

Grado en Estadística

Título: Uso de modelos lineales generalizados para analizar los datos de la Macroencuesta de Violencia contra la Mujer 2019

Autor: Maria Freixedas Morató

Director: Klaus Langohr

Departamento: Departamento de Estadística e Investigación Operativa

Convocatoria: Junio 2021



UNIVERSITAT DE
BARCELONA



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Facultat de Matemàtiques i Estadística



UNIVERSITAT DE
BARCELONA



UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Facultat de Matemàtiques i Estadística

TRABAJO DE FIN DE GRADO

Uso de modelos lineales generalizados para analizar los datos de la
Macroencuesta de Violencia contra la Mujer 2019

Grado en Estadística

Autor: Maria Freixedas Morató

Director: Klaus Langohr

Departamento: Departamento de Estadística
e Investigación Operativa

Convocatoria: Junio 2021

Agradecimientos

Agradecer a Klaus Langohr todo el empeño y dedicación que me ha ofrecido durante este último semestre además de todos los conocimientos aportados. A mis padres y hermano por ayudarme en los momentos más complicados y compartir los más felices.

Resumen

El objetivo de este trabajo es conocer cuáles eran las características más destacadas de las mujeres que sufrieron violencia de género por parte de su pareja en el año 2019 en España, así como las características la de los agresores. Para ello, se han utilizado distintos modelos estadísticos.

Los modelos estadísticos son una forma de aproximación a la realidad, ya que son herramientas utilizadas para observar la asociación que hay entre las variables que se quieren estudiar y el fenómeno de interés. También son útiles para hacer predicciones mediante formulas o ecuaciones matemáticas. La finalidad principal de los modelos de regresión logística y los modelos log-binomiales es ver cómo influyen las variables regresoras en la probabilidad de ocurrencia de un suceso. Los modelos Poisson cero-inflados son empleados en modelos de recuentos en los que los datos muestran más ceros de los que deberían.

Palabras clave: Modelos estadísticos, Modelos de regresión logística, Modelos log-binomiales, Modelos Poisson cero-inflados, Violencia de género.

Resum

L'objectiu d'aquest treball és conèixer quines eren les característiques més destacades de les dones que van patir violència de gènere per part de la seva parella en l'any 2019 a Espanya, així com les característiques dels agresors. Per a aixó, s'han utilitzat diferents models estadístics. Els models estadístics són una forma d'aproximació a la realitat, ja que es tracten d'eines utilitzades per observar l'associació que hi ha entre les variables que es volen estudiar i, el fenomen d'interès. També són útils per fer prediccions mitjançant fórmules o equacions matemàtiques. La finalitat principal dels models de regressió logística i els models log-binomials és veure com influeixen les variables regressores en la probabilitat de succés d'un esdeveniment. Els models Poisson zero-inflats, són utilitzats en models de recomptes en els quals les dades mostren més zeros haurien de mostrar.

Paraules Clau: Models estadístics, Models de regressió logística, Models log-binomials, Models Poisson zero-inflats, Violència de gènere.

Abstract

The objective of this work is to find out which were the most outstanding characteristics of women who suffered gender violence by their partner in the year 2019 in Spain, as well as the characteristics of the aggressors. For this, different statistical models have been used.

Statistical models are a way of approaching reality, since they are tools used to observe the association between the variables to be studied and, the phenomenon of interest. They are useful to make predictions using mathematical formulas or equations. The main purpose of logistic regression models and log-binomial models is to see how the regressing variables influence the probability of occurrence of an event. The zero-inflated Poisson models are used to count models in which the data show more zeros than they should.

Keywords: Statistical models, Logistic regression models, Log-binomial models, Zero-inflated Poisson models, Gender violence.

Índice general

1	Introducción	1
1.1	Motivación	1
1.2	Violencia de género	1
1.2.1	Violencia física	2
1.2.2	Violencia psicológica	3
1.2.3	Violencia sexual, abusos sexuales y acoso sexual	3
1.2.4	Violencia económica	3
1.3	Objetivo del trabajo	3
2	Gestión de la base de datos	5
2.1	Origen de los datos	5
2.2	Lectura de los datos	5
2.3	Variables de estudio	6
3	Modelos estadísticos	9
3.1	Modelo de regresión logística	9
3.1.1	Interpretación de los parámetros	10
3.1.2	Bondad de ajuste	11
3.1.3	Implementación en R	11
3.2	Modelo de regresión log-binomial	12
3.2.1	Interpretación de los parámetros	13
3.2.2	Bondad de ajuste	13
3.2.3	Implementación en R	13
3.3	Modelo Poisson cero-inflado	14
3.3.1	Interpretación de los parámetros	15
3.3.2	Bondad de ajuste	15
3.3.3	Implementación en R	15
4	Resultados	16
4.1	Análisis descriptivo	16
4.1.1	Parejas pasadas	16
4.1.2	Parejas actuales	23

4.2	Modelos estadísticos	31
4.2.1	Modelo de regresión logística	31
4.2.2	Modelo de regresión Log-binomial	35
4.2.3	Modelo Poisson cero-inflado	38
5	Discusión y conclusiones	44
	Bibliografía	48
A	Cuestionario Macro-encuesta de Violencia de Género contra la Mujer	49
B	Gráficos mencionadas	54
C	Bondad de ajuste de los modelos estadísticos	57
D	Código R	59

Índice de figuras

3.1	Función <i>glm</i> del paquete <i>stats</i> en R	12
3.2	Función <i>anova</i> del paquete <i>stats</i> en R	12
3.3	Función <i>zeroinfl</i> de la librería <i>pscl</i> en R	15
4.1	Gráfico de frecuencias del tipo de violencia Control Social	18
4.2	Gráfico de frecuencias del tipo de violencia Insultos o Humillaciones	19
4.3	Gráfico de frecuencias del tipo de Violencia Física	19
4.4	Gráfico de frecuencias del tipo de Violencia Sexual	20
4.5	Gráfico de mosaico del tamaño de municipio según si se ha sufrido o no violencia	21
4.6	Boxplot de la edad según si se ha sufrido o no violencia	21
4.7	Gráfico de mosaico de la situación laboral según si se ha sufrido o no violencia	22
4.8	Gráfico de mosaico del nivel de estudios según si se ha sufrido o no violencia	22
4.9	Gráfico de frecuencias del tipo de violencia Control Social	24
4.10	Gráfico de frecuencias del tipo de violencia Insultos o Humillaciones	25
4.11	Gráfico de frecuencias del tipo de Violencia Física	25
4.12	Gráfico de frecuencias del tipo de Violencia Sexual	26
4.13	Gráfico de mosaico del tamaño de municipio según si se ha sufrido o no violencia	27
4.14	Boxplot de la edad según si se ha sufrido o no violencia	27
4.15	Gráfico de mosaico de la situación laboral según si se ha sufrido o no violencia	28
4.16	Gráfico de mosaico del nivel de estudios según si se ha sufrido o no violencia	28
4.17	Boxplot de la edad de la pareja actual según si se ha ejercido o no violencia	29
4.18	Gráfico de mosaico del nivel de estudios de la pareja según si se ha ejercido o no violencia	30
4.19	Gráfico de mosaico de la situación laboral según si se ha sufrido o no violencia	30
4.20	<i>Odds Ratio</i> del modelo de regresión logística estimado	34
4.21	Riesgo Relativo del modelo de regresión Log-binomial	37
4.22	Función <i>dispersiontest</i> de la librería <i>AER</i> en R	38

Índice de tablas

2.1	VARIABLES DE INTERÉS PARA EL ESTUDIO	7
2.2	VARIABLES CORRESPONDIENTES AL TIPO DE VIOLENCIA	8
3.1	MATRIZ DE CONFUSIÓN	11
4.1	TABLA DE CONTINGENCIA ENTRE CONTROL SOCIAL E INSULTOS O HUMILLACIONES	17
4.2	TABLA DE CONTINGENCIA ENTRE INSULTOS O HUMILLACIONES Y VIOLENCIA FÍSICA	17
4.3	TABLA DE CONTINGENCIA ENTRE VIOLENCIA FÍSICA Y VIOLENCIA SEXUAL	18
4.4	VARIABLES DE INTERÉS PARA EL ESTUDIO DE PAREJAS PASADAS	20
4.5	TABLA DE CONTINGENCIA ENTRE CONTROL SOCIAL E INSULTOS O HUMILLACIONES	23
4.6	TABLA DE CONTINGENCIA ENTRE INSULTOS O HUMILLACIONES Y VIOLENCIA FÍSICA	23
4.7	TABLA DE CONTINGENCIA ENTRE VIOLENCIA FÍSICA Y VIOLENCIA SEXUAL	24
4.8	ANOVA PARA EL MODELO DE REGRESIÓN LOGÍSTICA PARA MUJERES QUE HAN SUFRIDO VIOLENCIA DE GÉNERO CON FUNCIÓN <i>glm</i>	32
4.9	MODELO DE REGRESIÓN LOGÍSTICA PARA MUJERES QUE HAN SUFRIDO VIOLENCIA DE GÉNERO CON FUNCIÓN <i>glm</i>	33
4.10	MODELO DE REGRESIÓN LOG-BINOMIAL PARA MUJERES QUE HAN SUFRIDO VIOLENCIA DE GÉNERO CON FUNCIÓN <i>glm</i>	36
4.11	MODELO DE POISSON CERO-INFLADO PARA MUJERES QUE HAN SUFRIDO VIOLENCIA DE GÉNERO CON FUNCIÓN <i>zeroinfl</i> CON LINK <i>logit</i> Y DISTRIBUCIÓN BINOMIAL	40
4.12	MODELO DE POISSON CERO-INFLADO PARA MUJERES QUE HAN SUFRIDO VIOLENCIA DE GÉNERO CON FUNCIÓN <i>zeroinfl</i> CON LINK <i>log</i> Y DISTRIBUCIÓN POISSON	42

Capítulo 1

Introducción

1.1. Motivación

La violencia de género es un tipo de violencia que lleva décadas afectando a mujeres en todo el mundo. Aun hoy en día, en pleno siglo XXI, sigue habiendo titulares de mujeres que han perdido la vida por esta causa: el simple hecho de ser mujer. Aun cuando se cree que ya se ha avanzado mucho sobre este tema y se espera que la sociedad ya no sea tan machista, la mayoría de personas tienen a su alrededor alguna persona que ha sufrido algún tipo de relación denominada tóxica, es decir, que ha sufrido control social, violencia psicológica y en casos más graves incluso violencia sexual o violencia física.

Después de la gran ola de feminismo que se está viviendo hoy en día por la igualdad de la mujer y, gracias a movimientos muy conocidos como fue el #metoo; muchas mujeres se han sentido fuertes para denunciar casos de violencia de género que han sufrido a lo largo de su vida. Aun así, se suele decir que la mayoría de agresiones que viven las mujeres son ejercidas por parte de su pareja o bien, por personas cercanas a ellas.

En el trabajo que se presenta a continuación se intentará describir qué características tenían las mujeres de España en el año 2019 que sufrieron algún tipo de violencia por parte de su pareja, así, se pretenderá analizar que factores hacen más vulnerable a una mujer al hecho de sufrir algún tipo de violencia de género ocasionada por su pareja.

1.2. Violencia de género

La violencia de género se ha constituido como un fenómeno invisible durante años, siendo una de las manifestaciones más claras de desigualdad, subordinación y de las relaciones de poder de los hombres sobre las mujeres. Ya en 1980, la II Conferencia Mundial sobre la Condición Jurídica y Social de la Mujer, establecía que la violencia contra las mujeres suponía el crimen más silenciado del mundo [1].

Es conocido como violencia de género cualquier acto violento o agresión, que se produce en una situación de desigualdad en un marco de un sistema de relaciones de dominación de los

hombres sobre las mujeres, que tenga o pueda tener como consecuencia un daño físico, sexual o psicológico. Este tipo de definiciones se aplican en casos en los que el género es la base de la violencia ejercida; es por eso que en este marco de violencia de género también se encuentra la violencia hacia personas del colectivo LGTB (lesbianas, gays, bisexuales, transgéneros y personas que no se ajustan a la norma heterosexual o a las categorías binarias de género tradicionales).

La violencia de género es una violación a los derechos humanos, no afecta solo a la salud física, sino que también afecta a la salud mental y puede traer consecuencias como autolesiones, aislamiento, depresión e incluso intentos de suicidio por parte de la víctima. La violencia de género es discriminación, arraigada a estereotipos y prejuicios dañinos contra las mujeres u otras personas que no encajan en la sociedad tradicional cis-hetero normativa. Esta es un obstáculo para llegar a la igualdad de género, la igualdad de derechos para todas las personas independientemente de su género, la igualdad a la visibilidad y a las mismas oportunidades de apoderamiento, responsabilidad y participación.

Hay factores culturales que legitiman la violencia de género tales como los estereotipos y prejuicios de género, las expectativas normativas de feminidad y masculinidad, la socialización del género, el conocimiento de la familia como privada y bajo la autoridad masculina.

Las tradiciones religiosas e históricas han sancionado el castigo físico de la mujer bajo la noción de derechos y propiedad de la mujer. Esta propiedad, a su vez, legitima el control sobre la sexualidad de las mujeres.

Ser víctima de violencia de género hoy en día aún se percibe en muchas culturas y sociedades como algo vergonzoso, por lo que sentirse débil, e incluso, muchas mujeres se sienten culpables de atraer la violencia contra ellas mismas a través de su comportamiento.

Aunque en la mayoría de países muchas de las formas de violencia de género están criminalizadas, las prácticas de la aplicación de la ley, en muchos casos, favorecen a los perpetradores, factor que ayuda a explicar los bajos niveles de confianza en las autoridades públicas y el bajo nivel de denuncia o simplemente, de manifestación a alguna persona cercana [2].

Tal y como describe la consejería de empleo e igualdad de la Xunta de Galicia [3] la violencia de género se puede clasificar en distintos tipos.

1.2.1. Violencia física

Este tipo de violencia incluye cualquier acto de fuerza contra el cuerpo de una mujer, con el efecto de producir un daño o una lesión física. Estos actos pueden ser desde un pellizco hasta una provocación de aborto. Aunque sea el tipo de maltrato más evidente, algunas veces no tiene por qué tener efectos visibles en el cuerpo.

Las consecuencias que pueden tener estas conductas para las mujeres pueden ser pánico, miedo, ansiedad, que puede llegar a ser extrema dado que la violencia se produzca de forma impredecible, humillación, vergüenza... [4]

1.2.2. Violencia psicológica

Abarca toda conducta, verbal o no verbal, que produzca desvalorización o sufrimiento, tales como: insultos “estás loca”, “ignorante”, “atrasada”, ...); menosprecios (“no sirves para nada”, “mala madre”, “eres incapaz de hacer algo bien”, ...); intimidaciones/amenazas (“te voy a matar”, “te voy a quitar a los niños”, “te voy a hundir”...); abuso de la autoridad (con acciones como: revisar el correo o tus cosas personales, ...); faltas de respeto (no respeta las necesidades, sentimientos, opiniones, manipula las palabras, te desautoriza delante de amigos, ...); exige obediencia (no deja estudiar, no deja que trabaje fuera de casa, le controla como se viste, ...) y un gran etcétera.

Este tipo de maltrato, a diferencia del anterior, es el más difícil de detectar, dado que es de gran sutileza. Aun así su duración a lo largo del tiempo perjudica de forma grave la estabilidad emocional, la autoestima y la personalidad de la víctima. Este sentimiento puede llevar a la mujer a sentirse culpable y a incrementar su compromiso en la relación dirigiendo a esta a los objetivos del maltratador. De tal forma que provoca un aislamiento, incomunicación, ruptura del apoyo social del entorno de la víctima y, una dependencia con el maltratador.

1.2.3. Violencia sexual, abusos sexuales y acoso sexual

La violencia y abusos sexuales se tratan de cualquier acto de naturaleza sexual forzada por el agresor o bien, no consentida por la víctima por medio de la fuerza o la intimidación.

En cambio el acoso sexual implica conductas que consisten en la solicitud de favores de naturaleza sexual, en las que el agresor se favorezca de una situación de superioridad tanto laboral, como docente o distinta de estas. Con la manifestación tanto verbal como táctica a la víctima relacionada con las expectativas que tenga está en el ámbito de la relación o bajo la promesa de una recompensa.

1.2.4. Violencia económica

Privación intencionada y no justificada legalmente de recursos para el bienestar físico o psicológico de la mujer y de sus hijos e hijas. El maltratador considera que la mujer no es competente y no administra bien el dinero y lo gasta en cosas innecesarias por lo que esta no puede tomar decisiones sobre el destino del dinero. Esto lleva a la víctima a una dependencia económica y a una carencia de recursos para tomar iniciativas para salir de la situación.

1.3. Objetivo del trabajo

El objetivo de este trabajo es saber cuáles eran las características más destacadas de esas mujeres que sufrieron violencia de género por parte de su pareja en el año 2019, año en el que se realizó la Macro-encuesta de Violencia de Género contra la Mujer. Estos datos son los que se analizarán en este trabajo final de grado.

Para llevar a cabo este estudio se realizarán tres tipos de modelos de regresión:

Primero de todo se realizará la construcción de un modelo de regresión logística, utilizado para variables respuesta distribuidas de forma binomial. El cual permite una estimación de los *odds ratio* asociadas a una variable de interés, en el caso de este estudio, si las mujeres sufrieron o no violencia de género.

El segundo modelo que se utilizará será el modelo de regresión log-binomial, este es una alternativa al modelo anterior para estudios de cohorte y transversales, es decir, estudios que analizan variables recopiladas en un período de tiempo sobre una muestra o población; en este caso, en el año de 2019 y, sobre las mujeres de entre 16 y 65 o más años de edad, de España.

Finalmente, se realizará un modelo de regresión de Poisson. Modelo en el que a diferencia de los casos anteriores, la variable respuesta es una variable de conteo o recuento con el número de sucesos. Por consiguiente, en este estudio esta variable se tratará del número de veces que ha sufrido violencia de género la mujer entrevistada. Para este caso, se observará si es posible la ejecución del modelo Poisson cero-inflado, utilizado en casos en los que se observan un gran número de ceros en la variable respuesta.

El presente Trabajo de Final de Grado se estructura de la siguiente forma:

1. **Gestión de la base de datos de la Macro-encuesta de Violencia de Género de España del año 2019:** Lectura de la base de datos, presentación de las variables de los modelos y, análisis estadístico de estas.
2. **Construcción, validación y capacidad predictiva de los modelos:** Construcción y descripción de los tres modelos finales descritos anteriormente, validación y, análisis de los residuos.
3. **Resultados:** Resultados obtenidos a partir de las estimaciones de los distintos modelos y, análisis de estos.
4. **Discusión y conclusiones:** Conclusiones a la que se han llegado después de la realización del estudio.

Capítulo 2

Gestión de la base de datos

2.1. Origen de los datos

La base de datos que se ha utilizado para la realización de este Trabajo de Final de Grado proviene del Centro de Investigaciones Sociológicas, *CIS*, un organismo autónomo dependiente del Ministerio de la Presidencia, cuya función principal es la de contribuir al conocimiento científico de la sociedad española.

Este dispone de un Banco de Datos a través del cual se ha obtenido la base de datos utilizada para este trabajo. Para la obtención de la base de datos se ha tenido que llenar una serie de cuestiones en las que se declaraba responsable de la explotación de estos datos, con el fin que iba a explotarlos, y que no se haría divulgación de estos sin tener un consentimiento previo de la CIS.

Después de firmar el consentimiento mencionado anteriormente, fue facilitada una carpeta comprimida que no solo contenía la base de datos sino, que también contenía 4 documentos *PDF*. El primero, de 67 páginas en el que se definen todas los códigos de la variables. El segundo de 43 páginas con el cuestionario que se llevó a cabo en el que además, vienen diferentes apuntes sobre cómo responder a las preguntas que pueden realizar las mujeres entrevistadas. El tercero, con unas tablas con los perfiles de las mujeres que habían sido entrevistada. Y finalmente, un documento con las diferentes tarjetas que se mostraban a la persona entrevistada.

2.2. Lectura de los datos

Por otra parte, la base de datos (que venía dentro de la carpeta comprimida) venía en formato *.sav*, formato para ser leído a través del programa *SPSS* (<https://www.ibm.com/products/spss-statistics>), el programa que se ha utilizado para la elaboración de este trabajo ha sido *R* y el entorno de trabajo ha sido *R-Studio* (<https://www.rstudio.com/>). Para poder leer esta base de datos a través de la herramienta *R-studio*, se ha utilizado la función *read_spss()* del paquete *haven* (Hadley Wickham and Evan Miller, 2020) [5].

2.3. Variables de estudio

La base de datos que se ha utilizado para la realización de este trabajo consta de 1121 variables y 9568 observaciones. Tratándose de una base de datos de tales dimensiones, se ha tenido que hacer un análisis previo de estas variables para seleccionar cuáles serían de utilidad para este estudio y, por lo tanto, qué variables se incluirían en el modelo.

