



**UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO**

---

---

**FACULTAD DE CIENCIAS**

**MODELO LOGÍSTICO PARA DESCRIBIR LAS  
MUERTES VIOLENTAS DE MUJERES EN  
CONDICIONES DE VULNERABILIDAD EN MÉXICO  
DEL AÑO 2015**

**T E S I S**

**QUE PARA OBTENER EL TÍTULO DE:**

**ACTUARIO**

**P R E S E N T A:**

**ANA CRISTINA SÁNCHEZ GARCÍA**



**DIRECTOR DE TESIS:  
DR. RICARDO RAMÍREZ ALDANA  
2018**

CD.MX.



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



1. Datos del alumno

Sánchez  
García  
Ana Cristina  
595-9284864  
Universidad Nacional Autónoma de  
México  
Facultad de Ciencias  
Actuaría  
309324609

2. Datos del tutor

Dr.  
Ricardo  
Ramírez  
Aldana

3. Datos del sinodal 1

Dra.  
Lizbeth  
Naranjo  
Albarrán

4. Datos del sinodal 2

M. en D.  
Ángel Jair  
Morales  
Eslava

5. Datos del sinodal 3

M. en C.  
José Antonio  
Flores  
Díaz

6. Datos del sinodal 4

M. en C.  
José Salvador  
Zamora  
Muñoz

7. Datos del trabajo escrito.

Modelo logístico para describir las muertes  
violentas de mujeres en condiciones de  
vulnerabilidad en México del año 2015.

116  
2018

# Agradecimientos

A mis padres. Porque no importa las veces que falle, nunca tendré un apoyo más sincero y desinteresado que el de ustedes, apoyándome y alentándome a nunca darme por vencida.

A mi hermana. Que siempre se preocupa por mi, y por que a través de lo que tu miras me ayudas entender y disfrutar de las maravillas de la vida.

A mis amigos. Por todos los momentos compartidos.

A mis maestros. Por haberme formado y por su excelente vocación. En especial a mi asesor de tesis, el Dr. Ricardo Ramirez, por su infinita amabilidad y ayuda en la elaboración de este trabajo. Además de a mis sinodales, por todo el tiempo brindado y sus valiosas correcciones.

A todas las personas que me han inspirado, y que gracias a eso logro terminar un capítulo más en mi vida.

# Índice general

<b>Introducción</b>	<b>1</b>
<b>1. Fuente de información y descripción de las variables</b>	<b>3</b>
1.1. Contrucción de la variable de respuesta . . . . .	4
1.2. Construcción de las variables explicativas . . . . .	7
<b>2. Violencia de género</b>	<b>12</b>
2.1. Contexto Histórico . . . . .	12
2.2. Violencia en México . . . . .	14
2.2.1. Compromisos internacionales . . . . .	14
2.3. Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2011 . . . . .	16
2.3.1. Tablas de contingencia . . . . .	17
2.3.2. Violencia en la comunidad . . . . .	19
2.3.3. Violencia dentro de la relación de pareja . . . . .	40
<b>3. Femicidio</b>	<b>47</b>
3.1. Introducción . . . . .	47

3.2. Femicidio en México . . . . .	48
<b>4. Modelo Logístico</b>	<b>53</b>
4.1. Introducción . . . . .	53
4.2. Modelos Lineales Generalizados . . . . .	54
4.3. Formulación del Modelo . . . . .	56
4.4. Estimación de los parámetros . . . . .	59
4.4.1. Modelo Desagregado . . . . .	59
4.4.2. Modelo agregado . . . . .	64
4.5. Parámetros significativos . . . . .	67
4.5.1. Prueba de Wald . . . . .	67
4.6. Medidas de bondad y ajuste . . . . .	69
4.6.1. Pruebas basadas en patrones de covariables . . . . .	70
4.6.2. Pruebas basadas en las probabilidades . . . . .	75
4.6.3. Interpretación de los parámetros . . . . .	79
4.7. Intervalos de confianza . . . . .	85
4.7.1. Medidas tipo $R^2$ . . . . .	86
4.7.2. Poder predictivo . . . . .	88
4.8. Diagnóstico de las observaciones . . . . .	94
4.8.1. Leverage . . . . .	95
4.8.2. Residuos . . . . .	96
4.8.3. Análisis gráfico . . . . .	97

4.9. Ligas Probit y Complementaria log-log . . . . .	104
<b>Conclusión</b>	<b>107</b>
<b>Anexo I: Código de graficas de análisis diagnóstico</b>	<b>109</b>
<b>Anexo II: Graficas de Poder predictivo</b>	<b>113</b>
<b>Bibliografía</b>	<b>114</b>



# Índice de figuras

2.1. Distribución marginal de la variable Lugar donde ocurrió la agresión, Pregunta 1. . . . .	22
2.2. Distribución marginal de la variable Persona que realizó la agresión, Pregunta 1. . . . .	22
2.3. Distribuciones condicionando a la variable Lugar donde ocurrió la agresión, sección Violencia en la comunidad, Pregunta 2 . . . . .	23
2.4. Distribución marginal de la variable Lugar donde ocurrió la agresión, Pregunta 2. . . . .	24
2.5. Distribución marginal de la variable Persona que realizó la agresión, Pregunta 2. . . . .	25
2.6. Distribuciones condicionando a la variable Lugar donde ocurrió la agresión, sección Violencia en la comunidad, Pregunta 2 . . . . .	26
2.7. Distribución marginal de la variable Lugar donde ocurrió la agresión, Pregunta 3. . . . .	27
2.8. Distribución marginal de la variable Persona que realizó la agresión, Pregunta 3. . . . .	27
2.9. Distribuciones condicionando a la variable Lugar donde ocurrió la agresión, sección Violencia en la comunidad, Pregunta 3 . . . . .	28
2.10. Distribución marginal de la variable Lugar donde ocurrió la agresión, Pregunta 4. . . . .	29

2.11. Distribución marginal de la variable Persona que realizó la agresión, Pregunta 4. . . . .	29
2.12. Distribuciones condicionando a la variable Lugar donde ocurrió la agresión, sección Violencia en la comunidad, Pregunta 4 . . . . .	30
2.13. Distribución marginal de la variable Lugar donde ocurrió la agresión, Pregunta 5. . . . .	31
2.14. Distribución marginal de la variable Persona que realizó la agresión, Pregunta 5. . . . .	31
2.15. Distribuciones condicionando a la variable Lugar donde ocurrió la agresión, sección Violencia en la comunidad, Pregunta 5 . . . . .	32
2.16. Distribución marginal de la variable Lugar donde ocurrió la agresión, Pregunta 6. . . . .	33
2.17. Distribución marginal de la variable Persona que realizó la agresión, Pregunta 6. . . . .	33
2.18. Distribuciones condicionando a la variable Lugar donde ocurrió la agresión, sección Violencia en la comunidad, Pregunta 6 . . . . .	34
2.19. Distribuciones condicionando a la variable Lugar donde trabajaba la mujer, sección Violencia laboral , Pregunta 1 . . . . .	36
2.20. Distribuciones condicionando a la variable Lugar donde trabajaba la mujer, sección Violencia laboral , Pregunta 2 . . . . .	37
2.21. Distribuciones condicionando a la variable Lugar donde trabajaba la mujer, sección Violencia laboral , Pregunta 3 . . . . .	38
2.22. Distribuciones condicionando a la variable Lugar donde trabajaba la mujer, sección Violencia laboral , Pregunta 4 . . . . .	39
2.23. Gráficas de las distribuciones marginales Frecuencia y Gravedad de la agresión, Pregunta 1. . . . .	41
2.24. Distribuciones condicionando a la variable Frecuencia con que ocurre la agresión, Pregunta 1. . . . .	42

2.25. Distribución marginal de las variables Frecuencia con que ocurre la agresión, Pregunta 2. . . . .	43
2.26. Distribuciones condicionando a la variable Frecuencia con que ocurre la agresión, Pregunta 2. . . . .	43
2.27. Distribución marginal de las variables Frecuencia con que ocurre la agresión, Pregunta 3. . . . .	44
2.28. Distribuciones condicionando a la variable Frecuencia con que ocurre la agresión, Pregunta 3. . . . .	45
2.29. Distribución marginal de las variables Frecuencia con que ocurre la agresión, Pregunta 4. . . . .	46
2.30. Distribuciones condicionando a la variable Frecuencia con que ocurre la agresión, Pregunta 4. . . . .	46
4.1. Ejemplo de gráfica de la función logística con parámetro positivo. . . .	58
4.2. Ejemplo de gráfica de la función logística con parámetro negativo. . . .	58
4.3. Gráfica que determina el punto de corte el cual maximiza el valor de la sensibilidad y especificidad. . . . .	91
4.4. Gráfica que determina el punto de corte óptimo, utilizando la función ROC de la paquetería Epi de R. . . . .	92
4.5. Gráfica que determina el punto de corte el cual maximiza el valor de la sensibilidad y especificidad. . . . .	93
4.6. Gráfica de diagnóstico del comportamiento del índice contra residuos. .	98
4.7. Gráfica de diagnóstico del comportamiento del predictor lineal contra los residuos de devianza estandarizados. . . . .	98
4.8. Gráfica de diagnóstico del comportamiento del predictor lineal contra residuos del modelo desagregado. . . . .	99
4.9. Gráfica de diagnóstico del comportamiento entre las proporciones estimadas contra las proporciones reales. . . . .	99

4.10. Gráfica de los residuos para cada variable del modelo. . . . .	100
4.11. Gráfica de diagnóstico del comportamiento del valor de $\Delta\chi^2$ . . . . .	101
4.12. Gráfica de diagnóstico del comportamiento de $\hat{\pi}_j$ contra $\Delta\chi^2$ . . . . .	102
4.13. Gráfica de diagnóstico del comportamiento de $\hat{\pi}_j$ contra $\Delta D_j^2$ . . . . .	103
4.14. Gráfica de diagnóstico del comportamiento en la distancia de Cook. . .	104

# Índice de cuadros

1.1. Lista de variables disponibles de las Estadísticas Vitales de Mortalidad	4
1.2. Variable de respuesta . . . . .	5
1.3. Variable que filtra los homicidios en estado vulnerable . . . . .	6
1.4. Variable que filtra los homicidios por edades . . . . .	7
1.5. Variable escolaridad con original y nueva categorización . . . . .	8
1.6. Nueva variable Escolaridad . . . . .	8
1.7. Variable Ocupación con original y nueva categorización . . . . .	8
1.8. Nueva variable Ingresos . . . . .	9
1.9. Variable estado civil con original y nueva categorización . . . . .	9
1.10. Nueva variable Estado Civil . . . . .	9
1.11. Variable Zona con original y nueva categorización . . . . .	10
1.12. Nueva variable Zona . . . . .	10
1.13. Variable sitio del hecho vital con original y nueva categorización . . . . .	11
1.14. Nueva variable Sitio . . . . .	11
2.1. Cronología de alguno de los principales eventos enfocados en la eliminación de todas las formas de discriminación y violencia en contra de las mujeres. . . . .	13

2.2. Tabla de Contingencia entre las variables X y Y . . . . .	17
2.3. Tabla de distribución conjunta entre las variables X y Y. . . . .	17
2.4. Tabla de distribuciones condicionales de Y dado cada nivel de X . . . . .	19
2.5. Tabla que resume el conjunto de preguntas consideradas para analizar la violencia de género fuera de la vida en pareja. . . . .	19
2.6. Variable Lugar donde Ocurrió la Agresión con original y nueva agrupación	20
2.7. Nueva Variable Lugar donde Ocurrió la Agresión . . . . .	20
2.8. Variable Persona que Realizó la Agresión con original y nueva agrupación	21
2.9. Nueva Variable Persona que Realizó la Agresión . . . . .	21
2.10. Tabla de distribución conjunta y marginal, sección Violencia en la comu- nidad, Pregunta 1. . . . .	22
2.11. Tabla de distribución condicional, sección Violencia en la comunidad, Pregunta 1. . . . .	23
2.12. Tabla de distribución conjunta y marginal, sección Violencia en la comu- nidad, Pregunta 2. . . . .	24
2.13. Tabla de distribución condicional, sección Violencia en la comunidad, Pregunta 2. . . . .	25
2.14. Tabla de distribución conjunta y marginal, sección Violencia en la comu- nidad, Pregunta 3. . . . .	26
2.15. Tabla de distribución condicional, sección Violencia en la comunidad, Pregunta 3. . . . .	28
2.16. Tabla de distribución conjunta y marginal, sección Violencia en la comu- nidad, Pregunta 4. . . . .	29
2.17. Tabla de distribución condicional, sección Violencia en la comunidad, Pregunta 4. . . . .	30
2.18. Tabla de distribución conjunta y marginal, sección Violencia en la comu- nidad, Pregunta 5. . . . .	31

2.19. Tabla de distribución condicional, sección Violencia en la comunidad, Pregunta 5. . . . .	32
2.20. Tabla de distribución conjunta y marginal, sección Violencia en la comunidad, Pregunta 6. . . . .	33
2.21. Tabla de distribución condicional, sección Violencia en la comunidad, Pregunta 6. . . . .	34
2.22. Variable Lugar donde trabajaba la mujer durante el último año. . . . .	35
2.23. Tabla de distribución condicional, sección Violencia laboral, Pregunta 1. . . . .	36
2.24. Tabla de distribución condicional, sección Violencia laboral, Pregunta 2. . . . .	37
2.25. Tabla de distribución condicional, sección Violencia laboral, Pregunta 3. . . . .	38
2.26. Tabla de distribución condicional, sección Violencia laboral, Pregunta 4. . . . .	39
2.27. Variable Frecuencia con que ocurre la agresión . . . . .	40
2.28. Variable Gravedad de la agresión . . . . .	40
2.29. Distribución marginal de las variables Frecuencia y Gravedad de la agresión, sección Violencia dentro de la relación de Pareja, Pregunta 1. . . . .	41
2.30. Tabla de distribución condicional, sección Violencia dentro de la relación de Pareja, Pregunta 1. . . . .	42
2.31. Distribución marginal de las variables Frecuencia y Gravedad de la agresión, sección Violencia dentro de la relación de Pareja, Pregunta 2. . . . .	42
2.32. Tabla de distribución condicional, sección Violencia dentro de la relación de Pareja, Pregunta 2. . . . .	43
2.33. Distribución marginal de las variables Frecuencia y Gravedad de la agresión, sección Violencia dentro de la relación de Pareja, Pregunta 3. . . . .	44
2.34. Tabla de distribución condicional, sección Violencia dentro de la relación de Pareja, Pregunta 3. . . . .	45
2.35. Distribución marginal de las variables Frecuencia y Gravedad de la agresión, sección Violencia dentro de la relación de Pareja, Pregunta 4. . . . .	45

2.36. Tabla de distribución condicional, sección Violencia dentro de la relación de Pareja, Pregunta 4. . . . .	46
3.1. Tabla con los estados que han declarado Alerta de género, según datos reportados por el Instituto Nacional de las Mujeres . . . . .	50
4.1. Resumen de pruebas individuales de significancia de las variables explicativas bajo prueba de Wald . . . . .	69
4.2. Resumen de pruebas individuales de significancia de las variables explicativas bajo diferencia de devianzas. . . . .	73
4.3. Tabla que muestra el valor del parámetro estimado correspondiente a la variable Zona y su exponencial . . . . .	80
4.4. Tabla que muestra el valor de los parámetros estimados correspondientes a la variable Nivel de escolaridad y su exponencial . . . . .	82
4.5. Tabla que muestra el valor de los parámetros estimados correspondientes a la variable Ingresos y su exponencial . . . . .	83
4.6. Tabla que muestra el valor de los parámetros estimados correspondientes a la variable Estado civil y su exponencial . . . . .	83
4.7. Tabla que muestra el valor de los parámetros estimados correspondientes a la variable Lugar de defunción y su exponencial . . . . .	84
4.8. Tabla de clasificación . . . . .	88
4.9. Cuadro que determina la capacidad de discriminación de un modelo. . .	94
4.10. Tabla que contiene el conjunto de elementos que tuvieron en valor absoluto los residuos de devianza estandarizados más grandes. . . . .	97
4.11. Tabla que compara los resultados de distintas medidas obtenidas entre las diferentes funciones ligas. . . . .	106



# Introducción

El objetivo principal, es dar una descripción aproximada del fenómeno conocido como feminicidio, ajustando un modelo logístico a datos obtenidos de las estadísticas vitales recopiladas por el organismo INEGI del año 2015. Lo anterior, identificando algunos de los principales factores de riesgo que determinan con mayor probabilidad, entre las mujeres de bajos recursos, que una mujer pueda morir asesinada.

Una de las principales motivaciones de esta investigación, es que a través de la identificación de estos factores de riesgo, se pueda entender de una mejor manera cuáles son las principales razones por lo que este fenómeno ha ido en aumento en los últimos años. Con la identificación de las variables que caracterizan de mejor manera a las mujeres asesinadas, se pueden identificar las principales vías de acción para combatirlos y reducirlos, pues aunque algunos estados de la república incluso han implementado mecanismo de emergencia, invirtiendo y puesto en práctica numerosos programas de concientización y erradicación, organizaciones especializadas en este tema han exteriorizado su preocupación por su falta de impacto, pues no se ataca de fondo el problema. El contenido de este trabajo comienza describiendo a la base de datos que fue seleccionada de las Estadísticas vitales del año 2015, pues como se mencionó en los objetivos el estudio está enfocado a las muertes de mujeres de bajos recursos, además de describir a las variables que formaron parte del estudio.

El segundo capítulo aborda el tema de violencia familiar, dando un contexto histórico de los avances hechos por diferentes organizaciones internacionales y tratados, haciendo énfasis en los ratificados por México. Además se analiza la Encuesta Nacional Sobre la Dinámica de las Relaciones en los Hogares del año 2011, para así describir los alcances de este problema en México.

El tercer capítulo da una introducción a la problemática del feminicidio, haciendo un pequeño análisis histórico de los esfuerzos hechos para definirlo, así como una descripción de su evolución en México. Además, directamente relacionada a la problemática anterior, se describe la llamada Alerta de Género, que es una iniciativa establecida en la Ley General de Acceso de las Mujeres a una Vida Libre de Violencia que es aplicada en casos extremos de violencia descontrolada en los estados, explicando sus planes de acción y sus contras.

El cuarto capítulo aborda todo el proceso de ajuste del modelo logístico a la base de datos correspondiente a las defunciones registradas en las Estadísticas Vitales del año 2015, tomando como variable de respuesta aquellas defunciones con presunción de homicidio, pues son la mejor aproximación que se tiene para medir los feminicidios. De esta manera, se describe detalladamente el modelo yendo desde su motivación y definición, hasta la estimación de los parámetros, para posteriormente evaluar qué tan adecuado es el modelo o los datos originales, por medio de medidas de bondad de ajuste y diagnóstico.

# Capítulo 1

## Fuente de información y descripción de las variables

Entre los objetivos principales de este trabajo se encuentra el poder caracterizar el fenómeno de los asesinatos violentos de un grupo muy particular de mujeres, que es el de aquellas consideradas de bajos recursos y cuyos rangos de edades vayan de entre los 15 a 84 años. Lo anterior, por medio del ajuste de un modelo logístico el cual describa las relaciones presentes entre este tipo de asesinatos, con algunas de las variables que han sido altamente mencionadas por los diferentes grupos y estudiosos en la materia de feminicidio.

Es importante mencionar que este ajuste de modelo logístico es sólo una aproximación al fenómeno conocido como feminicidio, pues la definición y los alcances de este tipo de asesinatos tiene dificultades tanto en su aplicación jurídica, por las diferentes tipos penales existentes para este delito, como en la recolección de información estadística, debido a errores u omisiones en el registro de defunción (Ciani, 2015).

La fuente de información se obtuvo de las Defunciones Generales y Fetales del 2015 (INEGI, 2015), a partir del certificado de defunción, el cual es uno de los instrumentos de captación de información más importante. Estos registros cuentan con una serie de variables las cuales se pueden dividir en tres grandes grupos, cuyo objetivo va primeramente en describir todo lo relacionado con la defunción, incluyendo variables como lugar de registro o la causa de defunción.

Posteriormente, para el caso en que la muerte haya sido causada por un accidente o haya sido violenta, se incluyen algunas variables que puedan describir, por ejemplo, el parentesco del agresor, o si se sospechaba de violencia familiar, variables que son vitales para el seguimiento de los feminicidios en México.

Por último, se incluyen variables que describen todo lo relacionado con el fallecido, como el estado conyugal, ocupación o nivel de escolaridad, entre otras características.

## 1.1. Contrucción de la variable de respuesta

El modelo logístico se ajustó a las defunciones registradas en 2015, que corresponde a los datos más recientes encontrados actualmente en las estadísticas vitales de INEGI. Estos registros constan de un total de 59 variables, las cuales se listan a continuación:

Estadísticas Vitales			
No.	Nombre Campo	No.	Nombre Campo
1	Entidad.Registro	31	Asistencia.Médica
2	Municipio Registro	32	Sitio donde Ocurrió la Defunción
3	Entidad Residencia	33	Certificada Por
4	Municipio Residencia	34	Nacionalidad
5	Tamaño Localidad Residencia	35	Derechohabencia
6	Clave Localidad de Residencia	36	Condición de embarazo
7	Entidad Ocurrencia	37	Causas relacionadas con embarazo
8	Municipio Ocurrencia	38	Hora de la defunción
9	Tamaño de Localidad Ocurrencia	39	Minuto de la defunción
10	Clave Localidad de Ocurrencia	40	Capítulo
11	Causa Defunción (Lista Detallada)	41	Grupo
12	Causa Defunción (Lista Mexicana)	42	Lista1
13	Sexo	43	Grupos lista mexicana
14	Edad	44	Violencia Familiar
15	Día Defunción	45	Área Urbano Rural
16	Mes Defunción	46	Edad agrupada
17	Año Defunción	47	Complicaron el embarazo
18	Día Registro	48	Día de Certificación
19	Mes Registro	49	Mes de Certificación
20	Año Registro	50	Año de Certificación
21	Día Nacimiento	51	Maternas
22	Mes Nacimiento	52	Lengua indígena
23	Año Nacimiento	53	Condición de Actividad económica
24	Ocupación	54	Parentesco del presunto agresor
25	Escolaridad	55	Entidad Ocurrencia de la lesión
26	Estado Conyugal	56	Municipio Ocurrencia de la lesión
27	Presunto	57	Clave Localidad de Ocurrencia de la lesión
28	Ocurrió Trabajo	58	Razón Materna
29	Sitio donde Ocurrió la Lesión	59	Distritos de Oaxaca
30	Necropsia		

Cuadro 1.1: Lista de variables disponibles de las Estadísticas Vitales de Mortalidad

Para el ajuste del modelo, se construyó una nueva variable de respuesta  $Y$ , tomando como referencia la causa de muerte establecida por la variable denominada Grupos lista mexicana. Ésta se recodificó de tal manera que la categoría  $Y = 1$  correspondiera a las muertes causadas por Agresiones u Otra violencia, tomando únicamente de las muertes clasificadas como Otra violencia aquellas incluidas: muerte por ahorcamiento, estrangulamiento y sofocación, disparo de armas de fuego, contacto traumático con objeto cortante y contacto traumático con objeto romo o sin filo, pues difícilmente un deceso con alguna de estas características no se relacionaría con homicidio. En cambio, en la categoría  $Y = 0$  se incluyen diversas causas de muerte natural pero que en general son curables y que atacan de manera desproporcionada a sectores vulnerables de la sociedad, lo anterior recordando que la población objetivo se concentra principalmente en aquellas mujeres de bajos recursos.

Es decir, se obtiene una nueva variable dicotómica que describe la condición de la muerte de la mujer, dividiéndola en muerte violenta, versus no lo fue, tomando además sólo a aquellas mujeres que puedan ser consideradas de bajos recursos. Siendo más específicos, la recodificación es como sigue:

Variable Lista de Causas		
Cve	Nuevo Cve	Descripción
001	0	Enfermedades infecciosas intestinales
002	0	Tuberculosis
004	0	Infecciones con modo de transmisión predominantemente sexual
005	0	Otras enfermedades infecciosas y parasitarias y efectos tardíos de las enfermedades infecciosas y parasitarias
021	0	Desnutrición y otras deficiencias nutricionales
032	0	Infecciones y otras enfermedades de las vías respiratorias superiores
033	0	Otras enfermedades del aparato respiratorio
043	0	Causas obstétricas directas
044	0	Parto único espontáneo
045	0	Causas obstétricas indirectas
046	0	Ciertas afecciones originadas en el periodo perinatal
E55	1	Agresiones
E56	1	Otra violencia

Cuadro 1.2: Variable de respuesta

Además de la recodificación anterior y obtener la variable de respuesta, fue necesario llevar a cabo dos filtros, los cuales aseguraran, en primer lugar, la condición de bajos recursos en las fallecidas, y otro que agrupara únicamente a aquellas mujeres cuyo rango de edad estuviera entre los 15 y 84 años de edad.

Para lograr el primer objetivo, se utilizó la variable Derechohabiencia, la cual especifica a qué institución de seguridad social, público o privado se encontraba afiliada la mujer al momento de su fallecimiento. Entonces, bajo el supuesto de que se podían considerar como un grupo vulnerable a aquellas mujeres que no tuvieran acceso a ningún servicio de salud, o se encontrarán afiliadas a alguno cuyo enfoque era para personas de bajos recursos, se eligieron las categorías Ninguna, Seguro Popular, IMSS Oportunidades y los No especificados, para separar a aquellas mujeres que se podían considerar de bajos recursos. El no contar con acceso a la salud y la seguridad social son algunas de las principales causas del empobrecimiento (Botello, 2016). Los filtros son los siguientes:

<b>Variable Derechohabiencia</b>	
<b>Cve</b>	<b>Descripción</b>
1	Ninguna
7	Seguro popular
9	IMSS oportunidades
99	Derechohabiencia no especificada

Cuadro 1.3: Variable que filtra los homicidios en estado vulnerable

La razón por la que se incluyó en el estudio a los no especificados fue bajo la suposición de que si no se pudo especificar si se pertenecía o no a alguna institución, seguramente no lo estaba o se era una persona de bajos recursos con poco acceso a estos beneficios. De lo contrario, de haber estado afiliada a algún organismo de salud se hubiera hecho el registro correctamente o se habría tenido algún historial médico, lo cual al no tenerse también se puede relacionar con la condición de vulnerabilidad o falta de recursos de la persona.

El segundo filtro hizo uso de la variable Edad agrupada, tomando las categorías cuyo rango abarcara las edades de entre los 15 años hasta los 84 años. La razón de esto se debió únicamente a las delimitaciones establecidas en los objetivos del trabajo, el cual consiste principalmente en ajustar un modelo logístico a las muertes violentas de mujeres especialmente vulnerables en edades cuyas agresiones puedan deberse a no cumplir cierto rol de género dentro del ámbito familiar o social, más encaminado a aquellas conductas morales impuestas sobre las mujeres adultas dentro de la familia y sociedad. Los filtros son los siguientes:

<b>Variable Edad agrupada</b>	
<b>Cve</b>	<b>Descripción</b>
8	De 15 a 19 años
9	De 20 a 24 años
10	De 25 a 29 años
11	De 30 a 34 años

Cve	Descripción
12	De 35 a 39 años
13	De 40 a 44 años
14	De 45 a 49 años
15	De 50 a 54 años
16	De 55 a 59 años
17	De 60 a 64 años
18	De 65 a 69 años
19	De 70 a 74 años
20	De 75 a 79 años
21	De 80 a 84 años

Cuadro 1.4: Variable que filtra los homicidios por edades

## 1.2. Construcción de las variables explicativas

Como se ha venido tratando, el fenómeno del feminicidio ha atravesado un largo camino en un esfuerzo de delimitarlo, caracterizarlo y definirlo, de tal manera que este pueda ser considerado como un tipo especial de asesinato violento, producto principalmente de un acto discriminatorio cuya motivación radica en razones de género y relaciones desiguales de poder (Lagarde, 2008; Toledo, 2009).

Para este trabajo, las variables que se utilizaron para describir al feminicidio en mujeres de bajos recursos fueron cinco. Las primeras cuatro se encargaron de describir algunas de las principales características de las fallecidas, como son su escolaridad, nivel de ingresos, estado conyugal, y tipo de zona donde residía la mujer, y una última dando mención del lugar donde ocurrió el hecho vital. Para fines prácticos y de interpretación, todas las variables incluídas en el modelo fueron recategorizadas.

### Variable Escolaridad

En el caso de la variable Escolaridad, de las doce categorías existentes cuyos niveles abarcaban desde Sin escolaridad, hasta nivel de Posgrado, se compactó la información y se crearon sólo tres niveles. Más específicamente, la nueva categorización se presenta en el Cuadro 1.5.

Variable Escolariad		
Cve	Nuevo Cve	Descripción
1	3	Sin escolaridad
2	3	Preescolar
3	3	Primaria incompleta
4	3	Primaria completa
5	2	Secundaria incompleta

Cve	Nuevo Cve	Descripción
6	2	Secundaria completa
7	2	Bachillerato o Preparatoria incompleto
8	2	Bachillerato o Preparatoria completo
9	1	Profesional
10	1	Posgrado
99	2	No especificado

Cuadro 1.5: Variable escolaridad con original y nueva categorización

Se generó entonces una nueva variable escolaridad con las características presentadas en el Cuadro 1.6.

Variable Escolaridad	
Nuevo Cve	Descripción
1	Superior
2	Media superior
3	Básica

Cuadro 1.6: Nueva variable Escolaridad

### Variable Ingresos

Al igual que en el caso de la variable Escolaridad, se redujo el número de categorías al pasar de catorce, a sólo tres, agrupándolas de tal manera que cada grupo reflejara la cantidad de ingresos generados por las distintas ocupaciones que ejercían las fallecidas. La nueva categorización se presenta en el Cuadro 1.7.

Variable Ocupación		
Cve	Nuevo Cve	Descripción
1	1	Funcionarios, directores y jefes
2	1	Profesionistas y técnicos
3	1	Trabajadores auxiliares en actividades administrativas
4	2	Comerciantes, empleados en ventas y agentes de ventas
5	2	Trabajadores en servicios personales y vigilancia
6	2	Trabajadores en actividades agrícolas, ganaderas, forestales, caza y pesca
7	2	Trabajadores artesanales
8	2	Operadores de maquinaria industrial, ensambladores, choferes y conductores de transporte
9	2	Trabajadores en actividades elementales y de apoyo
10	3	Busca trabajo
11	3	No trabaja
97		No aplica a menores de 5 años
98	3	Ocupaciones insuficientemente especificadas
99	3	Ocupación no especificada

Cuadro 1.7: Variable Ocupación con original y nueva categorización



Creando así una nueva variable Ingresos cuyas categorías se definieron en términos de los ingresos generados. La nueva variable se presenta en el Cuadro 1.8.

<b>Variable Ingresos</b>	
<b>Nuevo Cve</b>	<b>Descripción</b>
1	De medios a altos
2	De muy bajos a bajos
3	No genera ingresos

Cuadro 1.8: Nueva variable Ingresos

### **Variable Estado civil**

Para el caso de la variable Estado civil, de las ocho categorías existentes en la variable original, se creó una nueva con tres, recategorizando y creando los nuevos grupos en función de si al momento del fallecimiento las mujeres registraron o no tener una pareja sentimental.

<b>Variable Estado Civil</b>		
<b>Cve</b>	<b>Nuevo Cve</b>	<b>Descripción</b>
1	3	Soltera
2	1	Divorciada
3	1	Viuda
4	2	Unión libre
5	2	Casada
6	1	Separada
8		No aplica a menores de 12 años
9	2	Estado civil no especificado

Cuadro 1.9: Variable estado civil con original y nueva categorización

La nueva categorización se presenta en el Cuadro 1.10.

<b>Variable Estado civil</b>	
<b>Nuevo Cve</b>	<b>Descripción</b>
1	Separada o viuda
2	Casada o en unión libre
3	Soltera

Cuadro 1.10: Nueva variable Estado Civil

## Variable Zona

Para el caso de la variable Zona, cuyas categorías únicamente especificaban si la fallecida pertenecía a una clasificada como rural o urbana, mantuvo las dos categorías existentes, haciendo solo un intercambio de etiquetas entre las dos zonas.

Variable Zona		
Cve	Nuevo Cve	Descripción
1	1	Urbano
2	0	Rural
9	1	No especificado

Cuadro 1.11: Variable Zona con original y nueva categorización

La nueva variable se presenta en el Cuadro 1.12.

