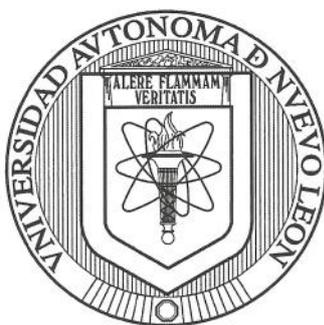


**UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN
FACULTAD DE ECONOMÍA
DIVISION DE ESTUDIOS DE POSGRADO**



**LOS ACCIDENTES VIALES EN MONTERREY
Y SUS DETERMINANTES**

Por

RODRIGO CRESPO ELIZONDO

**Tesis presentada como requisito parcial para
obtener el grado de Maestría en Economía con
Especialidad en Economía Industrial**

MAYO, 2008

Índice

I. Introducción.....	1
II. Antecedentes.....	7
II.I. Los accidentes y las compañías aseguradoras.....	7
II.II. Pronósticos de accidentes viales y análisis de temporalidad.....	8
II.III. Análisis de impacto de políticas públicas.....	9
II.IV. Los accidentes viales, sus determinantes y sus probabilidades.....	10
II.V. Otras metodologías e investigaciones.....	11
III. Hipótesis.....	13
IV. Marco teórico.....	14
IV.I. Modelos con variables dependiente dicotómica	15
IV.II. Estimaciones <i>robustas</i>	16
IV.III. El método <i>bootstrap</i>	19
V. Información.....	22
V.I. Las variables.....	23
V.II. Estadísticas descriptivas de la muestra	25
VI. Resultados.....	28
VI.I. Interpretación de resultados.....	33
VI.II. Comparación de resultados utilizando los modelos logit y probit.....	36
VI.III. Propiedades de las estimaciones.....	37
VI.III.I. Multicolinealidad.....	37
VI.III.II. Heteroscedasticidad.....	37
VI.III.IV. Sensibilidad de los resultados a cambios en el tamaño de la muestra.....	40
VI.IV. Distribución de probabilidades de acuerdo a la edad de los conductores	44
VII. Conclusiones.....	48

VII.I. Limitantes y futuras mejoras.....	49
VIII. Referencias.....	51
IX. Anexos	
IX.I. Anexo 1: Nota periodística: Son un caos estadísticas municipales.....	55
IX.II. Anexo 2: Encuesta.....	56
IX.III. Anexo 3: Matriz de correlación.....	57

Índice de cuadros.

Cuadro 1.1. Accidentes totales en el AMM.....	2
Cuadro 1.2. Accidentes totales por tipo de vehículo para el año 2005.....	3
Cuadro 1.3. Distribución de accidentes por mes.....	3
Cuadro 1.4. Distribución de accidentes por horario, para el año 2005.....	4
Cuadro 1.5. Accidentes por edad, para el año 2005 y autor particulares.....	4
Cuadro 1.6. Accidentes por género, para el año 2005 y autor particulares.....	4
Cuadro 3.1: Distribución de probabilidades del modelo MLP.....	14
Cuadro 5.1. Estadísticas descriptivas de la muestra.....	26
Cuadro 5.2. Frecuencia de variables dicotómicas.....	27
Cuadro 6.1. Ejemplo de nueve observaciones sin ponderar.....	28
Cuadro 6.2. Ejemplo de nueve observaciones ponderando por frecuencia.....	29
Cuadro 6.3. Resultados de los modelos Logit.....	32
Cuadro 6.4. Resultados del modelo (2) a niveles, momios e impacto de las variables dependientes en el cambio de probabilidades de sufrir un accidente vial.....	34
Cuadro 6.5 Resultados del modelo (2) por logit y probit.....	36
Cuadro 6.6. Resultados del modelo (2) utilizando estimaciones robustas.....	39
Cuadro 6.7. Resultados de los modelos logit inicial y final, utilizando estimaciones robustas.....	40
Cuadro 6.8. Resultados del <i>bootstrap</i> del modelo (2), sin ponderar, utilizando 500 y 1,000 repeticiones, y un tamaño de muestra $n = 300$	42
Cuadro 6.9. Resultados del <i>bootstrap</i> del modelo (2), sin ponderar, utilizando 500 y 1,000 repeticiones, y un tamaño de muestra $n = 200$	43
Cuadro 6.10. Resumen de los resultados de las pruebas “t” para comparar los coeficientes de las iteraciones y del modelo original.....	44
Cuadro 6.11. Resultados de los modelos utilizando edad al cuadrado	46
Cuadro 6.12 Correlación entre edad y edad al cuadrado.	46