Para empezar a descartar algunas variables que no serían necesarias con los objetivos que se habían planteado para este estudio, se realizó una lectura detallada del primero de los documentos *PDF* que se han descrito en el apartado anterior. Después de esta lectura detallada, se observó que la base de datos está dividida en cuatro grandes secciones. La primera, en la que se recopilan datos sobre los tipos de violencia y situaciones que ha podido vivir la víctima de violencia de género por parte de su pareja actual. En la segunda sección, preguntas similares a las de la sección anterior pero, esta vez, por violencia ejercida por parte de parejas pasadas que había tenido la víctima. Seguidamente, se detallan situaciones y tipos de violencia que ha podido sufrir la mujer entrevistada fuera del ámbito de pareja. Y por último se detallan episodios de acoso sexual que han podido sufrir las distintas víctimas.

De esta forma y, tal y como se ha explicado en apartados anteriores, se decidió tener en cuenta para el modelo solo aquellas mujeres que tenían pareja en el momento en el que se realizó la macro-encuesta, ya que en el caso de parejas pasadas no se podrá determinar en que momento de su vida sufrieron violencia ni por el número de parejas que la han sufrido. De esta manera se realizará un estudio transversal o de cohorte, para determinar las características de esas mujeres que estaban sufriendo violencia en el año 2019 por parte de su agresor, en este caso, su pareja.

Aunque solo se usaría en los modelos aquella población que tenía pareja en el momento del estudio, por los motivos que se han explicado anteriormente, tanto por el caso de parejas actuales como para el caso de parejas pasadas, las preguntas referidas al tipo de violencia y las situaciones que se pueden vivir dentro de esta, es el mismo, se decidió hacer un pequeño análisis estadístico para ver el comportamiento de los datos para las mujeres que habían sufrido alguna vez violencia por parte de alguna de sus parejas pasadas ya que, podría resultar de gran interés.

Después de descartar las dos últimas secciones de la base de datos, mencionadas anteriormente, se hizo un descarte de variables dentro del caso de violencia ejercida por la pareja actual y por las parejas pasadas. Las variables que se decidieron estudiar finalmente fueron las que se presentan en la Tabla 2.1.

Para la construcción de los modelos estadísticos es importante tener en cuenta que se han tratado como valores *missings* aquellas respuestas que no daban información a la respuesta, de esta forma los factores que se han eliminado han sido 8-N.S , 9-N.C, 0-N.P.

Variable	Descripción
TAMUNI	Tamaño del municipio
M0P0	Situación laboral de la entrevistada
SDE1	Edad de la persona entrevistada
ESTUDIOS	Nivel de estudios de la persona entrevistada
SDP2	Edad de la pareja actual
ESTUDIOSPAR	Nivel de estudios de la pareja actual
SDP6	Situación laboral de la pareja actual
SDP1	Sexo de la pareja actual
M0P17	Situación legal con la pareja actual
M0P16	Convivencia con la pareja actual
M0P18	Tiempo de relación con la pareja actual

Tabla 2.1: Variables de interés para el estudio

Cabe destacar que dentro de cada una de las secciones, que se han explicado al inicio de este apartado, se encuentran diferentes variables referidas al tipo de violencia de género. Cada uno de estos tipos de violencia estaba abarcado por distintas variables donde representaban alguna situación por la que podría haber pasado la víctima. Para poder proceder con un análisis centrado en el tipo de violencia y cómo afectan las variables descritas anteriormente a esta, se procedió a la recodificación de 8 distintas variables tanto para el caso de parejas actuales, como el caso de parejas pasadas.

Cuatro de las cuales representan un conteo de las distintas situaciones que ha podido vivir la víctima dentro de cada tipo de violencia. Las cuatro restantes, representan una variable binaria donde cero significa que no ha sufrido ninguna de las situaciones descritas y, uno que ha sufrido alguna de las situaciones descritas.

Estas variables fueron creadas uniendo todas aquellas variables que representaban las distintas situaciones que pudo haber vivido la mujer entrevistada es decir, en el caso de las variables tipo conteo, se utilizó la función `rowCounts()` del paquete `matrixStats` (Henrik Bengtsson, 2020) [6] para contar cuantas de las situaciones descritas había vivido la mujer. Para aquellas variables binarias se creó una función (*binaria*) que transforma las variables tipo conteo en variables dicotómicas, de tal forma que si se trataba de un cero dejaba el cero y si se trataba de un número superior lo sustituía por un uno.

Finalmente, se crearon dos variables más, una tipo conteo y una binaria, que representa la violencia total que ha sufrido la víctima, es decir se realizó una recodificación a partir de las variables mencionadas anteriormente.

Estas variables se muestran en la Tabla 2.2.

Variable	Descripción
control.social	Control social tipo conteo
insultos	Insultos y Humillaciones tipo conteo
viol.fisica	Violencia Fisica tipo conteo
viol.sexual	Violencia Sexual tipo conteo
control.social.b	Control social dicotómica
insultos.b	Insultos y Humillaciones dicotómica
viol.fisica.b	Violencia Fisica dicotómica
viol.sexual.b	Violencia Sexual dicotómica
violencia.b	Violencia total dicotómica
violencia.n	Violencia total tipo conteo

Tabla 2.2: Variables correspondientes al tipo de violencia

Capítulo 3

Modelos estadísticos

En este capítulo se tratará de explicar los modelos estadísticos que se utilizarán en este trabajo para analizar qué características de la víctima y del agresor, están asociadas con la violencia de género. Para realizar la explicación de estos modelos se seguirá la información del Trabajo de Final de Grado de Laura Julià (2020) [7] y la Tesis de Celia Mercedes Salcedo Poma del estudio “*Estimación de la ocurrencia de incidencias en declaraciones de pólizas de importación*” [8].

Los modelos estadísticos son una forma de aproximación a la realidad, ya que son herramientas utilizadas para observar la asociación que hay entre las variables que se quieren estudiar, variables explicativas, y el fenómeno de interés, variable respuesta. Además, también son útiles para hacer predicciones mediante fórmulas o ecuaciones matemáticas.

Los modelos lineales generalizados están constituidos por un componente aleatorio que se trata del vector de valores que toma la variable respuesta Y , un componente sistemático representado por η y una función de enlace que se relaciona con el predictor lineal, con el valor esperado de la respuesta condicionado al valor que toman las variables explicativas que se expresa como $g(\mu)$ donde μ es el valor esperado.

$$g(\mu_i) = \eta_i = \alpha + \beta_1 X_1 + \dots + \beta_m X_m$$

En este trabajo tan solo se estudiará la asociación que hay entre la variable respuesta y las variables de interés. Se utilizarán tres modelos estadísticos para ver dicha asociación entre variables. Los modelos que se estudiarán son el modelo de regresión logística, el modelo log-binomial y el modelo Poisson cero-inflado.

3.1. Modelo de regresión logística

El principal objetivo de los modelos de regresión logística es ver cómo influyen las variables regresoras en la probabilidad de ocurrencia de un suceso.

La variable respuesta o dependiente no se trata de una variable continua, sino de una variable binaria, donde generalmente toma valores 0 o 1; toma el valor 0 en caso de no ocurrencia del

suceso que se desea estudiar y, valor 1 en caso contrario. Las variables explicativas pueden ser cualitativas o cuantitativas.

Para poder presentarlo en forma lineal se realiza una sencilla transformación logarítmica, aquí es donde aparece la función de enlace *logit* utilizada para modelos de regresión logística. La expresión de un modelo de regresión logística vendría dada por:

$$\text{logit}(p_x) = \log\left(\frac{p_x}{1-p_x}\right) = \alpha + \beta_1 X_1 + \dots + \beta_m X_m$$

Esta expresión es equivalente a:

$$p_x = \frac{\exp(\alpha + \beta_1 X_1 + \dots + \beta_m X_m)}{1 + \exp(\alpha + \beta_1 X_1 + \dots + \beta_m X_m)}, \quad (3.1)$$

donde p_x es la probabilidad condicional para que la variable respuesta sea igual a 1. Es decir:

$$p_x = P(Y = 1|\mathbf{X}) = P(Y = 1|X_1, \dots, X_m)$$

La estimación de los parámetros en el modelo de regresión logística se realiza basándose en el criterio de máxima verosimilitud. Sea (y_i, x_i) con $i = 1, \dots, n$ una muestra de observaciones independientes, se puede expresar la función de máxima verosimilitud como:

$$L(\alpha, \beta|Y, \mathbf{X}) = \prod_{i=1}^n \frac{\exp(\alpha + \beta' \mathbf{x}_i)^{\delta_i}}{1 + \exp(\alpha + \beta' \mathbf{x}_i)}$$

donde $\beta' = (\beta_1, \dots, \beta_m)'$ y $\delta_i = 1$ si $Y_i = 1$ y 0 en caso contrario.

En el caso de usar variables regresoras categóricas con n categorías, será necesario incorporar $n-1$ variables indicadoras. Por lo tanto, el modelo *logit* para un modelo con m variables regresoras y una variable categórica sería:

$$\text{logit}(p_x) = \log\left(\frac{p_x}{1-p_x}\right) = \alpha + \beta_1 X_1 + \dots + \sum_{i=1}^{n-1} \beta_{ji} D_{ji} + \beta_m X_m$$

3.1.1. Interpretación de los parámetros

La interpretación de un modelo de regresión logística se trata de un incremento o disminución que se produce en el cociente entre $P(Y=1)/P(Y=0)$, también conocida como *odds*, con la expresión:

$$\frac{P(Y = 1)}{P(Y = 0)} = e^{\alpha + \beta_1 X_1 + \dots + \beta_m X_m}$$

Se debe tener en cuenta, que los coeficientes están expresados en forma de logaritmo, por lo que será necesario transformarlos, para evaluar de forma más fácil el efecto que tiene la variable explicativa sobre la probabilidad.

Por lo que el *odds ratio*, que se utiliza para cuantificar la asociación entre la variable respuesta y una variable explicativa, la cual, se empleará para la interpretación de los parámetros de un modelo de regresión logística, se calculará como:

$$OR_{x_i} = \frac{\text{odds}(Y = 1|X_1, \dots, X_i = 1, \dots, X_m)}{\text{odds}(Y = 1|X_1, \dots, X_i = 0, \dots, X_m)} = \frac{e^{\alpha + \beta_1 X_1 + \dots + \beta_i + \dots + \beta_m X_m}}{e^{\alpha + \beta_1 X_1 + \dots + \beta_m X_m}} = e^{\beta_i}$$

Un coeficiente positivo, al transformarlo, será mayor que 1 y, por lo tanto la *odds* del evento aumentará. Por el contrario, si un coeficiente es negativo, la transformación será menor a 1 y la *odds* disminuirá.

3.1.2. Bondad de ajuste

Hay distintas maneras de comprobar la bondad de ajuste de un modelo de regresión logística. Una de ellas es la matriz de confusión, se trata de una tabla 2×2 en la que se muestra la distribución de los objetos que pertenecen cuando $Y = 0$ e $Y = 1$ con la clasificación en cualquier de las dos categorías según su probabilidad estimada. Lo que se espera de esta matriz de confusión es que su porcentaje sea lo más alto posible en la diagonal, para poder esperar a que el modelo clasifica bien los individuos.

Valores reales	Valores predichos	
	Positivos	Negativos
Positivos	VP	FN
Negativos	FP	VN

Tabla 3.1: Matriz de confusión

Otra de las formas se trata del contraste de Hosmer-Lemeshow. Consiste en contar intervalo por intervalo el valor esperado y el valor observado para cada uno de los dos resultados posibles de la variable dependiente y a partir de ellas, las frecuencias esperadas y compararlas con las observadas mediante la prueba de χ^2 con $g - 2$ grados de libertad, se calcula como:

$$\chi_{HL}^2 = \sum_{i=1}^g \frac{(O_i - E_i)^2}{E_k} \sim \chi_{g-2}^2$$

El p-valor de este test será $P(\chi_{g-2}^2 \geq \chi_{HL}^2)$, el cual si es mayor a un nivel de significación α no habrá evidencias suficientes para rechazar la hipótesis nula y se podrá concluir que no hay evidencias de que el modelo no se ajuste correctamente a los datos, y por lo tanto, se podrá dar el modelo como bueno.

Por último, se puede saber si el modelo está bien ajustado a través de un análisis de los residuos de este, sabiendo que estos residuos se distribuyen aproximadamente como una Normal(0,1), si el modelo que se está ajustando es correcto. Estos residuos son los residuos estandarizados, los residuos studentizados y los residuos devianza.

3.1.3. Implementación en R

Para la implementación en R, del modelo que se desea estudiar se ha utilizado la función *glm* del paquete *stats* (R Core Team, 2020) [9], declarando la familia como binomial y link *logit*. De

la siguiente forma, donde *formula* sería el modelo que se desea estudiar, es decir la declaración de la variable respuesta y las variables explicativas que se desean investigar.

```
glm(formula, family = gaussian, data, weights, subset,
     na.action, start = NULL, etastart, mustart, offset,
     control = list(...), model = TRUE, method = "glm.fit",
     x = FALSE, y = TRUE, singular.ok = TRUE, contrasts = NULL, ...)
```

Figura 3.1: Función *glm* del paquete *stats* en R

En la construcción del modelo, se debe elegir qué variables serán útiles para explicar la variable dependiente, por consiguiente, se han seguido los pasos del libro de Hosmer, D. and S. Lemeshow (2000) [10]. *Applied Logistic Regression*. Estos pasos son:

1. Ajustar modelos logísticos sencillos para todas las variables de interés y se mantienen aquellas con un p-valor menor a 0,25.
2. Ajustar modelo con todas las variables retenidas en el paso anterior.
3. Las variables se pueden eliminar, si se cumplen las siguientes condiciones:
 - Si no son estadísticamente significativas.
 - Si el parámetro estimado es similar al del modelo simple.
 - Los coeficientes de las otras variables apenas cambian al eliminar esta variable del modelo.

Además también se puede usar la función *anova* del paquete *stats* (R Core Team, 2020) [9], para comprobar que los distintos modelos que se están construyendo siguiendo los pasos anteriores, realmente van mejorando o no. Se realiza un *anova* de la siguiente forma:

```
anova (modelo_1, modelo_2, test="Chisq")
```

Figura 3.2: Función *anova* del paquete *stats* en R

Donde el *modelo₁* se trata del modelo inicial y *modelo₂* el modelo una vez eliminadas las variables según los pasos descritos. Si el p-valor de este test *anova* es estadísticamente no significativo, el *modelo₂* mejorara al *modelo₁*, ya que con menos variables explica mejor la variable dependiente.

3.2. Modelo de regresión log-binomial

Se trata de un modelo lineal generalizado y, al igual que el modelo de regresión logística explicado anteriormente, la variable respuesta Y se trata de una variable aleatoria binaria. La diferencia que tiene este modelo respecto al anterior se encuentra en la función de enlace que se utilizará para la relación del predictor lineal η con el valor esperado de la respuesta, en este caso se utilizará como función de enlace el *logaritmo*.

El modelo log-binomial se define como:

$$\eta = g(p_x) = \log(p_x) = \alpha + \beta_1 X_1 + \dots + \beta_m X_m$$

Por lo que la probabilidad de respuesta positiva sería la que se muestra en la ecuación a continuación.

$$p_x = g^{-1}(\eta) = e^{\alpha + \beta_1 X_1 + \dots + \beta_m X_m}$$

En algunas ocasiones pueden encontrarse problemas en la estimación de los parámetros, ya que, $p_x \in (0, 1)$ y $e^{(\alpha + \beta_1 X_1 + \dots + \beta_m X_m)} \in \mathbb{R}^+$.

De la misma manera que en los modelos de regresión logística, la estimación de los parámetros se obtiene maximizando la función de verosimilitud de los datos observados:

$$L(\alpha, \beta | Y, X) = \prod_{i=1}^n \exp(\alpha + \beta' \mathbf{x}_i)^{\delta_i} \cdot [1 - \exp(\alpha + \beta' \mathbf{x}_i)]^{1 - \delta_i},$$

donde $\beta' = (\beta_1, \dots, \beta_m)'$ y $\delta_i = 1$ si $Y_i = 1$ y 0 en caso contrario.

3.2.1. Interpretación de los parámetros

Para estudios de cohortes y transversales se utiliza el riesgo relativo (RR) y el *prevalence ratio* (PR) para la interpretación de los parámetros de un modelo log-binomial. Estos valores se obtendrían de la misma forma.

$$RR_{x_i} (\text{o } PR_{x_i}) = \frac{P(Y = 1 | X_1, \dots, X_i = 1, \dots, X_m)}{P(Y = 1 | X_1, \dots, X_i = 0, \dots, X_m)} = \frac{e^{\alpha + \beta_1 X_1 + \dots + \beta_i X_i + \dots + \beta_m X_m}}{e^{\alpha + \beta_1 X_1 + \dots + \beta_m X_m}} = e^{\beta_i}$$

Lo que supone una ventaja del modelo log-binomial frente al modelo de regresión logística, ya que el riesgo relativo y el *prevalence ratio* son más sencillos de interpretar que el *odds ratio*.

3.2.2. Bondad de ajuste

Para comprobar la bondad de ajuste de un modelo log-binomial se utilizaría el test de Hosmer-Lemeshow descrito en el Apartado 3.1.2 anterior, aunque habría un problema al aproximar el estadístico χ_{HL}^2 con $g - 2$ grados de libertad, ya que, en este tipo de modelos la suma de los valores observados no es la misma que la suma de los valores esperados.

Sin embargo, los autores demostraron que estos valores variaban extremadamente cerca y que por lo tanto, si el modelo ajustado es bueno, el test de Hosmer-Lemeshow se aproxima de forma correcta a una χ_{HL}^2 con $g - 2$ grados de libertad.

3.2.3. Implementación en R

Para la realización de este modelo en este trabajo, se utilizará la función *glm* del paquete *stats* (R Core Team, 2020) [9] ya descrita anteriormente en la Figura 3.1. En este caso, pero, se utilizará *family = binomial (link = "log")*.

3.3. Modelo Poisson cero-inflado

Este tipo de modelos se utilizan en modelos de recuentos en los que los datos muestran más ceros de los que deberían según determinados modelos, como podrían ser los modelos Poisson. Esta gran cantidad de ceros que nos lleva a la utilización de este modelo se produce en sucesos que son poco frecuentes en unas observaciones y bastante frecuentes para otras.

Este tipo de modelos contendrán una proporción p_i de ceros estructurales y una proporción $(1 - p_i)$ de población que sigue una Poisson. Es decir:

$$P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i) \cdot e^{-\lambda} & \text{si } y_i = 0 \\ (1 - p_i) \cdot \frac{\lambda^{y_i} \cdot e^{-\lambda}}{y_i!} & \text{para } y_i > 0 \end{cases},$$

en que los vectores de los parámetros λ_i y p_i satisfacen las condiciones

$$\log(\lambda_i) = \mathbf{Z}_i \boldsymbol{\beta}$$

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1 - p_i}\right) = \mathbf{W}_i \boldsymbol{\gamma},$$

donde Z_i y W_i son los vectores de las variables explicativas [11].

Los ceros estructurales serían aquellos ceros que son inevitables mientras que los ceros aleatorios se trataría de aquellos ceros que ocurren debido al mecanismo de muestreo.

Esta probabilidad p_i puede establecerse como una constante o bien, puede ajustarse un modelo de respuesta binaria para poder predecirla. Normalmente, se combinan dos modelos. Un modelo binomial con *logit* (ya explicado en el Apartado 3.1) y, un modelo Poisson o Binomial negativa que predice la respuesta de los individuos en la segunda clase latente.

La estimación de los parámetros de un modelo cero inflado se obtiene por el método de la máxima verosimilitud. Primero de todo se debe definir una variable indicadora que mostrara si el valor observado se trata de un 0 o no:

$$d_i = \begin{cases} 1 & \text{si } y_i > 0 \\ 0 & \text{si } y_i = 0 \end{cases}$$

En consecuencia la log-versosimilitud será [12] :

$$l(\theta_1, \theta_2) = \sum_{i=1}^n (1 - d_i) \cdot \log[\pi(x_i, \theta_1) + (1 - \pi(x_i, \theta_1))f(0|x_i, \theta_2)] + \sum_{i=1}^n d_i \cdot \log[(1 - \pi(x_i, \theta_1))f(0|x_i, \theta_2)]$$

3.3.1. Interpretación de los parámetros

Teniendo en cuenta como se ha explicado en el apartado anterior que un modelo cero-inflado combina dos modelos, un modelo binomial con *logit* (ya explicado en el Apartado 3.1) y un modelo Poisson o Binomial negativa. Un signo positivo en la estimación del coeficiente del componente binario indica, que la probabilidad de un conteo mayor que cero aumenta. En cambio para el componente de conteos, la interpretación de los parámetros debe hacerse de la misma forma que se hace en los modelos Poisson o Binomial Negativa; de tal forma que un cambio de una unidad en el regresor llevará a un cambio proporcional al valor estimado de β_i [13].