Variable Zona	
Nuevo Cve	Descripción
0	Rural
1	Urbana

Cuadro 1.12: Nueva variable Zona

## Variable Sitio

Por último, para el caso de la variable Sitio la cual especificaba el sitio donde ocurrió el hecho vital. Se recodificó de tal manera que se pudiera evidenciar si las fallecidas tuvieron la oportunidad de tener asistencia médica antes de su muerte, o si los hechos sucedieron en el hogar o en algún sitio público. La nueva categorización se presenta en el Cuadro 1.13.

Variable Sitio del Hecho Vital		
Cve	Nuevo Cve	Descripción
1	1	Secretaría de salud
2	1	IMSS oportunidades
3	1	IMSS
4	1	ISSSTE
5	1	PEMEX
6	1	Secretaría de la defensa nacional
7	1	Secretaría de marina
8	1	Otra unidad pública
9	1	Unidad médica privada

<b>Cve</b>	<b>Nuevo Cve</b>	<b>Descripción</b>
10	2	Vía pública
11	3	Hogar
12	2	Otro lugar
99	3	No especificado

Cuadro 1.13: Variable sitio del hecho vital con original y nueva categorización

De tal manera que la nueva variable es como sigue:

<b>Variable Sitio</b>	
<b>Nuevo Cve</b>	<b>Descripción</b>
1	Hospital
2	Sitio público
3	Hogar

Cuadro 1.14: Nueva variable Sitio

Es importante notar que la categoría Hogar no presentó cambios en la nueva clasificación, lo anterior debido a que ésta es de gran importancia para estudiar los femi- cidios cometidos por la pareja o familiares, además de que por sí sola figura al nivel de las nuevas categorías formadas de las diferentes agrupaciones en cuanto a número de fallecidas.

# Capítulo 2

## Violencia de género

### 2.1. Contexto Histórico

Los esfuerzos que se han hecho a lo largo de los años para lograr el reconocimiento de los derechos y libertades de la mujer en un plano de igualdad han sido arduos y muy extensos, tomando relevancia en la agenda internacional a partir de la creación de la Comisión de la Condición Jurídica y Social de la Mujer (Olamendi, 2016).

Dicho órgano está dedicado principalmente a la promoción de la igualdad de género y el reconocimiento de sus derechos humanos, mediante la elaboración de normas y conferencias internacionales, cuyas resoluciones han representado grandes avances en la materia (ONU MUJERES, 2017)<sup>A</sup>.

Entre los principales logros de esta comisión se encuentra la construcción de la Convención sobre la Eliminación de Todas las Formas de Discriminación Contra la Mujer (CEDAW), el cual es uno de los tratados internacionales más ratificados en todo el mundo y cuyas funciones no sólo se han enfocado en reconocer los derechos de la mujer, sino que también han planteado planes de acción y recomendaciones generales los cuales ayudan a cumplir los compromisos hechos por los países miembros dentro de su sistema institucional y legislativo (ONU MUJERES, 2011).

Posteriormente, la Asamblea General de la Organización de las Naciones Unidas (ONU), hace la Declaración sobre la Eliminación de la violencia contra la Mujer, además de llevar a cabo la Cuarta Conferencia Mundial sobre la Mujer con sede en Beijing, cuya relevancia destaca entre las demás por ser un compendio de las tres conferencias anteriores (ONU MUJERES, 2017)<sup>B</sup>.

En dicha conferencia, se redacta una de las principales guías y planes de acción para los países comprometidos con alcanzar la igualdad de género, conocida como La Declaración y Plataforma de Acción de Beijing, cuyo seguimiento y revisión a los 189 países que la adoptaron se lleva a cabo por La Comisión de la Condición Jurídica y Social de la Mujer (ONU MUJERES, 2017)<sup>4</sup>.

Por otra parte, además de los esfuerzos hechos por la ONU, la Organización de los Estados Americanos (OEA) también realizó aportes significativos a la lucha por el reconocimiento e igualdad de los derechos humanos de las mujeres, principalmente con la creación de la Convención Interamericana para Prevenir, Sancionar y Erradicar la Violencia Contra la Mujer, también denominada Convención Belém do Pará. Ésta enuncia, entre otras cosas, los derechos y medidas internas que los estados parte deben cumplir para lograr la igualdad de género, además de las obligaciones que se comprometieron a cumplir con la convención para lograr los objetivos planteados en ésta (Olamendi, 2016).

A continuación se presenta una cronología resumida de algunos de los eventos más significativos referentes a los esfuerzos hechos para lograr el cese a la violencia de género:

<b>Cronología</b>	
<b>Año de Evento</b>	<b>Evento</b>
1946	Creación de la Comisión de la Condición Jurídica y Social de la Mujer (CSW)
1952	Creación de la Convención Sobre los Derechos Políticos de la Mujer
1957	Creación de la Convención sobre la Nacionalidad de la Mujer Casada
1975	Se lleva a cabo la Primera Conferencia Mundial de la Mujer con sede en Ciudad de México, México
1979	Convención Sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer (CEDAW)
1980	Se lleva a cabo la Segunda Conferencia Mundial de la Mujer con sede en Copenhague, Dinamarca
1985	Se lleva a cabo la tercera Conferencia Mundial de la Mujer con sede en Nairobi, Kenia
1993	Declaración Sobre la Eliminación de la Violencia Contra la Mujer por la Asamblea General de la ONU
1995	Convención Interamericana para Prevenir, Sancionar y Erradicar la Violencia Contra la Mujer (Belém do Pará)
1995	Se lleva a cabo la cuarta Conferencia Mundial de la Mujer con sede en Beijing, China, creando la Declaración y Plataforma de Acción de Beijing
2010	La Asamblea General de las Naciones Unidas crea ONU Mujeres

Cuadro 2.1: Cronología de alguno de los principales eventos enfocados en la eliminación de todas las formas de discriminación y violencia en contra de las mujeres.

## 2.2. Violencia en México

La violencia de género es uno de los actos discriminatorios más recurrentes y tolerados en todo el mundo, la cual impide el pleno goce de los derechos y libertades que cualquier persona tiene derecho a ejercer. Producto de un proceso histórico que ha permitido y enriquecido relaciones desiguales de poder entre hombres y mujeres, tanto en la vida de pareja como dentro del entorno familiar y social (Lagarde, 2008).

### 2.2.1. Compromisos internacionales

Como se mencionó anteriormente, existen varios tratados internacionales los cuales describen planes de acción dentro de los países miembro que los ratifican, sometiénndose así al escrutinio internacional y a las demandas hechas por estos organismos que verifican y dan seguimiento a los avances hechos por cada país. En el caso de México, al ratificar la Convención Sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer (CEDAW) y a la Convención Interamericana para Prevenir, Sancionar y Erradicar la Violencia Contra la Mujer (Belém do Pará), está obligado a ser evaluado y presentar los avances correspondientes a los cumplimiento de éstas (Olamendi, 2016).

Ambos tratados abordan cuatro secciones las cuales dan estructura a los compromisos hechos por cada nación, por lo que al ser México parte de éstas se ha obligado a cumplir los puntos que en ellas se aborden.

En primer lugar, se define el concepto discriminación contra la mujer, entendiéndolo como cualquier acto que abarca toda distinción, exclusión o restricción basada en el sexo, que impida ejercer tanto sus derechos como sus libertades fundamentales en los diferentes ámbitos de desarrollo, que abarca desde el social, político, económico, cultural y civil, y a partir del cual surge la violencia de género. Ésta se define como cualquier daño o sufrimiento físico, sexual o psicológico, que puede perpetrarse tanto dentro de la unidad familiar, como dentro de la comunidad o incluso por el Estado, y en cuyos casos extremos puede causar la muerte (ONU, 1981; AGOEA, 1995).

Posteriormente, se identifica el conjunto de derechos y libertades de las mujeres que los países miembro se comprometen a proteger. Gran parte de ellos ya fueron enunciados en 1948 en la Declaración Universal de los Derechos Humanos, pero debido a un análisis minucioso sobre las principales formas de discriminación ejercidas contra las mujeres, visto desde un análisis histórico, tanto la CEDAW como la Convención Belém do Pará vuelven a recalcar:

- I) Entre sus derechos fundamentales, se enuncia que toda mujer tiene el derecho a la vida y a la seguridad personal, gozando de libertad, libertad de asociación y libertad de creencias religiosas, además del derecho a no ser sometidas a torturas reconociéndose su dignidad inherente a su persona. Asimismo a que se respete su integridad tanto física, psíquica y moral, gozando de igualdad de protección ante la ley y de la ley.
- II) Entre sus derechos políticos se reconoce su derecho al voto en cualquier elección o referéndums públicos, además de garantizar, en igualdad de condiciones con el hombre, el poder ser elegidas en cualquier organización cuyos miembros sean objeto de elecciones. Asimismo, se reconoce el derecho a participar en cualquier formulación o ejecución de políticas públicas y poder ocupar cualquier cargo, tanto de la función pública, como en organizaciones no gubernamentales que se ocupen de la vida pública del país, tanto a nivel gubernamental, como en organizaciones internacionales en representación de su gobierno.
- III) Entre sus derechos civiles se reconoce que en igualdad de condiciones con el hombre la mujer pueda adquirir, cambiar o conservar su nacionalidad, enfatizando en el hecho de que la situación conyugal o la nacionalidad del esposo en caso de matrimonio, no cambie automáticamente la nacionalidad de la esposa, además de que ésta goze de los mismos derechos que el hombre con respecto a la nacionalidad de sus hijos.
- IV) Entre sus derechos económicos y sociales, y en igualdad de condiciones con el hombre, se establece que la mujer tiene derecho a gozar de prestaciones familiares, además de poder obtener préstamos bancarios, hipotecas o cualquier otra forma de crédito financiero. Asimismo se reconoce el derecho a poder participar en cualquier actividad de esparcimiento, deportes u cualquier otra actividad en el ámbito cultural.
- v) Además se enuncian una serie de derechos en el ámbito educativo y laboral.

El caso de educación, se recalca entre otras cosas:

- 1) Acceso a los mismos programas de estudios, exámenes y obtención de diplomas en todos los niveles de enseñanza.
- 2) Eliminar cualquier estereotipo que limite o encasille el desarrollo educativo de la mujer en todos los niveles y formas de enseñanza.
- 3) Mismas oportunidades para la obtención de becas o cualquier otra subvención.

En el caso de ocupación se destaca:

- 1) Reconocer al trabajo como un derecho inalienable para toda mujer.
- 2) Tener derecho al ascenso, estabilidad, mismo trato y misma remuneración por trabajo igual.

- 3) Derecho a la seguridad social, estabilidad y mismas prestaciones.
  - 4) Tener acceso a la formación profesional y al readiestramiento.
  - 5) Impedir discriminación por condiciones de embarazo o matrimonio.
  - 6) Implantar la licencia de maternidad con sueldo y prestaciones comparables sin riesgo a perder el empleo.
- VI) Se reconoce el derecho de cada mujer a elegir libre y responsablemente el número de hijos, además de poder tener acceso a información para poder ejercer este derecho.
- VII) En cuanto al matrimonio, se establece determinar una edad mínima para contraer matrimonio, además de:
- 1) Poder elegir libremente a su conyugue.
  - 2) Tener los mismos derechos personales y de propiedad.
  - 3) Gozar de los mismos derechos con la tutela, custodia y adopción de los hijos que su conyugue.

Posteriormente se describen el conjunto de deberes que los países se comprometen a cumplir, que van desde cambios en sus legislaciones internas que prohíban y penalicen cualquier forma de discriminación, hasta en la creación de programas de difusión y apoyo que ayuden a modificar patrones socioculturales que pueden llegar a marginar o encasillar la posición de la mujer dentro de la sociedad.

### **2.3. Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2011**

Con el fin de contextualizar la violencia de género en el país, se presenta un análisis que detalla la situación actual de los tipos de violencia, que impactan en la vida de la mujer como en la sociedad.

Para ello se utilizó la Encuesta Nacional sobre la Dinámica de las Relaciones de los Hogares de 2011 (INEGI, 2011), que permite describir la incidencia y las principales formas de discriminación de la mujer en México, tanto en los ámbitos familiar, laboral, educativo, patrimonial y social.



### 2.3.1. Tablas de contingencia

Las tablas de contingencia son ampliamente utilizadas en el área del análisis de datos categóricos, principalmente para describir la relación entre dos variables, así como la intensidad de esa asociación. Sean  $X$  y  $Y$  dos variables categóricas con  $I$  y  $J$  niveles, entonces la tabla de contingencia estará conformada por las frecuencias que representan al número de individuos que presentan alguna de las  $IJ$  combinaciones posibles formadas por ambas variables (Agresti, 2002).

La relación presente entre ambas variables puede ser descrita a través de su distribución conjunta, aunque también por medio de estas tablas se pueden obtener las distribuciones marginales de cada variable, así como la distribución de una de ellas, digamos la distribución de  $Y$ , dado un nivel específico de la variable  $X$ .

En el Cuadro 3.2 se presenta una tabla de contingencia bidimensional en donde  $n_{ij}$  representa el número de sujetos con las características  $i$  y  $j$  de las variables  $X$  y  $Y$  respectivamente.

Tabla de contingencia					
		Variable Y			
		Y=1	Y=2	...	Y=J
Variable X	X=1	$n_{11}$	$n_{12}$	...	$n_{1J}$
	X=2	$n_{21}$	$n_{22}$	...	$n_{2J}$
	⋮	⋮	⋮	⋮	⋮
	X=I	$n_{I1}$	$n_{I2}$	...	$n_{IJ}$

Cuadro 2.2: Tabla de Contingencia entre las variables  $X$  y  $Y$

#### Distribución Conjunta

Denotada por  $\pi_{ij}$ , ésta representa la probabilidad de que conjuntamente las variable  $X$  y  $Y$  tomen los valores  $i$  y  $j$  respectivamente. Es decir:

$$P(X = i, Y = j) = \pi_{ij}$$

Distribución Conjunta					
		Variable Y			
		Y=1	Y=2	...	Y=J
Variable X	X=1	$\pi_{11}$	$\pi_{12}$	...	$\pi_{1J}$
	X=2	$\pi_{21}$	$\pi_{22}$	...	$\pi_{2J}$
	⋮	⋮	⋮	⋮	⋮
	X=I	$\pi_{I1}$	$\pi_{I2}$	...	$\pi_{IJ}$

Cuadro 2.3: Tabla de distribución conjunta entre las variables  $X$  y  $Y$ .

Los estimadores de estas probabilidades están dados por:

$$\hat{\pi}_{ij} = \frac{n_{ij}}{n}$$

### Distribución Marginal

Denotadas por  $\pi_{i\cdot}$  y  $\pi_{\cdot j}$  a las probabilidades marginales de las variables X y Y respectivamente, para el caso de la primera la distribución marginal se interpreta como la probabilidad de que la variable X tome específicamente el valor de i, y para el caso de la segunda como la probabilidad de que Y tome el valor de j. Es decir:

$$P(X = i) = \pi_{i\cdot} = \sum_j \pi_{ij} = \sum_j \frac{n_{ij}}{n}$$

$$P(Y = j) = \pi_{\cdot j} = \sum_i \pi_{ij} = \sum_i \frac{n_{ij}}{n}$$

De tal manera que

$$\sum_i \pi_{i\cdot} = \sum_j \pi_{\cdot j} = \sum_i \sum_j \pi_{ij} = 1$$

### Distribución Condicional

En muy diversas aplicaciones al estudiar las relaciones existentes entre dos variables, uno puede estar interesado en determinar la distribución de cierta variable dado un nivel específico de la otra variable. Denotada por  $\pi_{j|i}$ , tenemos que:

$$P(Y = j|X = i) = \pi_{j|i} = \frac{\pi_{ij}}{\pi_{i\cdot}}$$

De tal manera que para cada nivel de la variable X se cumple:

$$\sum_j \pi_{j|i} = 1$$

En el cuadro 2.4 se muestra una tabla con las correspondientes distribuciones condicionales de Y para cada valor de X.

Distribución Condicional					
		Variable Y			
		Y=1	Y=2	...	Y=J
Variable X	X=1	$\pi_{1 1}$	$\pi_{2 1}$	...	$\pi_{J 1}$
	X=2	$\pi_{1 2}$	$\pi_{2 2}$	...	$\pi_{J 2}$
	⋮	⋮	⋮	⋮	⋮
	X=I	$\pi_{1 I}$	$\pi_{2 I}$	...	$\pi_{J I}$

Cuadro 2.4: Tabla de distribuciones condicionales de Y dado cada nivel de X

Para el análisis de la encuesta se subdividió la violencia de género en dos partes, la que se comete dentro de las relaciones de pareja y la que se comete por la comunidad o parientes, haciendo un énfasis especial de esta última a la violencia que se comete dentro del ámbito laboral.

### 2.3.2. Violencia en la comunidad

Para contextualizar la violencia que se ejerce dentro de la comunidad y parientes, se tomó una serie de preguntas con las cuales se puede medir la recurrencia de la violencia física, psicológica y sexual, además de localizar a los principales perpetradores y los lugares donde las mujeres se encontraron especialmente vulnerables.

La siguiente tabla enumera estas preguntas, añadiendo sus respectivas probabilidades estimadas de ocurrencia. Así, observamos que el tipo de violencia más recurrente sobre la mujer fue el de ser tocada o manoseada sin su consentimiento, con un valor de 0.1188. A continuación se presenta con mayor detalle cada una de ellas:

Violencia en la comunidad		
No.	Pregunta	Probabilidad Estimada
1	¿La han humillado o denigrado (la han hecho sentir menos o mal frente a otras personas)?	0.0731
2	¿La han tocado o manoseado sin su consentimiento?	0.1188
3	¿La han agredido físicamente?	0.0493
4	¿Le han hecho insinuaciones o propuestas para tener relaciones sexuales a cambio de algo?	0.0817
5	¿La han obligado a tener relaciones sexuales?	0.0121
6	¿Le han hecho sentir miedo de ser atacada o abusada sexualmente?	0.0759

Cuadro 2.5: Tabla que resume el conjunto de preguntas consideradas para analizar la violencia de género fuera de la vida en pareja.

Para cada pregunta, se creó una tabla de contingencia cuyas variables X y Y fueron tomadas de la sección **Ámbito Laboral, Escolar y Social**.

**Variable X:**

Para el caso de la variable X, se tomó la pregunta 2.8 que pedía señalar el lugar dónde se habían realizado la agresión. Así, se formó una nueva variable con tres categorías. A continuación se muestran ambas variables:

<b>Variable Lugar donde Ocurrió la Agresión</b>		
<b>Cve</b>	<b>Nuevo Cve</b>	<b>Descripción</b>
01	1	La calle
02	2	La escuela
03	2	La oficina
04	2	La fábrica o taller
05	3	La casa
06	1	El transporte público
07	1	El cine
08	1	El centro comercial
09	1	Otro

Cuadro 2.6: Variable Lugar donde Ocurrió la Agresión con original y nueva agrupación

<b>Variable Lugar donde Ocurrió la Agresión</b>	
<b>Nuevo Cve</b>	<b>Descripción</b>
1	Fuera de casa
2	Escuela o trabajo
3	Dentro de casa

Cuadro 2.7: Nueva Variable Lugar donde Ocurrió la Agresión

**Variable Y:**

Para el caso de la variable Y, se tomó la pregunta 2.7 que pedía señalar a la persona que había realizado la agresión. De modo que al igual que en el caso anterior, se formó una nueva variable con un total de siete categorías, las cuales se muestran a continuación:

Variable Persona que Realizó la Agresión		
Cve	Nuevo Cve	Descripción
01	1	Papá
02	1	Hermano
03	2	Suegro
04	2	Tío
05	2	Cuñado
06	2	Otro familiar
07	3	Patrón o jefe
08	4	Compañero de trabajo
09	3	Maestro
10	4	Compañero de la escuela
11	3	Director/Otra autoridad escolar
12	5	Desconocido
13	6	Vecino
14	7	Policía o militar
15	6	Amigos
16	7	Otro

Cuadro 2.8: Variable Persona que Realizó la Agresión con original y nueva agrupación

Variable Persona que Realizó la Agresión	
Nuevo Cve	Descripción
1	Familiar directo
2	Familiar no directo
3	Figura de autoridad en Escuela o Trabajo
4	Compañero en Escuela o Trabajo
5	Desconocido
6	Conocido
7	Otro

Cuadro 2.9: Nueva Variable Persona que Realizó la Agresión

### Pregunta 1

¿La han humillado o denigrado (la han hecho sentir menos o mal frente a otras personas)?

Se puede observar a partir de la tabla de distribución conjunta estimada, que de las mujeres que dijeron haber sido humilladas o denigradas, el mayor número de agresiones de este tipo se desarrollaron dentro de casa y fueron cometidas por familiares no directos, observándose una probabilidad estimada de 0.1698, seguido por las agresiones dentro del ámbito laboral o escolar, cometidas por compañeros y por figuras de autoridad, con un valor de 0.1356 y 0.1258 respectivamente. Posteriormente, se encuentra posicionada la violencia ocurrida dentro de casa por familiares directos y la ocurrida fuera de la casa hecha por conocidos, con probabilidades estimadas de 0.1219 y 0.1021.

El Cuadro 2.10, la Figura 2.1 y la Figura 2.2 presentan con mayor detalle cada una de estas probabilidades. Se tiene que de las mujeres que dijeron haber experimentado este tipo de violencia, la probabilidad de que haya sido humillada o denigrada por un familiar no directo es de 0.2338 y la probabilidad de haber sido humillada o denigrada dentro de su hogar es de 0.3710.

Distribución Conjunta y Marginal								
	Fam. Direc.	Fam. no Direc.	Aut. E.o T.	Comp. E.o T.	Desc.	Conocido	Otro	Marginal
Fuera de Casa	0.0282	0.0595	0.0163	0.0159	0.0890	<b>0.1021</b>	0.0119	0.3229
Esc. o Trab.	0.0027	0.0045	<b>0.1258</b>	<b>0.1356</b>	0.0122	0.0217	0.0036	0.3061
Dentro de casa	<b>0.1219</b>	<b>0.1698</b>	0.0054	0.0022	0.0053	0.0620	0.0045	<b>0.3710</b>
Marginal	0.1528	<b>0.2338</b>	0.1475	0.1536	0.1065	<b>0.1858</b>	0.0200	1.0000

Cuadro 2.10: Tabla de distribución conjunta y marginal, sección Violencia en la comunidad, Pregunta 1.

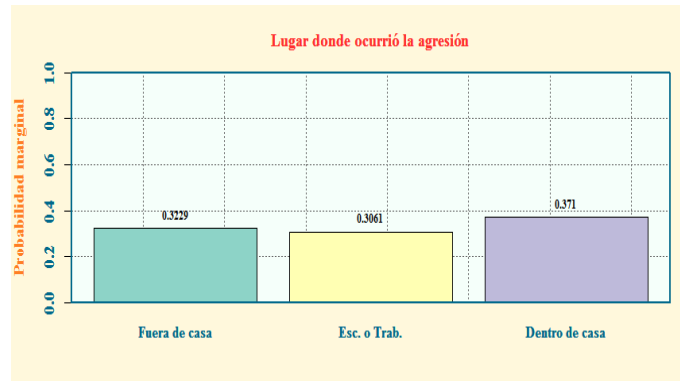


Figura 2.1: Distribución marginal de la variable Lugar donde ocurrió la agresión, Pregunta 1.

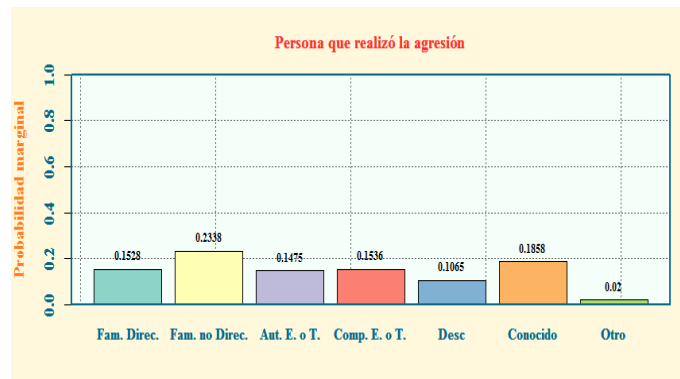


Figura 2.2: Distribución marginal de la variable Persona que realizó la agresión, Pregunta 1.

Para el caso de la distribución condicional, condicionamos a la variable Persona que realizó la agresión a el Lugar donde se realizó la agresión. Así, tenemos que las mujeres que sufrieron este tipo de violencia, la probabilidad de ser humillada o denigrada por un familiar no directo dado que la mujer se encontraba en su hogar es de 0.4576, y la probabilidad de ser humillada o denigrada por un conocido dado que la mujer se encuentra fuera de su hogar es de 0.3161. El cuadro 2.11 y la Figura 2.3 presentan con más detalle estas probabilidades.

Distribución Condicional							
	Fam. Direc.	Fam. no Direc.	Aut. E.o T.	Comp. E.o T.	Desc.	Conocido	Otro
Fuera de Casa	0.0873	0.1843	0.0506	0.0492	0.2755	<b>0.3161</b>	0.0370
Esc. o Trab.	0.0088	0.0147	0.4110	0.4429	0.0399	0.0710	0.0117
Dentro de casa	<b>0.3287</b>	<b>0.4576</b>	0.0145	0.0058	0.0143	0.1670	0.0121

Cuadro 2.11: Tabla de distribución condicional, sección Violencia en la comunidad, Pregunta 1.

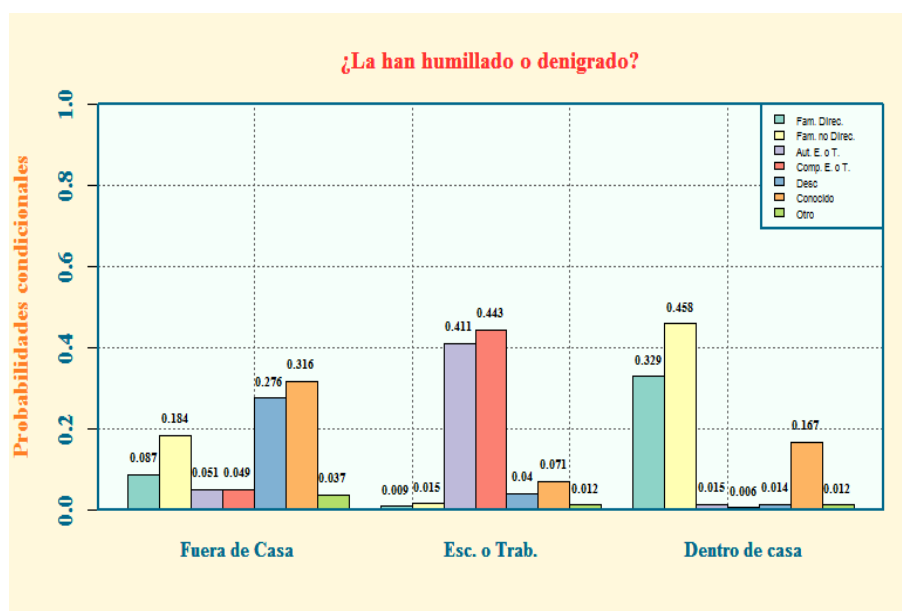


Figura 2.3: Distribuciones condicionando a la variable Lugar donde ocurrió la agresión, sección Violencia en la comunidad, Pregunta 2

En conclusión, se observa que la humillación es un tipo de violencia cometida en su mayoría por personas cercanas a la mujer, es decir, es cometida por familiares, conocidos y amigos, además, de que se comete más frecuentemente dentro del hogar.

## Pregunta 2

¿La han agredido físicamente?

De las mujeres que dijeron haber sido agredidas físicamente, se ve que el mayor número de agresiones se comete fuera de casa por desconocidos, seguido por los casos observados dentro de casa y que son cometidos por familiares directos, luego están por las agresiones físicas cometidas fuera de casa hechas por conocidos. Un dato relevante es que de las mujeres que fueron agredidas físicamente, la probabilidad de que haya ocurrido fuera de casa obtuvo el valor más alto, además de que las probabilidades de ser atacadas por un desconocido o por un conocido cercano resultaron ser muy similares, superando ambas a la violencia cometida por familiares directos y no directos. Los resultados se presentan en el Cuadro 2.12, la Figura 2.4 y Figura 2.5.

Distribución Conjunta y Marginal								
	Fam. Direc.	Fam. no Direc.	Aut. E.o T.	Comp. E.o T.	Desc.	Conocido	Otro	Marginal
Fuera de Casa	0.0302	0.0415	0.0048	0.0197	<b>0.2370</b>	<b>0.1382</b>	0.0077	<b>0.4792</b>
Esc. o Trab.	0.0015	0.0019	0.0146	0.0803	0.0112	0.0138	0.0016	0.1249
Dentro de casa	<b>0.1500</b>	0.1118	0.0017	0.0027	0.0160	0.1089	0.0048	0.3960
Marginal	<b>0.1817</b>	<b>0.1552</b>	0.0212	0.1026	<b>0.2641</b>	<b>0.2610</b>	0.0141	1.0000

Cuadro 2.12: Tabla de distribución conjunta y marginal, sección Violencia en la comunidad, Pregunta 2.

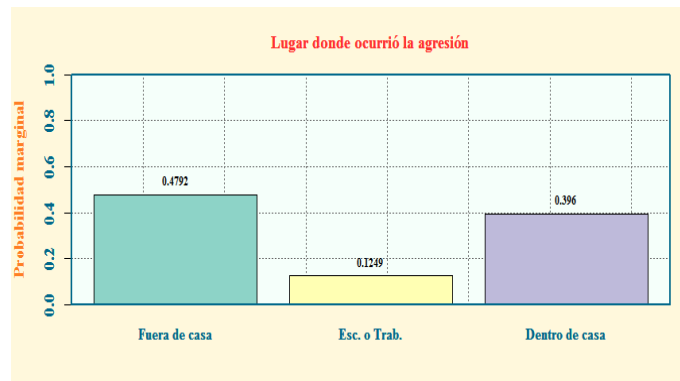


Figura 2.4: Distribución marginal de la variable Lugar donde ocurrió la agresión, Pregunta 2.



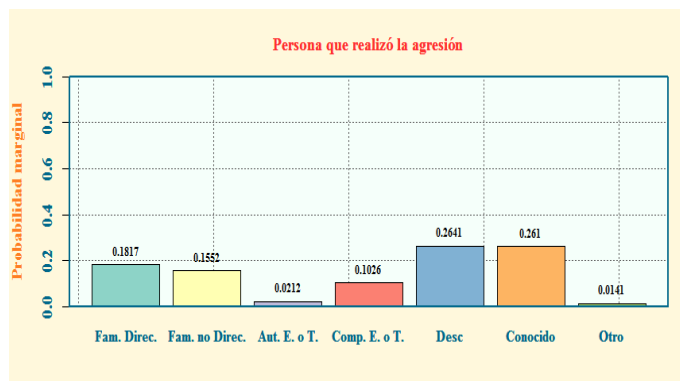


Figura 2.5: Distribución marginal de la variable Persona que realizó la agresión, Pregunta 2.

Para el caso de la distribución condicional, de las mujeres que dijeron haber sido agredidas físicamente, la probabilidad de ser violentadas por compañeros del entorno laboral o escolar dado que la mujer se encontraba en la escuela o trabajo fue de 0.6429, muy superior a la probabilidad de ser agredidas físicamente por un desconocido, dado que la mujer se encontraba en el trabajo o escuela, con un valor de 0.0896. Además, la probabilidad de ser agredidas físicamente por un conocido o un familiar no directo dado que la mujer se encontraba en el hogar es muy similar. El cuadro 2.13 y la Figura 2.6 presentan estas probabilidades.

Distribución Condicional							
	Fam. Direc.	Fam. no Direc.	Aut. E.o T.	Comp. E.o T.	Desc.	Conocido	Otro
Fuera de Casa	0.0631	0.0867	0.0100	0.0411	0.4946	0.2884	0.0161
Esc. o Trab.	0.0117	0.0149	0.1173	<b>0.6429</b>	<b>0.0896</b>	0.1109	0.0128
Dentro de casa	<b>0.3790</b>	<b>0.2824</b>	0.0044	0.0067	0.0403	<b>0.2751</b>	0.0121

Cuadro 2.13: Tabla de distribución condicional, sección Violencia en la comunidad, Pregunta 2.