Cuadro 6.13. Resultado de modelos utilizando variables de engaño por rangos de edad.....	47
--	----

Índice de gráficos

Gráfico 4.1 Funciones de densidad normal y logística.....	17
Gráfico 4.2 Distribuciones acumulativas logit y probit.....	18
Gráfico 6.1. Distribución logística de probabilidades de sufrir un accidente, por consumo de alcohol, de acuerdo a la edad.....	45

Los accidentes viales en Monterrey y sus determinantes.

Rodrigo Crespo Elizondo

Aprobación de la Tesis:

Asesor de la Tesis



DR. DANIEL FLORES CURIEL



DR. JULIO CÉSAR ARTEAGA GARCÍA



DR. ERICK RANGEL GONZÁLEZ



DR. JULIO CÉSAR ARTEAGA GARCÍA
Director de la División de Estudios de Posgrado
de la Facultad de Economía, UANL
Mayo, 2008

Agradecimientos

Quiero agradecer el constante y valioso apoyo del Dr. Daniel Flores. Su paciencia, generosidad y clarividencia fueron factores fundamentales para la conclusión de esta investigación. Al Dr. Julio Cesar Arteaga por su apoyo durante toda la maestría (incluso desde antes de comenzarla). Al Dr. Erick Rangel por sus comentarios, recomendaciones y puntos de vista siempre atinados. Al resto de la planta docente con quien tuve el gusto de convivir y aprender de ellos.

Además, quiero agradecer al CONACYT por el apoyo que me brindó durante el tiempo en el que cursé la maestría. A Magda Roque por su ayuda y amistad.

Finalmente, quiero agradecer a mi familia por su apoyo incondicional.

Rodrigo Crespo E.

I. Introducción

Los accidentes viales representan una pérdida en el bienestar social por diferentes motivos. Entre otros se pueden mencionar los siguientes: pérdida de capital humano en el caso de los accidentes fatales, pérdida material para las partes involucradas, pérdida en términos de tiempo, pérdida en la productividad laboral de quienes requieren reposo y un proceso de tratamiento, pérdida de infraestructura y el costo de tratamiento y hospitalización en que incurren los involucrados. Por ello, es importante analizar este tema desde el punto de vista económico.

En el área metropolitana de Monterrey (AMM), el tema de los accidentes viales se ha posicionado como un asunto prioritario y de gran relevancia tanto en la opinión pública, como en los medios de comunicación y en los mismos gobiernos estatal y municipal. Ejemplo de ello son las organizaciones no gubernamentales y las campañas en los medios de comunicación dedicadas a la concienciación de este problema. Estas organizaciones han sido creadas recientemente y cada vez tienen mayor presencia a nivel estatal, como es el caso de No a Conducir Ebrios (NACE) fundada en el año 2000.

El tema, a pesar de ser cada vez más relevante, carece de un análisis estadístico robusto. Actualmente, el análisis de este problema, en el mejor de los casos, se reduce al uso de la estadística descriptiva. Inclusive, hasta antes de la creación del Consejo Estatal de Vialidad y Tránsito (CETYV), no existía una entidad de gobierno que manejara la información estadística de accidentes viales en los municipios, sino que era cada una de las administraciones de los municipios la que creaba la información de manera aislada, desarticulada y, además, no se difundía ni se hacía pública.

En la actualidad, las administraciones municipales siguen siendo las encargadas de generar la información estadística de accidentes viales, reportando posteriormente la información al CETYV para su publicación. Sin embargo, en el sistema de creación y agregación de la información estadística se pueden presentar diversos problemas de distintas índoles como: manipulación de las estadísticas por cuestiones de evaluación del desempeño, falta de conocimiento y especialización por parte de los encargados de la recolección de la información y creación de bases de datos, rezagos en su publicación, limitaciones en la calidad de la información en términos de los registros de las características de los choques, entre otros.