3.3.2. Bondad de ajuste

Para poder percibir la bondad de ajuste de un modelo Poisson cero-inflado se deben graficar las predicciones del modelo respecto los residuos de Pearson que se calculan como:

$$e_i = \frac{y_i - \hat{p}_i}{\sqrt{\hat{p}_i(1 - \hat{p}_i)}}$$

Si el modelo es correcto, serán de media y varianza cero.

3.3.3. Implementación en R

Para la implementación de dicho modelo en *R* se puede hacer uso de la función *zeroinfl*, del paquete *pscl* (Achim Zeileis, Christian Kleiber, Simon Jackman, 2020) [14], de la siguiente forma. Donde *formula* sería el modelo que se desea estudiar, es decir la declaración de la variable respuesta y las variables explicativas que se desean estudiar. Y *dist* se trataría de la distribución que se quiera utilizar, en el caso de este estudio se tratara de la distribución Poisson.

```
zeroinfl(formula, data, subset, na.action, weights, offset,
dist = c("poisson", "negbin", "geometric"),
link = c("logit", "probit", "cloglog", "cauchit", "log"),
control = zeroinfl.control(...),
model = TRUE, y = TRUE, x = FALSE, ...)
```

Figura 3.3: Función *zeroinfl* de la librería *pscl* en R

Capítulo 4

Resultados

En este capítulo se presentan los resultados que se han obtenido en el análisis descriptivo de las variables, descritas en el Apartado 2.3 de este trabajo, tanto para el caso de parejas actuales como el caso de parejas pasadas. Además también se presentarán los distintos modelos estimados, los resultados que se han obtenido, así como las interpretaciones de los parámetros de estos.

4.1. Análisis descriptivo

Antes de la elaboración de los distintos modelos lineales generalizados que se decidieron estudiar en este trabajo se realizó un análisis descriptivo con el fin de ver qué variables serían de interés para la estimación de los modelos. Además como ya se ha explicado en apartados anteriores (2.3) se disponen de distintos tipos de violencia. Para los que sería necesario saber si había grandes diferencias entre estos, ya que, si las diferencias fueran muy grande se hubieran estimado distintos modelos diferenciando el tipo de violencia como variable respuesta.

4.1.1. Parejas pasadas

En este apartado se presentará el análisis descriptivo de aquellas mujeres que han tenido o consideran que han tenido alguna vez pareja por lo que, quedan fuera de la población de interés aquellas personas que no han tenido nunca pareja y también las mujeres que han tenido una pareja y esta es la actual. De este modo, antes de esta eliminación, se disponía de una base de datos con un total de 9568 mujeres y a continuación se obtiene una población total de interés de 1118 mujeres.

Análisis de los distintos tipos de violencia

Con el fin de tener una clara visión de cómo tienen que ver los distintos tipos de violencia uno con otros, en otras palabras, si una violencia puede llevar implícita la causa de violencia mayor, se presenta a continuación un análisis bivalente entre los distintos tipos de violencia que ya se han descrito en el Apartado 2.3.

Insultos o Humillaciones	Control Social		Total
	No	Si	
No	480 (42,9 %)	132 (11,8 %)	612 (54,7 %)
Si	65 (5,8 %)	441 (39,4 %)	506 (45,3 %)
Total	545 (48,7 %)	573 (51,3 %)	1118 (100,0 %)

Tabla 4.1: Tabla de contingencia entre Control Social e Insultos o Humillaciones

En esta tabla en la que se cruzan los datos dicotómicos del tipo de violencia, se puede observar que en la población que se está estudiando ha habido 573 mujeres que han sufrido control social por alguna de sus parejas pasadas, contra 506 mujeres que han sufrido insultos o humillaciones. De estas 573 mujeres que han sufrido control social, 441 también han sufrido insultos o humillaciones. Por lo contrario, se puede observar que de las 573 mujeres que han sufrido algún episodio de control social 132 declara no haber sufrido ningún tipo de insulto o humillaciones.

También cabe destacar que de las 545 que no han sufrido control social, se perciben 65 que sí que han sufrido insultos y humillaciones.

Violencia Física	Insultos o Humillaciones		Total
	No	Si	
No	601 (53,8 %)	250 (22,4 %)	851 (76,1 %)
Si	11(1,0 %)	256 (22,9 %)	267 (23,9 %)
Total	612 (54,7 %)	506 (45,3 %)	1118 (100,0 %)

Tabla 4.2: Tabla de contingencia entre Insultos o Humillaciones y Violencia Física

A medida que se va avanzando en el nivel de violencia se observa una disminución de esta por ejemplo, en el caso de violencia física se puede observar como 267 ha sufrido alguna vez por parte de una de sus parejas pasadas violencia física. Cabe destacar que de las mujeres que han sufrido insultos o humillaciones, un total de 506, el 50,6 % también han sufrido violencia física. Por otro lado, hay una minoría 1,8 % de mujeres que han sufrido violencia física sin haber sufrido antes insultos o humillaciones por parte de alguna de sus parejas pasadas.

Violencia Sexual	Violencia Física		Total
	No	Si	
No	780 (69,8 %)	134 (12,0 %)	914 (81,8 %)
Si	71 (6,4 %)	133 (11,9 %)	204 (18,2 %)
Total	851 (76,1 %)	267 (23,9 %)	1118 (100,0 %)

Tabla 4.3: Tabla de contingencia entre Violencia Física y Violencia Sexual

Finalmente, se muestra la violencia física y la violencia sexual que han sufrido las mujeres del estudio. Se puede observar como que de las mujeres que han sufrido violencia física, 267, el 49,8% de estas también han sufrido violencia sexual, por el contrario un 50,2% de las mujeres que han sufrido violencia física; no han sufrido violencia sexual.

Diferenciándose del caso que se ha analizado anteriormente, se observa que de entre las mujeres que no han sufrido violencia física; un 8,3% de estas ha sufrido violencia sexual por parte de alguna de sus parejas pasadas.

Con el fin de mostrar los tipos de violencia pero, con esta en forma numérica, es decir, el número de veces que las mujeres que han sido entrevistadas han sufrido algún tipo de violencia de las descritas anteriormente se muestran los siguientes gráficos de frecuencias.

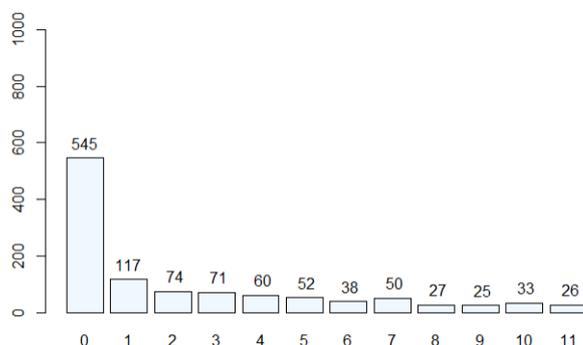


Figura 4.1: Gráfico de frecuencias del tipo de violencia Control Social

Cabe destacar observando este gráfico que 26 mujeres de las que han sido entrevistadas han sufrido un máximo de 11 situaciones de las descritas de tipo de control social por parte de al menos una de sus parejas pasadas. Mientras que 117 mujeres han sufrido uno de los maltratos descritos como control social.

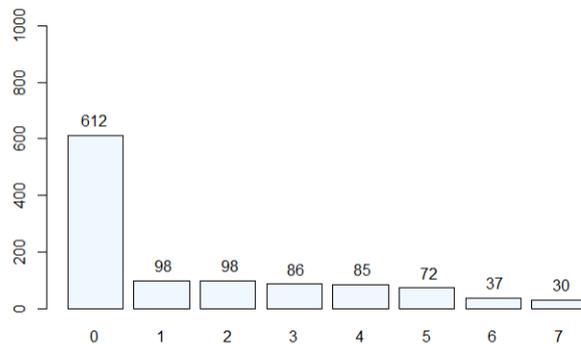


Figura 4.2: Gráfico de frecuencias del tipo de violencia Insultos o Humillaciones

Aunque el número de mujeres que no han sufrido este tipo de violencia por parte de alguna de sus parejas pasadas aumenta, es destacable que 30 entrevistadas han sufrido 7 de las situaciones descritas para este tipo de violencia y que, un total de 109 ha sufrido 5 o 6 situaciones de insultos o humillaciones, un total de 98 mujeres ha sufrido al menos una vez maltrato verbal por parte de su pareja actual.

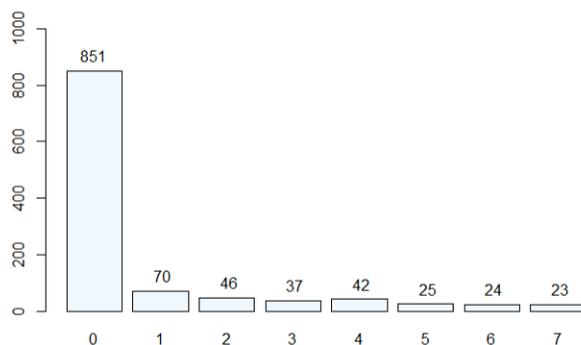


Figura 4.3: Gráfico de frecuencias del tipo de Violencia Física

Como ya se ha comentado anteriormente, el número de personas que han sufrido niveles de violencia mayores o mejor dicho, que en algunos casos se consideran más graves, va disminuyendo. Aun así es importante destacar que hay mujeres que han sufrido varias situaciones de violencia física, como por ejemplo, 23 mujeres han sufrido 7 situaciones de violencia física que se describían en la encuesta.

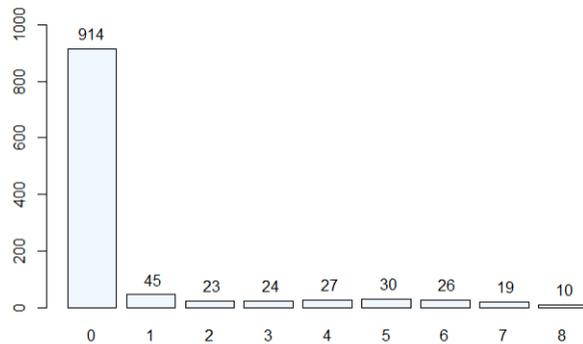


Figura 4.4: Gráfico de frecuencias del tipo de Violencia Sexual

La tendencia de sufrir este tipo de violencia sigue a la baja en comparación a las anteriores, aunque siempre es importante destacar que hay mujeres que han sufrido violencia sexual por parte de alguna de sus parejas pasadas, este tipo de violencia como ya se ha explicado anteriormente en el Apartado 1.2.3 consiste en la obligación de realizar prácticas sexuales sin consentimiento. 10 mujeres han sufrido este tipo de violencia en 8 de las situaciones descritas en el cuestionario.

Características de las mujeres que han sufrido violencia por alguna de sus parejas pasadas

En este apartado se presenta el análisis descriptivo de aquellas variables que se han considerado de interés para describir el perfil de las mujeres que han sufrido violencia de género por alguna de sus parejas pasadas.

Estas variables son:

Variable	Descripción
TAMUNI	Tamaño del municipio
SDE1	Edad de la persona entrevistada
MOP0	Situación laboral de la entrevistada
ESTUDIOS	Nivel de estudios de la persona entrevistada

Tabla 4.4: Variables de interés para el estudio de parejas pasadas

Cabe destacar que como no se observaron grandes diferencias analizando las siguientes variables según el tipo de violencia ejercida, a continuación se muestran los resultados con una variable en la que se recoge si la mujer ha sufrido o no violencia sin tener en cuenta que tipo de esta es.

En primer lugar se ha analizado la variable que describe el **tamaño del municipio** y, cómo está relacionado o no, a los distintos tipos de violencia ejercida sobre las mujeres por alguna de sus parejas pasadas.

Para este variable, en el tipo de municipio que más violencia se ejerce es en los que tienen un

tamaño de entre 50.001 a 100.000 habitantes con un 67,61 % de violencia. Mientras que en las grandes ciudades de más de 1.000.000 habitantes son en las que menos violencia se lleva a cabo.

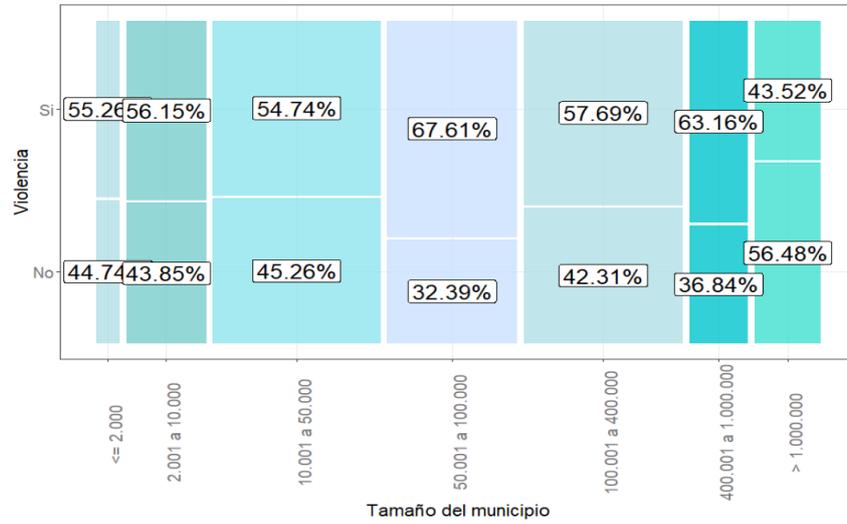


Figura 4.5: Gráfico de mosaico del tamaño de municipio según si se ha sufrido o no violencia

En el caso de la variable que describe la **edad** de la persona entrevistada, podemos observar como para todos los tipos de violencia, la mediana de la edad es más grande, acercándose a 50, para las mujeres que no sufren ningún tipo de violencia a diferencia de las mujeres que sí que sufren algún tipo de violencia, que son más jóvenes. Esto puede ser debido a que las mujeres que han tenido más de una pareja sean mujeres más jóvenes, ya que, en otra época no era normal tener más de una pareja o experimentar en este ámbito como lo es ahora. Esta podría ser una explicación a esta diferencia de edad entre las mujeres que sufren violencia.

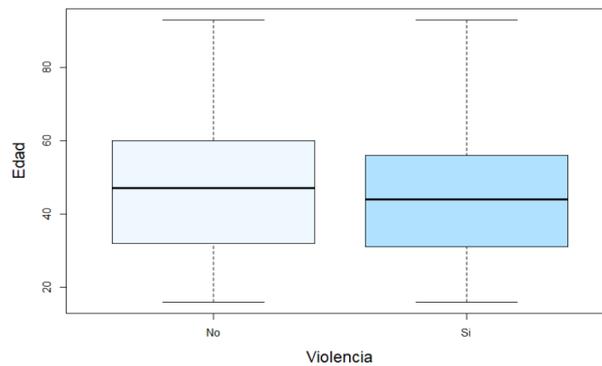


Figura 4.6: Boxplot de la edad según si se ha sufrido o no violencia

Diferenciando la violencia según la **situación laboral** en la que se encuentra la persona entrevistada se puede observar como en la situación laboral que más violencia se ejerce es en el caso de si las mujeres están paradas con un 70,44 % de violencia dentro de esta situación.

Aunque no es un valor tan grande como el anterior, cabe destacar que para el caso de trabajo domestico se ejerce dentro de este un 61,36 % de violencia

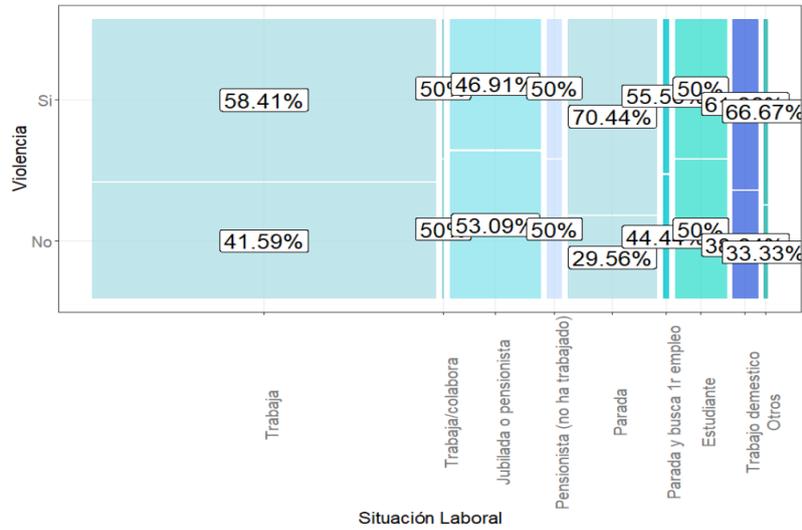


Figura 4.7: Gráfico de mosaico de la situación laboral según si se ha sufrido o no violencia

Otra de las variables que se ha decidido que podrían ser de interés para el estudio que se está llevando a cabo ha sido el **nivel de estudios** de la mujer entrevistada.

Se puede ver como hay tres tipos de nivel de estudios que destacan por su frecuencia en los que se ejerce violencia de género, estos son sin estudios, secundaria 2ª etapa y F.P.

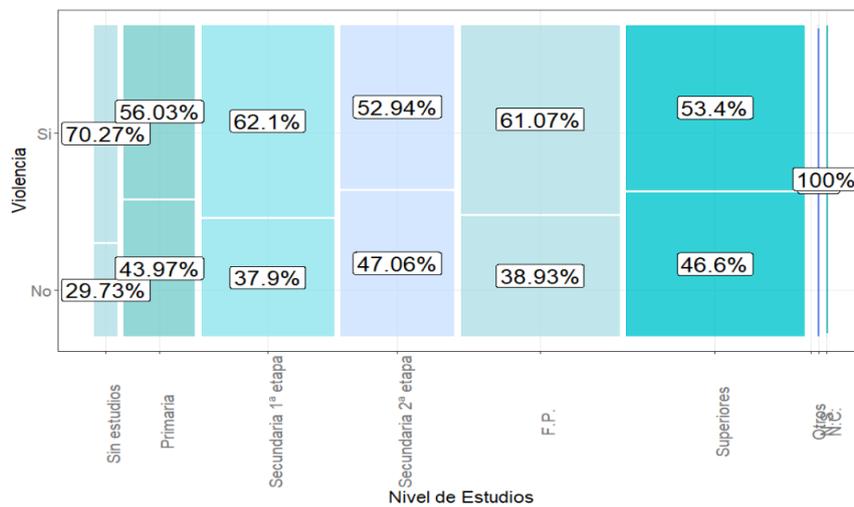


Figura 4.8: Gráfico de mosaico del nivel de estudios según si se ha sufrido o no violencia

4.1.2. Parejas actuales

A diferencia del Apartado 4.1.1 anterior, en este se encuentran recogidas aquellas mujeres que tenían en el momento en el que se realizó la encuesta pareja. De esta manera, quedan fuera de la población de interés aquellas mujeres que no tenían pareja. En consecuencia, se obtiene una población total de 6563 mujeres. A continuación se presenta el análisis descriptivo.

Análisis de los distintos tipos de violencia

Al igual que en el caso anterior de parejas pasadas, en primero lugar se muestra el análisis bivariante entre los distintos tipo de violencia.

Insultos o Humillaciones	Control Social		Total
	No	Si	
No	5644 (86,0 %)	397 (6,0 %)	6041 (92,0 %)
Si	184 (2,8 %)	338 (5,2 %)	522 (8,0 %)
Total	5828 (88,8 %)	735 (11,2 %)	6563 (100,0 %)

Tabla 4.5: Tabla de contingencia entre Control Social e Insultos o Humillaciones

En primer lugar haciendo un análisis bivariante entre el control social y los insultos o humillaciones, se puede observar como hay 552 mujeres que han sufrido insultos o humillaciones y 735 mujeres que han sufrido control social. Además de que hay 338 personas que han sufrido estos dos tipos de violencia.

Por otro lado, se puede observar que de entre las personas que han sufrido control social por parte de su pareja actual un 46 % también han recibido insulto o humillaciones.

En el caso de no haber recibido ningún tipo de control social, un 3,2 % de entre estas ha sufrido algún tipo de violencia verbal.

Violencia Física	Insultos o Humillaciones		Total
	No	Si	
No	6031 (91,9 %)	421 (6,4 %)	6452 (98,3 %)
Si	10 (0,2 %)	101 (1,5 %)	111 (1,7 %)
Total	6041 (92,0 %)	522 (8,0 %)	6563 (100,0 %)

Tabla 4.6: Tabla de contingencia entre Insultos o Humillaciones y Violencia Física

Observando la Tabla 4.6, se puede observar que de entre las mujeres que han sufrido insultos y humillaciones un 19,3 % también han sufrido algún tipo de violencia física. Por el contrario, se puede ver como hay una minoría 0,2 % de mujeres que nunca han sufrido insultos y humillaciones, pero que sí que han sufrido violencia física por parte de su parejas actual.

Finalmente, del total de 6563 mujeres, 522 han sufrido insultos y humillaciones y 111 han sufrido violencia física.

Violencia Sexual	Violencia Física		Total
	No	Si	
No	6373 (97,1 %)	69 (1,1 %)	6442(98,2 %)
Si	79 (1,2 %)	42 (0,6 %)	121 (1,8 %)
Total	6452 (98,3 %)	111 (1,7 %)	6563 (100,0 %)

Tabla 4.7: Tabla de contingencia entre Violencia Física y Violencia Sexual

Por último en el caso de violencia física y violencia sexual, se puede observar como que de entre las personas que han sufrido violencia física un 37,8% de ellas también ha sufrido algún tipo de violencia sexual, 42 mujeres.