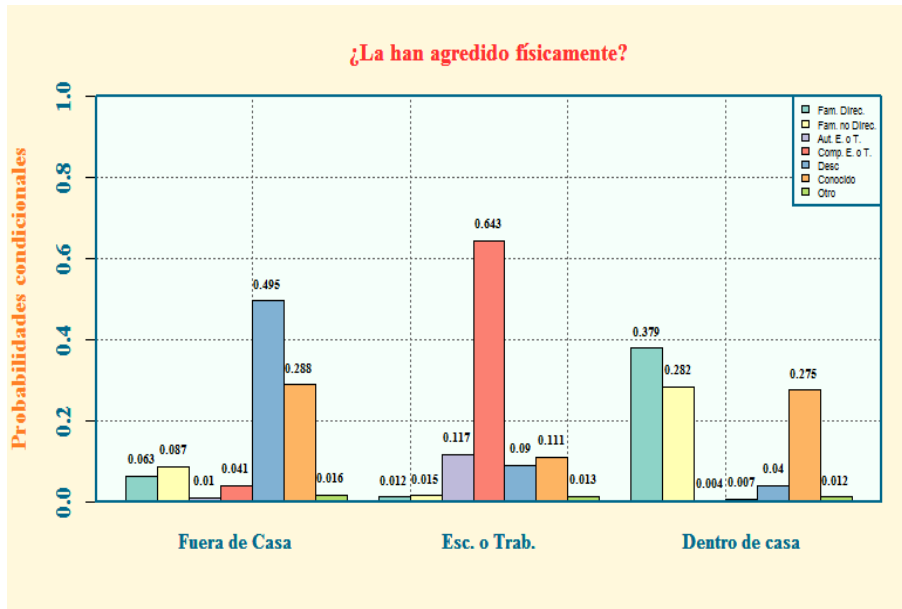


Figura 2.6: Distribuciones condicionando a la variable Lugar donde ocurrió la agresión, sección Violencia en la comunidad, Pregunta 2

### Pregunta 3

¿La han tocado o manoseado sin su consentimiento?

Las mujeres que dijeron haber experimentado este tipo de agresión, en su mayoría ocurrieron fuera de casa cuyos agresores fueron desconocidos, seguido por los que se cometieron dentro del hogar por familiares no directos. Es importante observar que el número de agresiones ocurridas en el primer caso está muy por encima del resto de los escenarios. Las probabilidades se muestran en el Cuadro 2.14, Figura 2.7 y Figura 2.8.

Distribución Conjunta y Marginal								
	Fam. Direc.	Fam. no Direc.	Aut. E.o T.	Comp. E.o T.	Desc.	Conocido	Otro	Marginal
Fuera de Casa	0.0056	0.0314	0.0127	0.0072	<b>0.6783</b>	0.0354	0.0100	<b>0.7806</b>
Esc. o Trab.	0.0005	0.0017	0.0267	0.0284	0.0107	0.0038	0.0011	0.0729
Dentro de casa	0.0304	<b>0.0828</b>	0.0019	0.0007	0.0100	0.0174	0.0034	0.1465
Marginal	0.0365	<b>0.1159</b>	0.0412	0.0363	<b>0.6990</b>	0.0566	0.0145	1.0000

Cuadro 2.14: Tabla de distribución conjunta y marginal, sección Violencia en la comunidad, Pregunta 3.

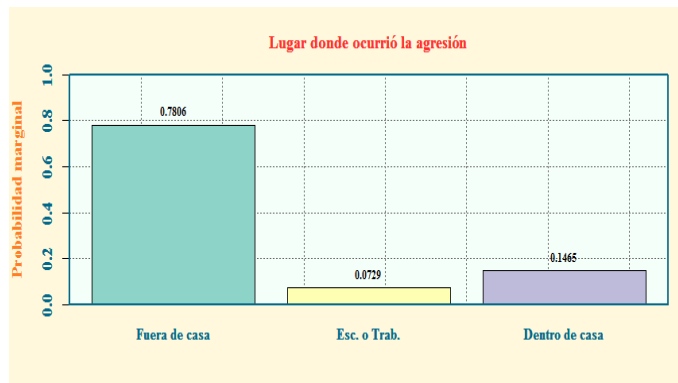


Figura 2.7: Distribución marginal de la variable Lugar donde ocurrió la agresión, Pregunta 3.

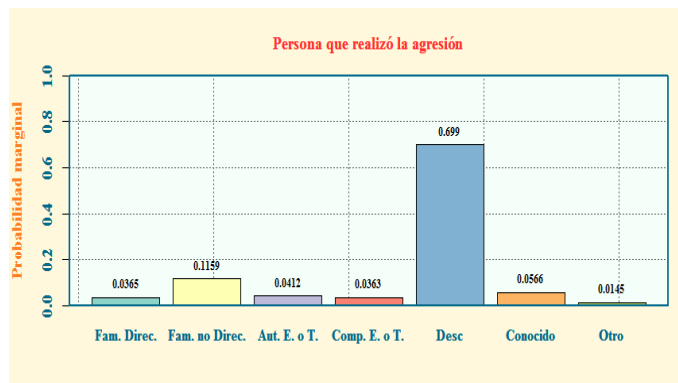


Figura 2.8: Distribución marginal de la variable Persona que realizó la agresión, Pregunta 3.

También se presentan las distribuciones condicionales, de tal forma que de los casos reportados, la probabilidad más alta que se observó fue la de ser tocada o manoseada por un desconocido dado que la mujer se encontraba fuera de casa, con un valor de 0.8690, seguida de la probabilidad de ser tocada o manoseada por un familiar no directo dado que la mujer se encontraba en su hogar, reportando un valor de 0.5650. Además, condicionando a los casos en que las agresiones ocurrieron en el lugar de trabajo o escuela, la probabilidad de ser agredida por un compañero o por una figura de autoridad fue muy parecida, siendo de 0.3894 y 0.3659 respectivamente.

Distribución Condicional							
	Fam. Direc.	Fam. no Direc.	Aut. E.o T.	Comp. E.o T.	Desc.	Conocido	Otro
Fuera de Casa	0.0072	0.0402	0.0163	0.0092	<b>0.8690</b>	0.0453	0.0129
Esc. o Trab.	0.0068	0.0235	<b>0.3659</b>	<b>0.3894</b>	0.1470	0.0523	0.0152
Dentro de casa	0.2075	<b>0.5650</b>	0.0128	0.0049	0.0682	0.1186	0.0230

Cuadro 2.15: Tabla de distribución condicional, sección Violencia en la comunidad, Pregunta 3.

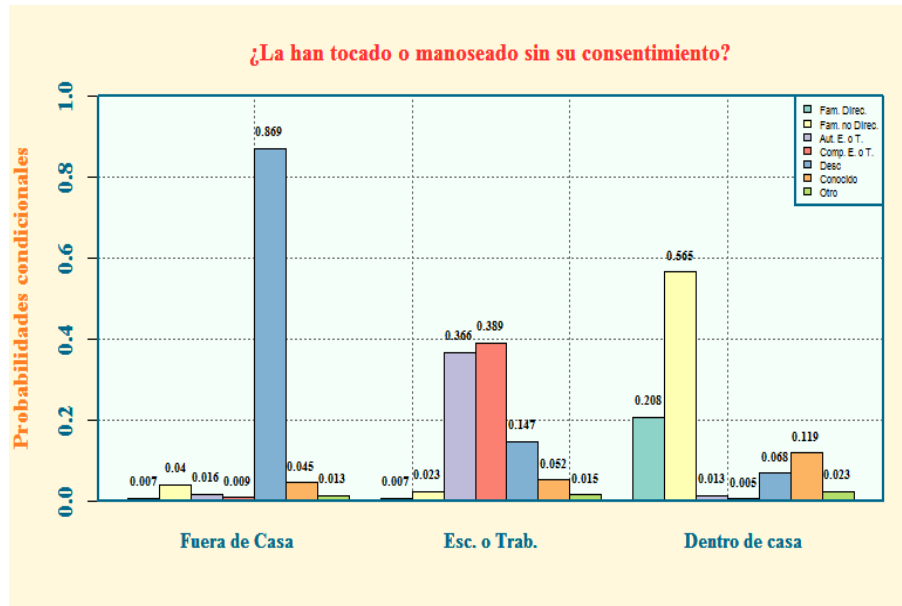


Figura 2.9: Distribuciones condicionando a la variable Lugar donde ocurrió la agresión, sección Violencia en la comunidad, Pregunta 3

#### Pregunta 4

¿Le han hecho insinuaciones o propuestas para tener relaciones sexuales a cambio de algo?

Este tipo de agresión predominó en el entorno fuera del hogar, cometida tanto por personas desconocidas como por personas que la mujer conocía bien. Seguida por el ámbito laboral o escolar, perpetrada principalmente por personas con un rango de autoridad sobre la mujer y por compañeros de trabajo. El cuadro 2.16 presenta la tabla de probabilidades conjuntas y marginales.

Distribución Conjunta y Marginal								
	Fam. Direc.	Fam. no Direc.	Auto. E.o T.	Comp. E.o T.	Desc.	Conocido	Otro	Marginal
Fuera de Casa	0.0028	0.0237	0.0259	0.0281	<b>0.2994</b>	<b>0.1245</b>	0.0139	<b>0.5183</b>
Esc. o Trab.	0.0003	0.0030	<b>0.1745</b>	<b>0.1368</b>	0.0323	0.0226	0.0071	0.3766
Dentro de casa	0.0073	0.0430	0.0031	0.0016	0.0118	0.0340	0.0042	0.1050
Marginal	0.0104	0.0697	<b>0.2035</b>	0.1665	<b>0.3435</b>	0.1812	0.0251	1.0000

Cuadro 2.16: Tabla de distribución conjunta y marginal, sección Violencia en la comunidad, Pregunta 4.

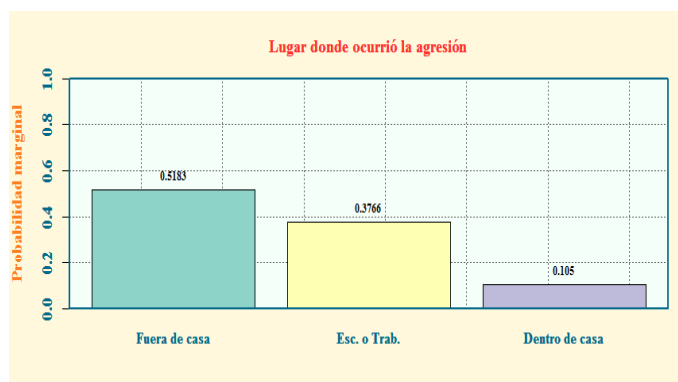


Figura 2.10: Distribución marginal de la variable Lugar donde ocurrió la agresión, Pregunta 4.

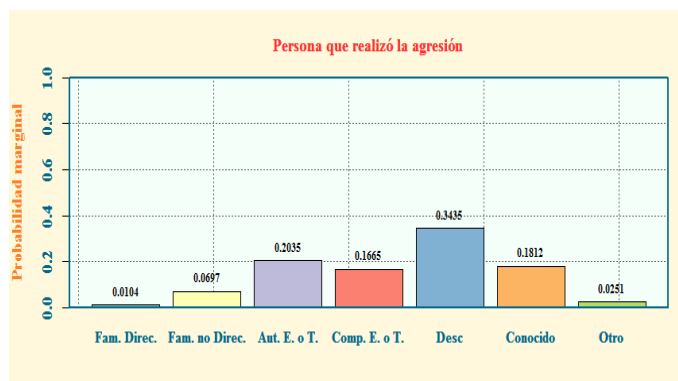


Figura 2.11: Distribución marginal de la variable Persona que realizó la agresión, Pregunta 4.

Para el caso de las probabilidades condicionales, un escenario interesante se observa cuando se condiciona el lugar de la agresión a dentro del hogar, donde se observa que de las mujeres que dijeron haber recibido esta propuesta, la probabilidad de ser agredida por un familiar no directo y un conocido, dado que la mujer se encontraba dentro de su hogar es de 0.4095 y 0.3239, respectivamente. Los resultados se muestran en el Cuadro 2.17 y la Figura 2.12.

Distribución Condicional							
	Fam. Direc.	Fam. no Direc.	Aut. E.o T.	Comp. E.o T.	Desconocido	Conocido	Otro
Fuera de Casa	0.0054	0.0457	0.0500	0.0542	<b>0.5776</b>	<b>0.2403</b>	0.0268
Esc. o Trab.	0.0009	0.0081	0.4632	0.3633	0.0857	0.0601	0.0188
Dentro de casa	0.0695	<b>0.4095</b>	0.0298	0.0153	0.1123	<b>0.3239</b>	0.0397

Cuadro 2.17: Tabla de distribución condicional, sección Violencia en la comunidad, Pregunta 4.

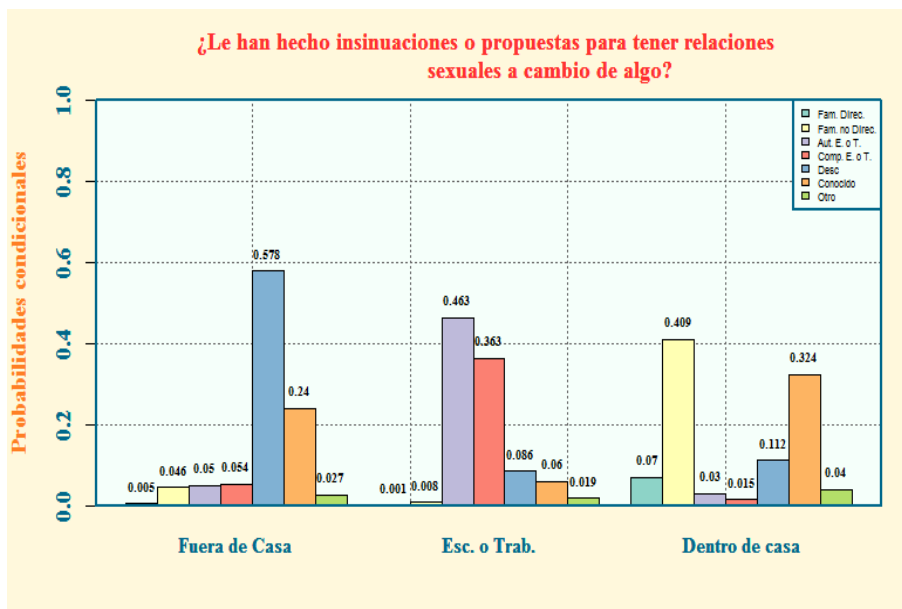


Figura 2.12: Distribuciones condicionando a la variable Lugar donde ocurrió la agresión, sección Violencia en la comunidad, Pregunta 4

En conclusión, el escenario más recurrente en el cual se realizan propuestas o insinuaciones para tener relaciones sexuales a cambio de algo fue fuera del hogar, seguido muy de cerca por las insinuaciones hechas en el entorno laboral o escolar, y donde los principales agresores son personas desconocidas y personas que son figuras de autoridad en el trabajo.

### Pregunta 5

¿Le han hecho sentir miedo de ser atacada o abusada sexualmente?

Nuevamente, de las mujeres que dijeron haber estado en esta situación, se observa que el mayor número de agresiones ocurrieron fuera del hogar por personas desconocidas, siendo en general el escenario fuera de casa el más propenso a este tipo de violencia, y cuyos principales responsables en cometerla fueron tanto personas desconocidas, como

conocidas pero sin ningún parentesco familiar, seguido de familiares pero no directos de la mujer. Los recuentos de cada caso y las probabilidades conjuntas y marginales se muestran en el Cuadro 2.18, Figura 2.13 y Figura 2.14.

Distribución Conjunta y Marginal								
	Fam. Direc.	Fam. no Direc.	Aut. E.o T.	Comp. E.o T.	Desc.	Conocido	Otro	Marginal
Fuera de Casa	0.0060	0.0380	0.0120	0.0125	0.5405	0.0857	0.0130	<b>0.7076</b>
Esc. o Trab.	0.0005	0.0015	0.0465	0.0405	0.0143	0.0080	0.0025	0.1137
Dentro de casa	0.0271	0.0792	0.0030	0.0010	0.0231	0.0402	0.0051	0.1787
Marginal	0.0336	<b>0.1186</b>	0.0616	0.0539	<b>0.5779</b>	<b>0.1338</b>	0.0206	1.0000

Cuadro 2.18: Tabla de distribución conjunta y marginal, sección Violencia en la comunidad, Pregunta 5.

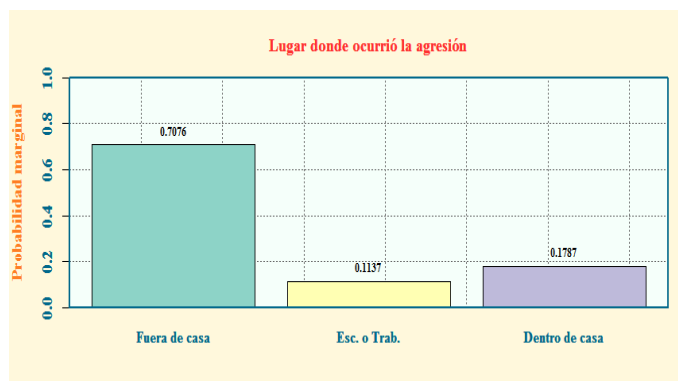


Figura 2.13: Distribución marginal de la variable Lugar donde ocurrió la agresión, Pregunta 5.

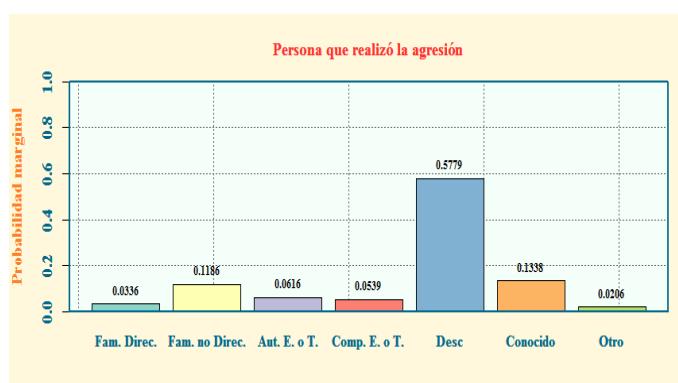


Figura 2.14: Distribución marginal de la variable Persona que realizó la agresión, Pregunta 5.

Por otro lado, al condicionar por ámbito laboral o escolar, se tiene que los principales responsables de estas agresiones son aquellas personas que eran figuras de autoridad sobre la mujer, seguida por sus compañeros y desconocidos. Los resultados se muestran en el Cuadro 2.19 y Figura 2.15.

Distribución Condicional							
	Fam. Direc.	Otro Fam.	Aut. E.o T.	Comp. E.o T.	Desconocido	Conocido	Otro
Fuera de Casa	0.0084	0.0536	0.0170	0.0177	0.7638	0.1211	0.0183
Esc. o Trab.	0.0046	0.0129	<b>0.4091</b>	<b>0.3559</b>	0.1255	0.0700	0.0221
Dentro de casa	0.1519	0.4432	0.0169	0.0053	0.1292	0.2250	0.0285

Cuadro 2.19: Tabla de distribución condicional, sección Violencia en la comunidad, Pregunta 5.

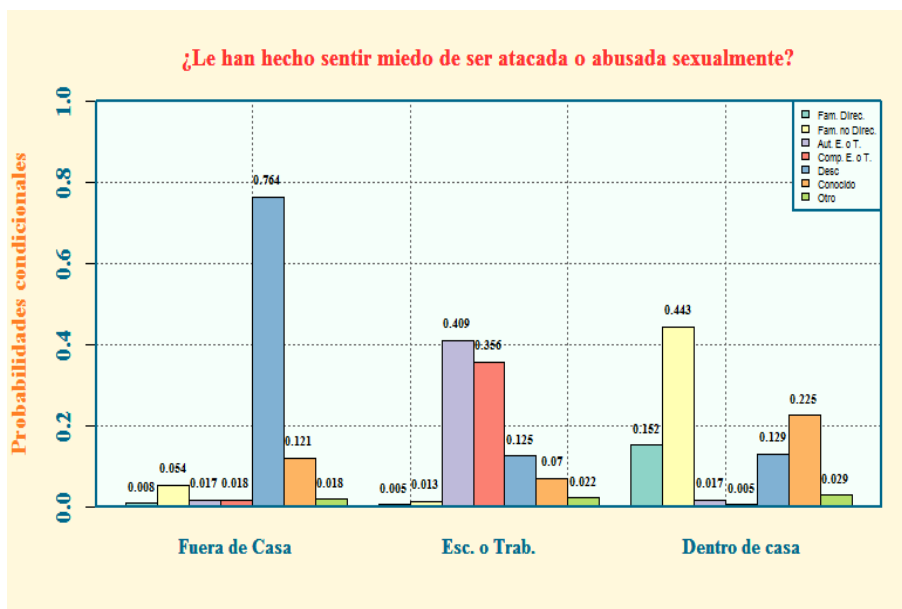


Figura 2.15: Distribuciones condicionando a la variable Lugar donde ocurrió la agresión, sección Violencia en la comunidad, Pregunta 5

## Pregunta 6

¿La han obligado a tener relaciones sexuales?

De las mujeres que afirmaron haber sido víctimas de abuso sexual, se ve que la probabilidad de ser abusada fuera del hogar o dentro del hogar es muy parecida, con un valor de 0.4523 y 0.4403, y cuyos principales responsables resultaron ser personas que eran conocidas por la mujer, seguido por desconocidos y familiares no directos, presentando una probabilidad de ocurrencia de 0.3216, 0.2556 y 0.2202, respectivamente.



Distribución Conjunta y Marginal								
	Fam. Direc.	Otro Fam.	Aut. E.o T.	Comp. E.o T.	Desc.	Conocido	Otro	Marginal
Fuera de Casa	0.0163	0.0553	0.0136	0.0152	0.1866	0.1437	0.0217	<b>0.4523</b>
Esc. o Trab.	0.0011	0.0027	0.0456	0.0293	0.0119	0.0146	0.0022	0.1074
Dentro de casa	0.0651	0.1621	0.0049	0.0087	0.0271	0.1632	0.0092	<b>0.4403</b>
Marginal	0.0824	<b>0.2202</b>	0.0640	0.0531	<b>0.2256</b>	<b>0.3216</b>	0.0331	1.0000

Cuadro 2.20: Tabla de distribución conjunta y marginal, sección Violencia en la comunidad, Pregunta 6.

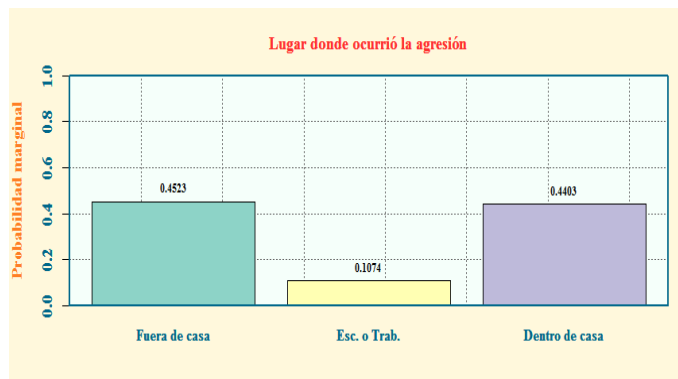


Figura 2.16: Distribución marginal de la variable Lugar donde ocurrió la agresión, Pregunta 6.

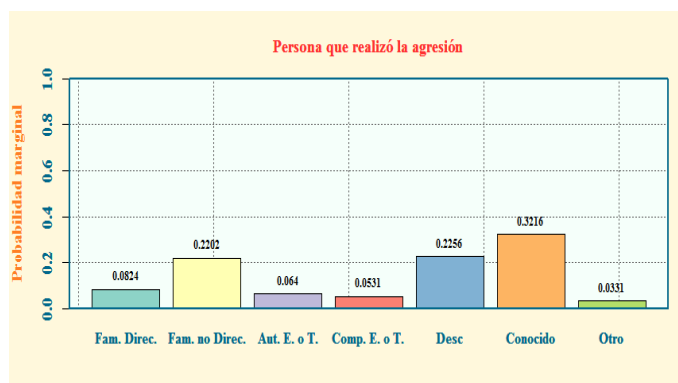


Figura 2.17: Distribución marginal de la variable Persona que realizó la agresión, Pregunta 6.

De las mujeres que afirmaron haber sido abusadas sexualmente, la probabilidad de que haya sido por un desconocido y un conocido con lazos no familiares dado que la agresión sucedió fuera de casa es muy parecida, con un valor de 0.4125 y 0.3177, respectivamente. Por otro lado, condicionando a que la agresión ocurrió en el entorno laboral o escolar, la probabilidad de haber sido abusada por una persona que era una

figura de autoridad dentro de la institución fue de 0.4242, la más alta observada en este escenario. Los resultados se muestran en el Cuadro 2.21 y Figura 2.18.

Distribución Condicional							
	Fam. Direc.	Otro Fam.	Aut. E.o T.	Comp. E.o T.	Desc.	Conocido	Otro
Fuera de Casa	0.0360	0.1223	0.0300	0.0336	<b>0.4125</b>	<b>0.3177</b>	0.0480
Esc. o Trab.	0.0101	0.0253	<b>0.4242</b>	0.2727	0.1111	0.1364	0.0202
Dentro de casa	0.1478	0.3682	0.0111	0.0197	0.0616	0.3707	0.0209

Cuadro 2.21: Tabla de distribución condicional, sección Violencia en la comunidad, Pregunta 6.

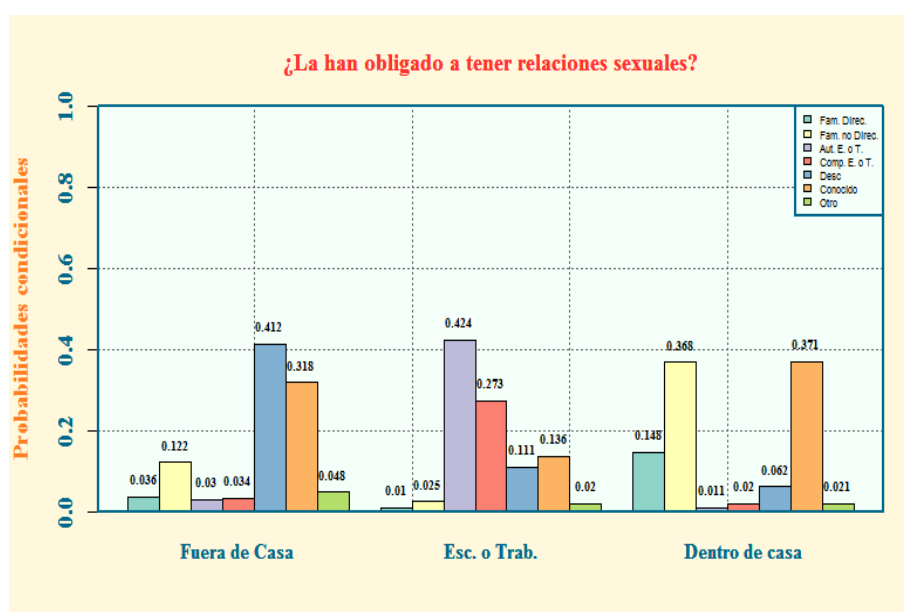


Figura 2.18: Distribuciones condicionando a la variable Lugar donde ocurrió la agresión, sección Violencia en la comunidad, Pregunta 6

## Violencia laboral

Una forma de violencia ejercida fuera del hogar es la que se comete dentro del entorno laboral, fundada a partir de estereotipos y costumbres que impiden alcanzar de manera plena el empoderamiento económico de la mujer. El negar la contratación, o negar las mismas condiciones y oportunidades de ascenso en situaciones iguales por condiciones de género, así como incurrir en descalificaciones injustificadas o humillaciones, además del acoso sexual, forman parte de algunas de las formas de violencia más reiteradas en este ámbito.

Para ilustrar la frecuencia de este tipo de violencia en algunos de los diferentes centros de trabajo, se tomaron una serie de preguntas de la subsección *Ámbito laboral*, haciendo nuevamente un análisis de tablas de contingencia similar al hecho en la sección anterior.

**Variable X:** Para el caso de la variable X, se tomó la pregunta 2.4, la cual registraba el lugar donde la mujer trabajaba en ese momento, o en su defecto el que trabajó pero únicamente durante el último año. La variable original y con la cual se hizo el análisis es:

<b>Variable Lugar de trabajo</b>	
<b>Cve</b>	<b>Descripción</b>
01	Una institución pública
02	Una empresa privada o banco
03	Un comercio
04	Una escuela
05	Una fábrica
06	Un taller
07	En la calle o vía pública
08	En una casa ajena (servicio doméstico)
09	En su propia casa
10	En el campo
11	Otro

Cuadro 2.22: Variable Lugar donde trabajaba la mujer durante el último año.

**Variable Y:** Para el caso de la variable Y, se tomó la pregunta 2.5, la cual consta a su vez de las cinco preguntas elegidas para estudiar las agresiones laborales, y cuyas respuestas únicamente responden a si experimentaron o no alguna de las situaciones planteadas.

### **Pregunta 1**

¿Le han pagado menos que a un hombre que desarrolla el mismo trabajo?

A continuación se muestra, al igual que en todo el análisis de violencia en la comunidad, una tabla de probabilidades condicionales. Así, observamos que los talleres, trabajos en servicios domésticos, el campo y fábricas, son los principales entornos donde se presentan las mayores probabilidades de ser discriminada, y recibir un pago menor por un trabajo igual al de un compañero hombre. Las probabilidades se muestran en el Cuadro 2.23 y la Figura 2.19:

Distribución Condicional		
	Sí	No
I. Pública	0.1037	0.8963
E. Privada o Banco	0.1380	0.8620
Comercio	0.1415	0.8585
Escuela	0.0988	0.9012
Fábrica	<b>0.1876</b>	0.8124
Taller	<b>0.2708</b>	0.7292
Vía P.	0.1589	0.8411
S. Dóm.	<b>0.1927</b>	0.8073
Propia casa	0.1331	0.8669
Campo	<b>0.1919</b>	0.8081
Otro	0.1406	0.8594

Cuadro 2.23: Tabla de distribución condicional, sección Violencia laboral, Pregunta 1.

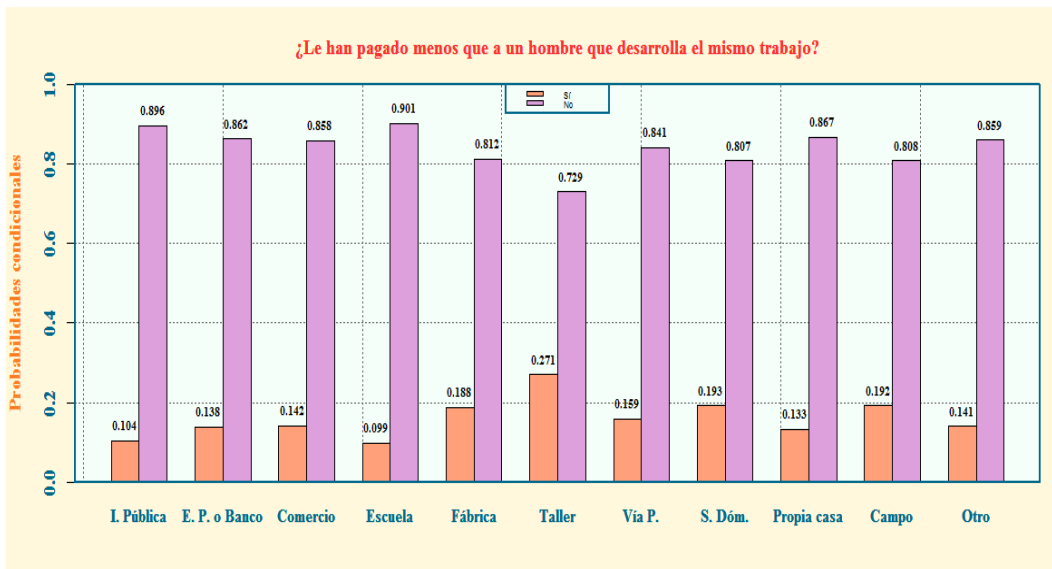


Figura 2.19: Distribuciones condicionando a la variable Lugar donde trabajaba la mujer, sección Violencia laboral, Pregunta 1

## Pregunta 2

¿Ha tenido menos oportunidad que un hombre para ascender?

Condicionando a que la mujer trabajaba en un taller, la probabilidad de ser discriminada es la más alta de entre todos los lugares de trabajo, con un valor de 0.2304, seguida por la probabilidad de ser discriminada dado que se trabajaba en una fábrica, con un valor de 0.2147. Por otro lado, la probabilidad más pequeña corresponde a la de ser discriminada dado que se trabajaba en una institución escolar, por una cantidad de 0.1049. En el Cuadro 2.24 y la Figura 2.10 se muestran las probabilidades para todos los escenarios:

Distribución Condicional		
	Sí	No
I. Pública	0.1329	0.8671
E. Privada o Banco	0.1478	0.8522
Comercio	0.1293	0.8707
Escuela	<b>0.1049</b>	0.8951
Fábrica	<b>0.2147</b>	0.7853
Taller	<b>0.2304</b>	0.7696
Vía P.	0.1513	0.8487
S. Dóm.	0.1973	0.8027
Propia casa	0.1286	0.8714
Campo	0.1840	0.8160
Otro	0.1467	0.8533

Cuadro 2.24: Tabla de distribución condicional, sección Violencia laboral, Pregunta 2.

