor sea la longitud del tramo recto anterior, mayor es el riesgo de que ocurra una salida de calzada, tal y como indica el *Odds-Ratio*, cuyo valor estimado es significativamente mayor que 1. Para traducir estos parámetros en una medida del incremento del riesgo se recurre al siguiente ejemplo:

$$\text{logit} = \frac{P}{1 - P} = e^{\beta_0} e^{\beta_1 X_1} e^{\beta_2 X_2} \dots e^{\beta_k X_k}$$

Si se aumenta en 50 metros, por ejemplo, la longitud del tramo recto anterior, manteniéndose constantes los valores del resto de variables, se llega a la siguiente expresión:

$$e^{\beta_1(X_1+50)} = e^{0,002 \cdot 50} e^{\beta_1 X_1} = 1,105 e^{\beta_1 X_1}$$

Por lo tanto, al incrementar 50 metros dicha longitud el riesgo de que ocurra una salida de calzada es aproximadamente un 10% mayor (OR = 1,105). Con cálculos similares tendríamos que un aumento de 300 metros supondría duplicar el riesgo (OR = 2,014).

- Cuando existe mediana divisoria en calzadas de doble sentido el riesgo de que suceda una salida de la misma se triplica, frente a si dicha calzada fuera de sentido único.

Por lo tanto, habrá que tener especial precaución en aquellas curvas e intersecciones de doble sentido con mediana divisoria, en las que el conductor disponga de una gran longitud recta anterior que le permita acelerar el vehículo a altas velocidades superando, en algunos casos, los límites legales.

### Modelo logístico final para las intersecciones

	β	Error estándar	Wald	gl.	Sig.	EXP(β)	95 % C. I. para EXP(β)	
							Inferior	Superior
Longitud del tramo recto	,003	,001	11,22	1	,001	1,003	1,001	1,004
Ancho de calzada	-,289	,122	5,629	1	,018	,749	,590	,951
Número de carriles	1,761	,413	18,15	1	,000	5,819	2,588	13,08
Constante	-1,730	,785	4,858	1	,028	,177		

Tabla 67. Modelo logístico final para las intersecciones.

- Al igual que cuando se analizan curvas e intersecciones en conjunto, a mayor longitud del tramo recto anterior mayor será el riesgo de que se produzca el desenlace estudiado. En concreto, si se retoma el mismo ejemplo que en el caso anterior, al aumentar la longitud en 50 metros se llega a la siguiente expresión:

$$e^{\beta_1(X_1+50)} = e^{0,003 \cdot 50} e^{\beta_1 X_1} = 1,162 e^{\beta_1 X_1}$$

Por lo tanto, el riesgo de que suceda una salida de la vía es un 16% mayor si la distancia recta que puede recorrer el vehículo hasta llegar a la intersección es 50 metros superior. Además, en este caso, si el valor se incrementa en 250 metros se obtiene un *Odds-Ratio* por encima de 2 (OR = 2,117), lo que indica que el riesgo bajo estas circunstancias es más del doble.

- En el análisis univariante realizado para la obtención del *Odds-Ratio* (apartado 5.3.) podía parecer que en las calzadas muy estrechas era más difícil que ocurriese este tipo de accidente (se incrementaba el riesgo para las categorías intermedias que comprenden las calzadas entre 6 y 8 metros). Sin embargo, en el modelo global para las intersecciones, esa característica parece estar compensada por otras variables, con lo que finalmente se trata de un factor de protección. En este modelo el valor del *Odds-Ratio* es significativamente menor que 1 (OR = 0,749), por lo que el riesgo se reduce en un 25% al incrementar un metro el ancho de la calzada.
- Por último, un número elevado de carriles para el sentido de la circulación es un factor de riesgo, ya que el OR toma un valor de 5,819 lo que se traduce en que al aumentar en una unidad el valor de esta variable se multiplica prácticamente por seis el riesgo de que ocurra una salida de la vía.

Por lo tanto, en aquellas intersecciones con una gran longitud del tramo recto anterior, calzada estrecha o con un ancho intermedio, y un elevado número de carriles para el mismo sentido habrá un mayor riesgo de que se produzca una salida de la vía.

### **Modelo logístico final para las curvas**

	$\beta$	Error estándar	Wald	gl.	Sig.	EXP( $\beta$ )	95 % C. I. para EXP( $\beta$ )	
							Inferior	Superior
Ancho de carril	1,623	,725	5,010	1	,025	5,066	1,224	20,97
Prioridad	21,78	15209	,000	1	,999	289343	,000	.
Constante	-6,618	2,799	5,589	1	,018	,001		

Tabla 68. Modelo logístico final para las curvas.

- En el caso de las curvas se puede ver que un carril amplio supone un factor de riesgo, ya que el valor estimado del OR es de 5,066, lo que indica que al aumentar en un metro el ancho de carril el riesgo se multiplica por cinco.

- Por otro lado, tal y como se puede ver en la tabla anterior, los estimadores para la variable “Prioridad” son muy inestables debido a que en la tabla cruzada para las curvas hay una celda sin ningún dato:

		Caso o control		Total
		Control	Caso	
Prioridad regulada por:	Semáforo	0 0,0%	6 31,6%	6 15,8%
	Ninguna (sólo norma)	19 100,0%	13 68,4%	32 84,2%
Total		19 100,0%	19 100,0%	38 100,0%

Tabla 69. Tabla cruzada de "Prioridad" frente a "Caso o control" para las curvas.

Sin embargo, aunque el ajuste del modelo sea inestable por cuestiones técnicas, parece un aspecto significativo y que debería tenerse en cuenta ya que, eligiendo al azar el grupo de curvas definidas como controles, no se encontró ninguna con semáforo. Así mismo, se trató de modelar la muestra sin esta variable, pero se observó que se perdía prácticamente un 10% de la clasificación. Por consiguiente, aquellas curvas en las que la regulación de la prioridad se realice mediante semáforo se considerarán como un factor de riesgo, aunque no sea posible evaluarlo bien cuantitativamente al tener una celda vacía en la tabla cruzada.

Esto no significa que el problema sea el semáforo, sino que podría deberse a que los usuarios de la vía no lo respetan adecuadamente. Otra posibilidad podría ser que el semáforo se hubiera colocado en esos lugares precisamente por ser considerados como puntos de riesgo. En ese caso, este hecho implicaría que la regulación con semáforo no es suficiente, por lo que no sería incompatible con la instalación de las barreras de contención adecuadas.

Por lo tanto, se concluye que en aquellas curvas en las que haya un ancho de carril especialmente amplio y cuya regulación de la prioridad se realice mediante semáforo, habrá que tomar las medidas oportunas para evitar que ocurran salidas de la vía.