Cabe destacar, que aunque no se haya sufrido nunca violencia física un 1,2% de mujeres ha sufrido violencia sexual por parte de su pareja. Esto llama la atención, ya que en total hay 121 mujeres que han sufrido violencia sexual, en conclusión un 62,2% de mujeres han sufrido violencia sexual sin haber sufrido violencia física.

Finalmente se muestran los distintos tipos de violencia según el número de veces que la mujer entrevistada ha sufrido violencia de género.

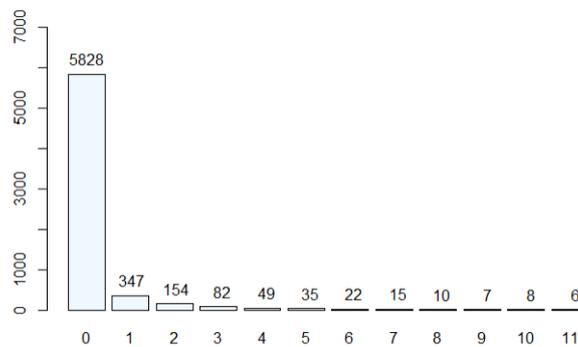


Figura 4.9: Gráfico de frecuencias del tipo de violencia Control Social

Aunque una clara mayoría no haya sufrido nunca control social por parte de su pareja actual, 347 mujeres han sufrido alguna de las situaciones descritas para este tipo de violencia y, 6 mujeres han sufrido 11 de las situaciones descritas.

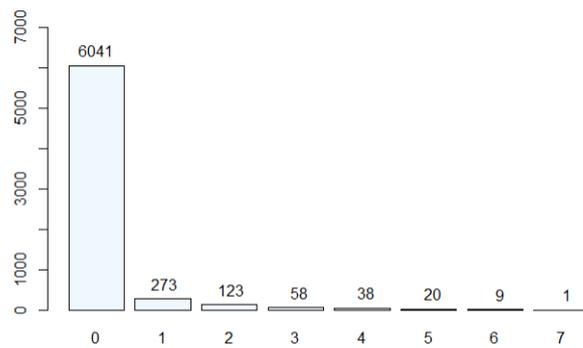


Figura 4.10: Gráfico de frecuencias del tipo de violencia Insultos o Humillaciones

Como ya se ha comentado en otros apartados, a medida que el nivel de violencia se agrava el número de mujeres que la sufren disminuye. No obstante, 273 mujeres han sufrido algún insulto o humillación por parte de su agresor y un total de 249 mujeres han pasado por al menos alguna situación de violencia verbal descritas en la encuesta.

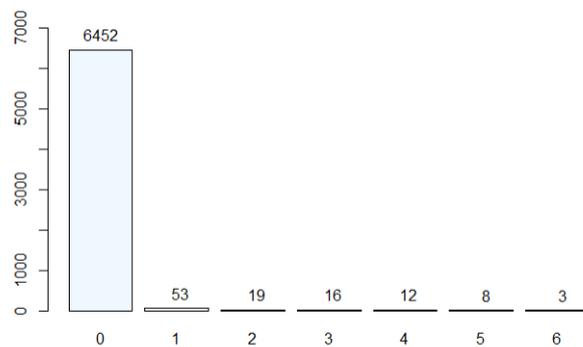


Figura 4.11: Gráfico de frecuencias del tipo de Violencia Física

Un total de 6452 mujeres no ha sufrido nunca violencia física, pero 53 mujeres han sufrido una de las situaciones descritas de violencia física. Y aunque sea una minoría, 3 mujeres han sufrido 6 de las situaciones descritas.

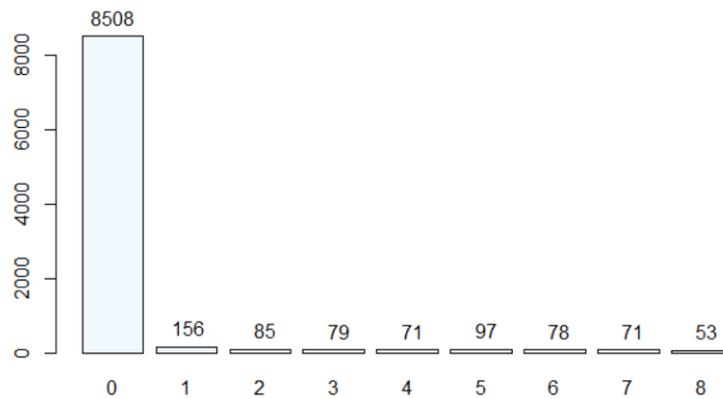


Figura 4.12: Gráfico de frecuencias del tipo de Violencia Sexual

Finalmente, hay un número más elevado de mujeres que ha sufrido alguna vez violencia sexual, 53 de las entrevistadas ha pasado por alguna de las situaciones descritas y por lo tanto ha recibido violencia sexual por parte de su pareja actual.

Características de las mujeres que han sufrido violencia por parte de su pareja

A continuación se muestran las variables que se han considerado de interés para tener una visión general de las características de las mujeres que sufren violencia de género así, como las de su agresor o pareja. Estas variables son las mismas que las descritas en la Tabla 2.1 del apartado 2.3 de este trabajo, ya que estas son las mismas que se han utilizado como variables regresoras en los modelos lineales que se mostraran más adelante.

De la misma forma que se ha hecho en el apartado anterior, solo se mostrarán los gráficos para la variable de violencia, sin diferenciar dentro de esta.

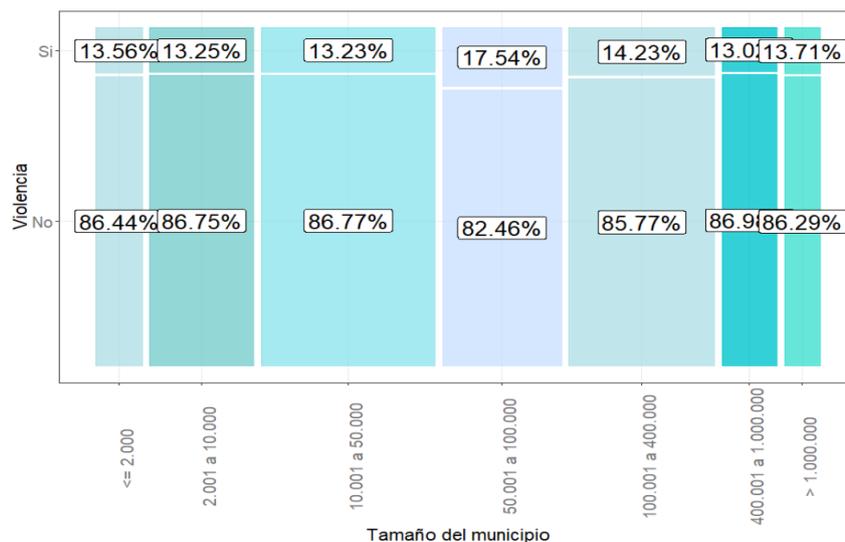


Figura 4.13: Gráfico de mosaico del tamaño de municipio según si se ha sufrido o no violencia

Se puede observar que en los tipos de **municipio según su tamaño** que más violencia se produce son en los de entre 50.001 a 100.000 habitantes con un 17,54% de violencia, aunque los demás tipos de municipio no distan mucho de este valor.

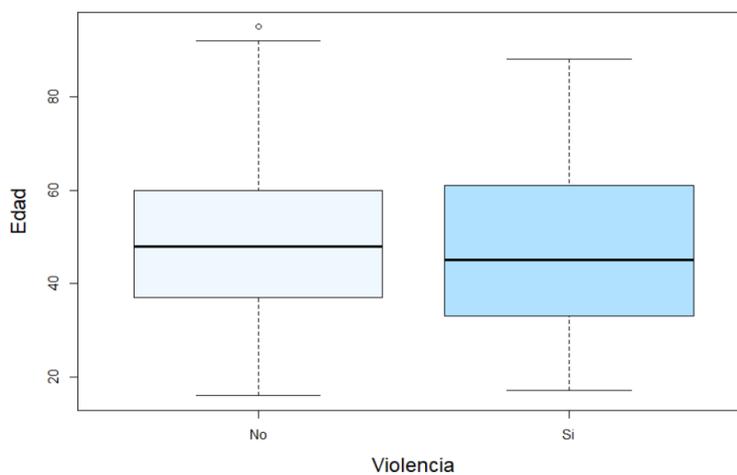


Figura 4.14: Boxplot de la edad según si se ha sufrido o no violencia

Viendo la mediana de la **edad** de las mujeres que se han entrevistado para este estudio esta en torno a los 50 años. Aunque se puede notar que hay una ligera disminución en la media de la edad para las mujeres que han sufrido violencia de género.

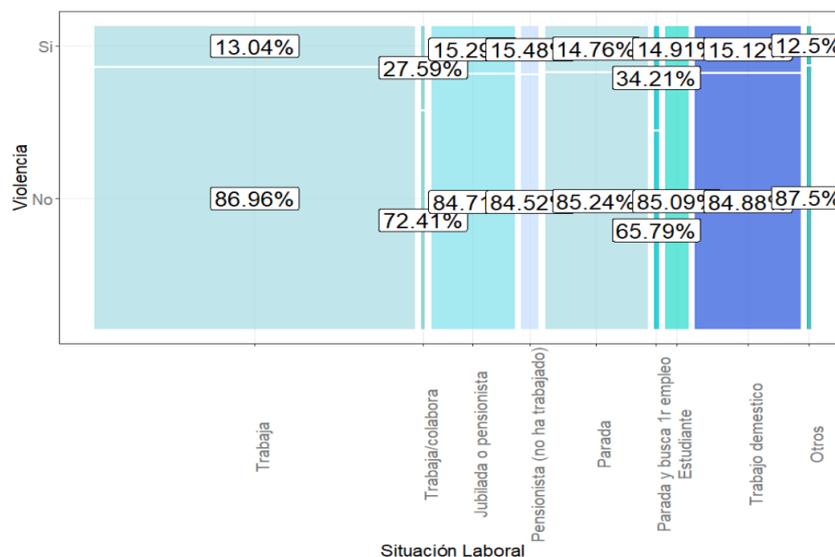


Figura 4.15: Gráfico de mosaico de la situación laboral según si se ha sufrido o no violencia

En primer lugar, cabe destacar que en los gráficos de mosaico la anchura de las barras explica el tamaño de población que se encuentra en ese factor, es decir, en este caso se observa como el factor trabaja es mucho más amplio que todos los demás; esto quiere decir que la mayoría de las mujeres que se han entrevistado trabajaban en el momento en el que se realizó la entrevista.

Aunque la mayoría de mujeres trabajen, se tiene que destacar que para aquellas mujeres que trabajan o colaboran en el negocio familiar se ejerce un 27,59% de violencia dentro de estas, al igual, que en el caso de mujeres que están paradas y buscan su primer empleo se ejerció un 34,21% de maltrato.

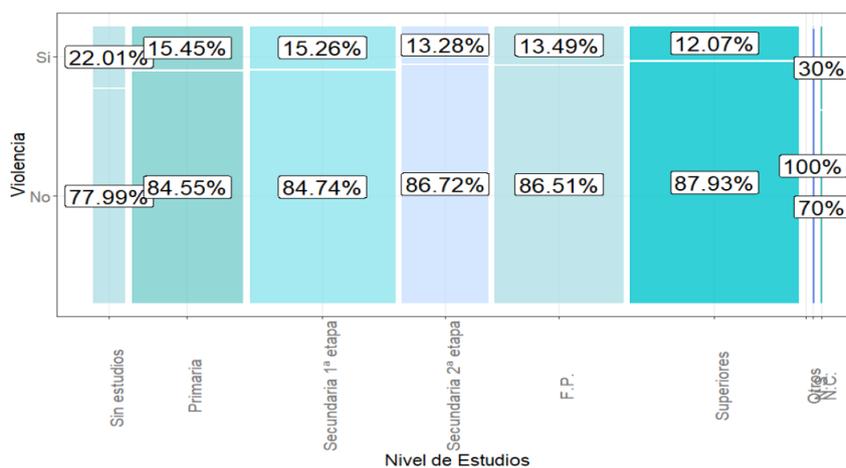


Figura 4.16: Gráfico de mosaico del nivel de estudios según si se ha sufrido o no violencia

Para el caso de **nivel de estudios** ha sido de gran interés ver como a medida que se va avanzando en el nivel de estudios superiores la violencia que se ejerce dentro de este es menor.

Es decir, en el caso de que las mujeres no tengan estudios la violencia que se ejerce sobre ellas es de un 22,01 % mientras que, la mujeres que disponen de estudios superiores el maltrato que reciben dentro de estas es de un 12,07 %.

Características de las parejas que han ejercido violencia de género

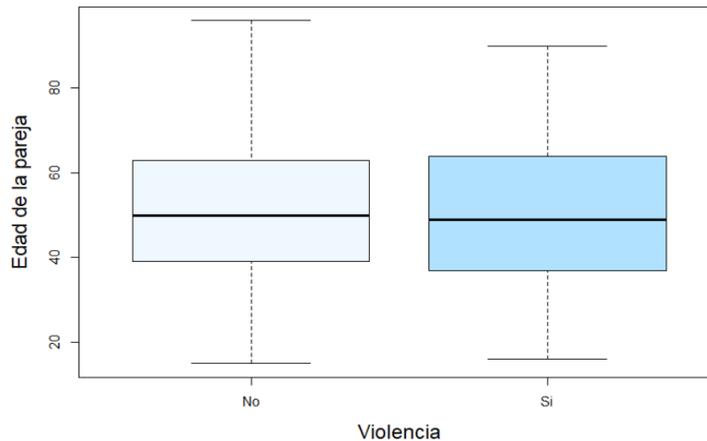


Figura 4.17: Boxplot de la edad de la pareja actual según si se ha ejercido o no violencia

En la Figura 4.17 se muestra la **edad de las parejas**. Se puede ver como los agresores que han ejercido o no violencia suelen tener la misma edad, entre los 50 y los 60 años.

Cuando se analizó esta variable por los distintos tipos de violencia fue muy destacable, que en el caso de violencia sexual los agresores son mucho más mayores, superando la edad de 60 años y acercándose a la de 65 mientras que los que no ejercen violencia de este tipo están entorno a los 45 y 50 años. (ver Apéndice A)

Esto puede ser debido a que la educación sexual en épocas más tardías que la actual era mucho más escasa a la que recibe hoy en día, ahora que ha habido más movimientos como el de “#metoo” dando visibilidad a que las violaciones, incluso en el ámbito de pareja existen, y puede ser que las personas más mayores se alejen de este tipo de movimientos y oleadas.



Figura 4.18: Gráfico de mosaico del nivel de estudios de la pareja según si se ha ejercido o no violencia

En este caso pasa lo mismo que pasaba con las mujeres que recibían violencia, a medida que el **agresor** se acerca más a **niveles de estudios** superiores la violencia que se ejerce es menor. De entre los que no tienen estudios un 23,81 % ejerce violencia, por el contrario, en el caso de tener estudios superiores un 10,36 % la realiza.

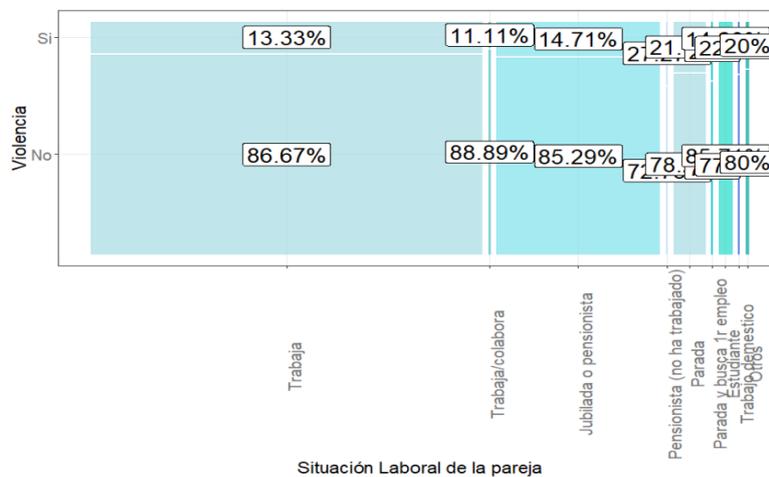


Figura 4.19: Gráfico de mosaico de la situación laboral según si se ha sufrido o no violencia

De la misma forma que pasaba con las mujeres entrevistadas, para la **situación laboral de la pareja**, la mayoría de parejas que tienen esas mujeres también trabajan, dentro de estos un 13,33 % ejerce violencia sobre su pareja. Aun así, se puede apreciar que un 21 % de violencia se ejerce si el agresor está parado.

Finalmente también se pudo observar como un 96,31 % de las personas que ejercen violencia de género son **hombres**. La mayoría de actos de violencia de genero se llevan a cabo en situaciones

de una relación **vínculos legales**, viven en **distintos domicilios** y el tiempo de relación es de entre **7 y 10 años**. (Apéndice A)

4.2. Modelos estadísticos

En este apartado se mostrarán los distintos modelos estadísticos ajustados para conocer mejor las características de aquellas mujeres que sufrieron violencia de género por parte de su pareja en el año 2019. Así como las expresiones de los distintos modelos, una breve interpretación de sus parámetros y los resultados de bondad de ajuste.

4.2.1. Modelo de regresión logística

Se pretende estudiar las características que presentan las mujeres que han sufrido alguna vez violencia de género en su relación de pareja actual. Con este fin, se ajusta un modelo con la variable *violencia.b* (variable que muestra si una mujer ha sufrido o no violencia de género) como variable respuesta y se precisan las variables mostradas en la Tabla 2.1 como variables explicativas.

Después de seguir los pasos del libro de D.Hosmer y S.Lemeshow [10], se ha obtenido el siguiente modelo de regresión logística, cuya expresión estimada se definiría tal y como se mostraba en el Apartado (3.1) del presente trabajo.

$$\text{logit}(\hat{p}_x) = \log\left(\frac{\hat{p}_x}{1 - \hat{p}_x}\right) = \alpha + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_{27} X_{27}$$

Donde X_1, \dots, X_{27} corresponden a las variables que se estudian y $\beta_1, \dots, \beta_{27}$ a los valores estimados obtenidos. En el caso de las variables categóricas se muestran $n - 1$ variables ficticias, es decir $n - 1$ niveles que tiene el factor de esta variable.

Alternativamente, se podría mostrar la fórmula anterior en función de la probabilidad estimada de respuesta positiva, con la Fórmula (3.1) del apartado (3.1. Regresión Logística) del presente trabajo.

Seguidamente, se muestran los valores obtenidos para el test *Anova* del modelo estimado para poder observar que variables influyen más en el hecho de que una mujer pueda padecer violencia de género.

Variable	LR Chisq	Df	p-valor
Tamaño del municipio	12,065	6	0,061
Situación laboral	19,018	8	0,015
Edad	23,053	1	1,576e-06
Estudios	1,426	5	0,921
Edad de la pareja	8,502	1	0,004
Estudios de la pareja	28,420	6	7,829e-05

Tabla 4.8: Anova para el modelo de regresión logística para mujeres que han sufrido violencia de género con función *glm*

Se han destacado en negrita los p-valores significativos usando un nivel de significación igual a 0,05, estos se han obtenido para las variables que describen la situación laboral, la edad de la persona entrevistada así como, la edad de la pareja y los estudios de la pareja. De este modo, estas son las variables que pueden estar relacionadas en que una mujer sufra violencia de género.

En la Tabla (4.9) que se muestra a continuación se presenta el modelo final que se ha obtenido conjuntamente con la estimación de los parámetros, el error estándar de estos, el valor Z , los p-valores y finalmente, los *Odds Ratio (OR)* asociados a cada una de las variables.