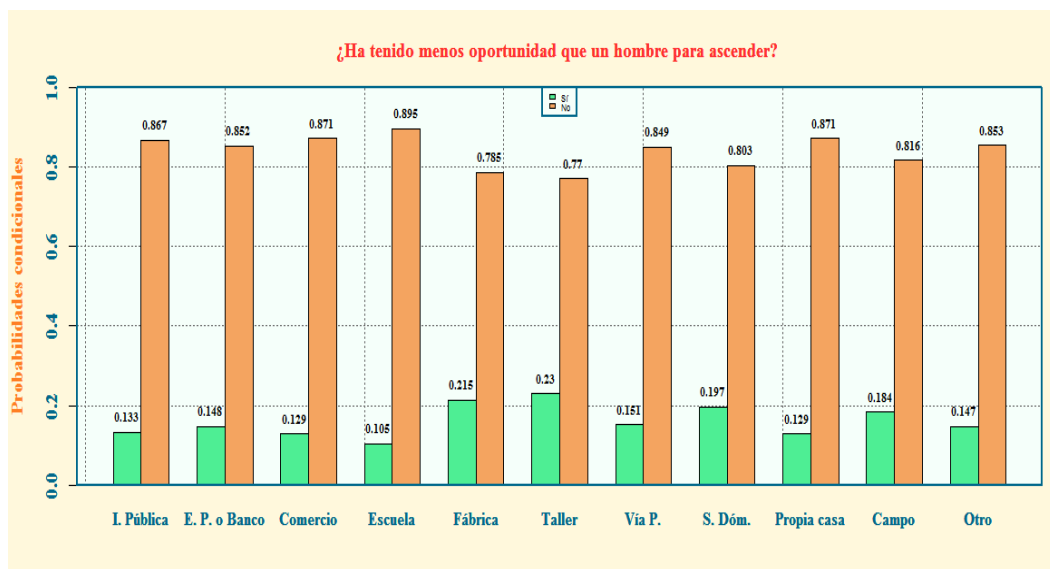


Figura 2.20: Distribuciones condicionando a la variable Lugar donde trabajaba la mujer, sección Violencia laboral, Pregunta 2

### Pregunta 3

¿Ha recibido menos prestaciones que un hombre?

Observamos que las probabilidades de que una mujer reciba menos prestaciones que un compañero hombre son menores a los escenarios anteriormente planteados, de modo que es un tipo de violencia menos recurrente. Sin embargo, nuevamente se observa que la probabilidad de recibir menos prestaciones dado que su entorno laboral es una fábrica, dando servicios domésticos o trabajando en el campo, son las más altas, con los valores de 0.1613, 0.1647 y 0.1376, respectivamente. En cuanto a los demás escenarios, los resultados se muestran en el Cuadro 2.25 y la Figura 2.21:

Distribución Condicional		
	Sí	No
I. Pública	0.0791	0.9209
E. Privada o Banco	0.0918	0.9082
Comercio	0.0973	0.9027
Escuela	0.0695	0.9305
Fábrica	<b>0.1203</b>	0.8797
Taller	0.1613	0.8387
Vía P.	0.1077	0.8923
S. Dóm.	<b>0.1647</b>	0.8353
Propia casa	0.1018	0.8982
Campo	<b>0.1376</b>	0.8624
Otro	0.1182	0.8818

Cuadro 2.25: Tabla de distribución condicional, sección Violencia laboral, Pregunta 3.

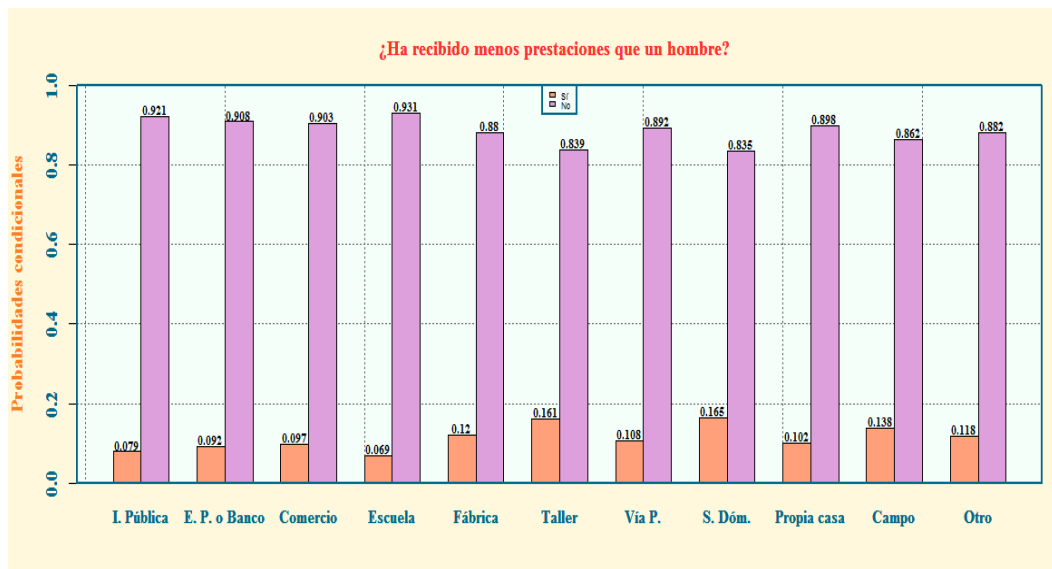


Figura 2.21: Distribuciones condicionando a la variable Lugar donde trabajaba la mujer, sección Violencia laboral, Pregunta 3

## Pregunta 4

¿En su trabajo le han pedido la prueba de embarazo?

En cuanto al requisito de la prueba de embarazo, la probabilidad de que a una mujer que trabaja en una fábrica le pidan esta prueba es la más alta de entre todos los espacios de trabajo evaluados, con un valor de 0.1799, seguido por las escuelas y las instituciones públicas, con probabilidades de 0.1163 y 0.1139.

Distribución Condicional		
	Sí	No
I. Pública	<b>0.1139</b>	0.8861
E. Privada o Banco	0.0897	0.9103
Comercio	0.0573	0.9427
Escuela	<b>0.1163</b>	0.8837
Fábrica	<b>0.1799</b>	0.8201
Taller	0.0509	0.9491
Vía P.	0.0386	0.9614
S. Dóm.	0.0292	0.9708
Propia casa	0.0406	0.9594
Campo	0.0202	0.9798
Otro	0.0643	0.9357

Cuadro 2.26: Tabla de distribución condicional, sección Violencia laboral, Pregunta 4.

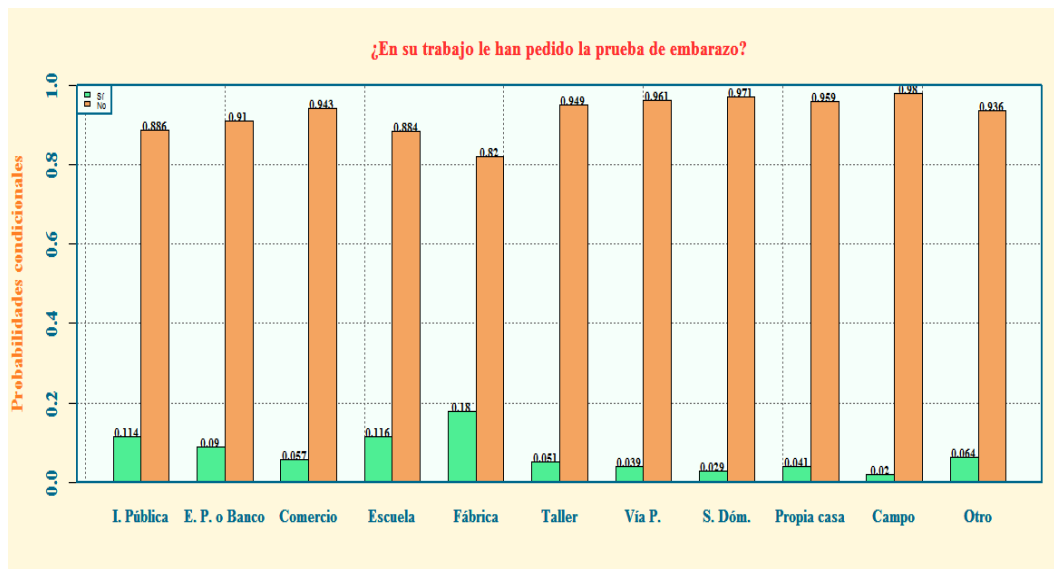


Figura 2.22: Distribuciones condicionando a la variable Lugar donde trabajaba la mujer, sección Violencia laboral , Pregunta 4

### 2.3.3. Violencia dentro de la relación de pareja

Desde los orígenes de la lucha por el cese de cualquier tipo de violencia en contra de las mujeres, debido a su alta tolerancia social e histórica, la violencia que se comete dentro de la relación de pareja ha sido una de las más analizadas (ONU MUJERES, 2017)<sup>B</sup>.

Para examinar esta problemática, se propuso estudiar la respuesta a una serie de preguntas, nuevamente por medio de tablas de contingencia y probabilidades conjuntas, las cuales visualizan las frecuencias con que este tipo de agresiones ocurrieron dentro de la última relación sostenida por la mujer entrevistada, así como la opinión que se tiene con respecto a su situación.

Para cada pregunta se consideraron únicamente los casos en que la persona entrevistada había sufrido la situación planteada, tomando como variable X a la frecuencia con que las agresiones ocurrieron u ocurren dentro de la última relación de pareja, y como variable Y a la opinión que éstas mismas tienen con respecto a la situación que padecen o padecieron. De esta manera, el principal objetivo consistió en analizar la percepción que tienen las mujeres víctimas de violencia ante las diferentes situaciones planteadas y su frecuencia. Las variables fueron tomadas de la sección Relación actual, correspondientes a las preguntas 6.1 y 6.2. A continuación se muestra la categorización:

#### Variable X

Variable X	
1	Muchas veces?
2	Pocas veces?
3	Una vez?

Cuadro 2.27: Variable Frecuencia con que ocurre la agresión

#### Variable Y

Variable Y	
1	Muy grave
2	Grave
3	Sin importancia

Cuadro 2.28: Variable Gravedad de la agresión



## Pregunta 1

¿La ha avergonzado, menospreciado o humillado?

Podemos observar que de las mujeres que han sido avergonzadas, menospreciadas o humilladas, la probabilidad de que sea un acto frecuente es de 0.4770, y la probabilidad de que se considere un suceso sin importancia es de 0.1884. El Cuadro 2.29 y la Figura 2.23 presentan las probabilidades de los demás escenarios.

Distribución Marginal X		
Muchas veces	Pocas veces	Una vez
<b>0.4770</b>	0.4106	0.1125
Distribución Marginal Y		
Muy grave	Grave	Sin importancia
0.3219	0.4897	<b>0.1884</b>

Cuadro 2.29: Distribución marginal de las variables Frecuencia y Gravedad de la agresión, sección Violencia dentro de la relación de Pareja, Pregunta 1.

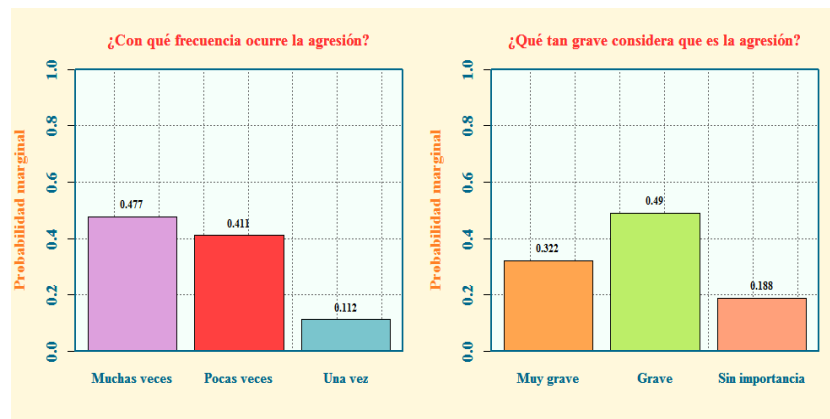


Figura 2.23: Gráficas de las distribuciones marginales Frecuencia y Gravedad de la agresión, Pregunta 1.

Por otro lado, condicionando a la frecuencia en que ocurre la agresión, observamos que la probabilidad de que la mujer violentada considere el suceso sin importancia dado que ocurre muchas veces es de 0.1104. Sin embargo, en general, se observa que sin importar la frecuencia con que son agredidas, las mujeres consideran que la situación que viven o vivieron es grave, y sólo la consideraron muy grave en el caso en que la agresión era demasiado frecuente. Las probabilidades se muestran en el Cuadro 2.30 y Figura 2.24.

Distribución Condicional			
	Muy grave	Grave	Sin importancia
Muchas veces	0.4598	0.4299	<b>0.1104</b>
Pocas veces	0.1908	0.5657	0.2435
Una vez	0.2161	0.4660	0.3178

Cuadro 2.30: Tabla de distribución condicional, sección Violencia dentro de la relación de Pareja, Pregunta 1.

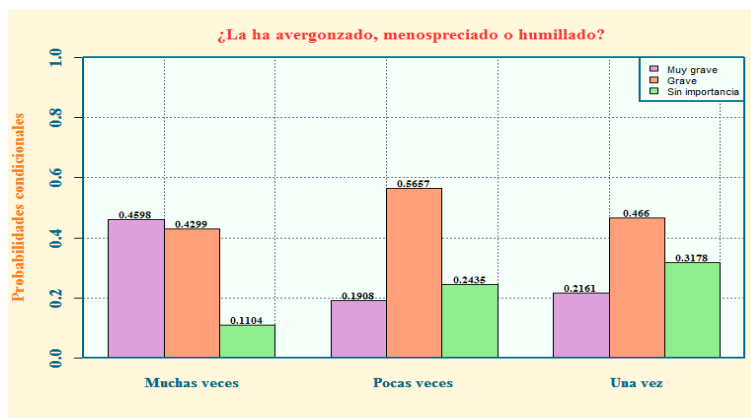


Figura 2.24: Distribuciones condicionando a la variable Frecuencia con que ocurre la agresión, Pregunta 1.

## Pregunta 2

¿Le ha hecho sentir miedo?

En este caso, observamos que de las parejas que dijeron haber sentido miedo por sus respectivas parejas, la probabilidad de que sea un acto frecuente es de 0.5345, y en general, con una probabilidad de 0.1870 la persona agredida lo consideran un suceso sin importancia. A continuación se presentan las probabilidades marginales.

Distribución Marginal X		
Muchas veces	Pocas veces	Una vez
<b>0.5345</b>	0.3846	0.0810
Distribución Marginal Y		
Muy grave	Grave	Sin importancia
0.3779	0.4351	<b>0.1870</b>

Cuadro 2.31: Distribución marginal de las variables Frecuencia y Gravedad de la agresión, sección Violencia dentro de la relación de Pareja, Pregunta 2.

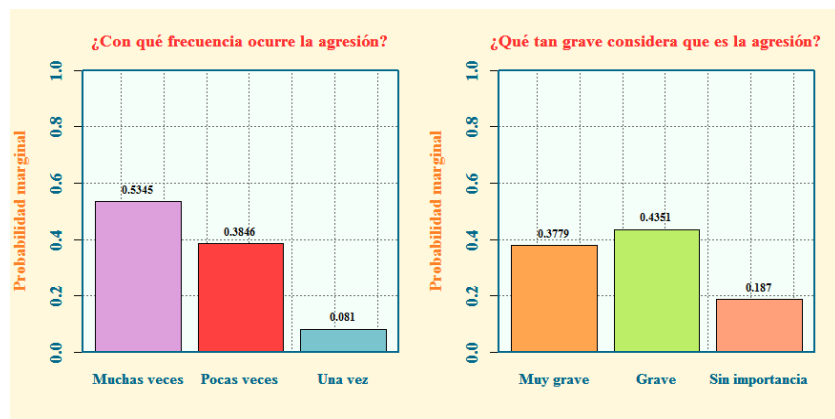


Figura 2.25: Distribución marginal de las variables Frecuencia con que ocurre la agresión, Pregunta 2.

Condicionando a la frecuencia con que ocurre la agresión, en los casos en que esta situación ha ocurrido pocas veces o sólo una vez, con probabilidades relativamente altas consideran que el haber sentido miedo por sus parejas es una situación sin importancia, con valores de 0.2809 y 0.3401 respectivamente. Los resultados se presentan a continuación.

Distribución Condicional			
	Muy grave	Grave	Sin importancia
Muchas veces	0.5309	0.3729	0.0962
Pocas veces	0.1942	0.5249	<b>0.2809</b>
Una vez	0.2404	0.4196	<b>0.3401</b>

Cuadro 2.32: Tabla de distribución condicional, sección Violencia dentro de la relación de Pareja, Pregunta 2.

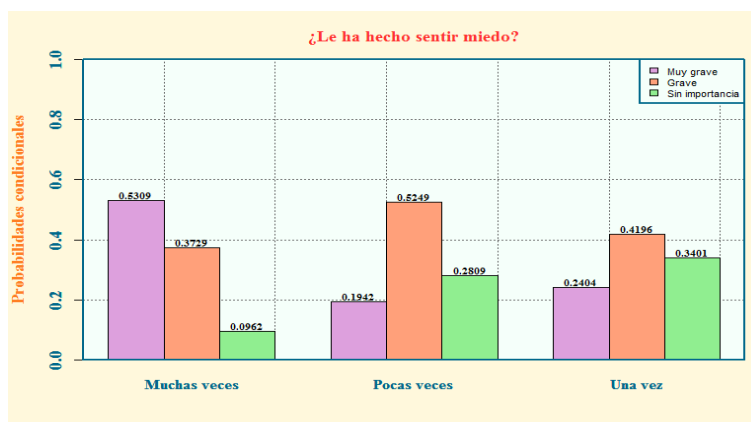


Figura 2.26: Distribuciones condicionando a la variable Frecuencia con que ocurre la agresión, Pregunta 2.

### Pregunta 3

¿La ha amenazado con algún arma (cuchillo, navaja, pistola o rifle)?

De las mujeres que dijeron haber sido amenazadas con algún arma por sus respectivas parejas, la probabilidad de que además ésta sea una práctica frecuente es de 0.3888. Por otro lado, también se observa que este hecho en general se considera de especial gravedad entre las encuestadas, pues la probabilidad de que el hecho lo consideren sin importancia tienen un valor de 0.0551.

Distribución Marginal X		
Muchas veces	Pocas veces	Una vez
<b>0.3888</b>	0.3003	0.3109
Distribución Marginal Y		
Muy grave	Grave	Sin importancia
<b>0.6948</b>	0.2501	<b>0.0551</b>

Cuadro 2.33: Distribución marginal de las variables Frecuencia y Gravedad de la agresión, sección Violencia dentro de la relación de Pareja, Pregunta 3.

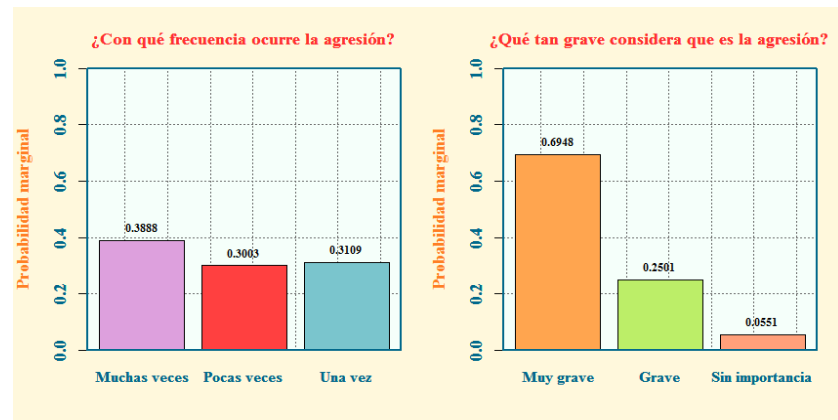


Figura 2.27: Distribución marginal de las variables Frecuencia con que ocurre la agresión, Pregunta 3.

Con respecto a las distribuciones condicionales, independientemente de la frecuencia con que ocurre la agresión, en general la opinión es que esta situación debe ser considerada en extremo grave. Además, aproximadamente sólo tres de cada cien mujeres que son amenazadas con un arma por su conyugue, consideran que la situación no tiene importancia, dado que ocurre muchas veces. El Cuadro 2.34 y la Figura 2.28 presentan las demás probabilidades:

Distribución Condicional			
	Muy grave	Grave	Sin importancia
Muchas veces	<b>0.7674</b>	0.2047	<b>0.0279</b>
Pocas veces	<b>0.6227</b>	0.2957	0.0816
Una vez	<b>0.6735</b>	0.2629	0.0636

Cuadro 2.34: Tabla de distribución condicional, sección Violencia dentro de la relación de Pareja, Pregunta 3.

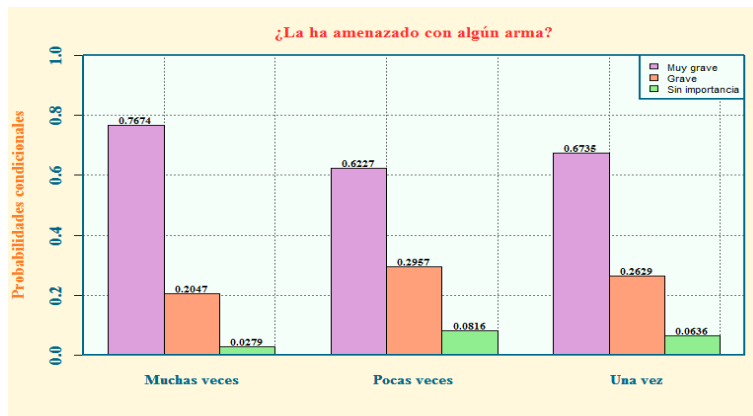


Figura 2.28: Distribuciones condicionando a la variable Frecuencia con que ocurre la agresión, Pregunta 3.

#### Pregunta 4

¿Le ha prohibido trabajar o estudiar?

Podemos observar que de las mujeres que han intentado trabajar y han sido detenidas por sus parejas, con una probabilidad de 0.5478 este suceso ocurre muy frecuentemente, no siendo solo un hecho aislado. Además, la probabilidad de que se considere muy grave es muy pequeña, con un valor de 0.2761. El Cuadro 2.35 y la Figura 2.29 muestran la tabla de contingencia y las probabilidades condicionales de ambas variables.

Distribución Marginal X		
Muchas veces	Pocas veces	Una vez
<b>0.5478</b>	0.3150	0.1373
Distribución Marginal Y		
Muy grave	Grave	Sin importancia
<b>0.2761</b>	0.3976	0.3264

Cuadro 2.35: Distribución marginal de las variables Frecuencia y Gravedad de la agresión, sección Violencia dentro de la relación de Pareja, Pregunta 4.

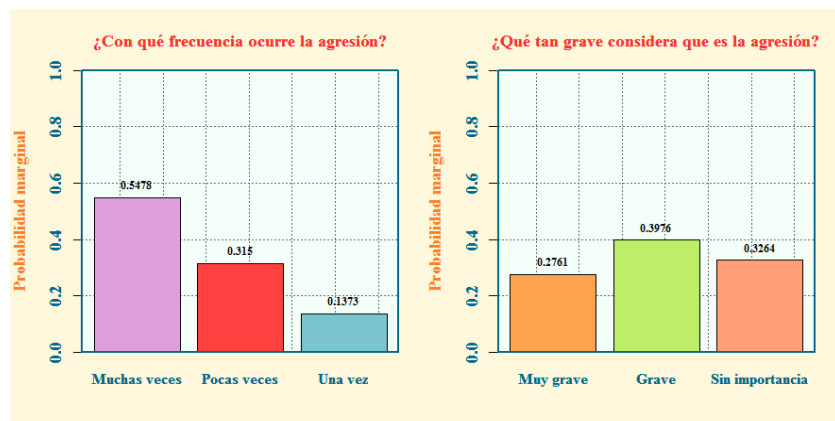


Figura 2.29: Distribución marginal de las variables Frecuencia con que ocurre la agresión, Pregunta 4.

Al condicionar la opinión de las mujeres discriminadas a la frecuencia con que ocurre esta agresión, vemos que el considerar un hecho sin importancia el que tu pareja te prohíba trabajar presenta probabilidades muy altas, además de que sólo con una probabilidad de 0.3731, de las mujeres que han sufrido esta restricción consideran que la falta es muy grave, dado que ha ocurrido de manera muy frecuente dentro de su relación en pareja.

Distribución Condicional			
	Muy grave	Grave	Sin importancia
Muchas veces	<b>0.3731</b>	0.3851	<b>0.2418</b>
Pocas veces	0.1565	0.4493	<b>0.3943</b>
Una vez	0.1633	0.3287	<b>0.5080</b>

Cuadro 2.36: Tabla de distribución condicional, sección Violencia dentro de la relación de Pareja, Pregunta 4.

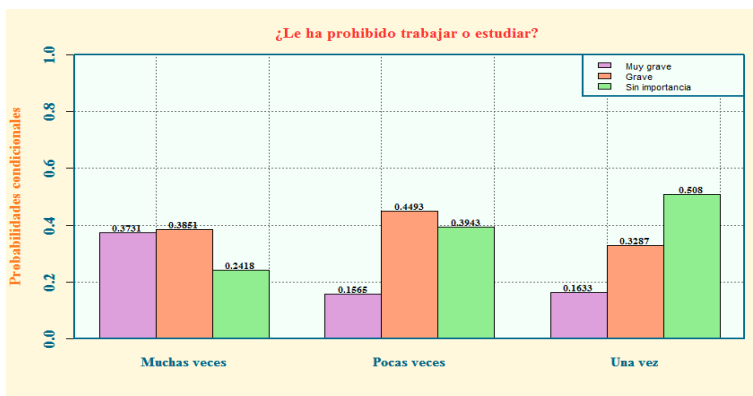


Figura 2.30: Distribuciones condicionando a la variable Frecuencia con que ocurre la agresión, Pregunta 4.

# Capítulo 3

## Feminicidio

### 3.1. Introducción

Existe una categoría de asesinatos violentos de mujeres que por sus características particulares, tanto defensores como organizaciones de los derechos humanos han pedido evidenciar y combatir, pues sus alcances y los orígenes que en ellos se localizan, como normas sociales y tradiciones, los hacen evitables si se decide combatir estos patrones socioculturales que desvalorizan a la mujer (Lagarde, 2008; ONU, 2011;).

El término feminicidio surge como una extensión al término femicidio propuesto por Diana Russell y Jane Caputi, el cual hace referencia a todas aquellas muertes que se desarrollan de forma violenta, y cuyas razones están asociadas a su género. Así, el término feminicidio añade dos elementos que el término femicidio no señala con claridad, los cuales son la misoginia que acompañan estos crímenes, y de la responsabilidad que los estados comparten al no combatir y en menor o mayor grado tolerar e invisibilizar el problema, señalándose por lo tanto en ocasiones como un crimen de estado (Toledo, 2009).

Por lo anterior, es importante mencionar que no todos los asesinatos de mujeres que se desarrollan de forma violenta pueden ser considerados feminicidios, pues para ser considerados como tales éstos deben estar asociados al desarrollo histórico que de alguna manera minimiza, permite y genera cierta tolerancia social a comportamientos violentos y de discriminación, que como etapa final terminó atentando contra el derecho a la vida de la mujer (Ciani, 2015).

Por lo tanto, el trabajo que se ha llevado a cabo por todos aquellos interesados en visibilizar y sensibilizar sobre estas muertes ha sido extenso, pues muchas de las

ideas que siguen apoyando la subordinación de las mujeres aún se encuentran altamente arraigadas dentro del pensamiento general, apoyadas incluso por aquellas personas encargadas de juzgar e impartir justicia a las víctimas, caracterizándose por una falta de profesionalismo que incluso puede llegar a la descalificación y humillación de las personas afectadas (Lagarde, 2008; Olamendi, 2016).

Así es como lo han denunciado en general las víctimas de estos delitos, que reportan irregularidades en todo su proceso tanto de denuncia como de investigación, declarando incluso haber sido víctimas de amenazas e intimidación (Olamendi, 2016). Además de reportarse un manejo poco minucioso en cuanto a los reportes estadísticos, que va desde registros incompletos de la defunción, hasta la poca comunicación y concordancia entre las diferentes instituciones encargadas de proporcionar información referente a esta problemática, haciendo difícil estudiar y medir el problema (Lagarde, 2008; ONU, 2011).

## 3.2. Femicidio en México

El estudio del femicidio en México comenzó a partir de la difusión de los asesinatos de mujeres en Ciudad Juárez, exponiendo a partir del Informe Final de la Fiscalía Especial para la Atención de los Delitos Relacionados con los Homicidios de Mujeres en el Municipio de Juárez, Chihuahua, la brutalidad y la impunidad como que estos homicidios se caracterizaban (ONU, 2011).

Posteriormente, con la creación de la Comisión Especial para el Seguimiento de los Femicidios de la LIX legislatura, liderada por Marcela Lagarde, quien es una de las principales voceras del movimiento feminista en América Latina y que además acuñó el término femicidio, realiza una Investigación diagnóstica sobre la violencia Femicida en la República Mexicana, el primero en su tipo en analizar la incidencia de los asesinatos de mujeres no sólo de Ciudad Juárez, sino a nivel nacional (Lagarde 2008, ONU, 2011).

Tales resultados fueron los primeros en evidenciar que tales agresiones tenían una incidencia muy fuerte no sólo en Ciudad Juárez, sino que era un problema que varios estados debían enfrentar y comenzar a reconocer, pues se comprobó incluso que Chihuahua no era el estado con el problema mayor en cuanto a tasa de asesinato de mujeres. En orden decreciente, los estados que reportaron los valores más altos fueron (Lagarde, 2008):

- i) Nayarit
- ii) Oaxaca



- III) Estado de México
- IV) Guerrero
- v) Quintana Roo
- VI) Baja California
- VII) Chihuahua

Además del informe anterior, dicha comisión redactó uno de los principales instrumentos para prevenir, sancionar y erradicar la violencia contra las mujeres, conocida como la Ley General de Acceso de las Mujeres a una vida Libre de violencia (Lagarde, 2008). Dicha ley dio pie en su título II, Modalidades de la violencia, capítulo V, De la violencia feminicida y de la alerta de violencia de género contra las mujeres, a uno de los principales mecanismos de protección contra la violencia feminicida, y cuyo artículo 22 define a la alerta de género como el conjunto de acciones gubernamentales de emergencia para enfrentar y erradicar la violencia feminicida en un territorio determinado, ya sea ejercida por individuos o por la propia comunidad (LGAMVLV, 2007).

El principal objetivo de dicha alerta según el artículo 23 es el de garantizar la seguridad de mujeres y niñas, el cese de la violencia en su contra y/o eliminar las desigualdades producidas por una legislación o política pública que agravia sus derechos humanos (LGAMVLV, 2007). Los estados que han declarado la alerta son los siguientes (INMUJERES, 2017):

Estados que emitieron alerta de Género			
Estado	Fecha	No. Mun.	Municipios
Estado de México	31 de julio de 2015	11	Ecatepec de Morelos, Nezahualcóyotl, Tlalnepantla de Baz, Toluca de Lerdo, Chalco, Chimalhuacán, Naucalpan de Juárez, Tultitlán, Ixtapaluca, Valle de Chalco y Cuautitlán Izcallí
Morelos	10 de agosto de 2015	8	Cuautla, Cuernavaca, Emiliano Zapata, Jiutepec, Puente de Ixtla, Temixco, Xochitepec, y Yautepec
Michoacán	27 de junio de 2016	14	Morelia, Uruapan, Lázaro Cárdenas, Zamora, Apatzingán, Zitácuaro, Los Reyes, Pátzcuaro, Tacámbaro, Hidalgo, Huétamo, La Piedad, Sahuayo y Maravatío.
Chiapas	18 de noviembre de 2016	7	Comitán de Domínguez, Chiapa de Corzo, San Cristóbal de las Casas, Tapachula, Tonalá, Tuxtla Gutiérrez y Villaflores Requiriendo acciones específicas para la región de los Altos de Chiapas

Estados que emitieron alerta de Género			
Estado	Fecha	No. Mun.	Municipios
Nuevo León	18 de noviembre de 2016	5	Apodaca, Cadereyta Jiménez, Guadalupe, Córdoba, Juárez y Monterrey
Veracruz	23 de noviembre de 2016	11	Boca del Río, Coatzacoalcos, Las Choapas, Martínez de la Torre, Minatitlán, Orizaba, Poza Rica de Hidalgo, Tuxpan, Veracruz y Xalapa.
Sinaloa	31 de marzo de 2017	5	Ahome, Culiacán, Guasave, Mazatlán y Navolato.
Colima	20 de junio de 2017	5	Colima, Coquimatlán, Cuauhtémoc, Tecomán y Villa de Álvarez.
San Luis Potosí	21 de junio de 2017	6	Ciudad Valles, Matehuala, San Luis Potosí, Soledad de Graciano Sánchez, Tamazunchale y Tamián.
Guerrero	22 de junio de 2017	8	Acapulco de Juárez, Ayutla de los Libres, Chilpancingo de los Bravo, Coyuca de Catalán, Iguala de la Independencia, José Azueta, José Azueta, y Tlapa de Comonfort.
Quintana Roo	7 de julio de 2017	3	Benito Juárez, Cozumel y Solidaridad. Requiriendo acciones específicas para Lázaro Cárdenas.