# CONCLUSIONES Y LÍNEAS FUTURAS DE TRABAJO



## CONCLUSIONES FINALES

Las cifras de accidentes muestran que se producen, anualmente, alrededor de 4000 salidas de calzada en entorno urbano dentro del territorio nacional (según datos de la DGT), con un balance medio de 60 víctimas mortales al año. Por todo ello, surge la necesidad de estudiar qué medidas podrían ayudar a reducir el número de víctimas para tratar de avanzar en la consecución del objetivo “cero lesionados”, aspecto principal en el que se centran las diferentes políticas de seguridad vial a nivel nacional y europeo.

Por otro lado, a pesar de que existen algunos proyectos que han analizado las bases de datos existentes con el fin de caracterizar los escenarios donde se producen las salidas de calzada, no existen estudios que hayan detallado las características geométricas y de tráfico en las dos ubicaciones más frecuentes donde se produce este tipo de accidente en zona urbana: intersecciones y curvas, en las que ocurren más del 40% y del 20% de las salidas de la vía, respectivamente.

Por estos motivos, resulta de máxima relevancia la realización de un trabajo de investigación que trate de definir qué factores geométricos incrementan el riesgo de que suceda una salida de calzada en intersecciones o curvas urbanas. Todo ello con el objetivo final de ayudar a las administraciones titulares de las vías a que determinen de manera fácil y rápida qué intersecciones o tramos curvos deben estar protegidos mediante las adecuadas barreras de contención para evitar una posible salida de calzada, así como los daños derivados de la misma sobre los usuarios más vulnerables (peatones, ciclistas y motociclistas).

### Conclusiones sobre las bases de datos

Para lograr estos objetivos se realizó un estudio caso-control, definiendo como “casos” aquellas intersecciones y curvas urbanas de la ciudad de Valladolid en las que se produjo alguna salida de calzada entre los años 2009 y 2014; y definiendo como “controles” las intersecciones y curvas en las que no hubo este tipo de accidente en dicho periodo. Para la obtención de la muestra de “casos” se emplearon dos bases de datos de accidentes de tráfico: en primer lugar, ARENA, con la información recogida por la DGT respecto a los accidentes con víctimas; y en segundo lugar, la elaborada para el desarrollo del PIMUSSVA (Plan Integral de Movilidad Urbana Sostenible y Segura de Valladolid). Tras comparar la información obtenida en ambas, se llegaron a las siguientes conclusiones:

- Se han detectado numerosas inconsistencias a la hora de codificar algunas de las variables categóricas, de modo que, para algunos accidentes, los valores almacenados no coinciden con los reales una vez estudiados individualmente en profundidad a partir de los expedientes que elabora la Policía Municipal de Valladolid.
- Por otro lado, existe una cierta ambigüedad en cuanto a la tipología de accidente entre ambas bases de datos, ya que el mismo accidente aparece en múltiples casos codificado de una forma distinta en cada una de ellas.
- En la base de datos ARENA resulta difícil localizar adecuadamente cada accidente, ya que únicamente se almacena el nombre de la calle en la que sucedió, lo que impide saber en cuál de todas las intersecciones o curvas de esa calle se produjo realmente. En algunas situaciones se introduce la calle principal y, además, dos calles cercanas de referencia entre las que se produjo el accidente. Sin embargo, esa referencia resulta poco válida porque, en muchos casos, entre esas dos calles existen varias intersecciones o curvas, con lo que no termina de precisarse con exactitud la ubicación del suceso. Este hecho es especialmente relevante en un estudio como el que nos ocupa, en el que se pretende analizar las características geométricas de las ubicaciones donde ocurren los accidentes, lo cual se ve dificultado por la falta de información exacta de estos escenarios.

Teniendo en cuenta estas cuestiones, parece recomendable sugerir que se refuerce la formación del personal de recogida de los datos, para evitar estas ambigüedades y lograr una mayor uniformidad en la toma de información. Así mismo, sería conveniente realizar un esfuerzo por recoger variables adicionales que puedan ser relevantes, con el fin de aprovechar al máximo las bases de datos en estudios como el que se ha desarrollado.

## Conclusiones sobre los modelos finales

Para el análisis de los datos se empleó la metodología de regresión logística sobre tres posibles conjuntos de datos: las intersecciones por un lado, las curvas por otro, y las intersecciones y curvas a la vez. Estos tres análisis pretendieron determinar de qué forma se obtendrían unos resultados mejores.

Finalmente, resultó mejor analizar por separado tanto las intersecciones como las curvas, ya que las variables explicativas, resultantes de los modelos logísticos

obtenidos con los métodos de selección de variables, eran diferentes para los dos posibles escenarios; es decir, la probabilidad de que ocurra una salida de calzada se explica con diferentes factores en curvas y en intersecciones.

Por consiguiente, aquellos aspectos más relevantes son los que se recogen a continuación:

---

## INTERSECCIONES

Las variables que afectan de forma relevante al riesgo de ocurrencia de una salida de calzada en las intersecciones son las siguientes:

- Si se incrementa la longitud del tramo recto anterior a la intersección, permaneciendo constantes el valor del resto de las variables, el riesgo de que suceda este tipo de accidente es mayor. Por ejemplo, si se incrementa esta longitud en 50 metros, el riesgo es un 10% mayor; mientras que si aumenta en 100 metros es un 22% mayor.
- Por otro lado, el ancho de calzada también es un factor de influencia, ya que si éste se incrementa en un metro, manteniendo constantes las otras variables, el riesgo disminuye un 25%. Este resultado entra en aparente contradicción con lo obtenido en los análisis individuales de las variables explicativas que mostraban que, si la calzada tiene un ancho de entre 6 y 8 metros, el riesgo aumenta frente a las calzadas más estrechas de menos de 6 metros. Sin embargo, ese resultado no está considerando el efecto del resto de variables incluidas en el modelo (a la vez que varía la anchura de la calzada pueden cambiar también las otras variables explicativas, solapándose los efectos de dichas variaciones). Así pues, habrá que tener especial precaución en las calzadas estrechas o con un ancho intermedio (hasta 8 metros) para evitar una salida de la vía.
- El número de carriles también es determinante a la hora de que ocurra este tipo de accidente. Cuando se añade un carril, manteniéndose constantes el resto de las variables, el riesgo se multiplica prácticamente por seis.

En consecuencia, en aquellas intersecciones en las que haya una gran longitud del tramo recto anterior, un ancho de calzada estrecho o intermedio (hasta 8 metros) y un elevado número de carriles en relación con el ancho de calzada, habrá que tener especial precaución y considerar la implantación de barreras de contención de los vehículos motorizados.

Para entender el papel de todos los factores de riesgo encontrados cabe destacar que, en todos los casos, podrían estar apuntando al papel de la velocidad inadecuada de los vehículos como verdadero factor de riesgo subyacente. Efectivamente, es razonable pensar que circular por tramos rectos muy largos y con varios carriles para el sentido de la circulación, es una circunstancia que puede favorecer el hecho de que muchos conductores circulen a velocidades elevadas. Si a estas circunstancias se añade un ancho de calzada insuficiente para dicha velocidad y número de carriles, la combinación de factores produce un incremento del riesgo de ocurrencia del evento.