Variable	$\hat{\beta}$	$s.e(\hat{\beta})$	Z	p-valor	\widehat{OR}
Constante	-0,457	0,302	-1,514	0,130	0,633
Tamaño del municipio (categoría referencia: ≤ 2.000)					
De 2.001 a 10.000	-0,004	0,170	-0,023	0,982	0,996
De 10.001 a 50.000	0,004	0,159	0,026	0,98	1,004
De 50.001 a 100.000	0,326	0,163	1,993	0,046	1,385
De 100.001 a 400.000	0,176	0,163	1,079	0,280	1,192
De 400.001 a 1.000.000	0,096	0,194	0,497	0,619	1,101
Más de 1.000.000	0,233	0,214	1,089	0,276	1,262
Situación laboral de la mujer entrevistada (categoría referencia: Trabaja)					
Trabaja o colabora	1,053	0,423	2,488	0,013	2,867
Jubilada o pensionista	0,399	0,142	2,807	0,005	1,491
Pensionista	0,224	0,255	0,878	0,380	1,251
Parada y ha trabajado	0,024	0,107	0,229	0,819	1,025
Parada y primer empleo	0,899	0,354	2,539	0,011	2,456
Estudiante	-0,186	0,211	-0,880	0,379	0,831
Trabajo doméstico	0,14	0,119	1,172	0,241	1,15
Otra situación	-0,013	0,543	-0,023	0,981	0,987
Edad de la mujer entrevistada					
Edad mujer	-0,038	0,008	-4,855	1,21e-06	0,963
Nivel de estudios de la mujer entrevistada (categoría de referencia: Sin estudios)					
Primaria	-0,17	0,189	-0,899	0,369	0,844
Secundaria 1ª etapa	-0,185	0,202	-0,915	0,360	0,831
Secundaria 2ª etapa	-0,232	0,220	-1,053	0,292	0,793
F.P.	-0,246	0,212	-1,158	0,247	0,782
Superiores	-0,225	0,218	-1,034	0,300	0,798
Edad de la pareja					
Edad pareja	0,022053	0,007510	2,937	0,003	1,022
Nivel de estudios de la pareja (categoría de referencia: Sin estudios)					
Primaria	-0,471	0,180	-2,612	0,009	0,625
Secundaria 1ª etapa	-0,633	0,192	-3,292	0,001	0,531
Secundaria 2ª etapa	-0,827	0,209	-3,966	7,31e-05	0,437
F.P.	-0,699	0,201	-3,471	0,001	0,497
Superiores	-1,035	0,209	-4,959	7,09e-07	0,355
Otros	0,898	1,431	0,628	0,530	2,456

Tabla 4.9: Modelo de regresión logística para mujeres que han sufrido violencia de género con función *glm*

Para poder validar el modelo se han calculado los p-valores asociados a los estadísticos de Pearson y de devianza, los valores obtenidos han sido 1 y 0,356 respectivamente. En los dos casos significativos por lo tanto, se acepta el modelo estimado como válido. También se ha realizado el contraste de Hosmer-Lemeshow y el p-valor obtenido ha sido 0,0713 no significativo con un nivel de confianza del 5 %, por lo tanto, se acepta la validación del modelo.

Se muestran los parámetros significativos al 5% resaltados en negrita, estos resultan ser relevantes para explicar el sufrimiento o no de violencia de género de las mujeres que fueron entrevistadas. Por lo que, se puede concluir que si la mujer trabaja o colabora de manera habitual en el negocio familiar, es jubilada o pensionista o esta parada y busca su primer empleo son características de riesgo para sufrir violencia de género respecto a la categoría de referencia de cada una, cómo también lo es su edad. Por otro lado, la edad de la pareja, si la pareja tiene estudios de primaria, secundaria primera etapa, secundaria segunda etapa, FP, o superiores son características de riesgo respecto las categorías de referencia que tienen los agresores de violencia de género en el contexto de pareja.

Atendiendo a los valores de los OR obtenidos, se puede observar que si la edad de la pareja aumenta, la *odds* de que la mujer sufra violencia de género también lo hará. Además, se puede destacar que, la *odds* de que la mujer sufra violencia de género es 2,23 veces mayor si esta trabaja o colabora de manera habitual en el negocio familiar respecto a la categoría de referencia, en este caso sí trabaja.

A continuación se muestran los *odds ratio* obtenidos en el modelo de forma gráfica, aquellos que están coloreados de un azul claro son aquellos mayores a 1, en caso contrario, los que están coloreados de un color azul más oscuro son aquellos *odds ratio* menores a 1.

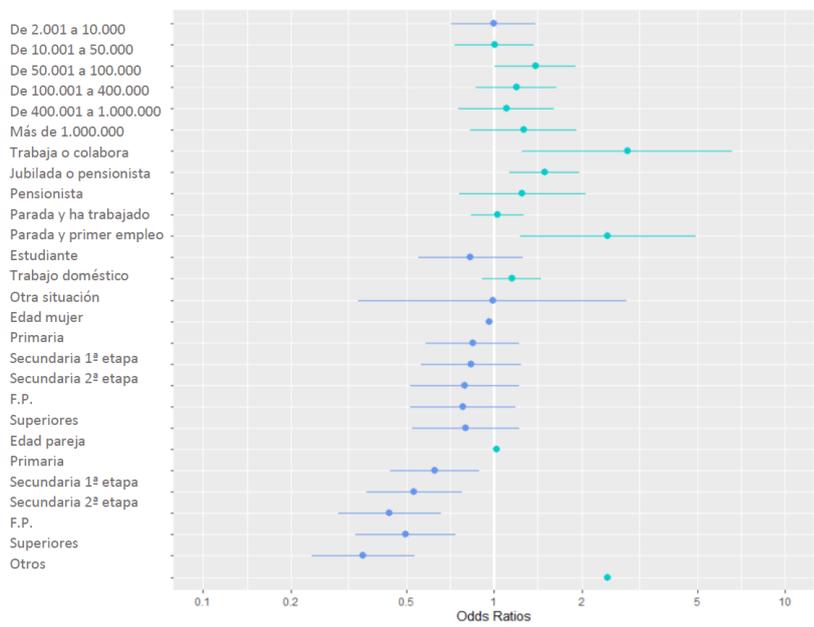


Figura 4.20: Odds Ratio del modelo de regresión logística estimado

4.2.2. Modelo de regresión Log-binomial

A continuación se presenta la estimación del modelo de regresión log-binomial para, al igual que en el caso anterior, estudiar las características que presentan las mujeres que han sufrido alguna vez violencia de género.

Como indicaba Laura Julià en su Trabajo de Final de Grado (2020) [7], en el caso de estimar un modelo de regresión log-binomial con la función *glm* se debe especificar un valor inicial de la variable constante con un -4 y 0 para las variables explicativas.

Para ajustar este modelo se han utilizado las mismas variables que se habían obtenido en el modelo de regresión logística una vez seguidos los pasos del libro de Hosmer, D. and S. Lemeshow (2000) [10].

La expresión del modelo de regresión log-binomial sería:

$$\eta = g(\hat{p}_x) = \log(\hat{p}_x) = \alpha + \hat{\beta}_1 X_1 + \dots + \hat{\beta}_{27} X_{27}$$

Donde X_1, \dots, X_{27} corresponden a las variables que se estudian, de la misma forma que en el caso anterior. Para aquellas variables categóricas se muestran $n - 1$ variables ficticias. Y $\hat{\beta}_1, \dots, \hat{\beta}_{27}$ las estimaciones obtenidas

En la Tabla 4.10 que se muestra seguidamente, se muestra el modelo obtenido, la estimación de los parámetros de este, los errores estándares, los valores Z , los p-valores y, por último el riesgo relativo (RR) estimado.

Variable	$\hat{\beta}$	$s.e(\hat{\beta})$	Z	p-valor	\widehat{RR}
Constante	-0,879	0,246	-3,572	3,55e-04	0,415
Tamaño del municipio (categoría referencia: ≤ 2.000)					
De 2.001 a 10.000	-0,007	0,144	-0,046	0,963	0,993
De 10.001 a 50.000	-0,003	0,135	-0,025	0,980	0,997
De 50.001 a 100.000	0,257	0,137	1,878	0,06	1,293
De 100.001 a 400.000	0,146	0,138	1,057	0,291	1,157
De 400.001 a 1.000.000	0,074	0,165	0,448	0,655	1,077
Más de 1.000.000	0,197	0,180	1,094	0,274	1,218
Situación laboral de la mujer entrevistada (categoría referencia: Trabaja)					
Trabaja o colabora	0,804	0,313	2,574	0,01	2,236
Jubilada o pensionista	0,323	0,12	2,699	0,007	1,381
Pensionista	0,193	0,214	0,901	0,367	1,213
Parada y ha trabajado	0,034	0,09	0,375	0,707	1,034
Parada y primer empleo	0,651	0,234	2,786	0,005	1,917
Estudiante	-0,137	0,174	-0,774	0,439	0,872
Trabajo doméstico	0,123	0,099	1,248	0,212	1,131
Otra situación	0,018	0,460	0,040	0,968	1,018
Edad de la mujer entrevistada					
Edad mujer	-0,032	0,006	-5,417	6,05e-08	0,969
Nivel de estudios de la mujer entrevistada (categoría de referencia: Sin estudios)					
Primaria	-0,091	0,15	-0,611	0,541	0,913
Secundaria 1ª etapa	-0,09	0,159	-0,565	0,572	0,914
Secundaria 2ª etapa	-0,129	0,176	-0,734	0,463	0,879
F.P.	-0,149	0,169	-0,881	0,378	0,862
Superiores	-0,140	0,175	-0,803	0,422	0,869
Edad de la pareja					
Edad pareja	0,019	0,006	3,278	0,001	1,019
Nivel de estudios de la pareja (categoría de referencia: Sin estudios)					
Primaria	-0,4	0,138	-2,901	0,004	0,67
Secundaria 1ª etapa	-0,543	0,147	-3,699	2e-04	0,581
Secundaria 2ª etapa	-0,708	0,163	-4,343	1,41e-05	0,492
F.P.	-0,591	0,156	-3,792	1e-04	0,554
Superiores	-0,883	0,164	-5,371	7,83e-08	0,414
Otros	0,511	0,714	0,715	0,475	1,666

Tabla 4.10: Modelo de regresión Log-binomial para mujeres que han sufrido violencia de género con función *glm*

Para la validación del modelo de regresión log-binomial se ha llevado a cabo el contraste de Hosmer-Lemeshow, el p-valor obtenido en este contraste ha sido 0,457, no significativo a un nivel de confianza del 5%. Por consiguiente, se considera el modelo como bueno.

Los parámetros significativos se han resaltado, de la misma forma que en el modelo anterior en negrita.

De esta forma las variables que se consideran relevantes para la explicación del modelo son la edad de la mujer entrevistada, si la mujer trabaja o colabora de manera habitual en el negocio familiar, si es jubilada o pensionista (anteriormente ha trabajado) o bien, está parada y busca su primer empleo. También se considera relevante para el modelo la edad de la pareja así como, los estudios que tiene la pareja.

Por lo que corresponde a los valores obtenidos del riesgo relativo estimado, se puede concluir que para dos mujeres que tengan las mismas características, la probabilidad de sufrir violencia de género por parte de su pareja es 1,019 (fijándonos en el valor de la relación relativa estimada asociada a la edad de la pareja) mayor para las mujeres que tengan parejas con una unidad mayor de edad.

En la Figura 4.21 que se muestra a continuación, se presentan de forma gráfica los valores obtenidos del riesgo relativo del modelo estimado. De igual manera que en el caso anterior, aquellos riesgos relativos que están coloreados de un azul claro son aquellos mayores a 1, en caso contrario, los que están coloreados de un color azul más oscuro son menores a 1.

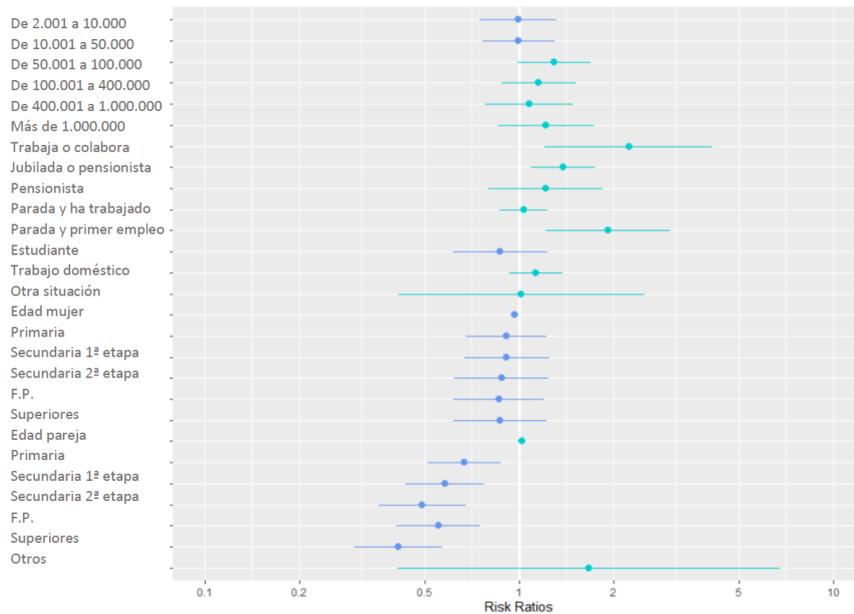


Figura 4.21: Riesgo Relativo del modelo de regresión Log-binomial

4.2.3. Modelo Poisson cero-inflado

Finalmente, se presenta el modelo Poisson cero-inflado. A diferencia de los modelos presentados anteriormente se selecciona la variable que muestra el número de veces que una mujer ha sufrido violencia como variable respuesta mientras que las variables explicativas seguirán siendo las mismas que en los dos casos anteriores.

Para este último modelo se han seguido los mismos pasos que se siguieron en el trabajo del primer cuatrimestre, del cuarto curso, de la asignatura de Modelos Lineales Generalizados del grado de Estadística (2020-21).

Primero de todo se ha estimado un modelo Poisson para poder demostrar que existía sobre dispersión con la función *dispersiontest* del paquete *AER* (Christian Kleiber and Achim Zeileis (2008)) [15], para poder justificar la utilización del modelo Poisson cero-inflado. A continuación se muestra la utilización de la función mencionada.

```
dispersiontest(object, trafo = NULL, alternative = c("greater", "two.sided", "less"))
```

Figura 4.22: Función *dispersiontest* de la librería *AER* en R

En el caso de añadir *trafo=1* se estará comprobando si existe relación lineal, en cambio, si se utiliza *trafo=2* se estará comprobando la relación cuadrática. Por otra parte, si *alternative = "less"* corresponde a la "underdispersion", si *alternative = "greater"* corresponde a la sobre dispersión y, finalmente, *alternative = "two.sided"* a cualquiera de las dos.

En el modelo que se está estimando se han obtenido unos p-valores $< 2, 2^{-16}$ tanto en el caso de la relación cuadrática como el de la relación lineal, por lo tanto es estadísticamente significativa y se puede concluir que existe sobre dispersión en el modelo.

A continuación se muestra el modelo Poisson cero-inflado obtenido, cabe destacar que no se ha incluido en el modelo la variable correspondiente a los estudios de la pareja ni si la persona convive o no con su pareja, ya que, los modelos cero-inflados son muy sensibles y difíciles para que la maximización de la función de verosimilitud converja. En el caso de añadir las variables mencionadas anteriormente se producían problemas por esto mismo y se decidió no tenerla en cuenta para este modelo.

Como ya se había explicado en el Apartado 4.2.3 del presente trabajo, el modelo Poisson cero-inflado combina dos modelos un modelo binomial con link *logit* y, un modelo Poisson.

Primeramente se muestran en la Tabla 4.11 los resultados obtenidos para el modelo Poisson cero-inflado con distribución **Binomial** y link **logit**.

Variable	$\hat{\beta}$	$s.e(\hat{\beta})$	Z	p-valor	\widehat{OR}
Constante	1,421	0,457	3,109	0,002	4,14
Tamaño del municipio (categoría referencia: ≤ 2.000)					
De 2.001 a 10.000	0,006	0,172	0,034	0,973	1,006
De 10.001 a 50.000	-0,011	0,161	-0,066	0,947	0,989
De 50.001 a 100.000	-0,293	0,165	-1,772	0,076	0,746
De 100.001 a 400.000	-0,116	0,165	-0,704	0,481	0,891
De 400.001 a 1.000.000	-0,046	0,196	-0,232	0,816	0,956
Más de 1.000.000	-0,122	0,216	-0,565	0,572	0,885
Edad de la mujer entrevistada					
Edad mujer	0,032	0,008	3,935	8,32e-05	1,033
Nivel de estudios de la mujer entrevistada (categoría de referencia: Sin estudios)					
Primaria	0,442	0,168	2,634	0,008	1,556
Secundaria 1ª etapa	0,579	0,175	3,310	0,001	1,784
Secundaria 2ª etapa	0,716	0,191	3,740	2e-04	2,046
F.P.	0,656	0,186	3,535	4e-04	1,927
Superiores	0,752	0,182	4,123	3,73e-05	2,121
Situación laboral de la mujer entrevistada (categoría referencia: Trabaja)					
Trabaja o colabora	-1,665	0,64	-2,603	0,009	0,189
Jubilada o pensionista	-0,396	0,152	-2,605	0,009	0,673
Pensionista	-0,236	0,260	-0,906	0,365	0,79
Parada y ha trabajado	0,003	0,11	0,032	0,975	1,004
Parada y primer empleo	-0,887	0,381	-2,326	0,02	0,412
Estudiante	0,091	0,256	0,354	0,723	1,095
Trabajo doméstico	-0,109	0,123	-0,883	0,377	0,897
Otra situación	0,034	0,548	0,062	0,951	1,034
Edad de la pareja					
Edad pareja	-0,023	0,008	-2,916	0,004	0,977
Sexo de la pareja					
Mujer	0,214	0,404	0,529	0,597	1,238
Situación laboral de la pareja (categoría referencia: Trabaja)					
Trabaja o colabora	0,257	0,765	0,337	0,736	1,294
Jubilada o pensionista	-0,036	0,145	-0,251	0,802	0,964
Pensionista	-0,783	0,692	-1,131	0,258	0,457
Parada y ha trabajado	-0,4	0,152	-2,636	0,008	0,670
Parada y primer empleo	-0,181	0,724	-0,249	0,803	0,835
Estudiante	-0,152	0,339	-0,449	0,653	0,859
Trabajo doméstico	-0,473	0,823	-0,575	0,565	0,623
Otra situación	-0,475	0,478	-0,993	0,321	0,622

Variable	$\hat{\beta}$	$s.e(\hat{\beta})$	Z	p-valor	\widehat{OR}
Situación legal con la pareja actual (categoría referencia: Casada)					
Pareja de hecho registrada	-0,027	0,202	-0,136	0,892	0,973
Pareja sin vínculos legales	-0,193	0,122	-1,580	0,114	0,824
Tiempo de relación con la pareja actual (categoría referencia: Menos de 6 meses)					
Entre 6 meses y 1 año	-0,384	0,403	-0,955	0,34	0,681
Más de 1 año menos de 4	-0,465	0,351	-1,324	0,185	0,628
Entre 4 y 6 años	-0,337	0,363	-0,929	0,353	0,714
Entre 7 y 10 años	-0,692	0,359	-1,928	0,054	0,501
Entre 11 y 20 años	-0,443	0,361	-1,226	0,220	0,642
Entre 21 y 30 años	-0,399	0,378	-1,057	0,290	0,671
Más de 30 años	-0,32	0,39	-0,820	0,412	0,726

Tabla 4.11: Modelo de Poisson cero-inflado para mujeres que han sufrido violencia de género con función *zeroinfl* con link *logit* y distribución Binomial

Por último se presenta en la Tabla 4.12 el segundo modelo por el que esta compuesto el Modelo cero-inflado estimado, un modelo con distribución **Poisson** y link **log**.

Variable	$\hat{\beta}$	$s.e(\hat{\beta})$	Z	p-valor	$\exp(\hat{\beta})$
Constante	1,054	0,228	4,620	3,84e-06	2,868
Tamaño del municipio (categoría referencia: ≤ 2.000)					
De 2.001 a 10.000	-0,05	0,0872	-0,568	0,57	0,952
De 10.001 a 50.000	-0,108	0,0817	-1,316	0,188	0,898
De 50.001 a 100.000	-0,052	0,082	-0,628	0,530	0,95
De 100.001 a 400.000	0,072	0,082	0,873	0,383	1,074
De 400.001 a 1.000.000	0,019	0,099	0,189	0,85	1,019
Más de 1.000.000	0,159	0,108	1,470	0,142	1,172
Edad de la mujer entrevistada					
Edad mujer	-0,007	0,004	-1,891	0,059	0,993
Nivel de estudios de la mujer entrevistada (categoría de referencia: Sin estudios)					
Primaria	-0,024	0,073	-0,330	0,742	0,976
Secundaria 1ª etapa	0,004	0,080	0,048	0,96	1,004
Secundaria 2ª etapa	0,075	0,091	0,824	0,410	1,078
F.P.	-0,24	0,089	-2,700	0,007	0,787
Superiores	-0,374	0,089	-4,206	2,60e-05	0,688
Situación laboral de la mujer entrevistada (categoría referencia: Trabaja)					
Trabaja o colabora	-1,126	0,386	-2,916	0,004	0,324
Jubilada o pensionista	-0,147	0,076	-1,934	0,053	0,864
Pensionista	-0,071	0,124	-0,575	0,566	0,931
Parada y ha trabajado	0,022	0,059	0,370	0,711	1,022
Parada y primer empleo	-0,027	0,170	-0,158	0,874	0,973
Estudiante	-0,202	0,149	-1,356	0,175	0,817
Trabajo doméstico	0,171	0,061	2,820	0,005	1,187
Otra situación	0,336	0,244	1,374	0,16938	1,399
Edad de la pareja					
Edad pareja	0,007	0,003	1,874	0,061	1,007
Sexo de la pareja					
Sexo Mujer	-0,324	0,2686	-1,207	0,227	0,723
Situación laboral de la pareja (categoría referencia: Trabaja)					
Trabaja o colabora	0,090	0,404	0,223	0,823	1,094
Jubilada o pensionista	0,132	0,074	1,795	0,073	1,141
Pensionista	0,420	0,272	1,542	0,123	1,522
Parada y ha trabajado	0,001	0,078	0,011	0,991	1,001
Parada y primer empleo	0,608	0,268	2,270	0,023	1,836
Estudiante	-0,732	0,233	-3,148	0,002	0,481
Trabajo doméstico	0,144	0,408	0,354	0,724	1,155
Otra situación	0,405	0,19	2,134	0,033	1,499

Variable	$\hat{\beta}$	$s.e(\hat{\beta})$	Z	p-valor	$\exp(\hat{\beta})$
Situación legal con la pareja actual (categoría referencia: Casada)					
Pareja de hecho registrada	0,005	0,119	0,041	0,967	1,005
Pareja sin vínculos legales	0,355	0,064	5,528	3,23e-08	1,426
Tiempo de relación con la pareja actual (categoría referencia: Menos de 6 meses)					
Entre 6 meses y 1 año	-0,200	0,21	-0,957	0,339	0,818
Más de 1 año menos de 4	-0,148	0,182	-0,811	0,417	0,862
Entre 4 y 6 años	-0,007	0,187	-0,036	0,971	0,993
Entre 7 y 10 años	0,060	0,183	0,328	0,743	1,062
Entre 11 y 20 años	0,011	0,186	0,058	0,953	1,011
Entre 21 y 30 años	0,011	0,197	0,055	0,956	1,011
Más de 30 años	0,404	0,202	2,002	0,045	1,498

Tabla 4.12: Modelo de Poisson cero-inflado para mujeres que han sufrido violencia de género con función *zeroinfl* con link *log* y distribución Poisson

Para poder validar el modelo se debe ver como se comportan los residuos de Pearson, como se trata de un modelo Poisson cero-inflado no es posible graficar directamente los residuos de Pearson, así pues, estos fueron graficados de manera manual, frente a los valores predichos.