Cuadro 3.1: Tabla con los estados que han declarado Alerta de género, según datos reportados por el Instituto Nacional de las Mujeres

Sin embargo, uno de los principales problemas que se denuncia aún después de haber declarado la ley anteriormente citada, tiene que ver con la falta de acceso y la aplicación de los mecanismos de justicia, pues aún continúa la desacreditación y la idea de que existen problemas mucho más urgentes que deben ser atendidos, de tal manera que se sigue minimizando la situación hasta incluso llegar a la desvalorización de quienes defienden al feminicidio como un delito grave (Olamendi, 2016).

Tal invisibilidad alcanza a los impartidores de justicia de todos los niveles, los cuales frecuentemente no han protegido ni han prevenido de la manera adecuada a las mujeres que potencialmente pueden llegar a ser víctimas. Además, los procesos de investigación en numerosas ocasiones se encuentran con irregularidades, debido principalmente a la poca diligencia y corrupción con que se llevan a cabo cada uno de ellos (olamendi, 2016).

Uno de los principales casos que evidenciaron esta situación y que llevaron a una sentencia por parte de la Corte Interamericana de los Derechos Humanos en contra de México es el conocido caso González y Otras vs El estado Mexicano, también conocido como el caso del Campo Algodonero, en el cual México fue juzgado por la muerte de las jóvenes Claudia Ivette González, Esmeralda Herrera Monreal y Laura Berenice Ramos Monárrez, cuyos cuerpos fueron encontrados en un campo algodón en Ciudad Juárez en 2001.

Según el análisis hecho por Patricia Olamendi (Olamendi, 2016), se condenó a México por el incumplimiento de la Convención Americana de los Derechos Humanos, violando los artículos:

- I) Cuarto, referente al derecho a la vida.
- II) Quinto, referente al derecho a la integridad personal.
- III) Octavo, referente a las garantías judiciales.
- IV) Diecinueve, referente a los derechos de la niña.
- V) Veinticinco, referente a la protección judicial.

Además de la violación del artículo 7 de la convención Belém do Pará, por medio del cual se comprometen en adoptar políticas orientadas a prevenir, sancionar y erradicar todas las formas de violencia contra la mujer, y de no haber actuado con la debida prontitud en la etapa inicial de la investigación, presentando irregularidades en todo el proceso de investigación a través de un mal manejo de evidencias y fabricación de culpables. Asimismo, se señala el trato degradante hacia los familiares de las víctimas, exigiendo una reparación integral, y además de algunas medidas como:

- I) Reglas para conducir adecuadamente el proceso penal.
- II) Investigar y sancionar a los funcionarios responsables de no responder pronto y adecuadamente
- III) Estandarizar protocolos, criterios de investigación y manuales.
- IV) Creación de bases de datos de mujeres desaparecidas.
- V) Sensibilizar y proporcionar información tanto a funcionarios públicos, como a la sociedad en general sobre la discriminación en contra de la mujer.
- VI) Proporcionar una adecuada atención de los familiares de las víctimas.

Por otra parte, otra de las problemáticas que causó gran controversia tuvo que ver con la tipificación del feminicidio dentro de los códigos penales, tanto a nivel federal como local, ya que jurídicamente este podía clasificarse como un homicidio doloso, y no era justificado crear un tipo penal específico para este delito. Sin embargo, la creación de estadísticas y la difusión de esta problemática dentro de la sociedad son algunos de los puntos centrales dentro de las medidas de prevención establecidas tanto en Convención Belém do Pará y la CEDAW, y tipificar el feminicidio contribuía a cumplir ambos puntos (Ciani, 2015).

La razón de esto es que al tipificar este homicidio, se separan los homicidios de mujeres cuyas razones son producto del proceso histórico que disminuyó, toleró y contribuyó a la idea de que la vida de una mujer puede ser quitada bajo ciertas circunstancias, de aquellos homicidios violentos cuyo móvil no es por razones de género, creando estadísticas que permitan medir la magnitud del problema y crear políticas públicas mucho más

efectivas. Además de que al tipificarlo se reconoce la existencia de este tipo de homicidio, lo cual por mucho tiempo no se quiso admitir (Ciani, 2015).

Actualmente, en México la tipificación del feminicidio se encuentra definida en todos los códigos penales a nivel local y federal, a excepción de los estados Nayarit y Chihuahua (Angel, 2017). Sin embargo, tanto las sentencias, como las razones de género que se estipulan en cada código difieren entre cada estado, además de que su eficacia en cuanto al número de averiguaciones y sentencias ha sido ampliamente criticada, pues sólo de 2011 a 2015 se dictaron 169 sentencias en 9 de los 32 estados del país (Soto, 2014; Angel, 2017).

Entre las razones de género que más se mencionan corresponden a:

- I) Cuando exista o haya existido entre el activo y la víctima una relación de parentesco por consanguinidad o afinidad.
- II) La víctima presente signos de violencia sexual de cualquier tipo.
- III) A la víctima se le hayan infligido lesiones infamantes, degradantes o mutilaciones, previa o posterior a la privación de la vida.
- IV) Existan antecedentes de amenazas, acoso o lesiones del sujeto activo en contra de la víctima.
- V) A la víctima haya sido incomunicada.
- VI) Se haya realizado por violencia familiar.
- VII) El cuerpo de la víctima sea expuesto o arrojado en un lugar público.

# Capítulo 4

## Modelo Logístico

### 4.1. Introducción

Es de interés común para un gran número de ciencias y áreas de estudio establecer modelos matemáticos que permitan hacer una simplificación aproximada de la realidad, con base en un conjunto de variables que por su relación a dicho fenómeno o situación en estudio, nos ayuden a describir sus características generales, y determinar con base en las diferentes combinaciones en estas variables la intensidad de dicho fenómeno.

Algunos de los modelos más populares que se utilizan para tales fines son los llamados modelos lineales, cuyo nombre se debe a que estos modelos relacionan al fenómeno en estudio, llamado de aquí en adelante variable dependiente, con las variables explicativas por medio de una función, una que es lineal en los parámetros, incluyendo un error que es considerado aleatorio.

Una forma general que describe este tipo de modelos es:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon$$

Donde:

- $Y$  es la variable dependiente o fenómeno en estudio.
- $X_i$  es la  $i$ -ésima variable explicativa.

- $\beta_i$  es el  $i$ -ésimo parámetro desconocido.
- $\epsilon$  es el error aleatorio.

Es importante destacar que estos modelos están incluidos en dos de las generalizaciones más importantes de modelos cuyas aplicaciones, características específicas y limitaciones han sido extensamente estudiadas en la literatura debido a enorme variedad y potencial para estudiar diferentes fenómenos. Estos son los modelos lineales Generalizados y los Modelos Generalizados mixtos, donde los primeros surgen de la necesidad de explicar una variable de respuesta que no sea necesariamente normal, continua y de varianza contante, y cuya relación con las variables explicativas no sea lineal directamente, y los segundos de la necesidad de explicar además una variable de respuesta que permita cierta estructura de correlación en la variable respuesta, pues en los primeros las respuestas  $y_i$ 's se suponen independientes.

## 4.2. Modelos Lineales Generalizados

Los modelos lineales generalizados surgen de la necesidad de abordar diferentes fenómenos cuyas características particulares impiden tratarlos como una variable de respuesta que sea continua normal y de varianza constante. Esto puede ocurrir porque la variable de respuesta toma únicamente un subconjunto de la recta real, por ejemplo el subconjunto de los reales positivos  $R^+$ , o porque la variable en estudio es de tipo categórica, la cual a su vez puede ser nominal, cuyas categorías especifican únicamente la pertenencia del individuo a cierto grupo, y ordinal, cuyas categorías describen un orden el cual no puede ser alterado. La única condición que debe cumplir la variable que se desea explicar es que se crea que ésta pertenece a una familia muy particular de variables, conocida como la familia exponencial. Así, la condición que se debe verificar para asegurar que se está dentro de ella, es que la función de densidad de dicha variable puede expresarse de la siguiente manera (Dobson, 2002):

$$f(y, \theta) = s(y)t(\theta)e^{a(y)b(\theta)}$$

Donde:

- $t(\theta)$  es función de los parámetros.

- $s(y)$  es función de la muestra.
- $b(\theta)$  es función de los parámetros, también llamado parámetro natural en el caso en que  $a(y)=y$ .

Es así como generalizamos la aplicación de modelos lineales a situaciones mucho más diversas. Algunas de las variables que han sido ampliamente estudiadas en la literatura y que pertenecen a esta gran familia son:

- Bernoulli
- Binomial
- Poisson
- Geométrica
- Binomial negativa
- Exponencial
- Gamma

En general, este tipo de modelos identifica 3 componentes principales: Un *componente aleatorio*, que hace referencia a la variable de respuesta independiente en cada observación con función de densidad dentro de la familia exponencial. Un *componente sistemático*, el cual está compuesto por el conjunto de las variables explicativas y los parámetros, donde las primeras son valores fijos para cada individuo y los segundos son constantes desconocidas las cuales serán estimadas a partir de la muestra (McCullagh & Nelder, 1989). Ambos componentes están relacionados por medio de una función lineal, y la *función liga*, la cual es una función de la media de la variable respuesta  $g(E(Y))$ , que se caracteriza por ser una función monótona diferenciable (Lindsey, 1997). Esta función relaciona al componente sistemático con el componente aleatorio de tal forma que la relación entre ambos componentes sea efectivamente lineal.

Es debido a esta función liga que permite que la media de la variable respuesta no necesariamente guarde una relación directamente lineal con el componente sistemático, pues permite que una transformación de la media y no necesariamente la media, sea la que tenga una relación lineal con el componente sistemático.

Vemos entonces cómo el modelo de regresión y los modelos ANOVA (modelos con variables explicativas categóricas) y ANCOVA (modelos con variables continuas y categóricas) forman parte de los modelos lineales generalizados, al tener a la función

identidad como su función liga y a su componente aleatorio a la distribución normal, es decir:

$$g(E(y))=g(\mu)=\mu=\beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_pX_p$$

Donde:

- $y$  es el componente aleatorio.
- $g(E(y))$  es la función liga.
- $\beta_i$  es el  $i$ ésimo parámetro desconocido.
- $\beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_pX_p$  es el componente sistemático.

### 4.3. Formulación del Modelo

Modelar una variable cuyo dominio esté contenido en un conjunto formado únicamente por dos valores, es de un interés vital para un gran número de áreas de estudio, que abarca desde el campo de las ciencias sociales, en el que por ejemplo, se podría estar interesado en determinar cuáles son las variables que determinan si una persona (Si/No) ha considerado alguna vez unirse al crimen organizado.

También en determinar qué variables, más específicamente, qué niveles o valores de éstas, para el caso de variables discretas, determinaron si una persona (Si/No) logró terminar satisfactoriamente el nivel de bachillerato o nivel media superior.

Otras aplicaciones son por ejemplo en la medicina, donde se podría estar interesado en los diferentes factores que involucran (Si/No) la presencia o muerte por alguna enfermedad, o si una persona presentó (Si/No) remisión de la enfermedad en cuestión.

El modelo logístico forma parte de los modelos lineales generalizados, y nos permite modelar las situaciones planteadas anteriormente. Formalmente, para cada observación, su componente aleatorio corresponde a una variable Bernoulli( $\pi_i$ ), en la que cada  $\pi_i$  depende del valor en las variables explicativas observadas en cada individuo, y cuya esperanza, o alguna función de ella, es la que se tratará de modelar.



Para este fin, es importante observar que el dominio de  $\pi_i$  se encuentra dentro del intervalo  $[0,1]$ , por ser esta última una probabilidad. Como consecuencia de lo anterior, un modelo que relacione a esta probabilidad con sus variables explicativas de forma directa a una función lineal, como en el caso de regresión lineal simple, es poco factible, pues diferentes valores de las variables  $x$ 's pueden hacer que el modelo tome valores que superen la unidad, o hasta tomar valores negativos.

Fue así, que al buscar una función que se encontrara acotada dentro del intervalo  $[0,1]$ , se propuso a la función logística, que cumple esta propiedad.

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

Por lo que el modelo de regresión asociado corresponde al siguiente:

$$\begin{aligned} P(Y=1|\bar{x}) &= \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}} \\ &= E(Y|\bar{x}) \end{aligned}$$

Donde:

- $Y = 1$  corresponde a la categoría o respuesta que se quiere analizar.
- $\bar{x}$  es el vector de variables explicativas.
- $x_1, x_2, \dots, x_p$  corresponde a los valores de las variables explicativas para cada sujeto.
- $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ , corresponde a los parámetros a estimar en el modelo.

Considerando el caso en que tenemos una variable explicativa, la función logística tiene la siguiente forma dependiendo si el parámetro  $\beta_1$  es positivo o negativo, respectivamente:

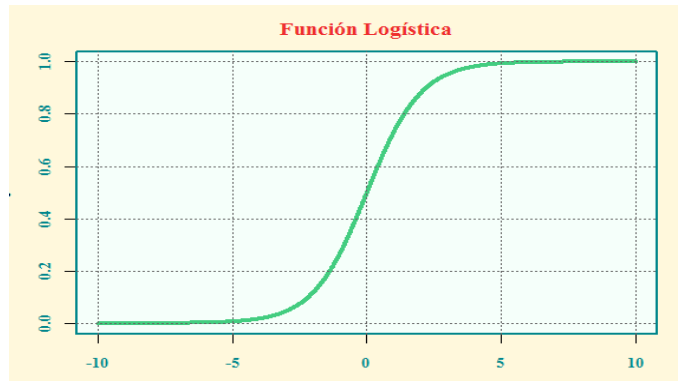


Figura 4.1: Ejemplo de gráfica de la función logística con parámetro positivo.

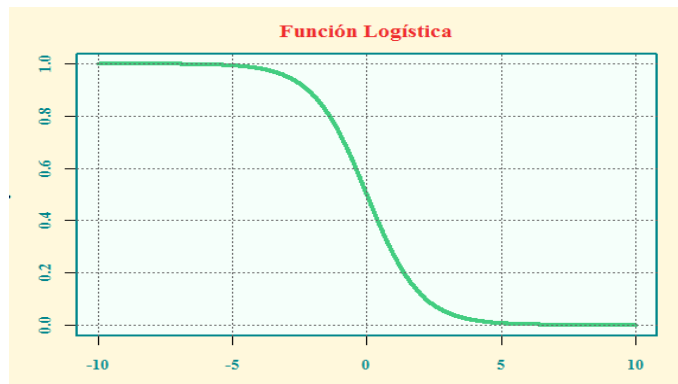


Figura 4.2: Ejemplo de gráfica de la función logística con parámetro negativo.

Vemos que para el primer caso, si  $\beta_1 > 0$ , y  $x$  es grande, entonces  $\pi_i$  tiende a 1, y si  $x$  es pequeño tiende a 0, pero si  $\beta_1 < 0$  y  $x$  es grande, entonces  $\pi_i$  tiende a 0, y si  $x$  es pequeña  $\pi_i$  tiende a 1. Además, se observan cambios mucho más pequeños en las probabilidades a cambios en la variable  $x$  en los extremos próximos al 0 y 1 que en medio de la función (Agresti, 2002).

Por otro lado, también podemos formular el modelo encontrando la función ligo que corresponde a esta distribución, y que hará que la relación entre nuestro componente aleatorio y su componente sistemático sea efectivamente una función lineal. Para ello expresamos a la función de densidad de manera que podamos encontrar su parámetro natural:

$$\begin{aligned}
 f(x : p) &= p^y(1 - p)^{1-y}1(y)_{0,1} \\
 &= (1 - p)\left(\frac{p}{1 - p}\right)^y1(y)_{0,1} \\
 &= (1 - p)e^{yLn\left(\frac{p}{1-p}\right)}1(y)_{0,1}
 \end{aligned}$$

Por lo que la transformación que habría que aplicar a la esperanza de la respuesta, conocida también como función logit, corresponde al logaritmo del cociente entre la probabilidad de presentar la respuesta  $Y = 1$ , contra la probabilidad de no presentar la respuesta  $Y = 1$ , es decir, contra la probabilidad de observar la respuesta  $Y = 0$ .

Así, el modelo que trataremos de ajustar corresponde al siguiente:

$$\text{Ln} \left( \frac{P(Y = 1|\bar{x})}{1 - P(Y = 1|\bar{x})} \right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

Donde:

- $\bar{x}$  es el vector de variables explicativas.

Este cociente tiene aplicaciones significativas en el área del análisis de datos categóricos, específicamente en la determinación del grado de asociación entre dos variables categóricas, y es conocido con el nombre de momio.

## 4.4. Estimación de los parámetros

Ya que el conjunto de las variables explicativas se consideran valores fijos para cada individuo, los parámetros que son desconocidos y que se requieren estimar mediante la muestra son el conjunto de las  $\beta_j$ , donde éstas describirán la relación entre las variables explicativas y la respuesta, dependiendo de los diferentes patrones presentes en los individuos en sus covariables. El método utilizado para estimarlos es el de máxima verosimilitud, el cual es uno de los más populares debido, entre otras cosas, a propiedades asintóticas que éstos poseen. A grandes rasgos, el método consiste en maximizar a la función de verosimilitud, o lo que es equivalente, a la log-verosimilitud, que no es más que la función de densidad conjunta del componente aleatorio visto como función de los parámetros a estimar, para así encontrar aquéllos que hacen más verosímil observar los datos muestreados de la variable respuesta.

### 4.4.1. Modelo Desagregado

Podemos considerar el modelo logístico tomando datos desagregados o agrupados. En el primer caso consideramos a  $y_i$ , con  $i = 1 \dots n$  como una variable Bernoulli( $\pi_i$ ) sobre cada individuo, de tal manera que la función de verosimilitud es como sigue:

$$L(\beta|y, x) = \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i}$$

Cuya log-verosimilitud es:

$$\begin{aligned} l(\beta|y, x) &= \sum_{i=1}^n y_i \text{Ln}(\pi_i) + (1 - y_i) \text{Ln}(1 - \pi_i) \\ &= \sum_{i=1}^n y_i \text{Ln}\left(\frac{\pi_i}{1 - \pi_i}\right) + \text{Ln}(1 - \pi_i) \\ &= \sum_{i=1}^n y_i \left(\sum_{j=0}^k \beta_j x_{ij}\right) + \text{Ln}\left(1 - \frac{e^{\sum_{j=0}^k \beta_j x_{ij}}}{1 + e^{\sum_{j=0}^k \beta_j x_{ij}}}\right) \\ &= \sum_{i=1}^n y_i \left(\sum_{j=0}^k \beta_j x_{ij}\right) - \text{Ln}\left(1 + e^{\sum_{j=0}^k \beta_j x_{ij}}\right) \\ &= \sum_{i=1}^n y_i \left(\sum_{j=0}^k \beta_j x_{ij}\right) - \sum_{i=1}^n \text{Ln}\left(1 + e^{\sum_{j=0}^k \beta_j x_{ij}}\right) \end{aligned}$$

Así, derivando la ecuación anterior, obtenemos las ecuaciones de verosimilitud:

$$\begin{aligned} \frac{\partial l}{\partial \beta_j} &= \sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n \left(\frac{x_{ij} e^{\sum_{j=0}^k \beta_j x_{ij}}}{1 + e^{\sum_{j=0}^k \beta_j x_{ij}}}\right) \\ &= \sum_{i=1}^n \left(y_i x_{ij} - \frac{x_{ij} e^{\sum_{j=0}^k \beta_j x_{ij}}}{1 + e^{\sum_{j=0}^k \beta_j x_{ij}}}\right) \\ &= \sum_{i=1}^n x_{ij} \left(y_i - \frac{e^{\sum_{j=0}^k \beta_j x_{ij}}}{1 + e^{\sum_{j=0}^k \beta_j x_{ij}}}\right) \\ &= \sum_{i=1}^n x_{ij} (y_i - \pi_i) = 0 \end{aligned}$$

Visto de forma matricial:

$$\frac{\partial l}{\partial \beta_j} = X'(Y - \Pi) = 0$$

Donde:

$$X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{pmatrix}$$

$$\Pi = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_{n-1} \\ \pi_n \end{pmatrix}$$

Cuyas soluciones son calculadas por métodos numéricos iterativos como el de Newton-Raphson (N-R), que requiere el cálculo de las segundas derivadas:

$$\begin{aligned} \frac{\partial^2 l}{\partial \beta_k \partial \beta_j} &= \frac{\partial}{\partial \beta_k} \sum_{i=1}^n x_{ij} (y_i - \pi_i) \\ &= -\frac{\partial}{\partial \beta_k} \sum_{i=1}^n (x_{ij} \pi_i) \\ &= -\frac{\partial}{\partial \beta_k} \sum_{i=1}^n x_{ij} \left( \frac{e^{\sum_{j=0}^k \beta_j x_{ij}}}{1 + e^{\sum_{j=0}^k \beta_j x_{ij}}} \right) \\ &= -\sum_{i=1}^n x_{ij} \left( \frac{x_{ik} e^{\sum_{j=0}^k \beta_j x_{ij}} (1 + e^{\sum_{j=0}^k \beta_j x_{ij}}) - x_{ik} e^{2 \sum_{j=0}^k \beta_j x_{ij}}}{(1 + e^{\sum_{j=0}^k \beta_j x_{ij}})^2} \right) \end{aligned}$$

$$\begin{aligned}
&= - \sum_{i=1}^n x_{ij} \left( \frac{x_{ik} e^{\sum_{j=0}^k \beta_j x_{ij}}}{(1 + e^{\sum_{j=0}^k \beta_j x_{ij}})^2} \right) \\
&= - \sum_{i=1}^n x_{ij} x_{ik} \left( \frac{e^{\sum_{j=0}^k \beta_j x_{ij}}}{(1 + e^{\sum_{j=0}^k \beta_j x_{ij}})} \right) \left( \frac{1}{1 + e^{\sum_{j=0}^k \beta_j x_{ij}}} \right) \\
&= - \sum_{i=1}^n x_{ij} x_{ik} (\pi_i) (1 - \pi_i)
\end{aligned}$$

Las cuales, al igual que en caso de las primeras derivadas, pueden ser vistas de forma matricial como sigue:

$$\frac{\partial^2 l}{\partial \beta_k \partial \beta_j} = X' V X$$

Donde:

$$V = \begin{pmatrix} \pi_1(1 - \pi_1) & 0 & \cdots & 0 \\ 0 & \pi_2(1 - \pi_2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \pi_n(1 - \pi_n) \end{pmatrix}$$

Y es a través de estas expresiones que hacemos uso del método de N-R, hayando así el valor al cual converge la sucesión de valores generados por el proceso iterativo que se describe a continuación, cuyo vector de parámetros final dará solución a las ecuaciones de verosimilitud.

$$\beta^{(t)} = \beta^{(t-1)} + (X' V X)^{-1} X' (Y - \Pi)$$

## Resultados

Volviendo al ajuste del modelo logístico, como se mencionó anteriormente en la descripción de las variables explicativas, fueron cinco las variables que se consideraron importantes por ser de interés en el fenómeno del feminicidio. Así, usando el lenguaje de programación R, ajustamos el modelo definitivo haciendo uso de la función *glm*, pues

permite ajustar modelos pertenecientes a la familia de los modelos lineales generalizados explicados anteriormente.

Entre sus componentes principales se encuentra en primer lugar el parámetro *formula*, por medio del cual se proporciona una descripción simbólica del modelo, detallando al conjunto de las variables explicativas y a la variable respuesta. Posteriormente *family*, cuya función es determinar la distribución del error y la función liga que se desea utilizar, y *data*, que establece la ubicación de las variables del modelo.

```
Mod5_Desag<-glm(Muerte~Nivel_Esc+Ingresos
+Estado_Civil
+ZonaR_U
+Lugar_Def, family=binomial(logit))
```

Por otro lado, los elementos que devuelve son diversos, pero por medio de la función *summary* se puede obtener un vistazo general del modelo, regresando el valor de los parámetros ajustados y su error estándar, junto con el valor del estadístico de Wald y su significancia. También proporciona el valor de la devianza, medida ampliamente utilizada en bondad y ajuste y cuyas propiedades se analizarán posteriormente, tanto del modelo nulo, el cual únicamente incluye al intercepto, como del modelo ajustado, junto con los grados de libertad de cada uno de ellos. El modelo es el siguiente, y la interpretación de cada uno de estos elementos se realizará en las secciones subsecuentes.

```
summary(Mod5_Desag)

##
## Call:
## glm(formula = Muerte ~ Nivel_Esc + Ingresos + Estado_Civil +
##      ZonaR_U + Lugar_Def, family = binomial(logit))
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.7062  -0.3732  -0.2353  -0.1318   3.1591
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.52130    0.16337 -21.554 < 2e-16 ***
## Nivel_Esc2    0.24041    0.10892   2.207 0.027301 *
## Nivel_Esc3   -1.40435    0.11304 -12.424 < 2e-16 ***
## Ingresos2     1.30672    0.15711   8.317 < 2e-16 ***
## Ingresos3    -0.05736    0.13961  -0.411 0.681199
```

```

## Estado_Civil2  1.16819    0.08052   14.508 < 2e-16 ***
## Estado_Civil3  1.49093    0.08605   17.326 < 2e-16 ***
## ZonaR_U1      0.24179    0.07149    3.382 0.000719 ***
## Lugar_Def2     3.87705    0.07875   49.232 < 2e-16 ***
## Lugar_Def3     0.86272    0.06247   13.811 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 15713.6  on 21265  degrees of freedom
## Residual deviance:  9686.7  on 21256  degrees of freedom
## AIC: 9706.7
##
## Number of Fisher Scoring iterations: 6

```

#### 4.4.2. Modelo agregado

Por otro lado, en el caso en que tenemos datos agregados, en vez de tomar  $n$  variables aleatorias Bernoulli ( $\pi_i$ ) para cada  $i = 1 \dots n$ , consideramos  $J$  variables aleatorias con distribución Binomial ( $m_j, \pi_j$ ) independientes, formadas a partir de las diferentes agrupaciones definidas por los patrones o arreglos diferentes que se hicieron presentes en los individuos, a partir de las  $k$  variables explicativas.

Así, la función de verosimilitud es:

$$L(\beta|y, x) = \prod_{j=1}^J \binom{m_j}{\pi_j} \pi_j^{y_j} (1 - \pi_j)^{m_j - y_j}$$

Cuya log-verosimilitud es:

$$\begin{aligned} l(\beta|y, x) &= \sum_{j=1}^J y_j \text{Ln}(\pi_j) + (m_j - y_j) \text{Ln}(1 - \pi_j) + \text{Ln} \left( \binom{m_j}{\pi_j} \right) \\ &= \sum_{j=1}^J y_j \text{Ln} \left( \frac{\pi_j}{1 - \pi_j} \right) + m_j \text{Ln}(1 - \pi_j) + \text{Ln} \left( \binom{m_j}{p_j} \right) \end{aligned}$$



$$\begin{aligned}
& \propto \sum_{j=1}^J y_j \left( \sum_{k=0}^K \beta_k x_{jk} \right) + m_j \text{Ln} \left( 1 - \frac{e^{\sum_{k=0}^K \beta_k x_{jk}}}{1 + e^{\sum_{k=0}^K \beta_k x_{jk}}} \right) \\
& = \sum_{j=1}^J y_j \left( \sum_{k=0}^K \beta_k x_{jk} \right) - m_j \text{Ln} \left( 1 + e^{\sum_{k=0}^K \beta_k x_{jk}} \right) \\
& = \sum_{j=1}^J y_j \left( \sum_{k=0}^K \beta_k x_{jk} \right) - \sum_{j=1}^J m_j \text{Ln} \left( 1 + e^{\sum_{k=0}^K \beta_k x_{jk}} \right)
\end{aligned}$$

Así, igual que en el caso de datos desagregados, obtenemos las ecuaciones de verosimilitud:

$$\begin{aligned}
\frac{\partial l}{\partial \beta_k} &= \sum_{j=1}^J x_{jk} y_j - \sum_{j=1}^J m_j \frac{e^{\sum_{k=0}^K \beta_k x_{jk}}}{1 + e^{\sum_{k=0}^K \beta_k x_{jk}}} \\
&= \sum_{j=1}^J x_{jk} (y_j - m_j \pi_j) = 0
\end{aligned}$$

Además de calcular las segundas derivadas requeridas para el método de N-R.

$$\frac{\partial^2 l}{\partial \beta_l \partial \beta_k} = - \sum_{j=1}^J m_j x_{jk} x_{jl} (\pi_j) (1 - \pi_j)$$

## Resultados

Al igual que en el caso de datos desagregados, obtenemos un vistazo general de los elementos del modelo por medio de la función *summary*, coincidiendo únicamente con el modelo desagregado en el valor de los parámetros ajustados y en todos los elementos que los acompañan en la sección de coeficientes, ya que tanto en ambas devianzas, como en el rango de los residuos y los grados de libertad presentaron una disminución considerable. La razón de ésto es que el primero ajustó un modelo tomando un total de 21266 elementos y el segundo sólo 131, que es el número total de patrones de covariables formados por las categorías de las variables que se hicieron presentes en los datos. Los resultados se presentan a continuación:

```

Mod5_Agreg <- glm(cbind(Si, No) ~ factor(Nivel_E)+factor(Ingresos)
                 +factor(Edo_C)
                 +factor(Zona)
                 +factor(Sitio), data = Datos_Agrup, family = binomial)
summary(Mod5_Agreg)

##
## Call:
## glm(formula = cbind(Si, No) ~ factor(Nivel_E) + factor(Ingresos) +
##      factor(Edo_C) + factor(Zona) + factor(Sitio), family = binomial,
##      data = Datos_Agrup)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -4.3302  -0.7254   0.0000   1.0454   3.1458
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -3.52130    0.16337  -21.554 < 2e-16 ***
## factor(Nivel_E)2  0.24041    0.10892   2.207 0.027301 *
## factor(Nivel_E)3 -1.40435    0.11304  -12.424 < 2e-16 ***
## factor(Ingresos)2  1.30672    0.15711   8.317 < 2e-16 ***
## factor(Ingresos)3 -0.05736    0.13961  -0.411 0.681199
## factor(Edo_C)2    1.16819    0.08052  14.507 < 2e-16 ***
## factor(Edo_C)3    1.49093    0.08605  17.326 < 2e-16 ***
## factor(Zona)2     0.24179    0.07149   3.382 0.000719 ***
## factor(Sitio)2    3.87705    0.07875  49.232 < 2e-16 ***
## factor(Sitio)3    0.86272    0.06247  13.811 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 6312.72  on 140  degrees of freedom
## Residual deviance:  285.77  on 131  degrees of freedom
## AIC: 671.95
##
## Number of Fisher Scoring iterations: 4

```

## 4.5. Parámetros significativos

Una vez que hemos estimado el valor de los parámetros del componente sistemático, es natural preguntarse qué parámetros aportan mayor información al modelo ajustado, y verificar que todas las variables explicativas que fueron elegidas para formar parte del modelo, efectivamente aportan nueva información y además suficiente como para ser incluidas, pues el mejor modelo para los propósitos individuales de cada investigador, no necesariamente es el que tenga el mayor número de variables. Dicho de otro modo, en general buscamos un modelo que sirva como generalización para describir un fenómeno en particular, y no uno que sólo describa la muestra que fue tomada del fenómeno a investigar.

Existen dos maneras de evaluar lo anterior. La primera de ellas es por medio de una prueba de hipótesis que hace provecho de las propiedades asintóticas de los estimadores máximos verosímiles, conocida como *Prueba de Wald*. La segunda, por otro lado, hace uso de modelos anidados, uno que contenga el parámetro a evaluar, llamémoslo  $M_1$ , y otro que no lo tenga, llamémoslo  $M_2$ , y cuyo desarrollo se revisará en la sección de bondad de ajuste en sentido global.