---

## CURVAS

En cuanto al análisis realizado para las curvas hay que destacar lo siguiente:

- El ancho de carril actúa como un factor de riesgo, ya que al incrementarse éste, manteniendo constantes el resto de variables, aumenta el riesgo de que suceda una salida de calzada. De hecho, si se incrementa en un metro el riesgo es aproximadamente cinco veces mayor.
- Por otro lado, respecto a aquellas curvas en las que la regulación de la prioridad se realiza mediante semáforo, se ha podido comprobar que el riesgo de ocurrir una salida de calzada es mayor que si hubiera una señal de STOP. Esto no significa que el problema sea el semáforo, sino que podría deberse a que los usuarios de la vía no lo respetan adecuadamente y aumentan, en ocasiones, la velocidad de circulación buscando poder alcanzarlo in extremis, antes de que se ponga en rojo. Otra posibilidad podría ser que el semáforo se hubiera colocado en esos lugares precisamente por ser considerados como puntos de riesgo. En ese caso, este hecho implicaría que la regulación con semáforo no es suficiente, por lo que no sería incompatible con la instalación de las barreras de contención adecuadas.

Por consiguiente, las curvas en las que habría un mayor riesgo de que ocurriese una salida de calzada serían aquellas con un carril amplio y cuya regulación de la prioridad se esté realizando a través de semáforo.

## LÍNEAS FUTURAS DE TRABAJO

Inicialmente se consideró incluir en el estudio la variable Intensidad Media Diaria (IMD) con objeto de determinar la influencia de la densidad de tráfico sobre las salidas de calzada durante el periodo de estudio. Sin embargo, los medidores presentes en la ciudad no abarcan muchos de los lugares incluidos en el análisis, lo que se tradujo en una significativa falta de información, por lo que se decidió no incluir esta variable en el modelo. A pesar de ello, como ya se ha mencionado en esta sección, se cree que la velocidad de circulación (parámetro que también se puede obtener a través de los aforos o medidores de IMD colocados en la ciudad) podría ser un aspecto relevante a considerar como factor desencadenante en la ocurrencia de este tipo de accidente. Por esta razón, se tratará de obtener esta información en un futuro para poder decidir sobre su influencia en las salidas de calzada.

El presente estudio se ha realizado con datos referentes a la ciudad de Valladolid, población de tamaño intermedio que podría ser identificada como ciudad tipo para, posteriormente y con las precauciones necesarias, tratar de extrapolar las conclusiones de este estudio a otras ciudades españolas. Por consiguiente, como líneas futuras de trabajo cabe la posibilidad de realizar este mismo análisis en ciudades de otra tipología, con diferente tamaño o distintas características urbanísticas y/o topográficas. El objetivo que se persigue es tener un mayor conocimiento sobre los factores que influyen en la ocurrencia de una salida de calzada y confirmar si los que se han detectado en esta investigación son comunes a todas las ciudades o si, por el contrario, dependen de las características específicas de cada una.

Por último, dado que los resultados obtenidos pueden ser relevantes en la literatura científica actual sobre accidentes urbanos, está prevista la redacción y posterior publicación de un artículo científico con los resultados del presente estudio.





# BIBLIOGRAFÍA



[1] Las principales cifras de la siniestralidad en vías urbanas. España 2013. (Dirección General de Tráfico)

[http://www.dgt.es/Galerias/seguridad-vial/estadisticas-e-indicadores/publicaciones/accidentes-urban/Siniestralidad\\_Urbana\\_2013\\_EE.pdf](http://www.dgt.es/Galerias/seguridad-vial/estadisticas-e-indicadores/publicaciones/accidentes-urban/Siniestralidad_Urbana_2013_EE.pdf)

[2] SafetyNet. European Road Safety Observatory.

[http://ec.europa.eu/transport/wcm/road\\_safety/erso/safetynet/content/safetynet.htm](http://ec.europa.eu/transport/wcm/road_safety/erso/safetynet/content/safetynet.htm)

[3] MAIDS (Motorcycle Accidents In Depth Study). European Association of Motorcycle Manufacturers.

<http://www.maids-study.eu/index.php?error=hastolog>

[4] Aprosys (Advanced Protection Systems). 6<sup>th</sup> RTD Framework Programme.

<http://www.transport-research.info/project/advanced-protection-systems>

[5] CARE. Base de datos de la Comunidad Europea sobre accidentes con víctimas.

[6] GIDAS (German In-Depth Accident Study).

[7] CCIS (Co-operate Crash Injury Study).

<http://www.dft.gov.uk/rmd/project.asp?intProjectID=10406>

[8] IRTAD (International Traffic Safety Data and Analysis Group).

<http://www.itf-oecd.org/IRTAD>

[9] Molinero, A.; Martín, O.; Perandones, J. M.; Pedrero, D.; Ocampo, M. A. "Accidentalidad e infraestructura vial en carreteras de Diputaciones Provinciales". Revista carreteras 169 (2010).

[10] Fernández, J. A. "Recomendaciones de instalación de sistemas de protección de motociclistas en las carreteras de la red regional de Castilla y León". Consejería de Fomento. Junta de Castilla y León (2007).

[11] Orden circular OC 35/2014 sobre criterios de aplicación de los Sistemas de Contención de Vehículos". Ministerio de Fomento.

[12] “Manual de recomendaciones de diseño, construcción, infraestructura, consecución y mantenimiento del carril bici”. DGT (2000).

[13] “Austroads Guide to Traffic – Bicycles”.

[14] “Specification for pedestrian restraint systems”. British Standard 7818:1995.

[15] “Desarrollo de un modelo concepto de barrera de segregación de espacios seguros para usuarios vulnerables en entornos urbanos”.

[16] Políticas sobre seguridad vial 2011 – 2020. Comisión Europea.

[17] Owusu, Daniel; Quinn, Megan; Sheng Wang, Ke. “Alcohol Consumption, Depression, Insomnia and Colorectal Cancer Screening: Racial Differences”.

[18] Benavides, F. G.; Giráldez, M. T.; Castejón, N.; Zaplana, M.; Delclós, J.; Benach, J.; Gimeno, D. “Análisis de los mecanismos de producción de las lesiones leves por accidentes de trabajo en la construcción en España”. Universitat Pompeu Fabra.

[19] Simpson, Sharon; Wadsworth, Emma; Moss, Susanna; Smith, Andrew. “Minor injuries, cognitive failures and accidents at work: incidence and associated features”. Occupational Medicine 55 (2005): 99 – 108.

[20] Stevenson, Mark; Jamrozik, Konrad; Burton, Paul. “A case-control study of childhood pedestrian injuries in Perth, Western Australia”. Journal of Epidemiology and Community Health 50 (1996): 280 – 287.

[21] de Irala Estévez, Jokin; Martínez González, Miguel Ángel; Seguí Gómez, María. “Epidemiología aplicada”. Ariel Ciencias Médicas (2004), 189 – 194.