En consecuencia se acepta la validación del modelo, debido a que muchos de los residuos se encuentran cerca de cero, aunque el modelo predice de forma incorrecta algunos valores, ya que, hay algunos residuos más elevados cuando el predictor lineal toma valores bajos.

Las variables relevantes para explicar el modelo, en la parte del modelo estimado con distribución binomial y link *logit*, son si la mujer entrevistada trabaja o colabora de manera habitual en el negocio familiar, si es jubilada o pensionista, o bien si está parada y busca su primer empleo, del mismo modo que su edad también es una variable relevante para la explicación del modelo así como, todos los tipos de estudios que pueda tener la mujer. Además también es destacado la edad de la pareja y finalmente, si está parada y ha trabajado antes.

Prestando atención a los valores que se han obtenido del **odds ratio** (OR) y dando como ejemplo la edad de la mujer, se puede concluir que si la edad aumenta, la **odds** de que la mujer sufra violencia de género también aumentará.

Para la parte del modelo estimado con distribución Poisson y link *log*, se puede concluir que si la mujer trabaja o colabora de manera habitual en el negocio familiar, es estudiante o bien ejerce trabajo doméstico no remunerado así como si tiene los estudios de F.P. o bien superiores son características relevantes. En cuanto a las factores propios de la pareja, relevantes para la explicación del modelo son su edad, si esta parado y busca su primer empleo, si es estudiante o bien, otra situación. Además también son factores importantes si la situación legal que tienen ambos es sin vínculos legales y si tienen una relación de más de 30 años.

Observando los valores obtenidos para la exponencial de los parámetros estimados, centrándonos en el resultado obtenido en el caso de que la mujer ejerza trabajo domestico el número de situaciones de maltrato que la mujer sufrirá aumentará 1,187 unidades respecto la categoría de referencia, si trabaja. De igual forma, si ambos tienen una relación sin vínculos legales el número de acontecimientos de violencia que podría sufrir aumentaría 1,426 unidades respecto si esta está casada.

Capítulo 5

Discusión y conclusiones

El trabajo presentado expone una estimación de distintos modelos estadísticos para la detección de las características más relevantes de las mujeres que sufrieron violencia de género en el año 2019. Primero de todo se ha ajustado un modelo de regresión logística utilizado para saber cómo están relacionadas las variables regresoras en la probabilidad de ocurrencia de un suceso, en el caso de este estudio, si la mujer ha sufrido violencia de género. A continuación se ha planteado un modelo de regresión log-binomial empleado para estudios con el mismo fin que los modelos de regresión logística. Ambos modelos se han ajustado con los mismos parámetros ya que la diferencia entre los dos rige en la función link, para los modelos de regresión logística la función link será *logit* en cambio, en el caso de los modelos log-binomial es *logaritmo*. Por lo que la interpretación de la medida de asociación entre el suceso y las variables regresoras (*odds ratio*) es complicada en el caso de modelos logísticos y puede llegarse a una estimación del riesgo relativo errónea. Y por último, se ha ajustado un modelo Poisson cero-inflado en el que la variable respuesta es una variable numérica y donde ésta, presenta más ceros de los que debería.

Para llevar a cabo este estudio se ha utilizado una base de datos del Centro de Investigaciones Sociológicas (CIS). Se disponía de 1121 variables y 9568 observaciones, por lo que se tuvo que hacer un análisis exhaustivo de esta base de datos para centrar la atención en el objetivo de este trabajo. Por este motivo, se crearon nuevas variables, otras se recodificaron y se eliminaron las que no se consideraban necesarias.

Antes de poder construir los modelos estadísticos se realizó un análisis descriptivo de las variables que se consideraron de interés para poder determinar si se deberían ajustar distintos modelos según el tipo de violencia o bien, si se podía crear una nueva variable que describiera si la mujer ha sufrido violencia o no, independientemente del tipo de esta. Después de realizar este análisis se concluyó que no había grandes diferencias que afectaran a las variables de interés según el tipo de violencia a la que había estado sometida la mujer, por eso, se crearon dos nuevas variables, una binomial y una tipo recuentos, para poder ajustar los modelos. Además, también se realizó un análisis descriptivo de aquellas mujeres que han sufrido alguna vez violencia de género por parte de alguna de sus parejas. Para estas, no se ha ajustado ningún tipo de modelo, ya que no se podría determinar en qué momento de su vida han estado expuestas a este tipo de

violencia, ni tampoco, por el número de parejas pasadas que la han padecido.

De este análisis descriptivo se han podido obtener resultados interesantes como que un 76,96 % de mujeres que han estado sometidas a alguna situación de control social también han sufrido insultos o humillaciones o bien, que de entre las mujeres que han padecido violencia física un 49,81 % también ha sufrido violencia sexual por parte de alguna de sus parejas pasadas. Del mismo modo que 23 mujeres que habían tenido pareja anteriormente, habían pasado por 7 de las situaciones descritas como violencia física.

Sin embargo, de entre las mujeres que sufrían control social por su pareja un 46 % también sufrió insultos o humillaciones. Mientras que de las mujeres que sufrieron violencia física, un 37,8 % también estuvieron expuestas a algún tipo de violencia sexual. Además 735 mujeres entrevistadas han pasado por al menos una de las situaciones descritas como control social.

Considerando las mujeres que tenían pareja en el momento en el que se realizó la encuesta los resultados del análisis descriptivo no han sido tan llamativos aunque no menos importantes. Esto podría explicarse, por el hecho de que si una mujer cuya pareja actual es su primera pareja, puede estar menos expuesta al hecho de sufrir violencia debido a que no habrá podido conocer más gente en el ámbito de relación de pareja y tampoco, cambiará de pareja si esta la trata bien.

Después de la estimación de los modelos estadísticos, en el caso de regresión logística y regresión log-binomial, se ha podido observar como ambos modelos han identificado las mismas variables que se relacionan con el hecho de que una mujer padezca violencia de género, a excepción que el modelo de regresión logística ha determinado que un tamaño de municipio de entre 50.001 a 100.000 habitantes puede estar relacionado con el sufrimiento de este tipo de maltrato mientras que el modelo log-binomial no lo ha identificado así. Observando los resultados obtenidos se puede concluir que la probabilidad de que una mujer sufra violencia de género aumenta si esta trabaja o colabora de manera habitual en el negocio familiar, es jubilada o pensionista o esta parada y busca su primer empleo. De igual manera, esta probabilidad también aumentará si la edad de la pareja también lo hace. Con los resultados obtenidos para la validación de ambos modelos, se puede concluir que el modelo log-binomial se ajusta mejor a los datos, por lo que en la práctica sería más conveniente escoger este modelo.

Por otro lado, teniendo en cuenta los resultados obtenidos para la estimación del modelo Poisson cero-inflado en la parte del modelo estimado con distribución Binomial, se concluye que la probabilidad de que una mujer sufra violencia de género aumenta cuando también lo hace la edad de esta. Del mismo modo, observando los resultados para el modelo estimado con distribución Poisson si la mujer realiza trabajo domestico no remunerado el número de situaciones de maltrato que vivirá aumentará en 1,187 unidades. De igual manera, si se trata de una pareja sin vínculos legales el número de acontecimientos de maltrato que vivirá la mujer aumentara 1,426 unidades y, en el caso de tener una relación de más de 30 años también hará que la cifra de situaciones de violencia de género aumente.

Con estos resultados se puede concluir finalmente que las características más relevantes que hacen más susceptible a que una mujer sufra violencia de género son si trabaja o colabora en el negocio familiar, si es jubilada o pensionista, si esta parada y busca su primer empleo, y la edad de su pareja.

Por último, es importante destacar que otras variables como el perfil psicológico y la personalidad del agresor, podrían haber sido más determinantes para saber que características llevan a una persona a ejercer este tipo de violencia sobre su pareja. Asimismo, sería fructífero seguir con este estudio y comparar los resultados obtenidos con los de la Macro-encuesta de Violencia de Género de España en el año 2020, año en el que tuvo lugar la pandemia ocasionada por el Covid-19 y es probable que esta situación haya llevado a un aumento de la violencia de género.

Recalcar que aunque la divulgación y concienciación en la sociedad sobre este tipo de maltrato cada vez es mayor, muchas mujeres siguen sufriendo violencia de género de forma diaria. Por este motivo, no se debe de dejar de prestar atención a situaciones como las descritas en este Trabajo de Final de Grado.

Bibliografía

- [1] *La violencia contra la mujer: definición - Delegación del Gobierno contra la Violencia de Género*. [Último acceso: 27.5.2021]. Ene. de 2020. URL: <https://violenciagenero.igualdad.gob.es/definicion/home.htm>.
- [2] *Violencia de género: Definición, tipos y causas | Grupo Atico34*. [Último acceso: 27.5.2021]. Dic. de 2020. URL: <https://protecciondatos-lopdc.com/empresas/violencia-de-genero>.
- [3] *¿Qué es la violencia de género? | Mujeres en Galicia*. [Último acceso: 27.5.2021]. Jun. de 2021. URL: <http://igualdade.xunta.gal/es/content/que-es-la-violencia-de-genero>.
- [4] *Fundación mujeres*. [Último acceso: 27.5.2021]. Abr. de 2008. URL: <http://www.fundacionmujeres.es/maletincoeducacion/pdf/CUAD5horiz.pdf>.
- [5] Hadley Wickham y Evan Miller. *haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. R package version 2.4.1. 2021. URL: <https://CRAN.R-project.org/package=haven>.
- [6] Henrik Bengtsson. *matrixStats: Functions that Apply to Rows and Columns of Matrices (and to Vectors)*. R package version 0.59.0. 2021. URL: <https://CRAN.R-project.org/package=matrixStats>.
- [7] Julià Melis, Laura. *El modelo de regresión log-binomial: una alternativa al modelo de regresión logística en estudios de cohortes y transversales*. [Último acceso: 25.06.2021]. Oct. de 2019. URL: http://diposit.ub.edu/dspace/bitstream/2445/142597/1/TFG_LauraJuli%c3%a0.pdf.
- [8] Salcedo Poma, Celia Mercedes. *Estimación de la ocurrencia de incidencias en declaraciones de pólizas de importación*. [Último acceso: 25.06.2021]. Ago. de 2004. URL: https://sisbib.unmsm.edu.pe/bibvirtualdata/Tesis/Basic/Salcedo_pc/enPDF/Cap2.PDF.
- [9] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2021. URL: <https://www.R-project.org/>.
- [10] D. Hosmer y S. Lemeshow. *Applied Logistic Regression*. 2.^a ed. John Wiley Sons, 2000.

- [11] José Manuel López Jiménez. *Análisis de la siniestralidad vial mediante modelos de datos de conteo*. [Último acceso: 25.06.2021]. Nov. de 2019. URL: <https://idus.us.es/bitstream/handle/11441/90014/L%C3%B3pez%20Jim%C3%A9nez%20Jos%C3%A9%20Manuel%20TFG.pdf?sequence=1&isAllowed=y>.
- [12] A. Colin Cameron y Pravin K. Trivedi. “Regression-based tests for overdispersion in the Poisson model”. En: *J. Econometrics* 46.3 (dic. de 1990), págs. 347-364. ISSN: 0304-4076. DOI: 10.1016/0304-4076(90)90014-K.
- [13] Martínez Calcaterra, Eloísa y Vaucher Silva, Pamela. *Una revisión de los modelos de conteo con excesos de ceros*. [Último acceso: 25.06.2021]. Sep. de 2020. URL: http://www.iesta.edu.uy/wp-content/uploads/2018/01/pasantia_martinez_voucher.pdf.
- [14] Achim Zeileis, Christian Kleiber y Simon Jackman. “Regression Models for Count Data in R”. En: *Journal of Statistical Software* 27.1 (jul. de 2008), págs. 1-25. ISSN: 1548-7660. DOI: 10.18637/jss.v027.i08.
- [15] Christian Kleiber y Achim Zeileis. *Applied Econometrics with R*. ISBN 978-0-387-77316-2. New York: Springer-Verlag, 2008. URL: <https://CRAN.R-project.org/package=AER>.

Apéndice A

Cuestionario Macro-encuesta de Violencia de Género contra la Mujer



*3235/0 MACROENCUESTA DE VIOLENCIA
CONTRA LA MUJER 2019*

*Estudio nº 3235
Septiembre 2019*

<u>Pregunta</u>	<u>Variable</u>	<u>Columnas</u>	<u>Título y categorías</u>
-	ESTUDIO	(1-4)	Código del estudio
-	REGISTRO	(5-10)	Número de registro
-	CUES	(11-15)	Nº de cuestionario
-	FECHAINI	(16-25)	Fecha de inicio de la entrevista
-	HORAINI	(26-33)	Hora de inicio de la entrevista
-	FECHAFIN	(34-43)	Fecha de fin de la entrevista
-	HORAFIN	(44-51)	Hora de fin de la entrevista
-	DURACION	(52-56)	Duración de la entrevista
-	CCAA	(57-58)	Comunidad autónoma 1.- Andalucía 2.- Aragón 3.- Asturias (Principado de) 4.- Balears (Illes) 5.- Canarias 6.- Cantabria 7.- Castilla-La Mancha 8.- Castilla y León 9.- Cataluña 10.- Comunitat Valenciana 11.- Extremadura 12.- Galicia 13.- Madrid (Comunidad de) 14.- Murcia (Región de) 15.- Navarra (Comunidad Foral de) 16.- País Vasco 17.- Rioja (La) 18.- Ceuta (Ciudad Autónoma de) 19.- Melilla (Ciudad Autónoma de)
-	PROV	(59-60)	Provincia (Código aparte)
-	MUN	(61-63)	Municipio (Código aparte)
-	DISTR	(64-66)	Distrito 0.- Anonimizada
-	SECCION	(67-69)	Sección 0.- Anonimizada
-	CAPITAL	(70)	Capital 1.- Capital de CC.AA. 2.- Capital de provincia 3.- Otros municipios
-	TAMUNI	(71)	Tamaño de municipio 1.- Menos o igual a 2.000 habitantes 2.- 2.001 a 10.000 habitantes 3.- 10.001 a 50.000 habitantes 4.- 50.001 a 100.000 habitantes 5.- 100.001 a 400.000 habitantes 6.- 400.001 a 1.000.000 habitantes 7.- Más de 1.000.000 habitantes
-	COORDINADOR	(72-76)	Coordinador/a 0.- Anonimizado
-	ENTREV	(77-81)	Nº de entrevistador/a 0.- Anonimizada
MOPO	MOPO	(82)	Situación laboral actual 1.- Trabaja 2.- Trabaja o colabora de manera habitual en el negocio familiar 3.- Jubilada o pensionista (anteriormente ha trabajado) 4.- Pensionista (anteriormente no ha trabajado) 5.- Parada y ha trabajado antes 6.- Parada y busca su primer empleo 7.- Estudiante 8.- Trabajo doméstico no remunerado 9.- Otra situación

<u>Pregunta</u>	<u>Variable</u>	<u>Columnas</u>	<u>Título y categorías</u>
MOP1	MOP1	(83)	Frecuencia de reunión con familiares o amigos 1.- Todos o casi todos los días 2.- Varias veces a la semana 3.- Varias veces al mes 4.- Una vez al mes 5.- Varias veces al año 6.- Una vez al año 7.- Menos de una vez al año 8.- Nunca 9.- N.C.
MOP2	MOP2	(84)	Persona de confianza con la que hablar 1.- Sí 2.- No 9.- N.C.
MOP3	MOP3	(85)	Familiares o amistades que le puedan acoger 1.- Sí 2.- No 9.- N.C.
MOP4	MOP4	(86)	Tenencia de hijos/as 1.- Sí 2.- No 9.- N.C.
MOP4a	MOP4A	(87)	Tenencia de hijos/as menores de edad 1.- Sí 2.- No 9.- N.C. 0.- N.P.
MOP4b	MOP4B	(88-89)	Número de hijos/as menores de edad 1.- 1 hijo/a 2.- 2 hijos/as 3.- 3 hijos/as 4-9.- De 4 a 9 hijos/as 10-98.- 10 y más hijos/as 99.- N.C. 0.- N.P.
MOP4c	MOP4C	(90-91)	Número de hijos/as menores de edad con quienes convive 0.- Ninguno 1.- 1 hijo/a 2.- 2 hijos/as 3.- 3 hijos/as 4-9.- De 4 a 9 hijos/as 10-96.- 10 y más hijos/as 99.- N.C. 97.- N.P.
MOP5	MOP5	(92)	Otra persona menor de edad con la que conviva 1.- Sí 2.- No 9.- N.C.
MOP5a	MOP5A	(93-94)	Número de personas menores de edad con las que convive 1.- 1 menor 2.- 2 menores 3.- 3 menores 4-9.- De 4 a 9 menores 10-98.- 10 y más menores 99.- N.C. 0.- N.P.
MOP6	MOP6	(95)	Estado de salud en el último año 1.- Muy bueno 2.- Bueno 3.- Regular 4.- Malo 5.- Muy malo 9.- N.C.