Entre los problemas más graves que se encuentran en la generación de esta información se halla el hecho de que cada uno de los municipios utiliza criterios distintos para el control estadístico de los accidentes, los choques relacionados al alcohol, las sanciones, etc. Esta situación ha sido denunciada en la prensa local¹. Además, únicamente hay datos de los accidentes reportados ante las autoridades.

Los datos generados bajo la situación previamente descrita pueden causar un sesgo en los estimadores independientemente de la metodología estadística. Por esta razón, es preferible la generación de información a través de la aplicación de encuestas. Sin embargo, este tipo de información tiene dos fuertes inconvenientes: su generación es tardada y es costosa. De cualquier forma es prudente analizar la información oficial existente. Asimismo, la generación de información resulta doblemente importante para el análisis de este tema: por un lado complementa la información oficial y, por otro lado, puede sustituirla cuando la información oficial no es confiable.

Como se puede apreciar en el cuadro 1.1, en los últimos años² se han experimentado alrededor de 70,000 accidentes viales anuales en el AMM. Además, del total de accidentes viales, alrededor del 20% de los casos genera lesiones. En el caso de las muertes por los accidentes viales, en los últimos años las cantidades varían alrededor de 100 a 120 casos.

Cuadro 1.1. Accidentes totales en el AMM

Año	Accidentes totales	Con lesionados	% del total	Accidentes con muertes	% del total
2002	69,779	14,233	20.4	122	0.17
2003	72,253	13,909	19.25	131	0.18
2004	71,043	15,091	21.24	103	0.14
2005	57,451	10,464	18.21	101	0.18

Fuente: CETYV (2006)

Con respecto a los accidentes por tipo de vehículos, en el cuadro 1.2 se observa que para la información acumulada de enero a septiembre de 2005, después de los automóviles particulares (con cerca de 45,000 accidentes; 7,000 lesionados y 76 muertos), los taxis representan los vehículos con mayor accidentes, con un total de 1,140.

¹ Ver en el anexo 1 la nota: "Son un caos estadísticas municipales" por Guerrero, Devra. Periódico El Norte, el 28 de mayo de 2007.

² La información presentada por el CETYV muestra únicamente el período a enero de 2002 a agosto de 2005.

Cuadro 1.2. Accidentes totales por tipo de vehículo para el año 2005

Tipo de automóvil	Accidentes	Lesionados	Muertos
Particular	44,893	7,102	76
Taxi	4,157	1,140	5
Transporte urbano	2,611	997	10
Transporte de carga	2,806	307	5
Otros	2,984	918	5

Fuente: CETYV (2006)

Con respecto a la distribución de los accidentes por mes, como se puede apreciar en el cuadro 1.3, octubre es el mes en el que se presentan más accidentes tanto de forma global como en promedio, seguido por los meses de mayo y marzo.

Cuadro 1.3. Distribución de accidentes por mes

Mes	Año			Total	Media mensual
	2003	2004	2005		
Enero	5,656	5,024	5,877	16,557	5,519.00
Febrero	6,214	5,411	5,610	17,235	5,745.00
Marzo	6,809	6,481	5,266	18,556	6,185.33
Abril	5,881	5,239	5,865	16,985	5,661.67
Mayo	6,627	6,034	6,161	18,822	6,274.00
Junio	6,436	5,962	5,532	17,930	5,976.67
Julio	5,860	6,003	5,420	17,283	5,761.00
Agosto	6,071	6,199	5,713	17,983	5,994.33
Septiembre	5,756	6,074	5,918	17,748	5,916.00
Octubre	6,217	6,552	6,089	18,858	6,286.00
Noviembre	5,417	6,307	-	11,724	5,862.00
Diciembre	5,309	5,757	-	11,066	5,533.00
Total	72,253	71,043	57,451	200,747	5,904.32
Media anual	6,021.08	5,920.25	5,745.10	66,915.67	

Fuente: CETYV (2006)

Con respecto a la distribución de los accidentes viales por hora, en el cuadro 1.4 se pueden observar una mayor cantidad por la tarde, específicamente en el rango de 14:00 a 18:00 horas. Sin embargo, estas estadísticas únicamente capturan los percances sin hacer distinción entre sus magnitudes, por lo que contar con información acerca de los accidentes fatales serviría para dimensionar este tipo de accidentes en los distintos horarios.