[22] Hautzinger, H.; Pastor, C.; Pfeiffer, M.; Schmidt, J. “Analysis Methods for Accident and Injury Risk Studies” (2007).

[23] Estudios caso – control (Boston University School of Public Health)

[http://sphweb.bumc.bu.edu/otlt/MPH-Modules/EP/EP713\\_Case-Control/EP713\\_Case-Control\\_print.html](http://sphweb.bumc.bu.edu/otlt/MPH-Modules/EP/EP713_Case-Control/EP713_Case-Control_print.html)

[24] Mann, C. J. "Observational research methods. Research design II: cohort, cross sectional and case control studies". *Emergency Medicine Journal* 20 (2003): 54 – 60.

[25] Lewallen, Susan; Courtright, Paul. "Epidemiology in Practice: Case-Control Studies". BC Centre for Epidemiologic & International Ophthalmology University of British Columbia.

[26] Wacholder, Sholom; Silverman, Debra; McLaughlin, Joseph; Mandel, Jack. "Selection of Control in Case-Control Studies". *American Journal of Epidemiology* 135.

[27] H. Keogh, Ruth; D. R. Cox. "Case – Control Studies". Cambridge University Press (2014).

[28] Pérez, A.; Kizys, R.; Manzanedo del Hoyo, L. M. "Regresión Logística Binaria". Proyecto e-math. UOC.

[http://www.uoc.edu/in3/emath/docs/T10\\_Reg\\_Logistica.pdf](http://www.uoc.edu/in3/emath/docs/T10_Reg_Logistica.pdf)

[29] Kleinbaum, David G.; Klein, Mitchel. "Logistic Regression". Springer.

[30] Hosmer, David W.; Lemeshow, Stanley. "Applied Logistic Regression". Wiley Interscience (1989).

[31] Izenman, Alan J. "Modern Multivariate Statistical Techniques". Springer (2008): 142-146.

[32] Chambers, J.; Eddy, W.; Härdle, W.; Sheather, S.; Tierney, L. "Local Regression and Likelihood". Springer (1999): 31 – 34.

[33] Burgueño, M. J.; García-Bastos, J. L.; González-Buitrago, J. M.; "Las curvas ROC en la evaluación de las pruebas diagnósticas".

<http://ferran.torres.name/download/shared/roc/ROC%20M1041707.PDF>

[34] López de Ullibarri, I.; Fernández, S. "Curvas ROC". *Cad Aten Primaria* 5 (1998): 229 – 235.

[http://test.fisterra.com/gestor/upload/guias/curvas\\_roc2.pdf](http://test.fisterra.com/gestor/upload/guias/curvas_roc2.pdf)

[35] Manual de Usuario Arena 2. (Dirección General de Tráfico)

[http://www.dgt.es/Galerias/seguridad-vial/estadisticas-e-indicadores/publicaciones/Manual\\_Usuariu\\_Aplicacion.pdf](http://www.dgt.es/Galerias/seguridad-vial/estadisticas-e-indicadores/publicaciones/Manual_Usuariu_Aplicacion.pdf)

[36] Manual de reconstrucción de accidentes de tráfico. CESVIMAP.

[37] Tablas de contingencia

[http://pendientedemigracion.ucm.es/info/socivmyt/paginas/D\\_departamento/materiales/analisis\\_datosyMultivariable/12contin\\_SPSS.pdf](http://pendientedemigracion.ucm.es/info/socivmyt/paginas/D_departamento/materiales/analisis_datosyMultivariable/12contin_SPSS.pdf)

[38] Moreno Jiménez, Antonio; Cañada Torrecilla, Rosa. “Sistemas y análisis de la información geográfica: manual de autoaprendizaje con ArcGIS”. Madrid: Rama (2007).

[39] Reyes Rodríguez, Antonio M. “Manual imprescindible AutoCAD 2016”. Madrid: Anaya Multimedia (2015).

[40] Martín, Quintín; Cabero, M<sup>a</sup> Teresa; del Rosario, Yanira. “Tratamiento estadístico de datos con SPSS”. Thomson (2008): 260 – 271.

[41] Cody, Ronald P. “An introduction to SAS”. Cary (NC): SAS Institute (2015).

[42] Der, Geoff; Everitt, Brian S. “Essential statistics using SAS: University edition”. Cary (NC): SAS Institute (2015).

[43] Hauer, Ezra. “Observational before-after studies in road safety”. Emerald (2007).