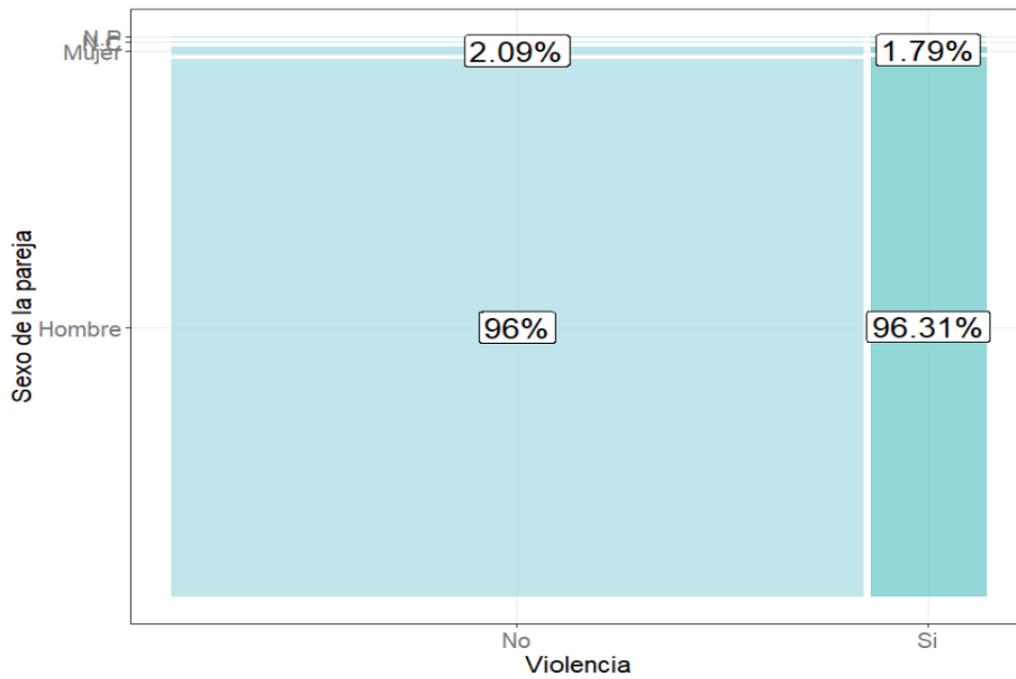
<u>Pregunta</u>	<u>Variable</u>	<u>Columnas</u>	<u>Título y categorías</u>
MOP7	MOP7_1, MOP7_2, MOP7_3, MOP7_4, MOP7_5, MOP7_6, MOP7_7, MOP7_8, MOP7_9	(96) a (104)	Dolencias que padece en el último año * Insomnio, falta de sueño * Fatiga permanente * Cambios de ánimo * Irritabilidad * Tristeza porque pensaba que no valía nada * Ganas de llorar sin motivos * Ansiedad o angustia * Ninguna de las anteriores * N.C. 1.- Menciona
MOP8	MOP8	(105)	Tenencia de certificado de discapacidad con grado igual o superior al 33% 1.- Sí 2.- No 9.- N.C.
MOP9	MOP9	(106)	Padecimiento y limitación de actividad de alguna dolencia, lesión o enfermedad que ha durado o durará más de 1 año 1.- Sí, y le limita gravemente 2.- Sí, pero no le limita gravemente 3.- No 9.- N.C.
MOP10	MOP10_1, MOP10_2, MOP10_3, MOP10_4, MOP10_5	(107) a (111)	Uso servicios sanitarios últimos 12 meses * Se ha visto obligada a quedarse algún día en la cama por motivos de salud * Ha acudido a un centro de salud, o médico general por motivos de su propia salud * Ha ingresado como paciente en un hospital al menos durante una noche (excluyendo partos o cesáreas) * Ha utilizado algún servicio de urgencias por algún problema o enfermedad suya * Ha visitado para Ud. misma a un/a psicólogo/a, psicoterapeuta o psiquiatra 1.- Sí 2.- No 9.- N.C.
MOP11	MOP11_1, MOP11_2, MOP11_3, MOP11_4, MOP11_9	(112) a (116)	Consumo de medicamentos en los últimos 12 meses * Tranquilizantes (para los nervios) * Antidepresivos (para mejorar el ánimo) * Analgésicos (calmante para los dolores) * Ninguno de los anteriores * N.C. 1.- Menciona
MOP12	MOP12	(117)	Tenencia pensamientos suicidas 1.- Sí 2.- No 9.- N.C.
MOP12a	MOP12A	(118)	Tenencia pensamientos suicidas en los últimos 12 meses 1.- Sí 2.- No 9.- N.C. 0.- N.P.
MOP13	MOP13	(119)	Intento de suicidio 1.- Sí 2.- No 9.- N.C. 0.- N.P.
MOP13a	MOP13A	(120)	Intento de suicidio en los últimos 12 meses 1.- Sí 2.- No 9.- N.C. 0.- N.P.
MOP14	MOP14	(121-122)	Nº de parejas que se han tenido a lo largo de la vida 0.- Ninguna 1.- 1 pareja 2.- 2 parejas 3.- 3 parejas 4-9.- De 4 a 9 parejas 10-98.- 10 y más parejas 99.- N.C.

<u>Pregunta</u>	<u>Variable</u>	<u>Columnas</u>	<u>Título y categorías</u>
MOP14a	MOP14A	(123)	Sexo de las parejas que se han tenido 1.- Solo hombres 2.- Solo mujeres 3.- Tanto hombres como mujeres 9.- N.C. 0.- N.P.
MOP15	MOP15	(124)	Tenencia de pareja 1.- Sí 2.- No 9.- N.C. 0.- N.P.
MOP16	MOP16	(125)	Convivencia con la pareja actual 1.- Sí, viven en el mismo domicilio 2.- Sí, pero por temporadas, de forma intermitente, los fines de semana 3.- No, viven en distintos domicilios 4.- Otros 9.- N.C. 0.- N.P.
MOP17	MOP17	(126)	Situación legal con la pareja actual 1.- Casada 2.- Pareja de hecho registrada 3.- Pareja sin vínculos legales 9.- N.C. 0.- N.P.
MOP18	MOP18	(127)	Tiempo de relación con la pareja actual (meses) 1.- Menos de 6 meses 2.- Entre 6 meses y 1 año 3.- Más de 1 año pero menos de 4 años 4.- Entre 4 y 6 años 5.- Entre 7 y 10 años 6.- Entre 11 y 20 años 7.- Entre 21 y 30 años 8.- Más de 30 años 9.- N.C. 0.- N.P.
M1P1	M1P1_0_1, M1P1_0_2, M1P1_0_3, M1P1_0_4, M1P1_0_5, M1P1_0_6, M1P1_0_7, M1P1_0_8, M1P1_0_9, M1P1_0_10, M1P1_0_11	(128) a (138)	Frecuencia con que se han sufrido diferentes comportamientos de la pareja actual: violencia psicológica (control social) * Trata o ha tratado de impedirle que vea a sus amigos o amigas * Trata o ha tratado de evitar que Ud. se relacione con su familia directa o parientes * Insiste o ha insistido en saber dónde está Ud. en cada momento * Le ignora o ha ignorado y le trata o ha tratado con indiferencia * Se enfada o se ha enfadado si habla con otro hombre o mujer * Sospecha o ha sospechado sin motivos que Ud. le es/era infiel * Espera o ha esperado que Ud. le pida permiso antes de ir por su cuenta a determinados sitios como por ejemplo un hospital o centro de salud, un centro cultural o deportivo, etc. * Se niega o se ha negado a darle dinero para los gastos del hogar cuando la pareja tiene/tenía dinero para otras cosas * Le impide o ha impedido tomar decisiones relacionadas con la economía familiar y/o realizar las compras de forma independiente * No le deja o no le ha dejado trabajar o estudiar fuera del hogar * Usa o ha usado su dinero o su tarjeta de crédito o pide préstamos a su nombre sin su consentimiento 1.- Frecuentemente 2.- Algunas veces 3.- Nunca 9.- N.C. 0.- N.P.

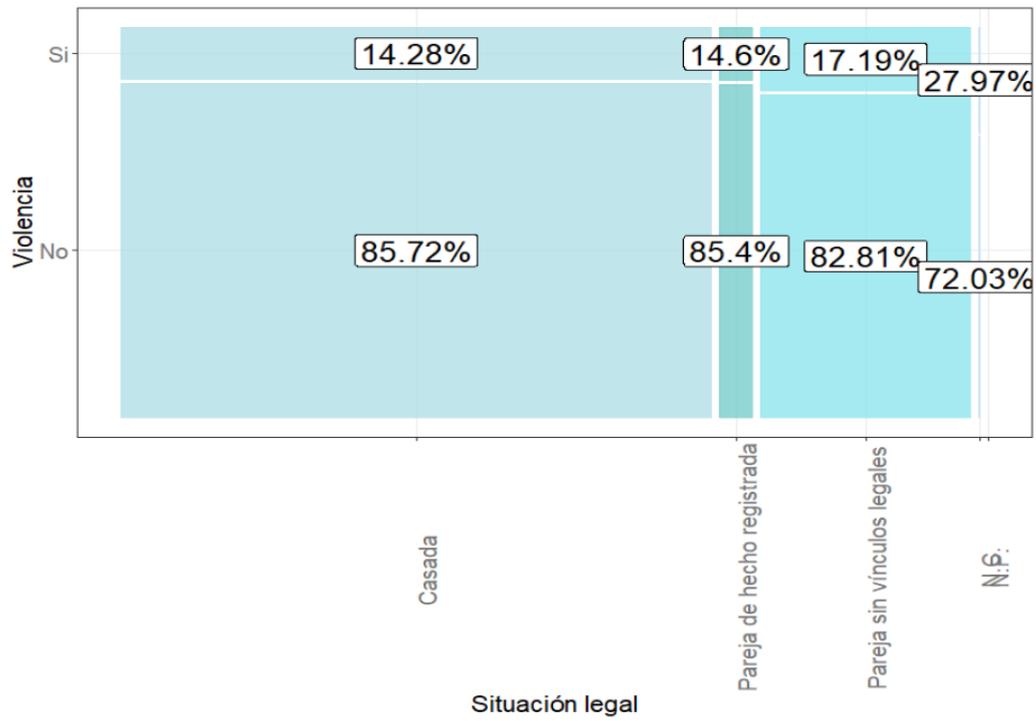
Apéndice B

Gráficos mencionadas

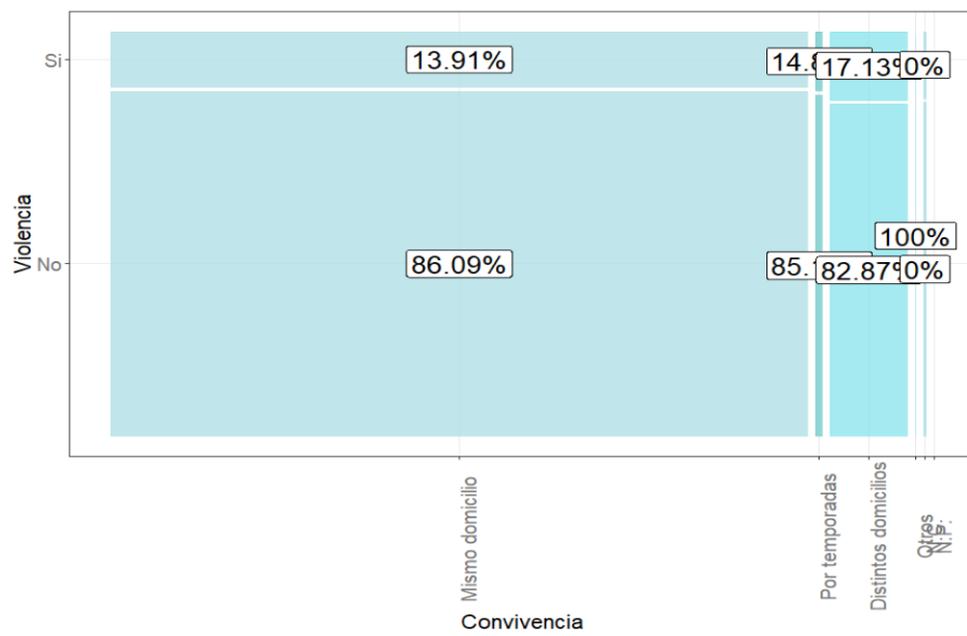
Sexo de la pareja



Situación legal con la pareja actual

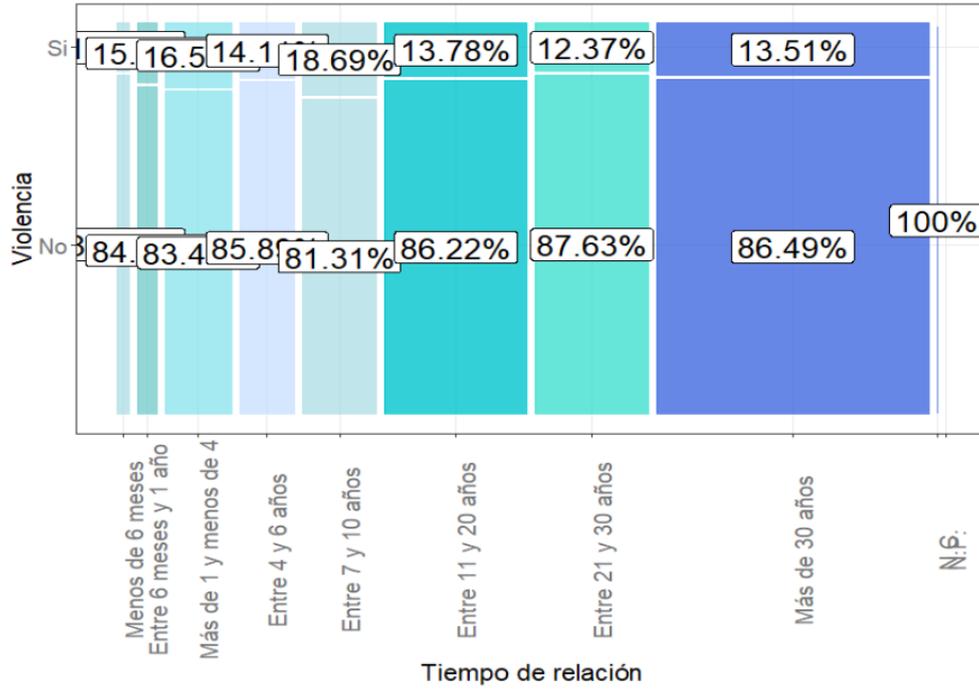


Convivencia con la pareja actual

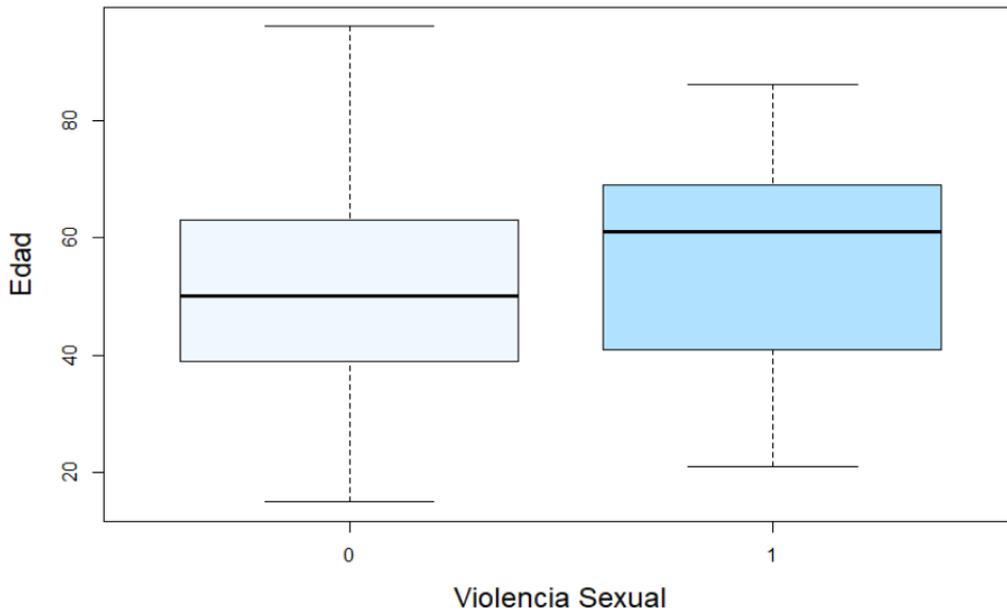


Tiempo de relación

Convivencia con la pareja actual



Edad de los agresores en violencia sexual



Apéndice C

Bondad de ajuste de los modelos estadísticos

Regresión logística (Test de Hosmer-Lemeshow)

Hosmer and Lemeshow Goodness-of-Fit Test

Call:

```
glm(formula = violencia.b ~ TAMUNI + MOPO + SDE1 + ESTUDIOS + SDP2 +  
     ESTUDIOSPAR, family = binomial(link = "logit"),  
     data = dd)  
ChiSquare  df      P_value  
14.42496   8      0.07133967
```

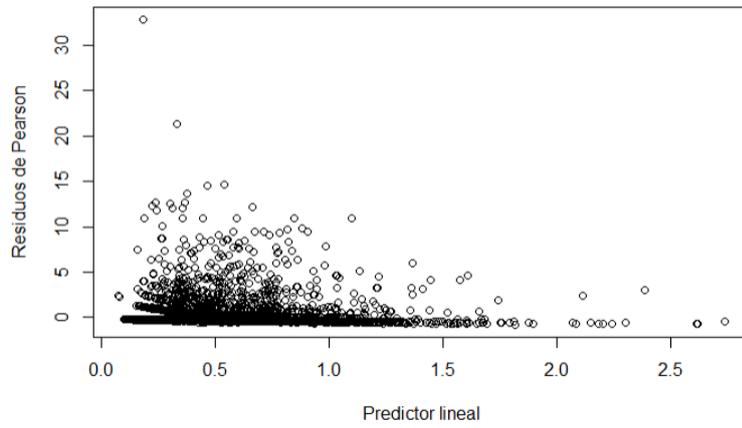
Regresión log-binomial (Test de Hosmer-Lemeshow)

Hosmer and Lemeshow Goodness-of-Fit Test

Call:

```
glm(formula = violencia.b ~ TAMUNI + MOPO + SDE1 + ESTUDIOS + SDP2 +  
     ESTUDIOSPAR, family = binomial(link = "log"),  
     data = dd, start = c(-4, rep(0, 27)))  
ChiSquare  df  P_value  
7.763256   8   0.4569312
```

Poisson Cero-inflado (Residuos de Pearson)



Apéndice D

Código R

A continuación se presenta el código R que se ha utilizado para la realización de este Trabajo de Final de Grado. Para la parte del análisis descriptivo se presentan sólo los códigos para el caso de parejas actuales ya que, en el caso de parejas pasadas se ha seguido el mismo procedimiento y, por lo tanto, solo cambian las variables. Además, para los códigos de los distintos modelos se presenta solo el modelo final y no todos los pasos propuestos por Hosmer-Lemeshow.

Librerías utilizadas

```
library(haven)
library(Epi)
library(descr)
library(matrixStats)
library(summarytools)
library(ggmosaic)
library(tidyverse)
library(ggpubr)
library(stats)
library(effects)
library(car)
library(fmsb)
library("ROCR")
library(pROC)
library("Epi")
library(vcdExtra)
library(AER)
library(foreign)
library(pscl)
library(sjPlot)
```

Lectura de datos en R

```
setwd("C:/TFG/scripts/Análisis uni-bivariante/PAREJA ACTUAL")
dd <- read_spss("C:/TFG/3235.sav")
dd <- as.data.frame(dd)
```

```
# Se transforman las variables a factor
n <- ncol(dd)
for (i in 1:n){
  dd[,i] <- as.factor(dd[,i])
}
```

Creación de las nuevas variables del tipo de violencia

```
# Función binaria
binaria <- function(x) {
  n <- length(x)
  v <- c()
  for (i in 1:n){
    if (x[i]==0){
      v <- c(v,0)
    }else{
      v <- c(v,1)
    }
  }
  v
}
```

```
# VIOLENCIA PSICOLOGICA - CONTROL SOCIAL
```

```
x <- c(summary(dd$M1P1_0_1)[1],summary(dd$M1P1_0_2)[1],summary(dd$M1P1_0_3)[1],
        summary(dd$M1P1_0_4)[1],summary(dd$M1P1_0_5)[1],summary(dd$M1P1_0_6)[1],
        summary(dd$M1P1_0_7)[1],summary(dd$M1P1_0_8)[1],summary(dd$M1P1_0_9)[1],
        summary(dd$M1P1_0_10)[1],summary(dd$M1P1_0_11)[1])
```

```
#pasar los NA a 0
```

```
haz.cero.na=function(x){
  ifelse(is.na(x),0,x)}
```

```
x =data.frame(sapply(dd[,60:70],haz.cero.na))
```

```
dd$viol.psico <- rowCounts(x==1) #contamos cuantas frecuentemente por filas
```

```
dd$viol.psico2 <- rowCounts(x==2) #contamos cuantas algunas veces por filas
```

```
y =data.frame(dd[,1122:1123]) # data frame de frecuentemente y algunas veces
```

```
dd$control.social <- rowSums(y) #sumamos los conteos de algunas veces y frecuentemente
```

```

## VIOLENCIA PSICOLOGICA - INSULTOS Y HUMILLACIONES
x <- c(summary(dd$M1P3_0_1)[1],summary(dd$M1P3_0_2)[1],summary(dd$M1P3_0_3)[1],
       summary(dd$M1P3_0_4)[1],summary(dd$M1P3_0_5)[1],summary(dd$M1P3_0_6)[1],
       summary(dd$M1P3_0_7)[1])

x <- data.frame(sapply(dd[,116:122],haz.cero.na)) #escogemos las columnas
                                                #que responden la pregunta

library(matrixStats)
dd$viol.fisica <- rowCounts(x==1) #contamos cuantas si por filas

## VIOLENCIA SEXUAL
x <- c(summary(dd$M1P5_0_1)[1],summary(dd$M1P5_0_2)[1],summary(dd$M1P5_0_3)[1],
       summary(dd$M1P5_0_4)[1],summary(dd$M1P5_0_5)[1],summary(dd$M1P5_0_6)[1],
       summary(dd$M1P5_0_7)[1],summary(dd$M1P5_0_8)[1])

x <- data.frame(sapply(dd[,139:146],haz.cero.na)) #escogemos las columnas
                                                # que responden la pregunta

library(matrixStats)
dd$viol.sexual <- rowCounts(x==1) #contamos cuantas si por filas

## VARIABLE BINARIA TOTAL DE VIOLENCIA
dd$violencia.b <- binaria (dd$violencia.n)
class(dd$violencia.b)
dd$violencia.b <- as.factor(dd$violencia.b)

## VARIABLE RECIENTOS TOTAL DE VIOLENCIA
t <- cbind(dd$control.social,dd$viol.fisica,dd$insultos,dd$viol.sexual)
dd$violencia.n <- rowSums(t)