Cuadro 1.4. Distribución de accidentes por horario, para el año 2005

Rango de horas	Año	
	2004	2005
2 - 6 AM	3,480	3,097
6 - 10 AM	10,004	9,590
10 - 2 PM	12,177	11,860
2 - 6 PM	14,512	14,038
6 - 10 PM	12,491	12,331
10 - 2 AM	6,699	6,093
Sin hora	50	28

Fuente: CETYV (2006)

La edad es uno de los determinantes de la cantidad de accidentes viales, de acuerdo con Islam y Mannering (2007) y Stutts, Huang y Hunter (2002), entre otros. En el cuadro 1.5 se puede observar que, para el año 2005, los automovilistas jóvenes tuvieron la mayor cantidad de choques en el AMM para automóviles particulares, especialmente en el rango de 26 a 35. De cualquier forma, se debe considerar que este último grupo de automovilistas representa un número mayor de conductores con respecto al rango inferior de edad.

Cuadro 1.5. Accidentes por edad del responsable y autos particulares, 2005

Rangos de edad	Accidentes	Heridos	Accidentes graves		
			Accidentes	% Total	Muertos
Hasta 17	941	215	136	14.45	3
De 18 a 25	9,161	1,742	1,141	12.45	19
De 26 a 35	13,155	2,168	1,423	10.82	21
De 36 a 45	8,814	993	665	7.54	5
De 46 a 55	3,601	510	351	9.75	2
De 56 o más	2,183	321	222	10.17	6
No se sabe	7,038	1,153	924	13.13	20
Total	44,893	7,102	4,862	10.83	76

Fuente: CETYV (2006)

Con respecto al género, en el cuadro 1.6 se observa para el año 2005 una considerablemente mayor proporción de accidentes, heridos y muertos causados por automovilistas masculinos; sin embargo, se debe considerar que la proporción de automovilistas masculinos también es más alta.

Cuadro 1.6. Accidentes por género del responsable y autos particulares, 2005

Género	Accidentes	Heridos	Muertos
Masculino	35,547	5,502	64
Femenino	5,908	1,057	2
Sin dato	3,438	543	10
Total	44,893	7,102	76

Fuente: CETYV (2006)

Como se mencionó anteriormente, no es la cantidad de información oficial la que puede conllevar un sesgo en su uso para la aplicación de métodos estadísticos, sino la calidad de la misma. Por esta razón se realiza un proceso de generación de información con base en una encuesta aplicada telefónicamente a 401 individuos.

A partir de dicha información se utiliza un modelo logit para calcular las probabilidades de sufrir, durante un año, al menos un accidente vial³ en el AMM, así como para probar la significancia estadística de las variables a considerar, en donde se abarcan las características de los individuos (sexo, edad, estado civil, nivel de educación), los hábitos de conducir (a través de variables proxis, como el uso del cinturón de seguridad), los hábitos de consumo de alcohol (independientemente de si conducen cuando manejan, como una conservadora aproximación), y las características de los automóviles (color, año, marca, entre otros). Posteriormente se utilizan distintas metodologías para analizar la robustez del modelo y la sensibilidad del mismo a cambios en el tamaño de la muestra.

La evidencia empírica muestra que el utilizar el teléfono celular mientras se conduce un automóvil aumenta las probabilidades de sufrir un accidente vial, controlando las demás variables. Lo mismo se encuentra cuando se trata de un conductor que se define como bebedor fuerte. Asimismo, la evidencia muestra que, controlando todas las demás variables, no existe una diferencia estadísticamente significativa en los conductores según su género. Adicionalmente se encuentra que la edad de los conductores tiene una relación significativa y negativa con las probabilidades de sufrir un accidente vial. Asimismo, cuando un conductor conduce un automóvil propio aumentan sus probabilidades de sufrir un accidente vial, con respecto a quienes no conducen un auto propio. Y finalmente, se encuentra que los coches marca Nissan muestran una probabilidad mayor de sufrir un accidente vial, controlando las demás variables.

La presente investigación se estructura de la siguiente manera. En el capítulo dos se establecen los antecedentes; en el capítulo tres se plantean las hipótesis. En el capítulo cuatro se hace una revisión del marco teórico. En el capítulo cinco se explica la información que se utilizará, la forma y la metodología que se emplea para obtenerla. En

³ Por sufrir un accidente vial se entiende el hecho de que un conductor se vea envuelto en un accidente, independientemente de que sea el responsable o no.

el capítulo seis se muestran los resultados. Y en el capítulo siete se establecen las conclusiones.