# ANEXOS



# ANEXO 1. CUESTIONARIO DE ARENA

CUESTIONARIO ESTADÍSTICO DE ACCIDENTES DE CIRCULACIÓN CON VÍCTIMAS																							
1. AÑO		2. NUMERO DE ACCIDENTE <small>(A rellenar por la Jefatura Provincial de Tráfico)</small>		HOJA		DIRECCIÓN GENERAL DE TRÁFICO		DE CONFORMIDAD CON LO DISPUESTO EN LOS ARTÍCULOS 10, 11 Y 19 DE LA LEY DE LA FUNCIÓN ESTADÍSTICA PÚBLICA, LAS PERSONAS QUE HAYAN INTERVENIDO EN UN ACCIDENTE ESTÁN OBLIGADAS A COLABORAR EN LA CUMPLIMENTACION DE ESTE CUESTIONARIO, CUYOS DATOS PERSONALES SE HALLAN AMPARADOS POR EL SECRETO ESTADÍSTICO.															
3. PROVINCIA		ISLA		8. TIPO DÍA L A F P		9. TOTAL VÍCTIMAS		10. MUERTOS		11. HERIDOS GRAVES		12. HERIDOS LEVES		13. VEHÍCULOS IMPLICADOS		18. HABITANTES (en miles) 1. <input type="checkbox"/> + de 100 2. <input type="checkbox"/> de 50 a 100 3. <input type="checkbox"/> de 5 a 50 4. <input type="checkbox"/> de 1 a 5 5. <input type="checkbox"/> - de 1		19. ZONA 1. <input type="checkbox"/> Carretera 2. <input type="checkbox"/> Urbana 3. <input type="checkbox"/> Travésia 4. <input type="checkbox"/> Variante					
4. MES		5. DÍA		6. HORA		7. DÍA SEMANA		14. RED CARRETERA		15. km.		16. SENTIDO		17. CÓDIGO MUNICIPIO		Nombre de la calle y número Entre calle y calle							
20. TIPO DE VÍA 1. <input type="checkbox"/> AUTOPISTA 2. <input type="checkbox"/> AUTOVÍA 3. <input type="checkbox"/> VÍA RÁPIDA 4. <input type="checkbox"/> VÍA CONVENCIONAL CON CARRIL LENTO 5. <input type="checkbox"/> VÍA CONVENCIONAL 6. <input type="checkbox"/> CAMINO VECINAL 7. <input type="checkbox"/> VÍA DE SERVICIO 8. <input type="checkbox"/> RAMAL DE ENLACE 9. <input type="checkbox"/> OTRO TIPO Número de carriles calzados .....				21. ANCHURA DE LA CALZADA 1. <input type="checkbox"/> MENOS DE 5,99 m. 2. <input type="checkbox"/> ENTRE 6 Y 6,99 m. 3. <input type="checkbox"/> DE 7 m. o MÁS				22. ANCHURA DEL CARRIL 1. <input type="checkbox"/> DE MÁS DE 3,75 m. 2. <input type="checkbox"/> DE 3,25 A 3,75 m. 3. <input type="checkbox"/> MENOS DE 3,25 m.				23. MARCAS VIALES 1. <input type="checkbox"/> INEXISTENTES O BORRADAS 2. <input type="checkbox"/> SÓLO SEPARACIÓN CARRILES 3. <input type="checkbox"/> SEPARACIÓN CARRILES Y BORDES 4. <input type="checkbox"/> SÓLO SEPARACIÓN DE BORDES				24. ARCÉN 1. <input type="checkbox"/> INEXISTENTE O IMPRACTICABLE 2. <input type="checkbox"/> MENOR DE 1,50 m. 3. <input type="checkbox"/> DE 1,50 A 2,49 m. 4. <input type="checkbox"/> DE 2,50 m. EN ADELANTE		25. ARCÉN PAVIMENTADO <input type="checkbox"/> SI <input type="checkbox"/> NO		26. ELEMENTOS DE SEGURIDAD DE LA VÍA 1. <input type="checkbox"/> MEDIANA ENTRE CALZADAS 2. <input type="checkbox"/> BARRERA DE SEGURIDAD 3. <input type="checkbox"/> PANELES DIRECCIONALES 4. <input type="checkbox"/> HITOS DE ARISTA 5. <input type="checkbox"/> CAPTAFAHOS			
27. FUERA DE INTERSECCIÓN 1. <input type="checkbox"/> RECTA 2. <input type="checkbox"/> CURVA SUAVE 3. <input type="checkbox"/> CURVA FUERTE SIN SEÑALIZAR 4. <input type="checkbox"/> CURVA FUERTE CON SEÑAL Y SIN VELOCIDAD SEÑALIZADA 5. <input type="checkbox"/> CURVA FUERTE CON SEÑAL Y VELOCIDAD SEÑALIZADA			28. Tipo 1. <input type="checkbox"/> EN T o Y 2. <input type="checkbox"/> EN X o + 3. <input type="checkbox"/> ENLACE DE ENTRADA 4. <input type="checkbox"/> ENLACE DE SALIDA 5. <input type="checkbox"/> GIRATORIA 6. <input type="checkbox"/> OTROS			29. Acondicionamiento 1. <input type="checkbox"/> NADA ESPECIAL 2. <input type="checkbox"/> SÓLO ISLETAS O PASO PARA PEATONES EN VÍA SECUNDARIA 3. <input type="checkbox"/> PASO PARA PEATONES O ISLETA EN CENTRO VÍA PRINCIPAL 4. <input type="checkbox"/> CARRIL CENTRAL DE ESPERA 5. <input type="checkbox"/> RAQUETA DE GIRO IZQUIERDA 6. <input type="checkbox"/> OTRO TIPO			30. PRIORIDAD REGULADA POR: 1. <input type="checkbox"/> AGENTE 2. <input type="checkbox"/> SEMÁFORO 3. <input type="checkbox"/> SEÑAL DE "STOP" 4. <input type="checkbox"/> SEÑAL DE "CEDA EL PASO" 5. <input type="checkbox"/> SÓLO MARCAS VIALES 6. <input type="checkbox"/> PASO PARA PEATONES 7. <input type="checkbox"/> OTRA SEÑAL 8. <input type="checkbox"/> NINGUNA (Sólo norma)			31. SUPERFICIE 1. <input type="checkbox"/> SECA Y LIMPIA 2. <input type="checkbox"/> UMBRÍA 3. <input type="checkbox"/> MOJADA 4. <input type="checkbox"/> HELADA 5. <input type="checkbox"/> NEVADA 6. <input type="checkbox"/> BARRILLO 7. <input type="checkbox"/> GRAVILLA SUELTA 8. <input type="checkbox"/> ACEITE 9. <input type="checkbox"/> OTRO TIPO											
32. LUMINOSIDAD 1. <input type="checkbox"/> PLENO DÍA 2. <input type="checkbox"/> CREPÚSCULO 3. <input type="checkbox"/> ILUMINACIÓN SUFICIENTE 4. <input type="checkbox"/> ILUMINACIÓN INSUFICIENTE 5. <input type="checkbox"/> SIN ILUMINACIÓN			33. FACTORES ATMOSFÉRICOS 1. <input type="checkbox"/> BUEN TIEMPO 2. <input type="checkbox"/> NIEBLA INTENSA 3. <input type="checkbox"/> NIEBLA LIGERA 4. <input type="checkbox"/> LLOVIZNANDO 5. <input type="checkbox"/> LLUVIA FUERTE 6. <input type="checkbox"/> GRANIZANDO 7. <input type="checkbox"/> NEVANDO 8. <input type="checkbox"/> VIENTO FUERTE 9. <input type="checkbox"/> OTRO			34. VISIBILIDAD RESTRINGIDA POR: 1. <input type="checkbox"/> EDIFICIOS 2. <input type="checkbox"/> CONFIGURACIÓN DEL TERRENO 3. <input type="checkbox"/> VEGETACIÓN 4. <input type="checkbox"/> FACTORES ATMOSFÉRICOS 5. <input type="checkbox"/> DESLUMBRAMIENTO 6. <input type="checkbox"/> POLVO O HUMO 7. <input type="checkbox"/> OTRA CAUSA 8. <input type="checkbox"/> SIN RESTRICCIÓN			35. OTRA CIRCUNSTANCIA 1. <input type="checkbox"/> PASO A NIVEL 2. <input type="checkbox"/> ESTRECHAMIENTO 3. <input type="checkbox"/> CAMBIO DE RASANTE 4. <input type="checkbox"/> FUERTE DESCENSO 5. <input type="checkbox"/> FIRME DESLIZANTE 6. <input type="checkbox"/> BADÉN 7. <input type="checkbox"/> ESCALÓN 8. <input type="checkbox"/> OBRAS 9. <input type="checkbox"/> BACHES 10. <input type="checkbox"/> INUNDACIÓN 11. <input type="checkbox"/> PERALTE INVERTIDO 12. <input type="checkbox"/> SIN CARRIL LENTO 13. <input type="checkbox"/> OTRA 14. <input type="checkbox"/> NINGUNA			36. SEÑALIZACIÓN DE PELIGRO <input type="checkbox"/> EXISTENTE <input type="checkbox"/> INEXISTENTE <input type="checkbox"/> INNECESARIA (no hay peligro)			37. ACERAS <input type="checkbox"/> SI <input type="checkbox"/> NO			38. ÁRBOLES <input type="checkbox"/> SIN FILA FILA A <input type="checkbox"/> m. DEL BORDE DE LA CALZADA			39. VISIBILIDAD DE LA SEÑALIZACIÓN VERTICAL (si existe) 1. <input type="checkbox"/> BUENA 2. <input type="checkbox"/> DEFICIENTE 3. <input type="checkbox"/> NULA		
40. TIPO DE ACCIDENTE 1. Colisión de vehículos en marcha 1.1. <input type="checkbox"/> FRONTAL 1.2. <input type="checkbox"/> FRONTOLATERAL 1.3. <input type="checkbox"/> LATERAL 1.4. <input type="checkbox"/> ALCANCE 1.5. <input type="checkbox"/> MÚLTIPLE O EN CARAVANA 2. Colisión vehículo-obstáculo en calzada 2.1. <input type="checkbox"/> VEHÍCULO ESTACIONADO 2.2. <input type="checkbox"/> VALLA DE DEFENSA 2.3. <input type="checkbox"/> BARRERA DE PASO A NIVEL 2.4. <input type="checkbox"/> OTRO OBJETO O MATERIAL 3. Atropello: 3.1. <input type="checkbox"/> PEATÓN SOSTENIENDO BICICLETA 3.2. <input type="checkbox"/> PEATÓN REPARANDO EL VEHÍCULO 3.3. <input type="checkbox"/> PEATÓN AISLADO O EN GRUPO 3.4. <input type="checkbox"/> CONDUCTOR DE ANIMALES 3.5. <input type="checkbox"/> ANIMAL CONDUCIDO O RESAÑO 3.6. <input type="checkbox"/> ANIMALES SUELTOS 4.1. <input type="checkbox"/> Vuelco en la calzada 4.2. <input type="checkbox"/> Salida de la calzada 4.3. <input type="checkbox"/> CHOQUE CON ÁRBOL O POSTE 4.4. <input type="checkbox"/> CHOQUE CON MURO O EDIFICIO 4.5. <input type="checkbox"/> CHOQUE CON CUNETTA O BORDILLO 4.6. <input type="checkbox"/> OTRO TIPO DE CHOQUE 4.7. <input type="checkbox"/> CON DESPEÑAMIENTO 4.8. <input type="checkbox"/> CON VUELCO 4.9. <input type="checkbox"/> EN LLANO 4.10. <input type="checkbox"/> OTRA 4.11. <input type="checkbox"/> Otro			41. CIRCULACIÓN 1. <input type="checkbox"/> FLUÍDA 2. <input type="checkbox"/> DENSA 3. <input type="checkbox"/> CONGESTIONADA 42. CIRCULACIÓN BAJO MEDIDAS ESPECIALES 1. <input type="checkbox"/> CARRIL REVERSIBLE 2. <input type="checkbox"/> HABILITACION ARCÉN 3. <input type="checkbox"/> OTRA MEDIDA 4. <input type="checkbox"/> NINGUNA MEDIDA																				
43. MATRICULA Y AÑO DE MATRICULACIÓN A <input type="text"/> B <input type="text"/> Dos últimas cifras del año <input type="text"/>			44. MARCA Y MODELO Color <input type="text"/>			45. VEHÍCULO A <input type="text"/>			46. VEHÍCULO B <input type="text"/>			47. DESCRIPCIONES											
48. POSIBLES FACTORES CONCURRENTES (Opinión del agente) 01. <input type="checkbox"/> Distracción 02. <input type="checkbox"/> Inexperiencia del conductor 03. <input type="checkbox"/> Alcohol o drogas 04. <input type="checkbox"/> Cansancio, sueño o enfermedad 05. <input type="checkbox"/> Velocidad inadecuada 06. <input type="checkbox"/> Infacción a norma de circulación 07. <input type="checkbox"/> Estado o condición de la vía 08. <input type="checkbox"/> Estado o condición de la señalización 09. <input type="checkbox"/> Tramo en obras 10. <input type="checkbox"/> Mal estado vehículo 11. <input type="checkbox"/> Avería mecánica 12. <input type="checkbox"/> Meteorología adversa 13. <input type="checkbox"/> Otro factor 14. <input type="checkbox"/> Sin opinión definida																							
SÍMBOLOS A UTILIZAR <input type="checkbox"/> Vehículo de 4 ó más ruedas <input type="checkbox"/> Vehículo de 2 ó 3 ruedas <input type="checkbox"/> Vehículo de tracción animal <input type="checkbox"/> Peatón <input type="checkbox"/> Animal <input type="checkbox"/> Obstáculo en calzada																							
65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81							
1																1							
2																2							
3																3							
4																4							
5																5							
6																6							