### 4.5.1. Prueba de Wald

En el caso de la prueba de Wald, la hipótesis a contrastar es:

$$H_0: \beta_j = 0 \qquad \text{vs} \qquad H_a: \beta_j \neq 0$$

Como  $\hat{\beta}_j \sim N(\beta_j, V(\hat{\beta}_j))$ , el estadístico de prueba utilizado para probar la significancia del parámetro a evaluar, considerando muestras grandes y bajo el supuesto de que la hipótesis nula es cierta, es:

$$Z = \frac{\hat{\beta}_j}{\widehat{SE}(\hat{\beta}_j)} \sim N(0, 1)$$

O alternativamente:

$$\left( \frac{\hat{\beta}_j}{\widehat{SE}(\hat{\beta}_j)} \right)^2 \sim \chi_{(1)}$$

Se rechaza la hipótesis de que el parámetro no es significativo si:

$$P(|Z| > w) < (1 - \alpha)$$

Donde:

- $Z$  es una variable aleatoria normal estándar.
- $w$  es el valor del estadístico.
- $\alpha$  es en nivel de significancia.

Cuyas varianzas de los estimadores en el caso de un ajuste multivariado son calculadas por medio de la inversa de la matriz de información observada (Hosmer & Lemeshow, 2000), denotémosla por  $\widehat{I}(\widehat{\beta}_j)$ , y cuyos elementos corresponden a las segundas derivadas parciales de la función de log-verosimilitud.

$$- \sum_{i=1}^n x_{ij} x_{ik} (\pi_i) (1 - \pi_i)$$

Visto de forma matricial, la información observada puede expresarse como:

$$\widehat{I}(\widehat{\beta}_j) = X'VX$$

Donde:

$$X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

$$V = \begin{pmatrix} \widehat{\pi}_1(1 - \widehat{\pi}_1) & 0 & \cdots & 0 \\ 0 & \widehat{\pi}_2(1 - \widehat{\pi}_2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \widehat{\pi}_n(1 - \widehat{\pi}_n) \end{pmatrix}$$

Y cuya varianza de cada estimador corresponde a la diagonal de la matriz:

$$\widehat{I}^{-1}(\widehat{\beta}_j)$$

## Resultados

Para cada parámetro del modelo final, y considerando un valor de significancia  $\alpha = 0.05$ , observamos que a excepción de la categoría 3 de la variable Ingresos, en todos los casos el valor de su *p-value* es mucho menor al valor de  $\alpha$ , por lo que rechazamos la hipótesis nula de que el coeficiente sea cero. Es decir, podemos concluir que a excepción del caso anteriormente mencionado, cada parámetro incluido aporta nueva información al modelo. Los resultados se muestran a continuación:

Prueba de Wald		
Variable	Valor z	Pr(> z )
Intercept	-21.554	< 2e-16 ***
Nivel_Esc2	2.207	0.027301 *
Nivel_Esc3	-12.424	< 2e-16 ***
Ingresos2	8.317	< 2e-16 ***
Ingresos3	-0.411	0.681199
Estado_Civil2	14.508	< e-16 ***
Estado_Civil3	17.326	< 2e-16 ***
ZonaR_U2	3.382	0.000719 ***
Lugar_Def2	49.232	< 2e-16 ***
Lugar_Def3	13.811	< 2e-16 ***

Cuadro 4.1: Resumen de pruebas individuales de significancia de las variables explicativas bajo prueba de Wald

## 4.6. Medidas de bondad y ajuste

Una vez que se han elegido las covariables que, con base en conocimientos previos de investigación y análisis hemos visto que guardan una relación funcional importante y de interés con la variable dependiente que se desea explicar, procedemos a evaluar el ajuste que estas variables elegidas dieron como modelo final.

La forma de hacerlo es nuevamente por medio de la herramienta estadística que conocemos como prueba de hipótesis, y cuyo procedimiento en general consiste en comparar dos modelos, uno que es el modelo de interés, contra uno de referencia llamado saturado, y el cual tiene tantos parámetros como datos tiene la base, de manera que reproduce perfectamente los datos.

El modelo que se desea evaluar se especifica en la hipótesis nula  $H_0$ , mientras que el modelo saturado es especificado en la hipótesis alternativa  $H_a$ . En el caso de ajuste global del modelo, las hipótesis a contrastar son las siguientes:

$$H_0: \quad \text{El modelo de interés} \quad \text{vs} \quad H_a: \quad \text{El modelo saturado ajusta} \\ \text{ajusta bien a los datos}$$

En este caso, lo que buscamos es no rechazar la hipótesis nula, pues que ajuste el modelo saturado no nos diría nada del fenómeno que en general queremos describir, ya que éste únicamente reproduce los datos muestrados.

La clasificación de estas pruebas puede dividirse en dos tipos, las basadas en los diferentes patrones de covariables presentes por las variables explicativas incluídas en el modelo elegido, y las pruebas basadas en las probabilidades estimadas. Las primeras sólo podrán ser utilizadas cuando se haya ajustado el modelo agrupado, y las segundas en el caso de que se haya ajustado el modelo individual.

#### 4.6.1. Pruebas basadas en patrones de covariables

##### Devianza

La devianza es un estadístico de prueba que hace uso de la transformación de menos dos veces el logaritmo del cociente generalizado de verosimilitudes, y cuyo uso es útil debido a que la distribución de este estadístico es totalmente conocida, pues corresponde a una distribución asintótica Ji Cuadrada cuyos grados de libertad son la diferencia entre el tamaño del espacio paramétrico del modelo saturado menos el espacio paramétrico del modelo bajo la hipótesis nula. En nuestro caso, corresponde a una Ji cuadrada con grados de libertad igual a la diferencia entre el número de patrones de covariables diferentes  $J$  menos el número de parámetros estimados  $p + 1$ .

El desarrollo del estadístico es como sigue:

$$\begin{aligned} G^2 &= -2Ln \left( \frac{L_m}{L_s} \right) \\ &= 2[Ln(L_s) - Ln(L_m)] \\ &= 2 \left[ \sum_{j=1}^J y_j Ln \left( \frac{y_j}{m_j} \right) + (m_j - y_j) Ln \left( 1 - \frac{y_j}{m_j} \right) \right] - \end{aligned}$$

$$\begin{aligned}
& 2 \left[ \sum_{j=1}^J y_j \text{Ln} \left( \frac{\hat{y}_j}{m_j} \right) + (m_j - y_j) \text{Ln} \left( 1 - \frac{\hat{y}_j}{m_j} \right) \right] \\
& = 2 \left[ \sum_{j=1}^J y_j \text{Ln} \left( \frac{y_j}{\hat{y}_j} \right) + (m_j - y_j) \text{Ln} \left( \frac{m_j - y_j}{m_j - \hat{y}_j} \right) \right]
\end{aligned}$$

Donde:

- $L_m$  es la verosimilitud del modelo propuesto.
- $L_s$  es la verosimilitud del modelo saturado.

Rechazando la hipótesis nula para valores de  $G^2$  grandes, pues esperaríamos que si el modelo ajusta bien, la log-verosimilitud de éste contra la log-verosimilitud del modelo saturado, que es el modelo que ajusta perfectamente, no fueran tan diferentes. Dicho formalmente, se rechaza la hipótesis nula de que el modelo elegido efectivamente describe los datos, a un nivel de significancia  $\alpha$ , cuando el valor del estadístico  $G^2$  es mayor al cuantil  $1 - \alpha$  de una distribución Ji-Cuadrada con  $J - (p + 1)$  grados de libertad.

Este estadístico es de gran utilidad y ampliamente utilizado para evaluar bondad de ajuste dentro del contexto de los Modelos lineales generalizados, sin embargo es mucho más utilizado para hacer comparaciones entre dos modelos diferentes que sean anidados. Aún así, las dos condiciones para utilizarlo nunca deben perderse de vista, pues este estadístico funciona únicamente para modelos que estén anidados y que ambos utilicen la misma función liga (Dobson, 2002).

## Resultados

Para el caso del modelo ajustado, observamos un estadístico con valor de 285.7728 y un p-value de 1.546541e-13, por lo que considerando un nivel de significancia  $\alpha=0.05$  se rechaza la hipótesis nula de que el modelo ajusta a los datos. Este resultado puede parecer contrario a lo que se esperaría observar, sin embargo, dado que se está comparando contra el modelo saturado, el resultado anterior puede deberse a que no se están contemplando el suficiente número de variables explicativas, por lo que el modelo podría considerarse una buena aproximación dependiendo del nivel de predicción y los resultados en el análisis de diagnóstico, pero no puede ser considerado igual de bueno que el modelo que ajusta perfectamente a los datos.

```

Estadistico_D<-Mod5_Agreg$deviance
Estadistico_D

## [1] 285.7728

Grados_L<-Mod5_Agreg$df.residual
Grados_L

## [1] 131

1-pchisq(Estadistico_D,Grados_L)

## [1] 1.546541e-13

```

Como se mencionó anteriormente, podemos realizar la prueba entre dos modelos que simplemente estén anidados, no siendo el más complejo de éstos dos necesariamente el modelo saturado, con la única condición de que el más complejo contenga todos los parámetros del otro modelo. Es así como se desarrolla el segundo método para probar significancia, sin embargo en lugar de probar la significancia de los parámetros individuales, se compara la significancia de las variables individuales. El procedimiento consiste en elegir dos modelos, uno que contenga la variable cuya significancia se quiera evaluar, y otro que no la tenga.

El desarrollo del estadístico para comparar dos modelos anidados, donde  $m_1$  está anidado a  $m_2$  es como sigue:

$$\begin{aligned}
 G_{m_1}^2 - G_{m_2}^2 &= -2Ln\left(\frac{L_1}{L_s}\right) + 2Ln\left(\frac{L_2}{L_s}\right) \\
 &= -2Ln\left(\frac{L_2}{L_1}\right)
 \end{aligned}$$

Donde:

- $L_2$  es la verosimilitud del modelo  $m_2$ .
- $L_1$  es la verosimilitud del modelo  $m_1$ .



Por lo que, si  $m_1$  tiene  $q$  número de parámetros y  $m_2$  tiene  $r$  parámetros, donde  $q > r$ . La distribución del estadístico anterior es Ji cuadrada, con  $(J-r) - (J-q) = q-r$  grados de libertad. Así, rechazamos la hipótesis nula de que el modelo  $m_2$ , que es el modelo con menos parámetros, ajusta igual de bien que el modelo  $m_1$ , cuando el valor del estadístico tome valores grandes, más específicamente, cuando el valor de éste supere el cuantil de una distribución Ji-cuadrada con  $q-r$  grados de libertad.

Un caso menos utilizado se presenta cuando probamos la significancia del parámetro de una variable continua o dicotómica bajo esta prueba, cuyo estadístico siempre se comparará contra el cuantil de una Ji cuadrada con 1 grado de libertad, rechazando la hipótesis de que el parámetro es cero cuando el valor del estadístico supere el valor del cuantil.

## Resultados

La siguiente tabla muestra los resultados de las pruebas de significancia de las variables incluídas en el modelo, mostrando en todos los casos un p-value menor al nivel de significancia  $\alpha=0.05$ , por lo que podemos concluir que tanto el nivel de escolaridad, ingresos, estado conyugal, tipo de zona rural o urbana donde residía la mujer y sitio donde ocurrió el hecho vital, son variables cuya relación con el fenómeno de asesinatos violentos permiten modelar la probabilidad de ser asesinada bajo esta característica.

Para cada variable se calculó el valor de cada estadístico como la diferencia entre la devianza del modelo completo de 5 variables y un modelo con todas las variables del anterior menos la variable a evaluar.

Significancia Individual			
Variable	Valor del Estadístico	Grados de Libertad	P-value
Nivel de escolaridad	786.23	2	$1.8730419 \times 10^{-171}$
Ingresos	246.46	2	$3.0390562 \times 10^{-54}$
Estado civil	363.22	2	$1.3429713 \times 10^{-79}$
Zona	11.71	1	$6.2266882 \times 10^{-4}$
Sitio	3151.28	2	0

Cuadro 4.2: Resumen de pruebas individuales de significancia de las variables explicativas bajo diferencia de devianzas.

Una forma alternativa de calcular este estadístico, es a través de la suma de cuadrados de los residuos de devianza (Hosmer & Lemeshow, 2000), definidos como sigue:

$$d(y_j, \hat{\pi}_j) = \pm \sqrt{2 \left[ y_j \text{Ln} \left( \frac{y_j}{m_j \hat{\pi}_j} \right) + (m_j - y_j) \text{Ln} \left( \frac{m_j - y_j}{m_j (1 - \hat{\pi}_j)} \right) \right]} \quad j = 1, \dots, J$$

Donde el valor  $\pm$  depende del signo de la resta  $y_j - m_j \hat{\pi}_j$ .

De modo que:

$$G^2 = \sum_{j=1}^J d(y_j, \hat{\pi}_j)^2$$

## Resultados

El procedimiento para el cálculo de la devianza y la prueba de ajuste global bajo esta variante se presenta a continuación, observado que coinciden con los obtenidos a partir de la primera definición del estadístico, rechazando nuevamente la hipótesis nula dado un p-value de 1.546541e-13.

```
E.residuos.devianza<-sum(residuals(Mod5_Agreg,type="deviance")^2)
E.residuos.devianza

## [1] 285.7728

Grados_L<-Mod5_Agreg$df.residual
1-pchisq(E.residuos.devianza,Grados_L)

## [1] 1.546541e-13
```

Estos residuos, en el caso de que se haya ajustado el modelo agrupado, son de gran interés en la sección de diagnóstico, junto con los residuos de Pearson que se verán en la siguiente sección. Por el momento, puede observarse que el estadístico de prueba  $G^2$  puede fragmentarse en residuos individuales correspondientes a cada patrón de covariables. Como el objetivo en nuestro caso es no rechazar la hipótesis nula  $H_0$ , que establece que el modelo propuesto ajusta a los datos, una señal de buen ajuste se vería reflejado en valores de residuos pequeños, pues cuando  $G^2$  no es muy grande, aceptamos la hipótesis nula.

## Ji-cuadrada

La prueba Ji-cuadrada hace uso de los residuos de Pearson, definidos para cada patrón de covariables como sigue (Hosmer & Lemeshow, 2000):

$$d(y_j, \hat{\pi}_j) = \frac{y_j - m_j \hat{\pi}_j}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}} \quad j = 1, \dots, J$$

De modo que el estadístico de prueba:

$$\chi^2 = \sum_{j=1}^J d(y_j, \hat{\pi}_j)^2$$

Bajo  $H_0$ , el anterior estadístico se distribuye como una Ji-cuadrada con  $J - (p + 1)$  grados de libertad, por lo que rechazamos la hipótesis nula cuando es mayor al cuantil  $(1 - \alpha)$  de dicha distribución.

## Resultados

Los cálculos correspondientes a esta prueba se presentan a continuación, obteniendo un estadístico de 309.1071 y un p-value de 2.220446e-16, concluyendo de la misma manera que en el caso de la prueba de devianza.

```
E.residuos.pearson<-sum(residuals(Mod5_Agreg,type="pearson")^2)
E.residuos.pearson

## [1] 309.1071

Grados_L<-Mod5_Agreg$df.residual
1-pchisq(E.residuos.pearson,Grados_L)

## [1] 2.220446e-16
```

### 4.6.2. Pruebas basadas en las probabilidades

Las pruebas basadas en los estadísticos de la Devianza y la Ji cuadrada que fueron abordados en la sección anterior, pueden tener problemas en su aplicación, pues sus estadísticos siguen una distribución asintótica Ji cuadrada con  $J - (p + 1)$  grados de

libertad únicamente cuando se ajusta el modelo agrupado. Además de que cuando los  $m_j$  son pequeños, es decir, cuando el número de sujetos en cada patrón de covariables  $j$  es pequeño, la prueba también puede tener dificultades en su aplicación.

Las pruebas cuyos estadísticos están basados en las probabilidades estimadas son utilizadas cuando los problemas anteriores se presentan, aplicadas irremediamente para probar la bondad de ajuste en sentido global cuando el modelo que se ajusta es del tipo desagregado, es decir, cuando se trabaja con una variable de respuesta del tipo Bernoulli( $\pi_i$ ).

### Hosmer-Lemeshow

La prueba de Hosmer-Lemeshow hace uso de las probabilidades estimadas por el modelo que se quiere evaluar, agrupándolas en  $g$  diferentes grupos de aproximadamente igual tamaño, haciendo una comparación de éstas contra los valores que realmente se observaron en la base trabajada. Si el ajuste es bueno, al igual que en una prueba Ji-cuadrada, se esperarían observar valores pequeños en el estadístico utilizado, pues significaría que las diferencias entre los verdaderos valores observados y los estimados por el modelo son pequeñas.

La forma en que podemos agrupar estas probabilidades son dos, la primera es agrupándolas ordenándolas primero de menor a mayor, y tomar los grupos de tal manera que el primero contenga las  $\frac{n}{g}$  probabilidades más pequeñas, y el último grupo contenga las  $\frac{n}{g}$  probabilidades más grandes. La segunda, por otro lado, consiste en dividir los grupos por medio de percentiles, de tal forma que el primer grupo contiene todas las probabilidades cuyo valor este contenido en el intervalo  $[0, \frac{1}{g}]$  y el último grupo las probabilidades que estén contenidas en el intervalo  $[\frac{g-1}{g}, 1]$  (Hosmer Lemeshow, 2000).

El estadístico utilizado es el siguiente:

$$\begin{aligned}\widehat{C} &= \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k} + \sum_{k=1}^g \frac{((n'_k - o_k) - n'_k(1 - \bar{\pi}_k))^2}{n'_k(1 - \bar{\pi}_k)} \\ &= \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)} (1 - \bar{\pi}_k + \bar{\pi}_k) \\ &= \sum_{k=1}^g \frac{(o_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)}\end{aligned}$$

Donde:

$$o_k = \sum_{j=1}^{c_k} y_j$$
$$\bar{\pi}_k = \sum_{j=1}^{c_k} \frac{m_j \hat{\pi}_j}{n'_k}$$

$n'_k$  es el número de elementos en el k-ésimo grupo.

$c_k$  es el número de patrones de covariables en el k-ésimo grupo.

Cuya distribución es Ji-cuadrada con  $g - 2$  grados de libertad cuando  $J = n$  y bajo la hipótesis nula de que el modelo es correcto. Rechazamos la hipótesis nula de que el modelo ajusta si:

$$\blacksquare P(\chi_{(g-2)} > x) < \alpha$$

Donde:

- $\chi$  es una variable aleatoria Ji-cuadrada con  $g - 2$  grados de libertad.
- $x$  es el valor del estadístico.
- $\alpha$  es en nivel de significancia.

En cuanto al número de grupos que deben elegirse, en general se considera que 10 son adecuados para casi cualquier modelo, además de que considerar menos de 6 hará en casi todos los casos aceptar la hipótesis de ajuste global. Asimismo, por medio de simulación se llegó a la conclusión de que agrupar por medio de cuantiles converge mejor a una Ji-cuadrada con G-2 grados de libertad que cuando se agrupa

## Resultados

Para el caso en que agrupamos de tal manera que el primer grupo contenga a las probabilidades más pequeñas y el último a las probabilidades más altas, hacemos uso de la función `fitted.values` y `cut`, las cuales primero extraen las probabilidades estimadas del modelo a evaluar para posteriormente agruparlas.

El desarrollo y cálculo del valor del estadístico se presenta a continuación, obteniendo un valor de 34.09524 y un p-value de 3.904365e-05. Por lo que un nivel de significancia de  $\alpha=0.05$ , se cuenta con evidencia suficiente para rechazar la hipótesis nula de que el modelo ajusta globalmente a los datos.

```

V_Ajustados<-fitted.values(Mod5_Desag)
Grupos1<-cut(V_Ajustados, breaks=10, include.lowest=TRUE)
Obs1<-xtabs(cbind(1 - Muerte, Muerte) ~ Grupos1)
Esp1<-xtabs(cbind(1-V_Ajustados,V_Ajustados)~Grupos1)
Est1<-sum(((Obs1 - Esp1)^2)/(Esp1))
Est1

## [1] 34.09524

pvalue1<-1-pchisq(Est1,10-2)
pvalue1

## [1] 3.904361e-05

```

Por otro lado, cuando consideramos los grupos de tal manera que agrupamos por percentiles, utilizamos la función *hoslem.test* de librería *ResourceSelection* de R. En este caso, se obtiene un estadístico con valor de 25.4948 y un p-value de 0.001281015, por lo que a un nivel de significancia de  $\alpha=0.05$  nuevamente se cuenta con evidencia suficiente para rechazar la hipótesis nula de ajuste global.

```

hoslem.test(Mod5_Desag$y, fitted(Mod5_Desag), g=10)

##
## Hosmer and Lemeshow goodness of fit (GOF) test
##
## data: Mod5_Desag$y, fitted(Mod5_Desag)
## X-squared = 25.495, df = 8, p-value = 0.001281

```

Esta prueba se considera algo inestable, pues los resultados obtenidos dependen en gran medida del número de grupos que se elijan, llegando incluso a conclusiones contrarias. Así se observa por ejemplo en el caso en que tomamos 15 grupos, obteniendo un estadístico de 20.862 y un p-value de 0.07571, contradiciendo el resultado anterior al no rechazar la hipótesis de ajuste global.

```

hoslem.test(Mod5_Desag$y, fitted(Mod5_Desag), g=15)

##

```

```
## Hosmer and Lemeshow goodness of fit (GOF) test
##
## data: Mod5_Desag$y, fitted(Mod5_Desag)
## X-squared = 20.862, df = 13, p-value = 0.07571
```

### 4.6.3. Interpretación de los parámetros

Una vez que se haya logrado ajustar un modelo, el siguiente paso consiste en dar una interpretación a los parámetros que fueron estimados, pues estos valores son los que describen las relaciones y el grado de impacto que tiene cada variable sobre el fenómeno estudiado. Esta interpretación se da en términos de cocientes de momios, y dependiendo si la variable es del tipo dicotómica, politómica o continua, la lectura de estas relaciones cambiará.

#### Variable dicotómica

El caso en que el modelo es simple y el parámetro estimado que se desea interpretar acompaña a una variable explicativa dicotómica, el logaritmo del cociente de los momios de la variable respuesta  $Y=1$ , cuando  $x = 1$  y  $x = 0$  respectivamente es:

$$\begin{aligned} \text{Ln} \left( \frac{\frac{P(Y = 1|x = 1)}{1 - P(Y = 1|x = 1)}}{\frac{P(Y = 1|x = 0)}{1 - P(Y = 1|x = 0)}} \right) &= \text{Ln} \left( \frac{P(Y = 1|x = 1)}{1 - P(Y = 1|x = 1)} \right) - \text{Ln} \left( \frac{P(Y = 1|x = 0)}{1 - P(Y = 1|x = 0)} \right) \\ &= \beta_0 + \beta_1(1) - \beta_0 - \beta_1(0) \\ &= \beta_1 \end{aligned}$$

Así, considerando el modelo simple, interpretamos a  $e^{\beta_1}$  como el cambio entre el momio de la respuesta  $Y = 1$  cuando  $x = 1$ , con respecto al momio de la respuesta  $Y = 1$  cuando  $x = 0$ .

Por otro lado, cuando el modelo es múltiple y se consideran fijos los valores de las demás variables explicativas, la interpretación es similar al caso en que el modelo es simple, siendo  $e^{\beta_j}$  el cambio entre el momio de la respuesta  $Y = 1$  cuando  $x_j = 1$ , con respecto al momio de la respuesta  $Y = 1$  cuando  $x_j = 0$ , siempre haciendo énfasis sobre la condición del resto de las variables explicativas fijas..

## Resultados

### Variable Zona

La única variable que es del tipo dicotómica es la que describe en qué tipo de zona residía la mujer, siendo rural o urbana las únicas dos opciones posibles. El cuadro 4.3 muestra el valor del parámetro estimado y su exponencial.

Coeficientes Variable Zona		
Comparación	Parámetro	Cociente de momios
Rural vs Urbana	0.242	1.274

Cuadro 4.3: Tabla que muestra el valor del parámetro estimado correspondiente a la variable Zona y su exponencial

Así, al haber ajustado un modelo múltiple, el valor de 1.274 se interpreta como el cambio entre el momio de la respuesta de morir de una forma violenta cuando la mujer vive en una zona rural, con respecto al momio de morir de una forma violenta cuando la mujer vive en una zona urbana, manteniendo el resto de las variables constantes.

Además, al ser 1.274 mayor a uno, se observa que la probabilidad de de morir de una forma violenta es mayor cuando la mujer vive en una zona urbana, que cuando vive en una zona rural.

### Variable politómica

Cuando el modelo es simple y el parámetro estimado acompaña a una variable explicativa politómica  $x$ , con  $x = 1, \dots, K$ , la lectura es similar al caso en el que la variable es dicotómica, pues el procedimiento requiere crear  $K - 1$  variables de este tipo, denominadas variables dummy. La construcción de estas variables es como sigue:

Variables Dummy					
Variable x	D <sub>1</sub>	D <sub>2</sub>	...	D <sub>K-1</sub>	D <sub>K</sub>
1	0	0	...	0	0
2	1	0	...	0	0
3	0	1	...	0	0
⋮	⋮	⋮	⋮	⋮	⋮
K-1	0	0	...	1	0
K	0	0	...	0	1



En este caso, las comparaciones se hacen entre las categorías existentes en la variable  $x$ , y otra de esta misma denominada *basal*, la cual es aquella que se identifica cuando todos sus valores son cero. En la tabla anterior corresponde a la categoría 1.

Así, siendo el logaritmo del cociente de los momios de la variable respuesta  $Y = 1$ , cuando  $x = k$  y  $x = 1$  como sigue:

$$\begin{aligned} \text{Ln} \left( \frac{\frac{P(Y = 1|x = k)}{1 - P(Y = 1|x = k)}}{\frac{P(Y = 1|x = 1)}{1 - P(Y = 1|x = 1)}} \right) &= \text{Ln} \left( \frac{P(Y = 1|x = k)}{1 - P(Y = 1|x = k)} \right) - \text{Ln} \left( \frac{P(Y = 1|x = 1)}{1 - P(Y = 1|x = 1)} \right) \\ &= \beta_0 + \beta_1(0) + \beta_2(0) + \dots + \beta_k(1) + \dots + \beta_K(0) \\ &\quad - \beta_0 + \beta_1(0) + \beta_2(0) + \dots + \beta_k(0) + \dots + \beta_K(0) \\ &= \beta_k \end{aligned}$$

Cuando el modelo es simple, para  $k=2, \dots, K$  se interpreta a  $e^{\beta_k}$  como el cambio entre el momio de la respuesta  $Y = 1$  cuando  $x = k$ , con respecto al momio de la respuesta  $Y = 1$  cuando  $x = 1$ , es decir, la categoría basal.

Por otro lado, en el caso en que se haya ajustado un modelo múltiple, la interpretación es similar al caso en que el modelo es simple, siendo  $e^{\beta_k}$  el cambio entre el momio de la respuesta  $Y = 1$  cuando  $x_j = k$ , con respecto al momio de la respuesta  $Y = 1$  cuando  $x_j = 1$ , nuevamente haciendo énfasis en la condición del resto de variables explicativas fijas.

## Resultados

Volviendo al modelo ajustado, de las variables incluídas cuatro son politómicas, las cuales describen el nivel de escolaridad de las mujeres fallecidas, su nivel de ingresos, su estado conyugal y el sitio donde fallecieron.

### Variable Nivel de escolaridad

Para el caso de la variable Nivel de escolaridad, se consideraron las categorías nivel de escolaridad Superior, Media superior y Básica, tomando como categoría *basal* al nivel de escolaridad Superior. El cuadro 4.4 muestra el valor de los parámetros estimados correspondientes a esta variable y su exponencial:

Coeficientes Variable Nivel de Escolaridad		
Comparación	Parámetro	Cociente de momios
Media S. vs Basal	0.24	1.272
Básica vs Basal	-1.404	0.246

Cuadro 4.4: Tabla que muestra el valor de los parámetros estimados correspondientes a la variable Nivel de escolaridad y su exponencial

Así, el valor de 1.272 se interpreta como el cambio entre el momio de la respuesta de morir de una forma violenta cuando la mujer tiene un nivel de escolaridad media superior, con respecto al momio de morir de una forma violenta cuando la mujer tiene un nivel de escolaridad superior, manteniendo el resto de las variables contantes. Por otro lado, 0.246, se interpreta como el cambio entre el momio de la respuesta de morir de una forma violenta cuando la mujer tiene un nivel de escolaridad básico, con respecto al momio de morir de una forma violenta cuando la mujer tiene un nivel de escolaridad superior, manteniendo el resto de las variables contantes.

Además, al ser 1.272 mayor a uno, se observa que la probabilidad de morir de una forma violenta es mayor cuando la mujer tiene un nivel de escolaridad medio superior, el cual abarca los grados desde la secundaria completa hasta bachillerato completo, que cuando su nivel de escolaridad es superior, que abarca los grados profesional y posgrado.

Asimismo, al ser 0.246 menor que uno, la probabilidad de morir de una forma violenta cuando la mujer tiene un nivel de escolaridad básico, que abarca los grados sin escolaridad hasta primaria completa, es menor que cuando su nivel de escolaridad es superior. En este caso, la interpretación resulta contraintuitiva, y se debe a que un número mucho mayor de casos de muerte natural se reportaron en la categoría nivel de escolaridad básico.

Por lo anterior, se da la impresión de que no es un factor de riesgo tener este grado de escolaridad, pues el número tan grande de muertes naturales no permite figurar a los asesinatos reportados de forma violenta.

### Variable Ingresos

Para el caso de la variable Ingresos, se consideraron las categorías con De medios a altos, De muy bajos a bajos y No genera ingresos, tomando como categoría *basal* la ocupación con ingresos De medios a altos. El cuadro 4.5 muestra el valor de los parámetros estimados correspondientes a esta variable y su exponencial:

Coeficientes Variable Ingresos		
Comparación	Parámetro	Cociente de momios
Muy bajos y Bajos vs Basal	1.307	3.694
No genera I. vs Basal	-0.057	0.944

Cuadro 4.5: Tabla que muestra el valor de los parámetros estimados correspondientes a la variable Ingresos y su exponencial

Así, el valor de 3.694 se interpreta como el cambio entre el momio de la respuesta de morir de una forma violenta cuando la mujer tiene ingresos de muy bajos a abajos, con respecto al momio de morir de una forma violenta cuando la mujer tiene un nivel de ingresos de medios a altos, manteniendo el resto de las variables constantes. Para el caso de 0.944, este valor resultó no ser significativo, por lo que se omite su interpretación.

Por otro lado, al ser 3.694 mayor a uno, se observa que la probabilidad de morir de una forma violenta es mayor cuando la mujer tiene ingresos de muy bajos a bajos, que cuando su nivel de ingresos es de medios a altos, según la agrupación que se explicó anteriormente en la sección 1.2.

### Variable Estado civil

Las categorías en este caso fueron: Separada o viuda, Casada o en Unión libre y Soltera, tomando como categoría *basal* el estado Separada o viuda. El cuadro 4.6 muestra el valor de los parámetros estimados correspondientes a esta variable y su exponencial:

Coeficientes Variable Estado civil		
Comparación	Parámetro	Cociente de momios
Casada o en U. libre vs Basal	1.168	3.216
Soltera vs Basal	1.491	4.441

Cuadro 4.6: Tabla que muestra el valor de los parámetros estimados correspondientes a la variable Estado civil y su exponencial

Por lo tanto, el valor de 3.216 se interpreta como el cambio entre el momio de la respuesta de morir de una forma violenta cuando la mujer estaba casada o en unión libre al momento de la muerte, con respecto al momio de morir de una forma violenta cuando la mujer era separada o viuda, manteniendo el resto de las variables constantes. Para el caso de 4.441, se interpreta como el cambio entre el momio de la respuesta de morir de una forma violenta cuando la mujer era soltera al momento de la muerte, con respecto al momio de morir de una forma violenta cuando la mujer era separada o viuda, manteniendo el resto de las variables constantes.

Además, al ser ambos cocientes mayores a uno, tenemos que la probabilidad de morir de una forma violenta cuando la mujer estaba casada o en unión libre, o era soltera, es mayor a la probabilidad de morir de una forma violenta cuando la mujer se encontraba separada o viuda.