II. Antecedentes

La literatura relacionada con los accidentes viales se puede dividir en dos grandes vertientes: los accidentes generales y los accidentes fatales. Además, se pueden agrupar en grandes líneas de investigación: los accidentes y las compañías aseguradoras; el análisis de impacto de políticas públicas; la política fiscal, su efecto en el consumo de alcohol y en los accidentes viales; los accidentes viales, sus determinantes y probabilidades; pronósticos de accidentes viales; entre otros. A continuación se revisa alguna de la literatura relevante por línea de investigación

II.1. Los accidentes y las compañías aseguradoras

Artís, Ayuso y Guillén (2002) analizan el comportamiento fraudulento de los asegurados por medio de reclamos bajo circunstancias “extrañas”. Utilizando información sobre los reportes o reclamos de seguros de automóviles en España, los autores demuestran el desempeño de los modelos de opción binaria para la detección de fraudes e implementan un modelo para corregir la clasificación errónea en respuesta a variables sobre el conductor (edad, sexo, si cuenta con licencia y su historial de accidentes viales), sobre el accidente (si sucedió de noche, si hubo testigos, la zona en que sucedió, si se llamó a la policía), entre otros. La utilidad de este modelo para las empresas aseguradoras es que permite estimar la proporción de reclamos fraudulentos que no son detectados; para el caso español concluyen que un 5% de los reclamos que son clasificados como honestos en realidad son fraudulentos.

Suponcic y Tennyson (1995) analizan el impacto de la regulación sobre las primas de las aseguradoras de automóviles en la estructura de tal mercado. Los autores establecen que la regulación restrictiva lo que provocan son desincentivos a la entrada al mercado por parte de posibles compañías aseguradoras, distorsiones en la cantidad ofrecida, disminuyendo la cantidad de compañías de bajos costos en el mercado y, finalmente, en una reducción del bienestar de los consumidores, contrario a lo que se busca en un principio con este tipo de políticas. El análisis se basa en regresiones de corte transversal

del número de firmas y las proporciones de mercado de diferentes características organizacionales.

Por otro lado, Dionne y Vanasse (1992) analizan la metodología que las compañías aseguradoras utilizan para asignar la prima de riesgo a los individuos: primeramente mediante una aproximación con base en una regresión de mínimos cuadrados ordinarios con variables relevantes del individuo, obteniendo así la prima base (PM). Después, utilizan un factor bonus-malus (FBM) que tasa el historial del conductor (premia cuando el conductor registra un año sin choque, y castiga cuando choca), y que se basa en computar factores sobre una distribución de Poisson. Los autores prueban con modelos “count data” que al utilizar modelos MCO, los cuales se basan en distribuciones normales, y factores *bonus-malus*, que a su vez utilizan distribuciones Poisson, se generan estimadores inconsistentes.

Spahar y Escolas (1982), utilizando información del año 1979 para la ciudad de Wyoming, analizan las características de los conductores y su diferencia estadística entre diferentes grupos, a través de pruebas Z, con el fin de avalar o rechazar el uso, por parte de las compañías aseguradoras, de las variables “sexo”, “edad” y “zona de residencia” como reflejos del riesgo potencial de los individuos, controlando la severidad y la frecuencia de los accidentes. Los hallazgos soportan la aproximación de las compañías aseguradoras a través de las variables descritas.

II.II Pronósticos de accidentes viales y análisis de temporalidad

Reaside (2004) analiza las tendencias de los accidentes viales fatales para el caso de Gran Bretaña; asimismo, desarrolla un modelo para el pronóstico de este tipo de accidentes, a través de modelos autoregresivos, integrados y de medias móviles (ARIMA). Estos modelos los utiliza para comparar las metas del gobierno con los pronósticos, concluyendo que las probabilidades de que se alcancen las metas trazadas son bajas.

Un artículo publicado por London: Her Majesty’s Stationery Office⁴ utiliza un modelo ARIMA para analizar los efectos de la temporabilidad en la serie de accidentes viales en el Reino Unido, concluyendo que los efectos más adversos en los accidentes fatales ocurren en los meses de noviembre y diciembre.

⁴ Citado por Harvey (1989).