## ANEXO 2. MAPA CON LOS MEDIDORES DE IMD EN VALLADOLID

---





# ANEXO 3. ÍNDICE DE TABLAS

---

## INTRODUCCIÓN

### CAPÍTULO 1. MÉTODOS ESTADÍSTICOS UTILIZADOS

Tabla 1. Tabla de contingencia de Caso/Control frente a Expuestos/No expuestos.....	15
Tabla 2. Interpretación del <i>Odds-Ratio</i> .....	18
Tabla 3. Comparación de los resultados de la prueba a evaluar frete al resultado real. ....	29

## PARTE I: OBTENCIÓN DE LA MUESTRA

### CAPÍTULO 2. BASES DE DATOS DE PARTIDA

Tabla 4. Descripción de las variables generales de la base de datos ARENA.....	40
Tabla 5. Descripción de las variables de la base de datos del PIMUSSVA. ....	44

### CAPÍTULO 3. OBTENCIÓN DE LOS CASOS Y LOS CONTROLES

### CAPÍTULO 4. VARIABLES DEL MODELO

Tabla 6. Resumen de variables finales. ....	63
---	----

## PARTE II: ANÁLISIS DE LOS DATOS

### CAPÍTULO 5. ANÁLISIS DESCRIPTIVO DE LAS VARIABLES Y *ODDS-RATIO*

Tabla 7. Tabla de frecuencias de "Tipo de intersección" para los casos. ....	68
Tabla 8. Tabla de frecuencias de "Tipo de intersección para los controles.....	68
Tabla 9. Tabla cruzada de "Tipo de intersección" frente a "Caso o control". ....	69
Tabla 10. <i>Odds-Ratio</i> de la variable "Tipo de intersección". ....	71
Tabla 11. Tabla cruzada de "Radio de curvatura codificado" frente a "Caso o control".....	75
Tabla 12. <i>Odds-Ratio</i> de la variable "Radio de curvatura codificado".....	75
Tabla 13. Tabla cruzada de "Anchura de calzada codificada" frente a "Caso o control". ....	79
Tabla 14. <i>Odds-Ratio</i> de la variable "Ancho de calzada codificado" para curvas e intersecciones. ....	80