### Variable Lugar de defunción

Las categorías para esta variable fueron Hospital, Sitio Público y Hogar, tomando como categoría *basal* Hospital. El cuadro 4.7 muestra nuevamente el valor de los parámetros estimados correspondientes a esta variable y su exponencial:

Coeeficientes Variable Lugar de defuncion		
Comparación	Parámetro	Cociente de momios
Sitio P. vs Basal	3.877	48.282
Hogar vs Basal	0.863	2.37

Cuadro 4.7: Tabla que muestra el valor de los parámetros estimados correspondientes a la variable Lugar de defunción y su exponencial

Así, el valor de 48.282 se interpreta como el cambio entre el momio de la respuesta de morir de una forma violenta en un sitio público, con respecto al momio de morir de forma violenta y lograr llegar a un hospital, manteniendo el resto de las variables contantes. Para el caso de 2.37, se interpreta como el cambio entre el momio de la respuesta de morir de una forma violenta en el hogar, con respecto al momio de morir de forma violenta y lograr llegar a un hospital, manteniendo el resto de las variables contantes.

Además, nuevamente al ser ambos cocientes mayores a uno, observamos que los decesos de forma violenta se dan en sitios públicos y alarmanamente dentro del hogar, y que en pocas ocasiones pueden llegar a un hospital y tener algún tipo de tratamiento.

### Variable continua

Por último, cuando el modelo es simple y el parámetro estimado corresponde a una variable explicativa continua, si el logaritmo del cociente de los momios de la variable respuesta  $Y=1$ , cuando  $X = x$  y  $X = x + 1$  es como sigue:

$$\begin{aligned}
Ln \left( \frac{\frac{P(Y = 1|X = x + 1)}{1 - P(Y = 1|X = x + 1)}}{\frac{P(Y = 1|X = x)}{1 - P(Y = 1|X = x)}} \right) &= Ln \left( \frac{P(Y = 1|X = x + 1)}{1 - P(Y = 1|X = x + 1)} \right) \\
&- Ln \left( \frac{P(Y = 1|X = x)}{1 - P(Y = 1|X = x)} \right) \\
&= \beta_0 + \beta_1(x + 1) - \beta_0 - \beta_1(x) \\
&= \beta_1
\end{aligned}$$

Entonces podemos interpretar a  $e^{\beta_1}$  como el cambio entre el momio de la respuesta  $Y = 1$  cuando  $X = x + 1$ , con respecto al momio de la respuesta  $Y = 1$  cuando  $X = x$ . Y al igual que en el caso de variables dicotómicas y politómicas, cuando el modelo es múltiple, se interpreta a  $e^{\beta_j}$  como el cambio entre el momio de la respuesta  $Y = 1$  cuando  $x_j = x$ , con respecto al momio de la respuesta  $Y = 1$  cuando  $x_j = x + 1$ , considerando el resto de las variables constantes.

## 4.7. Intervalos de confianza

Una vez que se han determinado los parámetros que son significativos para el modelo, es decir, una vez que verificamos que efectivamente exista una relación causal entre nuestra variable de respuesta y la explicativa que acompaña el parámetro estimado, es necesario establecer umbrales que nos permitan conocer con cierto grado de certeza, qué tanto puede variar esta estimación, y así realizar pruebas de estrés que nos ayuden a dar interpretaciones más completas.

Haciendo uso nuevamente de las propiedades asintóticas y de invarianza de los estimadores máximo verisímiles, el intervalo de confianza para cada parámetro estimado es como sigue:

$$\exp \left( \beta_j - z_{1-\frac{\alpha}{2}} \sqrt{Var(\hat{\beta}_j)} \right) < exp(\hat{\beta}_j) < \exp \left( \beta_j + z_{1-\frac{\alpha}{2}} \sqrt{Var(\hat{\beta}_j)} \right)$$

### Resultados

A continuación se presentan los intervalos de confianza para cada cociente de momios, dado un nivel de significancia  $\alpha = 0.05$ . Por lo que con una confianza del 95 %,

podemos asegurar que el verdadero valor de cada cociente analizado no rebasa los límites de los siguientes intervalos.

```
exp(confint.default(Mod5_Desag, level=.95))

##           2.5 %       97.5 %
## (Intercept)  0.02146157  0.04071741
## Nivel_Esc2   1.02729323  1.57441340
## Nivel_Esc3   0.19673316  0.30641930
## Ingresos2    2.71498968  5.02610059
## Ingresos3    0.71821201  1.24144704
## Estado_Civil2 2.74661651  3.76599588
## Estado_Civil3 3.75193118  5.25714996
## ZonaR_U1    1.10702371  1.46507188
## Lugar_Def2   41.37612186 56.33977409
## Lugar_Def3   2.09654033  2.67821463
```

#### 4.7.1. Medidas tipo $R^2$

Es importante mencionar que estas medidas de bondad de ajuste no tienen la misma interpretación que en el modelo de regresión estándar, pues en este caso  $R^2$  no representa la varianza explicada por el modelo. Además tampoco se suelen incluir en los informes de resultados, pues las personas acostumbradas a los resultados dados por el modelo de regresión estándar podrían mal interpretar sus resultados y creer que el ajuste del modelo no es adecuado. Su utilización radica principalmente en el proceso de la construcción del modelo, permitiendo comparar entre distintos modelos su nivel de ajuste (Hosmer & Lemeshow, 2000).

#### MacFadden

Esta medida hace uso de las funciones de log-verosimilitud, tanto del modelo propuesto que incluye las  $p$  variables y el intercepto, como del modelo que sólo incluye el intercepto. El valor máximo se alcanza cuando el modelo que se ajusta es el saturado, siendo  $R^2$  igual a 1. Se define como:

$$R^2 = \frac{l_0 - l_p}{l_0} = 1 - \frac{l_p}{l_0}$$

Donde:

- $l_0$  corresponde a la log-verosimilitud del modelo que contiene únicamente al intercepto.
- $l_p$  corresponde a la log-verosimilitud del modelo que contiene al intercepto más las  $p$  covariables respectivamente.

## Cox y Snell

Además de las funciones de log-verosimilitud tanto del modelo ajustado, como del modelo que incluye sólo el intercepto, para su cálculo se requiere del número total de patrones de covariables existentes en el modelo. Se define como:

$$R^2 = 1 - e^{-\frac{l_0 - l_p}{N}}$$

Donde:

- $N$  corresponde al número de patrones de covariables presentes en el modelo.

## Resultados

A continuación se presentan los resultados de las medidas MacFadden y Cox y Snell para el modelo con cinco variables, de tal forma que como se mencionó anteriormente su valor parece pequeño en comparación a las esperadas en un modelo de regresión.

```
MacFadden<-1-Mod5_Desag$deviance/Mod5_Desag$null.deviance
MacFadden

## [1] 0.3835488
```

```
N<-dim(Est_Vit)[1]
Dif<-Mod5_Desag$null.deviance-Mod5_Desag$deviance
CoxSnell<-1-exp(-Dif/N);CoxSnell

## [1] 0.2467873
```

### 4.7.2. Poder predictivo

Otro criterio que puede usarse para evaluar qué tan bien un modelo ajusta a los datos, es por medio de la evaluación de su poder predictivo, midiendo su capacidad de determinar correctamente cuando un individuo presentará la respuesta  $Y = 1$  o  $Y = 0$ . La forma de hacerlo es asignando a cada elemento de la base en alguna de las categorías anteriores, y a partir de un punto de corte  $c$  y sus probabilidades estimadas, clasificarlos en la categoría  $Y = 1$  cuando su probabilidad sea mayor a dicho punto  $c$ , y en la categoría  $Y = 0$  cuando su probabilidad sea menor a  $c$ . De esta manera, generamos una nueva variable aleatoria de la variable respuesta a partir del modelo ajustado, la cual será comparada contra los valores de la variable de respuesta observada originalmente.

#### Tablas de clasificación cruzada

Una de las formas en que podemos evaluar el poder predictivo es por medio de una tabla de clasificación cruzada, a través de la cual podemos comparar los resultados de la variable aleatoria observada  $Y$ , contra los de la variable respuesta obtenida a través del punto de corte  $c$  seleccionado anteriormente.

En general, si un modelo tiene un buen poder predictivo, se esperaría observar valores mucho más grandes sobre la diagonal de la tabla de clasificación, pues esas cillas corresponden al número de individuos cuyos valores en ambas variables aleatorias coincidieron, es decir, corresponden al número de individuos cuyos valores predichos coinciden con los valores que realmente ostentan.

La tabla de clasificación es como sigue:

Tabla de clasificación			
		Valores Predichos	
Valores Obs.		Y=0	Y=1
	Y=0	$n_{VN}$	$n_{FP}$
	Y=1	$n_{FN}$	$n_{VP}$

Cuadro 4.8: Tabla de clasificación



## Resultados

A continuación se presenta la tabla de clasificación del modelo eligiendo un punto de corte  $c = 0.5$ , observando que dado este valor, los resultados mejor clasificados corresponden a los casos de muerte natural, con un total de 346 falsos positivos. Además, dado que el total de falsos negativos supera a los verdaderos positivos, que son el objeto de estudio, podemos decir que el punto de corte  $c = 0.5$  no es muy adecuado. La combinación de algunas medidas y métodos que se presentarán posteriormente nos ayudará a determinar si existe o no un punto de corte que clasifique mejor a las muertes violentas.

```
Valores_P<-ifelse(Mod5_Desag$fitted.values>=0.5,1,0)
Tabla.clasificacion1<-table(Muerte,Valores_P)
Tabla.clasificacion1

##      Valores_P
## Muerte      0      1
##      0 18341   346
##      1  1300  1279
```

Las medidas utilizadas para determinar qué tan bien han sido clasificadas las observaciones, se expresan en forma de tasas. La proporción que mide qué tan bien se clasificaron a los individuos cuyo valor en su variable respuesta fue 1, se le conoce como sensibilidad, la cual se define como el cociente entre el total de sujetos cuyo valor en ambas variables respuesta, la real y la generada fue 1, es decir, el total de verdaderos positivos, contra estos mismos más el total de clasificados en la categoría 0, cuando el valor en su variable respuesta real fue 1, es decir, entre el total de verdaderos positivos más los falsos negativos (Kleinbaum & Klein, 2010).

$$\text{Sensibilidad} = \frac{V.Positivos}{V.Positivos + F.Negativos}$$

Cuyo complemento:

$$\text{Tasa F.Negativos} = \frac{F.Positivos}{V.Positivos + F.Negativos}$$

Por otro lado, la proporción que mide qué tan bien se clasificaron a los individuos cuyo valor en su variable respuesta fue 0, se le conoce como especificidad, la cual se define como el cociente entre el total de sujetos cuyo valor en ambas variables respuesta

fue 0, es decir, el total de verdaderos negativos, contra estos mismos más el total de clasificados en la categoría 1, cuando el valor en su variable respuesta real fue 0, es decir, entre el total de verdaderos negativos más los falsos positivos.

$$\textit{Especificidad} = \frac{V.Negativos}{V.Negativos + F.Positivos}$$

Cuyo complemento:

$$\textit{TasaF.Positivos} = \frac{F.Positivos}{V.Negativos + F.Positivos}$$

Por lo tanto, un modelo con un buen poder predictivo puede definirse como aquel en el que ambas medidas, la sensibilidad y especificidad, sean lo más cercanas a 1, y cuyo resultado, cabe mencionarlo, depende directamente de la elección de un punto de corte óptimo que maximice ambas medidas, ya que conforme se aumente el valor de este punto, el valor de la especificidad incrementa, pero con efectos negativos en la sensibilidad, pues este último disminuye. En cambio, conforme el punto de corte esté más próximo a cero, el valor de la sensibilidad aumentará, pero la especificidad se hará más pequeña (Kleinbaum & Klein, 2010).

## Resultados

Una alternativa para encontrar el punto de corte óptimo consiste en graficar a la sensibilidad y especificidad para todo punto de corte en el intervalo  $[0, 1]$ , y elegir aquel punto en el que ambas gráficas se cruzan. Una forma de hacerlo es por medio de la función *performance* de la paquetería *ROCR*, la cual puede calcular entre otras medidas y para cada punto de corte:

1. La sensibilidad, por medio del comando “*sens*”
2. La especificidad, por medio del comando “*spec*”
3. La tasa de verdaderos positivos, por medio del comando “*tpr*”
4. La tasa de verdaderos negativos, por medio del comando “*tnr*”
5. La tasa de falsos positivos, por medio del comando “*fpr*”
6. La tasa de falsos negativos, por medio del comando “*fnr*”
7. El área bajo la curva de ROC, por medio del comando “*auc*”

Para utilizar la función *performance*, es necesario crear un objeto de la clase *prediction*. El procedimiento se presenta a continuación, observándose que un punto de corte con valor de alrededor del 0.1 sería aquel que maximiza tanto el valor de la sensibilidad, como de la especificidad.

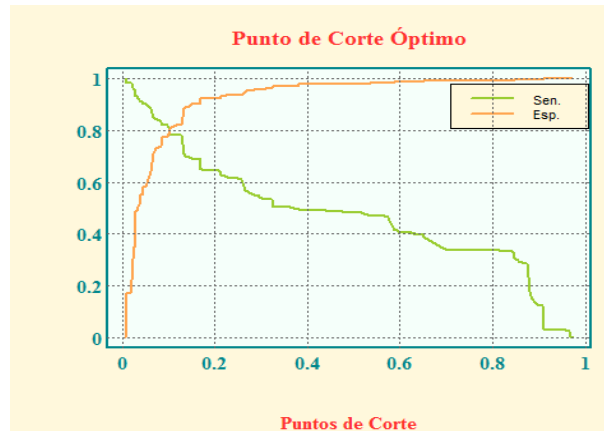


Figura 4.3: Gráfica que determina el punto de corte el cual maximiza el valor de la sensibilidad y especificidad.

A continuación se presenta la tabla de cruce correspondiente al punto de corte  $c = 0.1$ , obteniendo una sensibilidad con valor de 0.81 y una especificidad de 0.79, un resultado notoriamente mejor al obtenido con el punto de corte  $c = 0.5$ .

```
Valores_P<-ifelse(Mod5_Desag$fitted.values>=0.1,1,0)
Tabla.clasificacion2<-table(Muerte,Valores_P)
Tabla.clasificacion2
```

```
##      Valores_P
## Muerte      0      1
##      0 14810  3877
##      1   490  2089
```

```
Sensibilidad<-Tabla.clasificacion2[2,2]/(Tabla.clasificacion2[2,1]
+Tabla.clasificacion2[2,2])
```

```
Sensibilidad
```

```
## [1] 0.8100039
```

```
Especificidad<-Tabla.clasificacion2[1,1]/(Tabla.clasificacion2[1,1]
```

```

+Tabla.clasificacion2[1,2])
Especificidad

## [1] 0.7925296

```

Por otro lado, también se puede utilizar la función *ROC* de la paquetería *Epi*, la cual arroja un punto de corte óptimo y el valor de las medidas anteriormente citadas. El procedimiento se presenta a continuación, dando como resultado un punto *c* con valor de 0.110, y una sensibilidad y especificidad de 0.787 y 0.819 respectivamente. Además, también dibuja la curva de ROC y calcula el área debajo de ésta, cuya construcción e interpretación se explicará a continuación.

```

par(bg="cornsilk", mar=c(4.2,4,3,1), lwd=1.5, col.lab="firebrick1",
    font.axis=7,font.lab=7, col.axis="turquoise4")
ROC(form=Muerte~Nivel_Esc+Ocupacion
     +Estado_Civil+ZonaR_U
     +Lugar_Def, data=Estadisticas_V)
box(lwd = 2, col="turquoise4")
title("Punto de Corte Óptimo", font.main=7, cex.main=1.1,
      col.main="firebrick1")

```

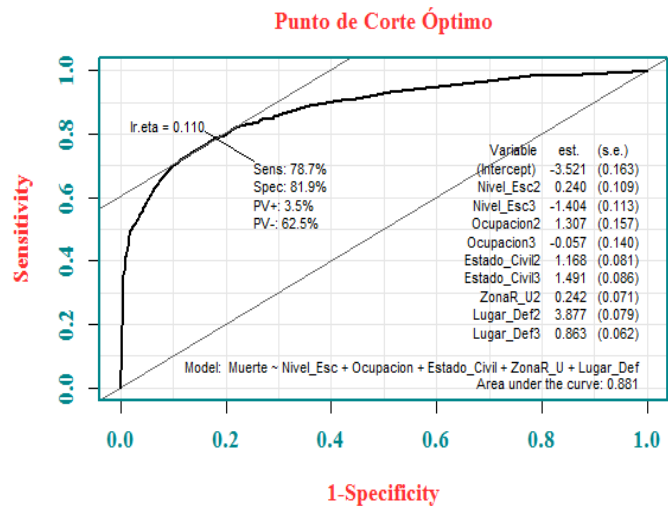


Figura 4.4: Gráfica que determina el punto de corte óptimo, utilizando la función ROC de la paquetería *Epi* de R.

## Curva de ROC

Otra alternativa para medir el poder predictivo es la llamada curva de ROC, la cual nos proporciona un panorama más general con respecto a la relación entre los distintos puntos de corte y el grado de sensibilidad y especificidad que producen cada uno de ellos. El procedimiento consiste en graficar a  $1-\text{Especificidad}$ , también conocida como *Tasa de Falsos Positivos*, contra la *Sensibilidad*, para todo punto de corte  $c$  entre el rango de  $[0, 1]$ .

La forma de la curva en general consiste en una figura cóncava cuyos valores se encuentran por encima de la función identidad, y que además crece mucho más rápidamente antes del punto de corte  $c=0.5$ , pues lo anterior aseguraría que tanto la especificidad como la sensibilidad tomarán valores altos. Para el caso de nuestro modelo, la curva es como sigue.

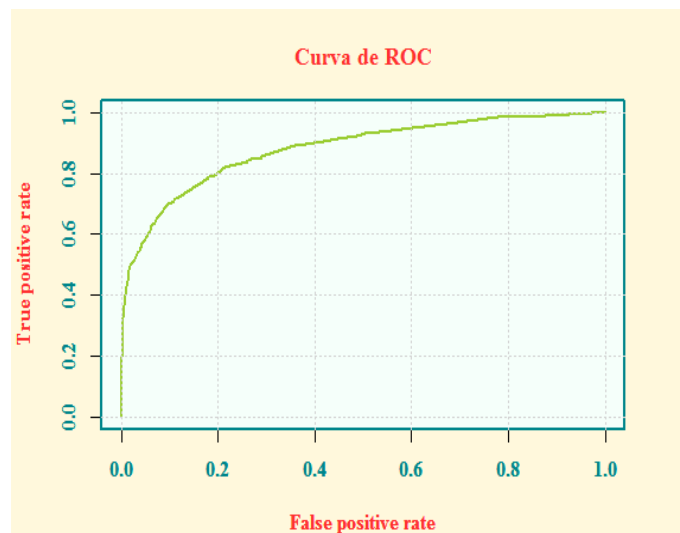


Figura 4.5: Gráfica que determina el punto de corte el cual maximiza el valor de la sensibilidad y especificidad.

Por otro lado, otra medida que también suele usarse para medir el poder predictivo de un modelo es aquella que mide el área bajo la curva de ROC, y la cual además permite comparar entre diferentes modelos el poder predictivo de forma estandarizada. El criterio a utilizar cuando se están comparando varios modelos consiste en elegir aquel cuya área bajo la curva sea mayor, pues necesariamente un área mayor bajo la curva es el resultado de una cuyos valores se encuentren muy por arriba de la función identidad, lo que significa que para todos los puntos de corte  $c$ , la tasa de verdaderos positivos es mucho mayor a la tasa de falsos positivos.

En general, el área no supera el valor de 1, pues sería el caso en que el modelo predice perfectamente cada caso, ni tampoco es menor a 0.5, pues es el caso en que la clasificación se da de manera aleatoria, pues la tasa de verdaderos positivos es la misma que la tasa de falsos positivos.

Cuando se toma un modelo, se suele utilizar la siguiente regla de decisión para determinar el grado predicción (Kleinbaum & Klein, 2010):

Capacidad de Discriminación	
Área	Descripción
0.9 - 1.0	Excelente Discriminación
0.8 - 0.9	Buena Discriminación
0.7 - 0.8	Razonable Discriminación
0.6 - 0.7	Pobre Discriminación
0.5 - 0.6	Discriminación fallida

Cuadro 4.9: Cuadro que determina la capacidad de discriminación de un modelo.

## Resultados

Habiendo elegido como punto de corte  $c = 0.110$ , y al tener un área bajo la curva de 0.8811868, se puede concluir que este modelo, el cual describe la relación entre las muertes violentas de mujeres en condiciones especialmente vulnerables con las variables Nivel de escolaridad, Ingresos, Estado Civil, tipo de Zona de residencia y Sitio donde ocurrió el hecho vital, aunque muestra evidencia de un ajuste pobre en comparación del modelo saturado, tiene un buen nivel de discriminación. Los resultados se muestran a continuación:

```
Pred<-prediction(fitted.values(Mod5_Desag), Muerte)
AUC<-performance(Pred, "auc")
AUC@y.values

## [[1]]
## [1] 0.8811868
```

## 4.8. Diagnóstico de las observaciones

Evaluar si el modelo que se ajustó se puede considerar adecuado al conjunto de datos es uno de los elementos más importantes en el proceso de modelado, pues es primordial verificar si los supuestos que definieron el modelo y bajo los cuales se trabajaron se cumplen sobre los datos trabajados. Además de eso, podría ocurrir que el modelo no

contiene variables explicativas que debería agregarse, o por ejemplo podría ocurrir que sobre un conjunto específico de los datos el modelo no ajusta adecuadamente.

Al proceso que evalúa lo anteriormente planteado se le conoce como diagnóstico, y el conjunto de técnicas del cual se conforma pueden basarse tanto en pruebas estadísticas formales, o en técnicas que involucran un análisis gráfico, donde la mayoría de éstas requieren de los elementos derivados de un modelo del tipo agregado.

### 4.8.1. Leverage

El leverage y los residuos son elementos fundamentales en el proceso de diagnóstico, pues de ellos se derivan un gran número de medidas que formarán parte del análisis gráfico para detectar algunas de las irregularidades que fueron anteriormente planteadas.

En el caso de regresión lineal, el leverage es una medida que ayuda a identificar elementos que influyen de manera significativa en los valores finales de los parámetros estimados, pues proporciona una medida la cual es proporcional a la distancia que existe entre el vector de valores observados sobre elemento  $j$  y su media.

Para el caso de regresión logística, la interpretación de este valor como una medida de distancia es un poco diferente en el sentido de que únicamente se puede tomar como tal cuando la probabilidad estimada para el elemento  $j$  se encuentre situada en el intervalo entre 0.1 y 0.9.

El leverage corresponde a la diagonal de la matriz que proporciona a los valores ajustados como la proyección de la variable respuesta dentro de espacio de covariables, la cual es denotada por  $H$ . En el caso de regresión logística, la matriz es como sigue:

$$H = V^{1/2}X(X'VX)^{-1}X'V^{1/2}$$

Donde:

$$V = \begin{pmatrix} m_1\hat{\pi}_1(1 - \hat{\pi}_1) & 0 & \cdots & 0 \\ 0 & m_2\hat{\pi}_2(1 - \hat{\pi}_2) & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & m_J\hat{\pi}_J(1 - \hat{\pi}_J) \end{pmatrix}$$

Por lo que cada elemento de la diagonal se puede expresar como:

$$\begin{aligned} h_j &= m_j \hat{\pi}_j (1 - \hat{\pi}_j) (\bar{x}_j') (X' V X)^{-1} (\bar{x}_j') \\ &= v_j b_j \end{aligned}$$

Donde:

- $b_j = (\bar{x}_j') (X' V X)^{-1} (\bar{x}_j')$
- $\bar{x}_j' = (1, x_{1j}, x_{2j}, \dots, x_{pj})$

## 4.8.2. Residuos

Los residuos se definen como aquellos valores que miden el nivel de semejanza entre el verdadero valor observado y el valor que fue ajustado por el modelo, por lo que dan una visión general de la falta de ajuste sobre ciertos elementos de la base. Algunos de éstos se listan a continuación:

**Residuos de Pearson:**

$$d_{Pj} = \frac{y_j - m_j \hat{\pi}_j}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}} \quad j = 1, \dots, J$$

**Residuos de Pearson Estandarizados:**

$$d_{PEj} = \frac{d_{Pj}}{\sqrt{1 - h_j}} \quad j = 1, \dots, J$$

**Residuos de Devianza:**

$$d_{Dj} = \pm \sqrt{2 \left[ y_j \text{Ln} \left( \frac{y_j}{m_j \hat{\pi}_j} \right) + (m_j - y_j) \text{Ln} \left( \frac{m_j - y_j}{m_j (1 - \hat{\pi}_j)} \right) \right]} \quad j = 1, \dots, J$$



### Residuos de Devianza Estandarizados:

$$d_{DEj} = \frac{d_{Dj}}{\sqrt{1 - h_j}} \quad j = 1, \dots, J$$

### Residuos de Verosimilitud:

$$d_{Lj} = \text{sgn}(y_j - \hat{y}_j) \sqrt{h_j(d_{PEj})^2 + (1 - h_j)(d_{DEj})^2} \quad j = 1, \dots, J$$

### 4.8.3. Análisis gráfico

Algunas de las gráficas con las cuales se puede comenzar a hacer el análisis y divisar problemas en el ajuste, son las siguientes:

### Índice vs Residuos:

Ayuda a localizar principalmente outliers o elementos alejados del resto de los datos, los cuales valen la pena ser revisados, pues pueden ser el resultado de un error de registro o un error de cálculo. Además de que permite observar algún patrón sistemático en los puntos lo cual puede ser evidencia de que el modelo no es correcto.

La gráfica se muestra en la Figura 4.6, y se consideran outliers aquellos elementos cuyo valor en valor absoluto sea mayor a 2. Así, en general observamos valores cuyos residuos se encuentran agregados dentro del rango permitido, siendo la categoría No genera ingresos de la variable Ingresos la categoría que más apareció entre los elementos cuyos residuos fueron mayores. La información se muestra en el Cuadro 4.10, que muestra los patrones de covariables y el valor de los residuos para cada uno de ellos.

Elementos mal ajustados						
Posición	Nivel.E	Ocup	Edo.C	Zona	Sitio	Residuo
1	2	2	2	2	1	-4.8373
2	3	3	3	2	3	-4.6526
3	2	3	3	1	1	3.2769
4	3	2	2	2	3	3.1556
5	2	3	1	2	2	-3.1225

Cuadro 4.10: Tabla que contiene el conjunto de elementos que tuvieron en valor absoluto los residuos de devianza estandarizados más grandes.

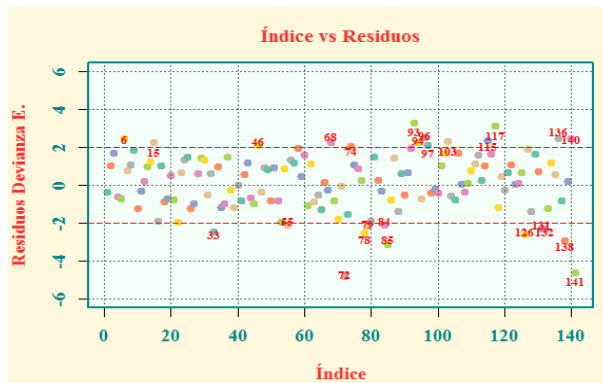


Figura 4.6: Gráfica de diagnóstico del comportamiento del índice contra residuos.

### Predictor lineal vs Residuos:

Al igual que en el caso anterior, ayuda a visualizar outliers o revelar alguna inconsistencia en el modelo. Observar patrones sistemáticos da evidencia de ello.

La gráfica del modelo se muestra en la Figura 4.7, observándose en este caso un pequeño patrón de líneas verticales de puntos, que como anteriormente se señaló son evidencia de algún error en el modelo. Éstos van desde una mala especificación del modelo, ya sea por falta de variables explicativas significativas o por una función liga inadecuada. En el caso en que trabajemos con variables continuas, podría ser evidencia de la necesidad de realizar alguna transformación de la variable.

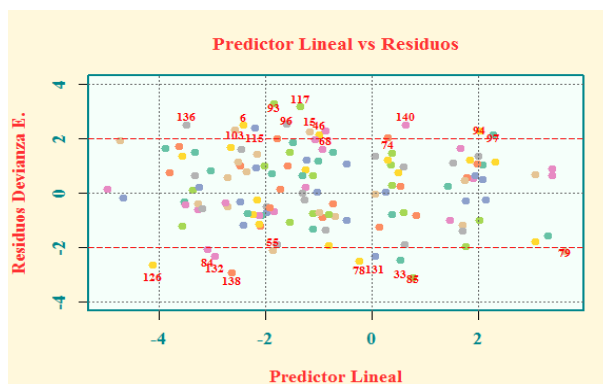


Figura 4.7: Gráfica de diagnóstico del comportamiento del predictor lineal contra los residuos de devianza estandarizados.

Para el caso en que se ajusta el modelo desagregado la gráfica en general muestra un patrón muy marcado, esto debido a que la variable de respuesta sólo puede tomar los valores de 0 o 1. Por lo tanto, no es fácil dar una interpretación a tal gráfica, por lo que no suele utilizarse por ser poco informativa. La gráfica se muestra en la Figura 4.8.

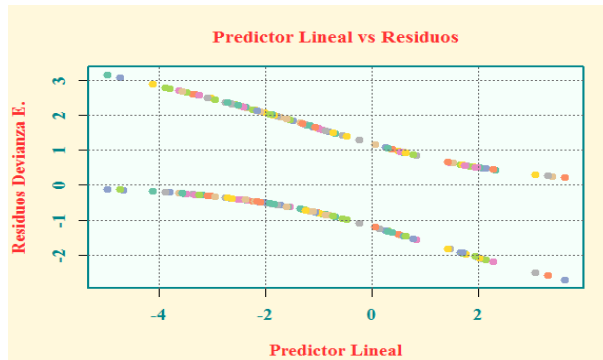


Figura 4.8: Gráfica de diagnóstico del comportamiento del predictor lineal contra residuos del modelo desagregado.

### Probabilidades estimadas vs Proporciones reales:

Otra manera de evaluar el ajuste del modelo consiste en graficar a las probabilidades estimadas contra las proporciones reales, de tal forma que cuando el ajuste es bueno se espera observar una gráfica cuya nube de puntos se encuentre cercana a la diagonal. La gráfica se muestra a continuación, observándose ligeramente la forma anteriormente planteada pero con algunos ajustes totalmente extremos.

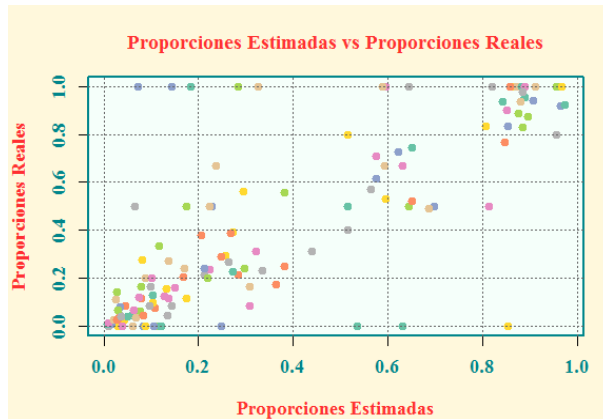


Figura 4.9: Gráfica de diagnóstico del comportamiento entre las proporciones estimadas contra las proporciones reales.

En el caso en que se trabaje con variables continuas, pueden graficarse los valores en dicha variable contra ambas proporciones, y observar la similitud entre ambas.

## Variables predictoras vs Residuos:

En este caso, graficar el valor de las variables predictoras contra los residuos nos permiten visualizar sobre qué valores de éstas se localizan los residuos más significativos, de tal manera que se pueda vislumbrar de manera general sobre qué patrones de covariables el modelo puede no ajustar correctamente. En este caso, utilizamos la función *residualPlots* de la paquetería *car*, la cual despliega para el caso de variables categóricas los boxplots de cada categoría de cada variable. Para que el modelo se considere correcto los boxplots deberán estar centrados alrededor de la recta  $y = 0$  y no deben de presentar ningún patrón.