II.III Análisis del impacto de políticas públicas

Cook, Ostermann y Sloan (2005) utilizan la metodología de datos de panel, con información de los últimos 30 años, para analizar los efectos de los impuestos sobre el alcohol en las tasas de mortalidad. Los resultados muestran que en el corto plazo los impuestos sobre el alcohol causan una reducción sobre la cantidad de alcohol consumida, así como de la mortalidad. Sin embargo, en el largo plazo los efectos de la reducción en el consumo de alcohol sobre la mortalidad son nulos.

Cohen y Aivan (2003) analizan el impacto de las políticas de uso del cinturón de seguridad por ley, tanto en su uso, como en la cantidad de accidentes viales fatales en los estados de EE-UU. Los autores reportan que esta política impacta estadística y negativamente al número de accidentes fatales; por un aumento en el uso del cinturón de seguridad del 1%, se estima que se salven 136 vidas al año. Por otro lado, los autores encuentran que esta política aumenta el número de conductores que utiliza el cinturón del 68% al 77%.

Por otro lado, Rhum (1995) analiza, bajo una metodología de datos de panel, los efectos de una serie de políticas públicas en los accidentes fatales en carretera. Utilizando información de 1982 a 1988, así como de cada uno de los estados de EE-UU, encuentra que la edad legal para consumir alcohol impacta fuerte, negativa y significativamente a los accidentes fatales de jóvenes entre 18 y 20 años; sin embargo, no sucede lo mismo con el resto de los automovilistas. Asimismo, la política más efectiva, de acuerdo con Rhum, son los impuestos al alcohol. Otras de las variables que demuestran ser estadísticamente significativas son el ingreso personal y la tasa de desempleo.

Chaloupka, Saffer y Grossman (1993) utilizan información anual (1982–1988) para 48 estados de los EE-UU (datos de panel), para analizar los accidentes fatales en función del ingreso real, velocidad, edad de los conductores, leyes de cinturón de seguridad obligatorio, tasa de desempleo, religión del conductor, edad mínima para el consumo de alcohol, impuestos al alcohol, y políticas de penalización a los conductores ebrios como cárcel a los conductores ebrios y servicio social. Los autores utilizan dos muestras: para los conductores en edades de 18 a 20 y para los conductores en general. Los resultados indican que las políticas más efectivas contra los accidentes fatales son los impuestos al alcohol y la acción legal en contra de conductores ebrios.

II.IV. Los accidentes viales, sus determinantes y probabilidades

Loureiro, Sachside y Moreira (2004) desarrollan un modelo probit para analizar las probabilidades de sufrir un accidente vial en la ciudad de Brasilia. Aplicaron una encuesta a 1,455 individuos en centros comerciales. Los autores demuestran que las variables que utilizaron como aproximaciones a las distracciones, como el uso del teléfono celular o fumar mientras conducen, son factores estadísticamente significantes en los accidentes vehiculares. Sin embargo, el proceso de muestreo está lejos de ser el óptimo.

Golob y Recker (2002) utilizan modelos multivariados no lineales para determinar el tipo de accidente en las avenidas más transitadas del sur de California. Los autores obtienen información estadística que demuestra que el tipo de accidente está fuertemente vinculado con la velocidad mediana del tráfico, así como por las variaciones temporales en la velocidad en los carriles interiores e izquierdos. Además, la evidencia empírica demuestra que los accidentes de percances múltiples son principalmente ocasionados por maniobras de movimientos de un carril a otro, y ocurren con más frecuencia en pavimentos mojados.

Levitt y Porter (1999) desarrollan un modelo basado en el teorema de Bayes para determinar la probabilidad de sufrir un accidente fatal. Los autores encuentran evidencia estadísticamente significativa que sugiere que los automovilistas con alcohol en las venas son por lo menos ocho veces más propensos a sufrir un accidente fatal. Al mismo tiempo, los conductores ebrios presentan una probabilidad de sufrir un accidente fatal al menos 15 veces mayor que los conductores sobrios. Además, los conductores masculinos, jóvenes y con un historial desfavorable sobre conducción, son también más propensos a chocar, aunque en una proporción menor que aquellos que conducen con alcohol en las venas.