Tabla 15. <i>Odds-Ratio</i> de la variable "Ancho de calzada codificado" para las intersecciones. ....	80
Tabla 16. <i>Odds-Ratio</i> de la variable "Ancho de calzada codificado" para las curvas.....	81
Tabla 17. Tabla cruzada de "Ancho de carril codificado" frente a "Caso o control".....	84
Tabla 18. <i>Odds-Ratio</i> de la variable "Ancho de carril codificado" para intersecciones y curvas. ....	85
Tabla 19. <i>Odds-Ratio</i> de la variable "Ancho de carril codificado" para las intersecciones..	86
Tabla 20. <i>Odds-Ratio</i> de la variable "Ancho de carril codificado" para las curvas. ....	86
Tabla 21. Tabla de frecuencias de "Número de carriles" para los casos.....	87
Tabla 22. Tabla de frecuencias de "Número de carriles" para los controles.....	87
Tabla 23. Tabla cruzada de "Número de carriles" frente a "Caso o control".....	88
Tabla 24. <i>Odds-Ratio</i> de la variable "Número de carriles" para intersecciones y curvas. ...	89
Tabla 25. <i>Odds-Ratio</i> de la variable "Número de carriles" para las intersecciones. ....	90
Tabla 26. <i>Odds-Ratio</i> de la variable "Número de carriles" para las curvas. ....	91
Tabla 27. Tabla de frecuencias de "Sentido" para los casos.....	91
Tabla 28. Tabla de frecuencias de "Sentido" para los controles.....	92
Tabla 29. Tabla cruzada de "Sentido" frente a "Caso o control".....	92
Tabla 30. <i>Odds-Ratio</i> de la variable "Sentido" para curvas e intersecciones.....	93
Tabla 31. <i>Odds-Ratio</i> de la variable "Sentido" para las intersecciones.....	94
Tabla 32. <i>Odds-Ratio</i> de la variable "Sentido" para las curvas.....	95
Tabla 33. Tabla cruzada de "Longitud del tramo recto anterior codificada" frente a "Caso o control".....	98
Tabla 34. <i>Odds-Ratio</i> de la variable "Longitud del tramo recto anterior" para curvas e intersecciones.....	99
Tabla 35. <i>Odds-Ratio</i> de la variable "Longitud del tramo recto anterior" para las intersecciones.....	100
Tabla 36. <i>Odds-Ratio</i> de la variable "Longitud del tramo recto anterior" para las curvas.	101
Tabla 37. Tabla cruzada de "Distancia al semáforo anterior codificada" frente a "Caso o control".....	104
Tabla 38. <i>Odds-Ratio</i> de la variable "Distancia al semáforo anterior codificada" para curvas e intersecciones.....	104
Tabla 39. <i>Odds-Ratio</i> de la variable "Distancia al semáforo anterior codificada" para las intersecciones.....	105
Tabla 40. <i>Odds-Ratio</i> de la variable "Distancia al semáforo anterior codificada" para las curvas. ....	105
Tabla 41. Tabla de frecuencias de "Pendiente" para los casos.....	106
Tabla 42. Tabla de frecuencias de "Pendiente" para los controles.....	106

Tabla 43. Tabla cruzada de "Pendiente" frente a "Caso o control" .....	106
Tabla 44. <i>Odds-Ratio</i> de la variable "Pendiente" para curvas e intersecciones.....	107
Tabla 45. <i>Odds-Ratio</i> de la variable "Pendiente" para las intersecciones.....	108
Tabla 46. <i>Odds-Ratio</i> de la variable "Pendiente" para las curvas.....	108
Tabla 47. Tabla de frecuencias de "Prioridad" para los casos.....	108
Tabla 48. Tabla de frecuencias de "Prioridad" para los controles.....	109
Tabla 49. Tabla cruzada de "Prioridad" frente a "Caso o control" .....	109
Tabla 50. <i>Odds-Ratio</i> de la variable "Prioridad" para curvas e intersecciones.....	110
Tabla 51. <i>Odds-Ratio</i> de la variable "Prioridad" para las intersecciones.....	111
Tabla 52. <i>Odds-Ratio</i> de la variable "Prioridad" para las curvas.....	112
Tabla 53. Tabla cruzada de "Prioridad" frente a "Caso o control" para las curvas.....	113

## **CAPÍTULO 6. MODELOS DE REGRESIÓN LOGÍSTICA**

Tabla 54. Resumen de las categorías de referencia para las variables cualitativas del modelo.....	116
Tabla 55. Tabla comparativa de posibles modelos logísticos para curvas e intersecciones.....	117
Tabla 56. Tabla comparativa de posibles modelos logísticos para las intersecciones.....	117
Tabla 57. Tabla comparativa de posibles modelos logísticos para las curvas.....	118
Tabla 58. Tabla resumen con los modelos a comparar con el análisis ROC.....	119

## **CAPÍTULO 7. SELECCIÓN Y VALIDACIÓN DEL MODELO**

Tabla 59. Estimación de contraste ROC para curvas e intersecciones.....	126
Tabla 60. Tabla de las áreas bajo la curva ROC de cada modelo para curvas e intersecciones.....	126
Tabla 61. Estimación de contraste ROC para las intersecciones.....	131
Tabla 62. Tabla de las áreas bajo la curva ROC de cada modelo para las intersecciones.....	132
Tabla 63. Estimación de contraste ROC para las curvas.....	137
Tabla 64. Tabla de las áreas bajo la curva ROC de cada modelo para las curvas.....	138
Tabla 65. Resumen de variables para el modelo final.....	138
Tabla 66. Modelo logístico final para curvas e intersecciones.....	139
Tabla 67. Modelo logístico final para las intersecciones.....	140
Tabla 68. Modelo logístico final para las curvas.....	141
Tabla 69. Tabla cruzada de "Prioridad" frente a "Caso o control" para las curvas.....	142



# ANEXO 4. ÍNDICE DE IMÁGENES

---

## INTRODUCCIÓN

Ilustración 1. Letalidad en vías urbanas en función del tipo de accidente con víctimas. Año 2013. [1] .....	4
---	---

## CAPÍTULO 1. MÉTODOS ESTADÍSTICOS UTILIZADOS

Ilustración 2. Procedimiento para un estudio caso-control prospectivo.....	11
Ilustración 3. Procedimiento para un estudio caso-control retrospectivo. ....	12
Ilustración 4. Procedimiento para un estudio de cohorte prospectivo. ....	12
Ilustración 5. Procedimiento para un estudio de cohorte retrospectivo.....	13
Ilustración 6. Leyenda para los esquemas de los procedimientos de los distintos diseños de estudios observacionales.....	13
Ilustración 7. Esquema de casos y controles para el estudio.....	19
Ilustración 8. Ajuste de una variable dependiente binaria con un modelo de regresión lineal [28]. ....	21
Ilustración 9. Tipo de función necesaria para la regresión con una variable dependiente binaria [28].....	22
Ilustración 10. Descripción Curva ROC. ....	31