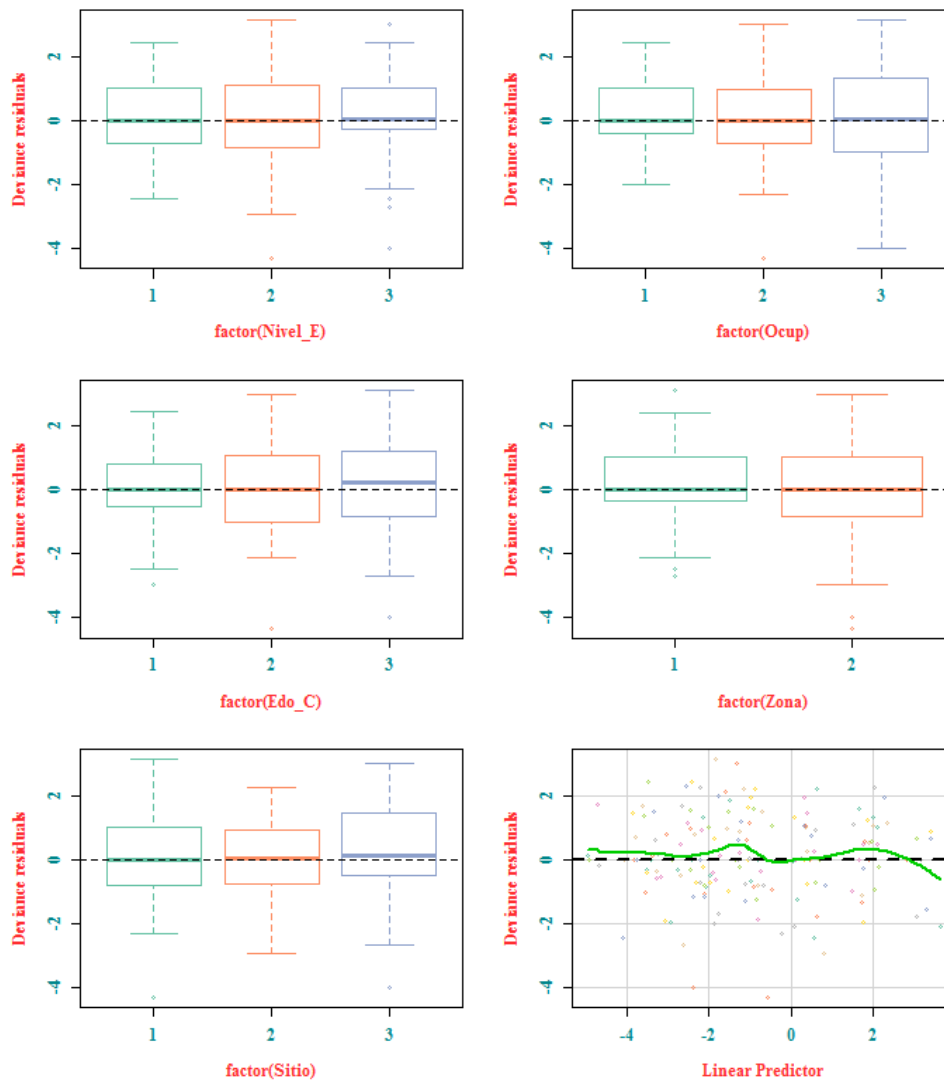


Figura 4.10: Gráfica de los residuos para cada variable del modelo.

## Índice vs $\Delta\chi_j^2$

El valor de  $\Delta\chi_j^2$  mide cuánto se reduce el valor del estadístico  $\chi^2$  obtenido del modelo original, al eliminar al elemento con patrón de covariables  $j$ . Una manera de medir la influencia del  $j$ -ésimo patrón de covariables sobre el ajuste completo es por medio de esta medida, de tal forma que conforme en valor de estos elementos incrementa, su impacto sobre el modelo se considera más significativa. Su aproximación se presenta a continuación:

$$\Delta\chi_j^2 = \frac{d_{Pj}^2}{1 - h_j} = d_{PEj}^2$$

La gráfica se muestra en la Figura 4.11, de tal forma que se consideran significativos a un nivel de significancia  $\alpha = 0.05$  aquellos elementos cuyo valor supera el valor de 4, pues cuando  $m_j$  es grande la distribución de  $\Delta\chi_j^2$  se aproxima a una Ji-cuadrada de un grado de libertad y  $\chi_{0.95}^2(1) = 3.84$  (Hosmer & Lemeshow).

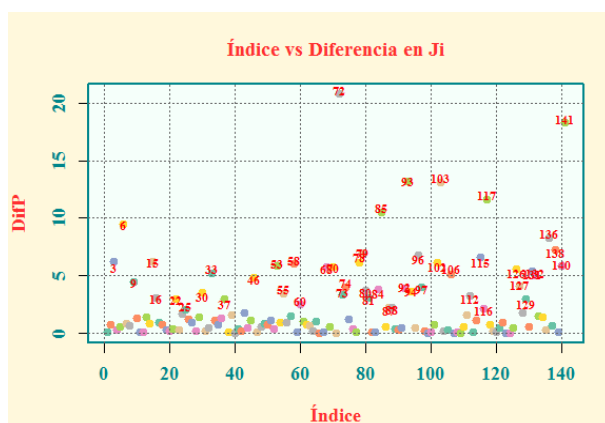


Figura 4.11: Gráfica de diagnóstico del comportamiento del valor de  $\Delta\chi_j^2$ .

## Probabilidades estimadas vs $\Delta\chi_j^2$

Cuando el número de patrones de covariables  $J$  se aproxima al valor de  $n$ , se puede graficar a  $\hat{\pi}_j$  contra el valor de  $\Delta\chi_j^2$ . Por lo tanto, esta gráfica puede usarse en el caso en que se ajusta un modelo de tipo desagregado, observándose dos funciones cuadráticas correspondientes a los casos en que  $y_j=1$  y  $y_j=0$ .

Aquellos puntos cuya gráfica es decreciente corresponden a los sujetos cuya respuesta fue  $y_j=1$ , y aquellos cuya gráfica es creciente corresponden a los sujetos cuya respuesta fue  $y_j=0$ , considerando mal ajustados aquellos puntos ubicados en los extremos superiores de la gráfica agrupada.

La gráfica se muestra en la Figura 4.12, y observamos que aquellos individuos cuya respuesta fue  $y_i=1$  tuvieron mayores dificultades en ser ajustados, pues los puntos más extremos se encuentran en la esquina superior izquierda.

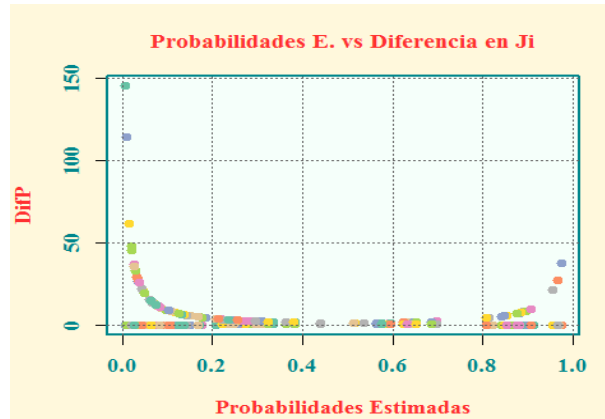


Figura 4.12: Gráfica de diagnóstico del comportamiento de  $\hat{\pi}_j$  contra  $\Delta\chi^2$ .

Por otro lado, existe otra medida que puede ser intercambiada en las dos últimas gráficas de diagnóstico, y cuya interpretación es análoga. Ésta mide el cambio en el valor del estadístico de la devianza y su aproximación está dada de la siguiente manera:

$$\Delta D_j^2 = d_{D_j}^2 + \frac{d_{P_j}^2 h_j}{1 - h_j} = \frac{d_{D_j}^2}{1 - h_j}$$

Como se muestra en la Figura 4.13, los valores de  $\Delta D_j^2$  en general son menores a los reportados por  $\Delta\chi_j^2$ .

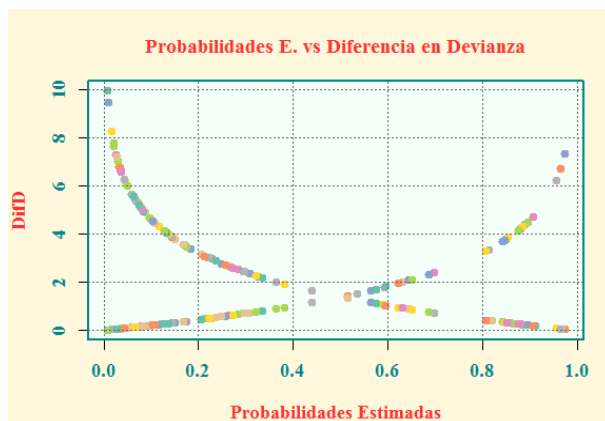


Figura 4.13: Gráfica de diagnóstico del comportamiento de  $\hat{\pi}_j$  contra  $\Delta D_j^2$ .

### Índice vs Distancia de Cook:

Un elemento de la base de datos se dice que es influyente si el valor final del modelo ajustado cambia de manera considerable si este es incluido o no, y localizarlos es importante pues deben ser analizados cuidadosamente y así determinar si estos datos son legítimos o son el producto de un mal registro. Una de las medidas propuestas para localizar a estos elementos es la distancia de Cook, la cual se obtiene estandarizando la diferencia entre el vector de parámetros estimados obtenido originalmente y el vector de parámetros estimados excluyendo el elemento con el patrón de covariable  $j$ . Para el caso de regresión logística se define como sigue:

$$\begin{aligned} \Delta\beta_j &= (\hat{\beta}_j - \hat{\beta}_{(-j)})'(X'VX)(\hat{\beta}_j - \hat{\beta}_{(-j)}) \\ &= \frac{d_P h_j^2}{1 - h_j} \\ &= \frac{d_P E_j h_j}{1 - h_j} \end{aligned}$$

Este valor mide principalmente el efecto que tiene un patrón de covariables específico sobre los parámetros finales del modelo de manera simultánea, y se consideran influyentes aquellos cuyo valor en su distancia de Cook es cercana al valor de 1 o mayor. La gráfica para el modelo se muestra en la Figura 4.14:

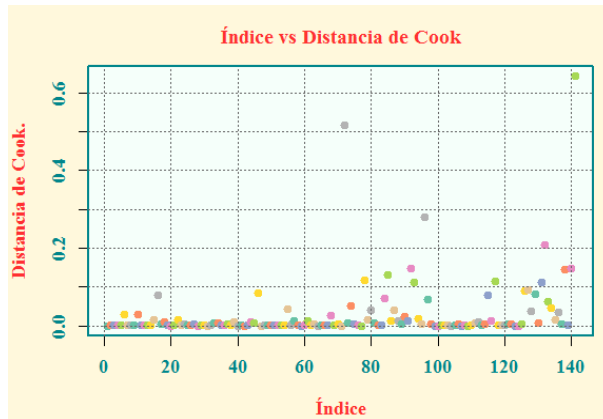


Figura 4.14: Gráfica de diagnóstico del comportamiento en la distancia de Cook.

## 4.9. Ligas Probit y Complementaria log-log

### Probit

Una justificación del modelo probit surge de la distribución de tolerancia empleada en toxicología para medir el efecto de dosificación de determinada toxina que un individuo puede soportar antes de la muerte. Sea  $T$  la distribución de tolerancia de dicho sujeto, que un sujeto muera dado un nivel de dosificación  $x$  es equivalente a el valor de la función de distribución de  $T$  evaluada en  $x$ . Es decir:

$$\pi(x) = P(Y = 1|X = x) = P(T \leq x)$$

Así, conocemos como modelo probit a aquel cuya distribución de tolerancia es una normal estándar  $N(0, 1)$ , de tal forma que:

$$\pi(\bar{x}) = \Phi(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \beta_{p+1} x_{p+1})$$

Donde:

- $\Phi$  denota la función de distribución de una normal estándar  $N(0, 1)$ .



Por lo que la función liga corresponde a la función inversa de la distribución de una normal estándar  $\Phi^{-1}$ , pues como se mencionó anteriormente es aquella que une al componente sistemático con el componente aleatorio de tal forma que la relación entre ambos sea lineal.

## Complementaria log-log

Las funciones logit y probit son funciones simétricas alrededor del 0.5, lo que hace que igualmente la curva de la respuesta sea simétrica alrededor del 0.5. Lo anteriormente señalado significa que la curva de la respuesta se aproxime a 0 a la misma tasa en que se aproxima a 1, por lo que la función complementaria log-log surge como alternativa a las dos anteriormente citadas en aquellos casos en que las probabilidades estimadas no se comporten de dicha manera.

La curva de la respuesta se caracteriza por aproximarse a cero muy lentamente, y acercarse a uno más bruscamente, y está dada de la siguiente manera:

$$\pi(\bar{x}) = 1 - \exp(-\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \beta_{p+1} x_{p+1}))$$

Por lo que la función liga queda determinada por  $\log[-\log(1 - \pi(x))]$ .

## Resultados

Hinkley (1985) propone una prueba para corroborar si la función liga es adecuada, la cual puede ser usada tanto para datos agregados como desagregados. Ésta consiste en crear una nueva variable definida como el cuadrado del predictor lineal, posteriormente se agrega y evalúa si es significativa en el modelo estudiado, y en caso de ser afirmativo, se tiene evidencia para creer que la función liga está mal especificada. El procedimiento se presenta a continuación. Al obtener un p-value de 0.6501193, no tenemos evidencia suficiente para pensar que la función liga es inadecuada.

```
VarExt<-predict(Mod5_Desag)
VarExt<-VarExt^2
ModeloLink<-glm(Muerte~Nivel_Esc+Ingresos+Estado_Civil +ZonaR_U
                +Lugar_Def+VarExt,family=binomial(logit))
DifDEv<-deviance(Mod5_Desag)-deviance(ModeloLink)
GradosL<-df.residual(Mod5_Desag)-df.residual(ModeloLink)
pvalue<-pchisq(deviance(Mod5_Desag)-deviance(ModeloLink),
               df.residual(Mod5_Desag)-df.residual(ModeloLink), lower=F)
pvalue
```

Además de la prueba anterior, el Cuadro 4.11 presenta el valor de las medidas de MacFadden y Cox y Snell para las tres diferentes funciones logit, además del valor del área bajo la curva de ROC, siendo el caso en que se ajustó la función logit aquel que obtuvo los mejores resultados.

Comparación entre funciones logit			
Función Liga	MacFadden	Cox y Snell	AUC ROC
Logit	0.3835	0.2468	0.8812
Probit	0.382	0.2459	0.8813
C.log-log	0.3763	0.2427	0.8805

Cuadro 4.11: Tabla que compara los resultados de distintas medidas obtenidas entre las diferentes funciones ligas.

Aunque la función logit fue la liga que obtuvo los mejores resultados, la diferencia entre éstos no es muy grande, por lo que al no rechazar que la liga logit sea inadecuada en la prueba propuesta anteriormente, no implica que otras ligas no puedan ajustar también al modelo. Es decisión del investigador y de sus propósitos particulares elegir entre varias funciones ligas que ajusten.

# Conclusión

A partir del modelo se concluye que entre las mujeres de bajos recursos, el grupo que presenta menos ataques de esta naturaleza lo caracterizan aquellas cuyo nivel de escolaridad es universitario o más. Así pues, verificamos nuevamente que el contar con un nivel de escolaridad superior, a pesar de que no necesariamente puede garantizar un buen nivel de vida, si te permite tener mejores herramientas que te ayudan a combatir este tipo de violencia.

Además, considerando el nivel de ingresos, estado civil y tipo de zona donde se reside, se observa que las mujeres menos violentadas son aquellas cuyo nivel de ingresos es medio, su estado conyugal es separada, divorciada o viuda, y cuya zona de residencia es rural. Por lo tanto, aunque no se puede determinar con exactitud quienes fueron los perpetradores del homicidio, si se observa que las mujeres que no contaban con una pareja al momento de la muerte presentaron menores probabilidades de ser asesinadas, lo que apunta y refuerza la idea de que una gran proporción de estos asesinatos se desarrollan dentro del entorno de la violencia en pareja.

En cuanto al sitio donde ocurrió el asesinato, vemos que en la mayoría de los casos las mujeres no alcanzan a llegar con vida al hospital, lo que hace sospechar de la naturaleza de violencia extrema que acompañan estos asesinatos, además de que contrario a la idea de que la violencia ocurre más frecuentemente en localidades y pueblos pequeños, debido a una falta de acceso a educación, vemos más bien que esta discriminación se presenta principalmente por patrones socioculturales y estereotipos, y que se hacen presentes independientemente del mayor acceso a la información que tienen las personas que viven en zonas urbanas.

Por otra parte, el análisis hecho a partir de la ENDIREH concluye que los familiares no directos son los principales abusadores dentro de la violencia dentro de la comunidad, la cual como se señaló anteriormente se analiza excluyendo la violencia dentro de la relación en pareja. Además, la violencia con connotaciones sexuales, que incluye insinuaciones, propuestas, y amenazas tanto directas como no directas, se desarrolla principalmente fuera del hogar y es cometida principalmente por desconocidos.

Sin embargo, el abuso sexual propiamente hablando se presenta casi en igual proporción tanto fuera como dentro del hogar, y cuando ocurre dentro del hogar es llevado a cabo principalmente por familiares no directos, seguido de amigos y vecinos.

En cuanto a la violencia dentro de la relación de pareja, observamos que cuando esta ocurre no suele ser ocasional, pues cuando la mujer afirma haber sido víctima y se le pregunta la frecuencia con que ocurre, una mayor proporción afirmó que ocurre de manera frecuente. Además, cuando se le pregunta lo grave que consideran que es la agresión, un dato interesante es que en casi todos los escenarios afirmaron que era muy grave, lo que indica que aún estando consientes de su situación, no son capaces de detener la violencia que sufren, ya sea por aceptación y razones socioculturales que ellas mismas terminaron por aceptar, o por el poco apoyo y atención que reciben al momento de denunciar o buscar apoyo.

Por lo tanto, entre las principales medidas para detener los asesinatos violentos, y en general cualquier otro tipo de violencia en contra de las mujeres, están el promover la educación en las mujeres a niveles superiores, pues esto les proporciona libertad económica y herramientas que las vuelven menos vulnerables en contra de cualquier tipo de abuso. Además de que se requiere una campaña de concientización que promuevan el respeto principalmente dentro del hogar, enfocada no solamente a la violencia que ocurre dentro de la relación de pareja, pues se observó que en general los principales abusadores fuera de ella son los propios familiares y personas cercanas a la mujer.

# Anexo I: Código de graficas de análisis diagnóstico

## 1.1 Gráfica de Índice vs residuos

```
par(bg="cornsilk", mar=c(4.2,4,3,1))
ResiduosD_E<-rstandard(Mod5_Agreg,type="deviance")
signif <- which(abs(ResiduosD_E) > 2)
plot(ResiduosD_E, xlab="", ylab="", xaxt="n", yaxt="n")
rect(-7, -7, 180, 7, col="mintcream")
par(new=TRUE)
plot(ResiduosD_E,ylim=c(-6,6),xlab="Índice", ylab="Residuos Devianza E.",
     main="Índice vs Residuos",col=brewer.pal(8,"Set2"), pch=19, lwd=2,
     font.main=7,font.lab=7,font.axis=7, cex.axis=1.1, cex.lab=1.1,
     cex.main=1.2, col.main="firebrick1",col.lab="firebrick1",
     col.axis="turquoise4")
text(signif, jitter(ResiduosD_E[signif], amount=0.5),
     label = signif, cex = .7, font=7, col="red")
box(lwd = 2, col="turquoise4");grid(col="grey40")
abline(h=(c(-2,2)), col="red", lwd=1.2, lty=2)
```

## 1.2 Gráfica de Predictor lineal vs Residuos: Modelo agregado

```
Predictor_A<-predict(Mod5_Agreg)
Predictor_A<-Predictor_A[-c(19,20,21,25,26,27,32,38,57,74,109,110,
                          111,113,114,115,116,117,121,122,125)]
ResiduosD_E<-rstandard(Mod5_Agreg,type="deviance")
par(bg="cornsilk", mar=c(4.2,4,3.5,1.5))
plot(Predictor_A,ResiduosD_E, xlab="", ylab="", xaxt="n", yaxt="n")
```

```

rect(-7, -7, 180, 7, col="mintcream")
par(new=TRUE)
Pred<-Predictor_A[signif <- which(abs(ResiduosD_E) > 2)]
plot(Predictor_A,ResiduosD_E,ylim=c(-4,4),xlab="Predictor Lineal",
      ylab="Residuos Devianza E.",main="Predictor Lineal vs Residuos",
      col=brewer.pal(8,"Set2"), pch=19, lwd=2, font.main=7,font.lab=7,
      font.axis=7, cex.axis=1.1, cex.lab=1.1, cex.main=1.2,
      col.main="firebrick1",col.lab="firebrick1", col.axis="turquoise4")
text(Pred, jitter(ResiduosD_E[signif], amount=0.5),
      label = signif, cex = .7, font=7, col="red")
box(lwd = 2, col="turquoise4"); grid(col="grey40")
abline(h=(c(-2,2)), col="red", lwd=1.2, lty=2)

```

### 1.3 Gráfica de Predictor lineal vs Residuos: Modelo desagregado

```

ResiduosD_E<-rstandard(Mod5_Desag,type="deviance")
Predictor_A<-predict(Mod5_Desag)
par(bg="cornsilk", mar=c(4.2,4,3.5,1.5))
plot(Predictor_A,ResiduosD_E, xlab="", ylab="", xaxt="n", yaxt="n")
rect(-7, -7, 180, 7, col="mintcream")
par(new=TRUE)
plot(Predictor_A,ResiduosD_E,xlab="Predictor Lineal",
      ylab="Residuos Devianza E.", col.main="firebrick1",
      main="Predictor Lineal vs Residuos",
      col=brewer.pal(8,"Set2"), pch=19, lwd=2, font.main=7,font.lab=7,
      font.axis=7, cex.axis=1.1, cex.lab=1.1, cex.main=1.2,
      col.lab="firebrick1", col.axis="turquoise4")
box(lwd = 2, col="turquoise4"); grid(col="grey40")

```

### 1.4 Gráfica de Probabilidades estimadas vs Probabilidades reales

```

ProporcionR<-Datos_Agrup$Si/(Datos_Agrup$Si+Datos_Agrup$No)
Predictor_A<-fitted.values(Mod5_Agreg)
par(bg="cornsilk", mar=c(4.2,4,3.5,1.5))
plot(Predictor_A,ProporcionR, xlab="", ylab="", xaxt="n", yaxt="n")
rect(-7, -7, 180, 7, col="mintcream")
par(new=TRUE)

```

```

plot(Predictor_A,ProporcionR,xlab="Proporciones Estimadas",
     ylab="Proporciones Reales",
     main="Proporciones Estimadas vs Proporciones Reales",
     col=brewer.pal(8,"Set2"), pch=19,lwd=2, font.main=7,
     font.lab=7,font.axis=7, cex.axis=1.1, cex.lab=1.1, cex.main=1.2,
     col.main="firebrick1",col.lab="firebrick1", col.axis="turquoise4")
box(lwd = 2, col="turquoise4")
grid(col="grey40")

```

## 1.5 Gráfica de Variables predictoras vs Residuos

```

par(font.main=7,font.lab=7,font.axis=7, cex.axis=1.1,
     cex.lab=1.1, cex.main=1.2, col.main="firebrick1",
     col.lab="firebrick1", col.axis="turquoise4", grid(col="grey40"),
     bg="white")
residualPlots(Modelo5_A, type = "deviance", cex = 0.6, tests = TRUE,
              col=brewer.pal(12,"Set2"), border=brewer.pal(12,"Set2"),
              pch=1.3)

```

## 1.6 Gráfica de Índice vs $\Delta\chi_j^2$

```

par(bg="cornsilk", mar=c(4.2,4,3.5,1.5))
RPE<-rstandard(Mod5_Agreg,type="pearson")
DifP<-(RPE^2)
Signif <- c(which(abs(DifP) > 2))
plot(DifP, xlab="", ylab="",xaxt="n", yaxt="n" )
rect(-7, -7, 165, 90, col="mintcream")
par(new=TRUE)
plot(DifP, xlab="Índice", ylab="DifP",
     main="Índice vs Diferencia en Ji",
     col=brewer.pal(8,"Set2"), pch=19, lwd=2, font.main=7,font.lab=7,
     font.axis=7, cex.axis=1.1, cex.lab=1.1, cex.main=1.2,
     col.main="firebrick1",col.lab="firebrick1", col.axis="turquoise4")
text(signif, jitter(DifP[signif], amount=0.5),
     label = signif, cex = .7, font=7, col="red")
box(lwd = 2, col="turquoise4")
grid(col="grey40")

```

## 1.7 Gráfica de Probabilidades estimadas vs $\Delta\chi_j^2$

```
RPE<-rstandard(Mod5_Desag,type="pearson")
DifP<-(RPE^2)
Predictor_A<-fitted.values(Mod5_Desag)
par(bg="cornsilk", mar=c(4.2,4,3.5,1.5))
plot(Predictor_A,DifP, xlab="", ylab="",xaxt="n", yaxt="n" )
rect(-7, -7, 180, 160, col="mintcream")
par(new=TRUE)
plot(Predictor_A,DifP, xlab="Probabilidades Estimadas", ylab="DifP",
     col=brewer.pal(8,"Set2"), pch=19, lwd=2, font.main=7,font.lab=7,
     main="Probabilidades E. vs Diferencia en Ji",
     font.axis=7, cex.axis=1.1, cex.lab=1.1, cex.main=1.2,
     col.main="firebrick1",col.lab="firebrick1", col.axis="turquoise4")
box(lwd = 2, col="turquoise4")
grid(col="grey40")
```

## 1.8 Gráfica de Probabilidades estimadas vs $\Delta D_j^2$

```
Hat<-hatvalues(Modelo5)
RD<-residuals(Modelo5,type="deviance")
RPE<-rstandard(Modelo5,type="pearson")
DifD<-(RD^2)+(RPE^2)*Hat
Predictor_A<-fitted.values(Modelo5)
par(bg="cornsilk", mar=c(4.2,4,3.5,1.5))
plot(Predictor_A,DifD, xlab="", ylab="",xaxt="n", yaxt="n" )
rect(-7, -7, 180, 160, col="mintcream")
par(new=TRUE)
plot(Predictor_A,DifD, xlab="Probabilidades Estimadas",
     ylab="DifD", col=brewer.pal(8,"Set2"),pch=19,lwd=2,
     font.main=7,font.lab=7,
     main="Probabilidades E. vs Diferencia en Devianza",
     font.axis=7, cex.axis=1.1, cex.lab=1.1, cex.main=1.2,
     col.main="firebrick1",col.lab="firebrick1", col.axis="turquoise4")
box(lwd = 2, col="turquoise4")
grid(col="grey40")
```



# Anexo II: Graficas de Poder predictivo

## 2.1 Código de gráfica que maximiza sensibilidad y especificidad

```
Pred<-prediction(fitted.values(Mod5_Desag),Muerte)
Perf1.1<-performance(Pred,"sens")
Perf1.2<-performance(Pred,"spec")

par(bg="cornsilk", mar=c(4.2,4,3,1), xaxt="n", yaxt="n", lwd=1.5)
plot(c(seq(0,1,0.1)), type="n", xaxt="n", yaxt="n", ylab="", xlab="")
rect(-1.5, -1.5, 1.5, 1.5, col="mintcream")
par(new=TRUE)
plot(Perf1.1, xlab="", ylab="", col="olivedrab3", lwd=2)
par(new=TRUE)
plot(Perf1.2,xlab="Puntos de Corte",ylab="", col="tan1", font.lab=7,
     col.lab="firebrick1", lwd=2)
legend(.71, .98,legend=c("Sen.", "Esp."), lty=c(1,1),
      col=c("olivedrab3", "tan1"), cex=0.7, lwd=2)
box(lwd = 2, col="turquoise4"); grid(col="grey40")
mtext(seq(0,1,0.2), side=2, las=1,at=seq(0,1,0.2), line=0.2,
      col="turquoise4", font=7)
mtext(seq(0,1,0.2), side=1, las=1,at=seq(0,1,0.2), line=0.2,
      col="turquoise4", font=7)
title("Punto de Corte Óptimo", font.main=7, cex.main=1.1,
      col.main="firebrick1")
```

# Bibliografía

Agresti, A. (2002). *Categorical Data Analysis*. Wiley, 2a edición.

AGOEIA (1995). Convención interamericana para prevenir, sancionar y erradicar la violencia contra la mujer “Convención de Belem do Para”. Belem, Brasil: Asamblea General de la Organización de Estados Americanos.

Angel, A. (2017, Mayo 10). En 12 estados no se investiga como feminicidio el asesinato de una mujer a manos de su pareja. *Animal Político*. Consultado en línea en <http://www.animalpolitico.com/2017/05/feminicidio-delito-tipificado-estados/>

Botello, B.E. (2016, Mayo 4). Falta de acceso a la salud y seguridad social, principales causas de empobrecimiento: Meade. *Cronica*. Consultado en línea en <http://www.cronica.com.mx/notas/2016/958980.html>

Castañas, R.J. (2013). *Regresión logística. Tratamiento computacional con R*. Tesis no publicada de Maestría. Universidad de Granada, España.

Ciani, I. (2015). Ponencia “El feminicidio en México”, auditorio Mario de la Cueva, UNAM. Consultado en línea en <https://www.youtube.com/watch?v=DpwtqakUEBc>.

Collett David. (2003). *Modelling binary data*. Chapman Hall/CRC, 2a edición.

Dobson, A. J. (2002). *An introduction to generalized linear models*. Chapman Hall/CRC, 2a edición.

Faraway, J. (2006). *Extending the linear model with r. Generalized linear, mixed effects and nonparametric regression models*. Chapman Hall/CRC, 1a edición.

Hinkley, D. V. (1989). *Transformation diagnostics for linear models*. Biometrika.

Hosmer, D.W., S. Lemeshow (2000). *Applied Logistic Regression*, Wiley, 2a edición.

INEGI (2011). Encuesta Nacional sobre la Dinámica de las Relaciones en los Hogares 2011. Consultado en línea en <http://www.beta.inegi.org.mx/proyectos/enchogares/especiales/endireh/2011/default.html>

INEGI (2015). Estadísticas Vitales. Defunciones Generales y Fetales 2015. Consultado en línea en <http://www.beta.inegi.org.mx/proyectos/registros/vitales/mortalidad/default.html?init=2>

INMUJERES(2017). Alerta de Violencia de Género contra las Mujeres (Acciones y Programas). Consultado en línea en: <https://www.gob.mx/inmujeres/acciones-y-programas/alerta-de-violencia-de-genero-contra-las-mujeres-80739>.

Kleinbaum, D. G., Klein M.(2010) Logistic Regression. A Self-Learning Text. Springer, 3a edición.

LGAMVLV (2007). Ley General de Acceso de las Mujeres a una Vida libre de Violencia. Ciudad de México, México: Diario Oficial de la Federación.

McCullagh, P., J. Nelder. (1989). Generalized linear models. Chapman Hall, London New York, 2a edición.

Lagarde, Y. M. (2008). Antropología, feminismo y política: Violencia feminicida y derechos humanos de las mujeres. DF, México: Congreso de antropología.

Lindsey, J. K. (1997) Applying generalized linear models. Springer, 1a edición.

Montgomery, Douglas C., Peck, Elizabeth A. (1992). Introduction to Linear Regression Analysis. John Wiley Sons, 2a edición.

Olamendi, P. (2016). Femicidio en México. DF, México: INMUJERES.

ONU (1981). Convención sobre la eliminación de todas las formas de discriminación contra la mujer. Nueva York, EUA: Organización de las Naciones Unidas.

ONU (2011). Femicidio en México, aproximación, tendencias y cambios 1985-2009. DF, México: ONU mujeres; INMujeres, Instituto Nacional de las Mujeres.

ONU MUJERES(2011). Convención sobre la Eliminación de todas formas de Discriminación contra la Mujer (CEDAW) . Consultado en línea en <http://mexico.unwomen.org/es/digiteca/publicaciones/2011/12/cedaw>.

ONU MUJERES (2017)<sup>A</sup>. Comisión de la Condición Jurídica y Social de la Mujer. Consultado en línea en <http://www.unwomen.org/es/csw>.

ONU MUJERES(2017)<sup>B</sup>. Conferencias mundiales sobre la mujer. Consultado en línea en <http://www.unwomen.org/es/how-we-work/intergovernmental-support/world-conferences-on-women>.

Ronald Christensen (1990). Log-linear models and logistic regression. Springer, 2a edición.

SEGOB (2016). La violencia Feminicida en México, aproximaciones y tendencias 1985-2014. DF, México: SEGOB; INMUJERES ONU Mujeres

Soto, E. A. (2014, Noviembre 20). Por mala tipificación, el feminicidio en México no se investiga. Cimac Noticias. Consultado en línea en <http://www.cimacnoticias.com.mx/node/68200>

Toledo, V.P. (2009). Feminicidio. DF, México: Oficina en México del Alto Comisionado; Naciones Unidas Derechos Humanos.