Saffer (1997) prueba la posible endogeneidad entre los accidentes fatales y la publicidad de bebidas alcohólicas a través de un modelo de dos ecuaciones: la primera representa las probabilidades de sufrir un accidente fatal y la segunda la demanda por publicidad, bajo la metodología de mínimos cuadrados en dos etapas (2LS). El autor utiliza ocho distintas regresiones, en donde en las primeras cuatro incluye variables independientes que miden las condiciones demográficas y económicas; mientras que en las otras utiliza únicamente la variable de publicidad, el precio del alcohol, el tamaño de la ciudad (dummy), variables regionales (dummy) y variables de tiempo (dummy). El autor

encuentra que, para el caso de los EE-UU, la publicidad sí es una variable positiva y significativamente relacionada a los accidentes fatales. Sin embargo, encuentra que el precio del alcohol es el determinante más importante de los accidentes fatales, con los cuales mantiene una relación positiva.

II.V. Otras metodologías e investigaciones

Algunos investigadores utilizan otro tipo de metodologías para abordar este fenómeno urbano, que en muchas ocasiones se reduce al uso de simulaciones con el fin de obtener información y utilizarla para llegar a conclusiones a través del uso de la estadística inferencial.

Por ejemplo, la organización no gubernamental estadounidense AAA Foundation for Traffic Safety's destinada a desarrollar investigación, proponer políticas públicas y educar y concienciar a los estadounidenses en asuntos de tráfico y tránsito, ha desarrollado numerosos estudios basados en simulaciones. Las simulaciones que llevan a cabo con hardwares y softwares permiten que los investigadores sometan a los conductores a pruebas virtuales de manejo, bajo condiciones establecidas y de las cuales se logra obtener una gran cantidad de información. En otras palabras, se trata de pruebas experimentales, en donde los conductores observan un video y se espera que respondan a él, moviendo el volante y controles que simulan a los de un automóvil.

McKnight y Mcknight (1991) desarrollan un análisis estadístico utilizando información derivada de simulaciones. Las simulaciones constan en el análisis de 151 sujetos participantes, los cuales observan un video de manejo de 25 minutos en un simulador virtual. En el video se presentan 47 situaciones en donde se espera que los conductores respondan y ejerzan un movimiento en los controles. Los hallazgos del estudio demuestran que todas las distracciones aumentan de manera significativa tanto las situaciones en las cuales el conductor puede fallar al responder y el tiempo de respuesta del mismo. Asimismo, encuentran que cuando utilizan el teléfono celular, las probabilidades de sufrir un accidente aumentan de acuerdo con intensidad de la conversación: si se trata de una plática casual, la distracción se asemeja a la que se presenta cuando el conductor entabla una conversación con algún pasajero. Sin embargo, cuando las llamadas requieren mayor concentración, como es el caso de conversaciones

de negocios, las probabilidad de chocar aumentan. Además, las probabilidades aumentan también entre los adultos mayores que utilizan el teléfono celular.

Un punto criticable de estas metodologías es que la selección de los participantes no permite que se represente a la población general, y que, por lo tanto, las conclusiones no se puedan generalizar. Sin embargo, permite recaudar información más exacta como el tipo de llamadas que se realizan, los segundos que tardan los conductores en reaccionar ante situaciones de peligro, entre otras.

Haigney y Taylor (1998) también utilizan simulaciones para analizar la incidencia del uso del teléfono celular sobre la probabilidad de sufrir accidentes viales. Sin embargo, únicamente sugieren que el uso del celular genera variaciones en la conducta, las cuales son asociadas a mayores riesgos.

III. Hipótesis

Las hipótesis que se plantean en la presente investigación son:

1. Los hábitos de manejo tienen un impacto significativo en las probabilidades de sufrir un accidente vial. Los hábitos negativos se resumen en distracciones que aumentan las probabilidades de sufrir un accidente vial en el Área Metropolitana de Monterrey.
2. Los hábitos de consumo de alcohol son determinantes de las probabilidades de sufrir un accidente vial. Por lo tanto, una persona abstemia se espera que tenga menores probabilidades de sufrir un accidente vial con respecto a los bebedores medios y fuertes.
3. Controlando las demás variables, existe una diferencia significativa en las probabilidades de sufrir un accidente vial de acuerdo con el género del conductor.

IV. Marco teórico

IV. I. Modelos con variables dependientes dicotómicas.

De acuerdo con Gujarati (2003), el análisis econométrico de variables dependientes dicotómicas se ha abordado de tres formas distintas: bajo el modelo lineal de probabilidad (MLP), el modelo logit y el modelo probit.