## PARTE I: OBTENCIÓN DE LA MUESTRA

### CAPÍTULO 2. BASES DE DATOS DE PARTIDA

### CAPÍTULO 3. OBTENCIÓN DE LOS CASOS Y LOS CONTROLES

Ilustración 11. Esquema de la obtención de los casos.....	50
Ilustración 12. Localización de los casos en la cartografía de Valladolid.....	51
Ilustración 13. Esquema de la cuadrícula para la obtención de los controles.....	52
Ilustración 14. Formato para la selección de los controles en la cartografía en AutoCAD. ...	53
Ilustración 15. Localización de los controles en la cartografía de Valladolid. ....	54
Ilustración 16. Localización de los casos y los controles en la cartografía de Valladolid.....	54

## CAPÍTULO 4. VARIABLES DEL MODELO

Ilustración 17. Ejemplo de intersección en Y.....	56
Ilustración 18. Ejemplo de intersección en T.....	56
Ilustración 19. Ejemplo de intersección en X.....	57
Ilustración 20. Ejemplo de intersección en cruz.....	57
Ilustración 21. Ejemplo de glorieta partida.....	57
Ilustración 22. Esquema para la medida del radio de curvatura. ....	58
Ilustración 23. Ejemplo de medidor de IMD en una intersección. ....	61

## PARTE II: ANÁLISIS DE LOS DATOS

### CAPÍTULO 5. ANÁLISIS DESCRIPTIVO DE LAS VARIABLES Y ODDS-RATIO

Ilustración 24. Gráfico de barras agrupadas de "Tipo de intersección" respecto a "Caso o control".....	70
Ilustración 25. Histograma de "Radio de curvatura" para los casos.....	72
Ilustración 26. Histograma de "Radio de curvatura" para los controles. ....	72
Ilustración 27. <i>Box-plot</i> múltiple de "Radio de curvatura" frente a "Caso o control".....	73
Ilustración 28. Gráfico de barras de la variable "Radio de curvatura codificado".....	74
Ilustración 29. Histograma de "Ancho de calzada" para los casos. ....	76
Ilustración 30. Histograma de "Ancho de calzada" para los controles.....	77
Ilustración 31. <i>Box-plot</i> múltiple de "Ancho de calzada" frente a "Caso o control". ....	77
Ilustración 32. Gráfico de barras de la variable "Anchura de calzada codificada".....	78
Ilustración 33. Histograma de "Ancho de carril" para los casos.....	82
Ilustración 34. Histograma de "Ancho de carril" para los controles. ....	82
Ilustración 35. <i>Box-plot</i> múltiple de "Ancho de carril" frente a "Caso o control".....	83
Ilustración 36. Gráfico de barras de la variable "Ancho de carril codificado".....	84
Ilustración 37. Gráfico de barras agrupadas de "Número de carriles" frente a "Caso o control".....	89
Ilustración 38. Gráfico de barras agrupadas de "Sentido" frente a "Caso o control". ....	93
Ilustración 39. Histograma de "Longitud del tramo recto anterior" para los casos.....	95
Ilustración 40. Histograma de "Longitud del tramo recto anterior" para los controles.....	96
Ilustración 41. <i>Box-plot</i> múltiple de "Longitud del tramo recto anterior" frente a "Caso o control". ....	96

Ilustración 42. Gráfico de barras de la variable "Longitud del tramo recto anterior codificada".....	98
Ilustración 43. Histograma de "Distancia al semáforo anterior" para los casos. ....	101
Ilustración 44. Histograma de "Distancia al semáforo anterior" para los controles. ....	102
Ilustración 45. <i>Box-plot</i> múltiple de "Distancia al semáforo anterior" frente a "Caso o control". ....	102
Ilustración 46. Gráfico de barras de la variable "Distancia al semáforo anterior codificada". ....	103
Ilustración 47. Gráfico de barras agrupadas de "Pendiente" frente a "Caso o control". ....	107
Ilustración 48. Gráfico de barras agrupadas de "Prioridad" frente a "Caso o control". ....	110

## **CAPÍTULO 6. MODELOS DE REGRESIÓN LOGÍSTICA**

### **CAPÍTULO 7. SELECCIÓN Y VALIDACIÓN DEL MODELO**

Ilustración 49. Curva ROC del modelo completo para curvas e intersecciones. ....	123
Ilustración 50. Curva ROC del modelo intermedio para curvas e intersecciones ....	124
Ilustración 51. Curva ROC del modelo hacia delante para curvas e intersecciones.....	125
Ilustración 52. Comparación de curvas ROC para curvas e intersecciones. ....	127
Ilustración 53. Curva ROC del modelo completo para las intersecciones. ....	128
Ilustración 54. Curva ROC del modelo intermedio para las intersecciones.....	130
Ilustración 55. Curva ROC del modelo hacia atrás para las intersecciones. ....	131
Ilustración 56. Comparación de los tres modelos para las intersecciones. ....	132
Ilustración 57. Curva ROC del modelo completo para las curvas. ....	134
Ilustración 58. Curva ROC del modelo intermedio para las curvas.....	135
Ilustración 59. Curva ROC del modelo hacia delante para las curvas. ....	136
Ilustración 60. Comparación de los tres modelos para las curvas. ....	137



# ANEXO 5. ÍNDICE DE ECUACIONES

---

## INTRODUCCIÓN

### CAPÍTULO 1. MODELOS ESTADÍSTICOS UTILIZADOS

Ecuación 1. Odds de exposición en casos. ....	16
Ecuación 2. Odds de exposición en controles. ....	16
Ecuación 3. Odds-Ratio. ....	16
Ecuación 4. Intervalo de confianza para el Odds-Ratio.....	16
Ecuación 5. Chi cuadrado de Haenszel-Mantel. ....	17
Ecuación 6. Transformación de Odds-Ratio en probabilidad.....	17
Ecuación 7. Modelo de probabilidad lineal.....	20
Ecuación 8. Función logística.....	23
Ecuación 9. Modelo logístico. ....	23
Ecuación 10. Definición de <i>odds</i> . ....	24
Ecuación 11. Definición del modelo Logit.....	24
Ecuación 12. Definición de Odds-Ratio en regresión logística. ....	25
Ecuación 13. Sensibilidad.....	29
Ecuación 14. Especificidad.....	30
Ecuación 15. Valor predictivo del resultado positivo.....	30
Ecuación 16. Valor predictivo del resultado negativo. ....	30
Ecuación 17. Valor global de resultados válidos. ....	30

## PARTE I: OBTENCIÓN DE LA MUESTRA

### CAPÍTULO 2: BASES DE DATOS DE PARTIDA

### CAPÍTULO 3: OBTENCIÓN DE LOS CASOS Y LOS CONTROLES

### CAPÍTULO 4: VARIABLES DEL MODELO

Ecuación 18. Medida del radio de curvatura [28].....	58
Ecuación 19. Intensidad Media Diaria (IMD).....	61