```

Análisis descriptivo

```

##CONTROL SOCIAL vs. INSULTOS Y HUMILLACIONES
dd$`Control Social`<- dd$control.social.b
dd$`Insultos/Humillaciones`<- dd$insultos.b
dd$`Insultos/Humillaciones` <- as.factor(dd$`Insultos`)
dd$`Control Social` <- as.factor(dd$`Control Social`)
levels (dd$`Insultos/Humillaciones`) <- c("No","Si")
levels (dd$`Control Social`) <- c("No","Si")
dd$control.social.b <- as.factor(dd$control.social.b)

```

```

dd$insultos.b <- as.factor(dd$insultos.b)
print(ctable(x =dd$`Insultos/Humillaciones` , y = dd$`Control Social`, prop = "t"),
      method = "render")

## INSULTOS Y HUMILLACIONES vs. VIOLENCIA FISICA
dd$`Violencia Fisica`<- dd$viol.fisica.b
dd$`Insultos/Humillaciones`<- dd$insultos.b
dd$`Insultos/Humillaciones` <- as.factor(dd$`Insultos`)
dd$`Violencia Fisica` <- as.factor(dd$`Violencia Fisica`)
levels (dd$`Insultos/Humillaciones`) <- c("No","Si")
levels (dd$`Violencia Fisica`) <- c("No","Si")
print(ctable( y = dd$`Insultos/Humillaciones`,x = dd$`Violencia Fisica`, prop = "t"),
      method = "render")

## VIOLENCIA FÍSICA vs. VIOLENCIA SEXUAL
dd$viol.fisica.b <- as.factor(dd$viol.fisica.b)
dd$viol.sexual.b <- as.factor(dd$viol.sexual.b)
dd$`Violencia Fisica`<- dd$viol.fisica.b
dd$`Violencia Sexual`<- dd$viol.sexual.b
print(ctable(x = dd$`Violencia Sexual`, y = dd$`Violencia Fisica`, prop = "t"),
      method = "render")

## GRÁFICOS DE BARRAS DE LA VARIABLE DE RECUEENTOS DE VIOLENCIA
t <- summary(as.factor(dd$control.social))
bp1 <- plot(as.factor(dd$control.social),ylim=c(0,7000),col="aliceblue")
text(bp1,t,format(t), cex=1,pos=3)

t <- summary(as.factor(dd$insultos))
bp1 <- plot(as.factor(dd$insultos),ylim=c(0,7000),col="aliceblue")
text(bp1,t,format(t), cex=1,pos=3)

t <- summary(as.factor(dd$viol.fisica))
bp1 <- plot(as.factor(dd$viol.fisica),ylim=c(0,7000),col="aliceblue")
text(bp1,t,format(t), cex=1,pos=3)

t <- summary(as.factor(dd$viol.sexual))
bp1 <- plot(as.factor(dd$viol.sexual),ylim=c(0,7000),col="aliceblue")
text(bp1,t,format(t), cex=1,pos=3)

```

```

## GRAFICOS DE MOSAICO
## TAMAÑO MUNICIPIO

dd1 <- data.frame (dd$violencia.b,dd$TAMUNI)

levels(dd1$dd.violencia.b) <- c("No","Si")
levels(dd1$dd.TAMUNI) <- c("<= 2.000 ", "2.001 a 10.000", "10.001 a 50.000", "50.001 a
100.000", "100.001 a 400.000", "400.001 a 1.000.000", "> 1.000.000")
cbbPalette <- c("powderblue", "darkslategray3", "cadetblue2", "lightsteelblue1",
"powderblue1", "turquoise3", "turquoise", "royalblue", "royalblue")

p1<-ggplot(data=dd1)+
  geom_mosaic(aes(x=product(dd.violencia.b,dd.TAMUNI
),fill=dd.TAMUNI))+  scale_fill_manual(values = cbbPalette)

p1d<- ggplot_build(p1)$data %>% as.data.frame() %>% filter(.wt > 0)
head(p1d)
compt_perc=function(x){
  d=c(x,1)-c(0,x)
  d[-length(d)]
}
x=tapply(p1d$ymax,factor(p1d$fill,levels=unique(p1d$fill)),compt_perc)
x=unlist(x)
p1d$percentage=paste0(round(100*x,2),"%")
p2<-p1 +
  geom_label(data = p1d,
             aes(x = (xmin + xmax)/2,
                 y = (ymin + ymax)/2,
                 label = percentage),size = 7)
p2<-p2+ylab("Violencia")+xlab("Tamaño del municipio")+theme_bw()
p2<-p2+theme(legend.position="none")
p2 <- p2+  theme(axis.text.x = element_text(color="gray46",
                                             size=15, angle=90), axis.title.x = element_text(size=rel(1.5)) ,
                axis.text.y = element_text( color="gray46",
                                             size=15, angle=0), axis.title.y = element_text(size=rel(1.5)))
p2

```

```
## SITUACIÓN LABORAL
```

```
dd1 <- data.frame (DD$violencia.b,DD$MOP0)
```

```
levels(dd1$dd.violencia.b) <- c("No","Si")
```

```
levels(dd1$dd.MOP0) <- c("Trabaja ", "Trabaja/colabora",
```

```
"Jubilada o pensionista", "Pensionista (no ha trabajado)", "Parada",
```

```
"Parada y busca 1r empleo", "Estudiante", "Trabajo demestico", "Otros")
```

```
cbbPalette <- c("powderblue", "darkslategray3", "cadetblue2", "lightsteelblue1",
```

```
"powderblue", "turquoise3", "turquoise", "royalblue","lightseagreen")
```

```
library(ggmosaic)
```

```
p1<-ggplot(data=dd1)+
```

```
  geom_mosaic(aes(x=product(dd.violencia.b,dd.MOP0  
) ,fill=dd.MOP0))+  scale_fill_manual(values = cbbPalette)
```

```
p1d<- ggplot_build(p1)$data %>% as.data.frame() %>% filter(.wt > 0)
```

```
head(p1d)
```

```
compt_perc=function(x){
```

```
  d=c(x,1)-c(0,x)
```

```
  d[-length(d)]
```

```
}
```

```
x=tapply(p1d$ymax,factor(p1d$fill,levels=unique(p1d$fill)),compt_perc)
```

```
x=unlist(x)
```

```
p1d$percentage=paste0(round(100*x,2),"%")
```

```
p2<-p1 +
```

```
  geom_label(data = p1d,  
             aes(x = (xmin + xmax)/2,  
                 y = (ymin + ymax)/2,  
                 label = percentage),size = 7)
```

```
p2<-p2+xlab("Situación Laboral")+ylab("Violencia")+theme_bw()
```

```
p2<-p2+theme(legend.position="none")
```

```
p2 <- p2+  theme(axis.text.x = element_text(color="gray46",  
                                             size=15, angle=90), axis.title.x = element_text(size=rel(1.5)),  
                axis.text.y = element_text( color="gray46",  
                                             size=15, angle=0), axis.title.y = element_text(size=rel(1.5)))
```

```
p2
```

```

## NIVEL DE ESTUDIOS

dd1 <- data.frame (dd$ESTUDIOS,dd$violencia.b)

levels(dd1$dd.violencia.b) <- c("No","Si")
levels(dd1$dd.ESTUDIOS) <- c("Sin estudios ", "Primaria", "Secundaria 1ª etapa",
"Secundaria 2ª etapa", "F.P.", "Superiores", "Otros", "N.S.", "N.C.")
cbbPalette <- c("powderblue", "darkslategray3", "cadetblue2", "lightsteelblue1",
"powderblue", "turquoise3", "turquoise", "royalblue", "lightseagreen")

p1<-ggplot(data=dd1)+
  geom_mosaic(aes(x=product(dd.violencia.b,dd.ESTUDIOS
),fill=dd.ESTUDIOS))+ scale_fill_manual(values = cbbPalette)

pid<- ggplot_build(p1)$data %>% as.data.frame() %>% filter(.wt > 0)
head(pid)
compt_perc=function(x){
  d=c(x,1)-c(0,x)
  d[-length(d)]
}

x=tapply(pid$ymax,factor(pid$fill,levels=unique(pid$fill)),compt_perc)
x=unlist(x)
pid$percentage=paste0(round(100*x,2),"%")

p2<-p1 +
  geom_label(data = pid,
             aes(x = (xmin + xmax)/2,
                 y = (ymin + ymax)/2,
                 label = percentage),size = 7)
p2<-p2+ylab("Violencia")+xlab("Nivel de Estudios")+theme_bw()
p2<-p2+theme(legend.position="none")
p2 <- p2+ theme(axis.text.x = element_text(color="gray46",
                                             size=15, angle=90), axis.title.x = element_text(size=rel(1.5)),
               axis.text.y = element_text( color="gray46",
                                             size=15, angle=0), axis.title.y = element_text(size=rel(1.5)))

p2

```

```

## SEXO DE LA PAREJA

dd1 <- data.frame (dd$violencia.b,dd$SDP1)

levels(dd1$dd.violencia.b) <- c("No","Si")
levels(dd1$dd.SDP1) <- c("Hombre","Mujer","N.C","N.P")
cbbPalette <- c("powderblue", "darkslategray3", "cadetblue2", "lightsteelblue1",
"powderblue", "turquoise3", "turquoise", "royalblue","royalblue")

p1<-ggplot(data=dd1)+
  geom_mosaic(aes(x=product(dd.SDP1,dd.violencia.b
),fill=dd.violencia.b))+  scale_fill_manual(values = cbbPalette)

p1d<- ggplot_build(p1)$data %>% as.data.frame() %>% filter(.wt > 0)
head(p1d)
compt_perc=function(x){
  d=c(x,1)-c(0,x)
  d[-length(d)]
}

x=tapply(p1d$ymax,factor(p1d$fill,levels=unique(p1d$fill)),compt_perc)
x=unlist(x)
p1d$percentage=paste0(round(100*x,2),"%")

p2<-p1 +
  geom_label(data = p1d,
            aes(x = (xmin + xmax)/2,
               y = (ymin + ymax)/2,
               label = percentage),size = 7)
p2<-p2+ylab("Sexo de la pareja")+xlab("Violencia")+theme_bw()
p2<-p2+theme(legend.position="none")
p2 <- p2+  theme(axis.text.x = element_text(color="gray46",
                                             size=15, angle=0), axis.title.x =element_text(size=rel(1.5)),
               axis.text.y = element_text( color="gray46",
                                             size=15, angle=0), axis.title.x =element_text(size=rel(1.5)))
p2

```

```
## SITUACIÓN LEGAL
```

```
dd1 <- data.frame (dd$violencia.b,dd$MOP17)
```

```
levels(dd1$dd.violencia.b) <- c("No","Si")
```

```
levels(dd1$dd.MOP17) <- c("Casada ", "Pareja de hecho registrada",  
"Pareja sin vínculos legales", "N.C.", "N.P.")
```

```
cbbPalette <- c("powderblue", "darkslategray3", "cadetblue2", "lightsteelblue1",  
"powderblue", "turquoise3", "turquoise", "royalblue", "royalblue")
```

```
p1<-ggplot(data=dd1)+
```

```
  geom_mosaic(aes(x=product(dd.violencia.b,dd.MOP17),fill=dd.MOP17)) +  
  scale_fill_manual(values = cbbPalette)
```

```
p1d<- ggplot_build(p1)$data %>% as.data.frame() %>% filter(.wt > 0)
```

```
head(p1d)
```

```
compt_perc=function(x){  
  d=c(x,1)-c(0,x)  
  d[-length(d)]  
}
```

```
x=tapply(p1d$ymax,factor(p1d$fill,levels=unique(p1d$fill)),compt_perc)
```

```
x=unlist(x)
```

```
p1d$percentage=paste0(round(100*x,2),"%")
```

```
p2<-p1 +
```

```
  geom_label(data = p1d,  
            aes(x = (xmin + xmax)/2,  
                y = (ymin + ymax)/2,  
                label = percentage),size = 7)
```

```
p2<-p2+xlab("Violencia")+ylab("Situación legal")+theme_bw()
```

```
p2<-p2+theme(legend.position="none")
```

```
p2 <- p2+ theme(axis.text.x = element_text(color="gray46",  
                                             size=15, angle=90),axis.title.x = element_text(size=rel(1.5)),  
               axis.text.y = element_text( color="gray46",  
                                             size=15, angle=0),axis.title.y = element_text(size=rel(1.5)))
```

```
p2
```

```

## TIEMPO DE RELACIÓN

dd1 <- data.frame (dd$violencia.b,dd$MOP18)

levels(dd1$dd.violencia.b) <- c("No","Si")
levels(dd1$dd.MOP18) <- c("Menos de 6 meses","Entre 6 meses y 1 año",
"Más de 1 y menos de 4", "Entre 4 y 6 años","Entre 7 y 10 años",
"Entre 11 y 20 años","Entre 21 y 30 años","Más de 30 años","N.C.,""N.P.")
cbbPalette <- c("powderblue", "darkslategray3", "cadetblue2", "lightsteelblue1",
"powderblue", "turquoise3", "turquoise", "royalblue","royalblue","lightskyblue3")

library(ggmosaic)
p1<-ggplot(data=dd1)+
  geom_mosaic(aes(x=product(dd.violencia.b,dd.MOP18),fill=dd.MOP18))+
  scale_fill_manual(values = cbbPalette)

p1d<- ggplot_build(p1)$data %>% as.data.frame() %>% filter(.wt > 0)
head(p1d)
compt_perc=function(x){
  d=c(x,1)-c(0,x)
  d[-length(d)]
}

x=tapply(p1d$ymax,factor(p1d$fill,levels=unique(p1d$fill)),compt_perc)
x=unlist(x)
p1d$percentage=paste0(round(100*x,2),"%")

p2<-p1 +
  geom_label(data = p1d,
             aes(x = (xmin + xmax)/2,
                 y = (ymin + ymax)/2,
                 label = percentage),size = 7)
p2<-p2+xlab("Tiempo de relación")+ylab("Control Social")+theme_bw()
p2<-p2+theme(legend.position="none")
p2 <- p2+ theme(axis.text.x = element_text(color="gray46",
                                             size=15, angle=90),axis.title.x = element_text(size=rel(1.5)),
               axis.text.y = element_text( color="gray46",
                                             size=15, angle=0),axis.title.y = element_text(size=rel(1.5)),)
p2

```

```

## CONVIVENCIA CON LA PAREJA

dd1 <- data.frame (dd$violencia.b,dd$MOP16)

levels(dd1$dd.violencia.b) <- c("No","Si")
levels(dd1$dd.MOP16) <- c("Mismo domicilio ","Por temporadas",
"Distintos domicilios", "Otros","N.C. ","N.P.")
cbbPalette <- c("powderblue", "darkslategray3", "cadetblue2", "lightsteelblue1",
"powderblue", "turquoise3", "turquoise", "royalblue","royalblue")

p1<-ggplot(data=dd1)+
  geom_mosaic(aes(x=product(dd.violencia.b,dd.MOP16),fill=dd.MOP16))+
  scale_fill_manual(values = cbbPalette)

p1d<- ggplot_build(p1)$data %>% as.data.frame() %>% filter(.wt > 0)
head(p1d)
compt_perc=function(x){
  d=c(x,1)-c(0,x)
  d[-length(d)]
}

x=tapply(p1d$ymax,factor(p1d$fill,levels=unique(p1d$fill)),compt_perc)
x=unlist(x)
p1d$percentage=paste0(round(100*x,2),"%")

p2<-p1 +
  geom_label(data = p1d,
             aes(x = (xmin + xmax)/2,
                 y = (ymin + ymax)/2,
                 label = percentage),size = 7)
p2<-p2+xlab("Convivencia")+ylab("Violencia")+theme_bw()
p2<-p2+theme(legend.position="none")
p2 <- p2+ theme(axis.text.x = element_text(color="gray46",
                                             size=15, angle=90),axis.title.x = element_text(size=rel(1.5)),
               axis.text.y = element_text( color="gray46",
                                             size=15, angle=0),axis.title.y = element_text(size=rel(1.5)),)

p2

```

```
## EDAD
```

```
dd$SDE1 <- as.numeric(dd$SDE1)
dd1 <- data.frame(dd$dd.SDE1 ,dd$violencia.b)
levels(dd1$dd.violencia.b) <- c("No","Si")

boxplot(dd1$.dd.SDE1~dd1$dd.violencia.b,
        xlab = "Violencia", ylab = "Edad", col= c("aliceblue","lightskyblue1"),
        cex.axis = 1, cex.lab = 1.5)
```

Modelos estadísticos

```
# SE ELIMINAN LOS VALORES DE 8-N.S, 9-N.C, 0-N.P
is.na(dd$SDE1) <- dd$SDE1 %in% c(99)
is.na(bbdd$ESTUDIOS) <- dd$ESTUDIOS %in% c(8,9)
is.na(dd$SDP2) <- dd$SDP2 %in% c(99,0)
is.na(dd$ESTUDIOPAR) <- dd$ESTUDIOPAR %in% c(8,9,0)
is.na(dd$SDP6) <- dd$SDP6 %in% c(0)
is.na(dd$SDP1) <- dd$SDP1 %in% c(9,0)
is.na(dd$MOP17) <- dd$MOP17 %in% c(9,0)
is.na(dd$MOP16) <- dd$MOP16 %in% c(9,0)
is.na(dd$MOP18) <- dd$MOP18 %in% c(9,0)
```

```
# SE ELIMINAN TODOS LOS MISSINGS
dd <- na.omit(dd)
```

Modelo de regresión logística

```
model_logit_3 <- glm(violencia.b~TAMUNI+MOPO+SDE1+ESTUDIOS+SDP2+ESTUDIOPAR
                    ,data=dd,family=binomial(link="logit"))
anova(model_logit_3,test="LRT") #test anova de las devianzas
summary(model_logit_3)

# Anova del modelo
Anova(model_logit_3)

# Test de Naglkerke
Nagelkerker2(model_logit_3)
```

```

# Estadísticos de la Devianza y de los Residuos de Pearson y su respectivo p-valor
D <- model_logit_3$deviance
1 - pchisq(D,model_logit_3$df.residual)
X2P <- sum(resid(model_logit_3,type="pearson")^2)
1 - pchisq(X2P,model_logit_3$df.residual)

# Curva de ROC
dadesroc<-prediction(predict(object=model_logit_3,type="response",newdata = dd)
                      ,as.numeric(dd$violencia.b))
par(mfrow=c(1,2))
plot(performance(dadesroc,"err"))# Error
plot(performance(dadesroc,"tpr","fpr"))
roc.perf <- performance(dadesroc,"auc",fpr.stop=0.05)
pred<- predict(model_logit_3, type="response")
auc(roc(factor(as.numeric(dd$violencia.b)),pred))

# Prueba de Hosmer & Lemeshow
HLtest(model_logit_3)

# Estimación de las odds ratio (OR)
exp(coef(model_logit_3))

# Grafico de las odds ratio
plot_model(model_logit_3,colors = c("cornflowerblue","cyan3"),axis.lim = c(0.1,10))

Modelo log-binomial
logbinomial1 <- glm(violencia.b~TAMUNI+MOP0+SDE1+ESTUDIOS+SDP2+ESTUDIOSPAR
                   ,data=dd,family=binomial(link="log"), start=c(-4, rep(0,27)))
summary(logbinomial1)

# Prueba de Hosmer & Lemeshow
HLtest(logbinomial1)

# Estimación del riesgo relativo (RR)
plot_model(logbinomial1,colors = c("cornflowerblue","cyan3"),axis.lim = c(0.1,10))

```

Modelo Poisson cero-inflado

```
# MODELO POISSON
model_pois <- glm(violencia.n~TAMUNI+SDE1+ESTUDIOS+MOPO+SDP2+ESTUDIOSPAR+SDP6+SDP1
                 +MOP17+MOP16+MOP18
                 ,data=dd,family=poisson)

# Comprobación sobre-dispersión
dispersiontest(model_pois,trafo=1)
dispersiontest(model_pois,trafo=2)

# Se construye el modelo cero-inflado.
mod_Z_I <- zeroinfl(violencia.n~TAMUNI+SDE1+ESTUDIOS+MOPO+SDP2
                   #+ESTUDIOSPAR
                   +SDP1+SDP6+MOP17
                   #+MOP16
                   +MOP18|ddTAMUNI+SDE1 +ESTUDIOS+MOPO+SDP2+SDP1
                   #+ESTUDIOSPAR
                   +SDP6+MOP17
                   #+MOP16
                   +MOP18
                   ,data=dd, dist ="poisson")

summary(mod_Z_I)

# Residuos de Pearson
x<- predict(mod_Z_I, type="response")
y<- residuals(mod_Z_I, "pearson")
plot(x,y, xlab="Predictor lineal", ylab="Residuos de Pearson")

# Error cuadrático medio
predZBN<- predict(mod_Z_I, type="response")
errors<- predZBN-dd$violencia.n
(EQM<- sqrt(sum(errors^2)/nrow(dd)))

# Estimación para la interpretación de los parámetros
exp(summary(mod_Z_I)$coefficients[[1]][,1])
exp(summary(mod_Z_I)$coefficients[[2]][,1])
```