El modelo MLP cuya variable dependiente, como ya se mencionó, es dicotómica, se puede interpretar como una probabilidad condicional, es decir: $E(Y=1 | X)$. Así, un modelo MLP sencillo se podría expresar de la siguiente forma:

$$E(Y=1 | X_1) = \beta_1 + \beta_2 X_1$$

De forma tal que la variable dependiente enfrenta una distribución de probabilidades Bernoulli, como se muestra en el cuadro 3.1.

Cuadro 3.1: Distribución de probabilidades del modelo MLP

Y_t	Probabilidad
0	$1-P_t$
1	P_t
Total	1

Fuente: Gujarati (2005)

De aquí que: $E(Y_i) = 0(1-P_t) + 1(P_t) = P_t$, y por lo tanto la ecuación inicial se puede expresar de la siguiente manera:

$$E(Y=1 | X_1) = \beta_1 + \beta_2 X_1 = P_t$$

De esta forma, la esperanza condicionada de que el valor esperado de Y puede ser interpretado como la probabilidad condicionada de Y.

Los problemas asociados a los modelos MPL son los siguientes:

- No normalidad en el término de error: dada la distribución de la variable dependiente, el término de error también sigue una distribución Bernoulli. Este problema puede ser ignorado cuando se cuenta con muestras grandes, ya que cuando eso sucede los estimadores de Mínimos Cuadrados Ordinarios (MCO) seguirán una distribución normal. Además, bajo este modelo, las estimaciones MCO parecen ser insesgadas.

- Heteroscedasticidad. Los estimadores MCO bajo este modelo presentarán el problema de no ser insesgados, sin embargo, este problema puede ser tratado como cualquier otro modelo MCO con este problema.
- No se cumple la restricción de que la probabilidad del valor condicional de Y se encuentre entre cero y uno. Se presenta el problema de encontrar probabilidades negativas o mayores a uno, lo cual contradice la teoría estadística.
- R² no es una buena medida de bondad de ajuste. Dada la naturaleza dicotómica de la variable dependiente, es muy probable que el valor de la R² sea muy inferior a uno.

Los modelos logit y probit no enfrentan el problema de probabilidades menores a cero y mayores a uno, y por lo tanto, normalmente son más utilizados. Estos modelos difieren únicamente por la función de distribución acumulativa que se utiliza para definir las probabilidades.

El caso del modelo logit supone una función de distribución logística. Parte del modelo de MLP, sin embargo, ahora la probabilidad condicionada será de la siguiente forma:

$$P_t = E(Y=1 | X_t) = 1 / [1 + e^{- (\beta_1 + \beta_2 X_t)}]$$

Que puede re escribirse de la siguiente forma:

$$P_t = E(Y=1 | X_t) = 1 / [1 + e^{-Z_t}] = e^{Z_t} / (1 + e^{Z_t})$$

Expresión conocida como la función de distribución logística. Donde:

$$Z_t = (\beta_1 + \beta_2 X_t)$$

De igual forma:

$$1 - P_t = 1 / 1 + e^{Z_t}$$

Y por lo tanto,

$$(P_t / 1 - P_t) = (1 + e^{Z_t}) / (1 + e^{-Z_t}) = e^{Z_t}$$

Por lo que la parte izquierda es simplemente una razón de probabilidades.

Por lo tanto, si Y representa la probabilidad de que una persona sufra un accidente vial con respecto a que no lo sufra, y encontramos que P_i = 0.2, quiere decir que las probabilidades son una a cinco.

Es importante notar que si al modelo de razones de probabilidades se le aplica logaritmo natural, se convierte en lineal tanto en la razón de probabilidades, como en los parámetros. A un modelo así se le conoce como L (logit):

$$L = \ln(P_i/1 - P_i) = Z_i = \beta_1 + \beta_2 X_1$$

Para analizar cómo las probabilidades P_i se ven afectadas por alternativas o características de los individuos en el caso de los modelos de probabilidad lineal se considera que:

$$P_i = \mathbf{X}'\boldsymbol{\beta} = \beta_1 + \beta_2 X_{i2} + \dots + \beta_k X_k$$

$$dP_i / dX_{ik} = \beta_k$$

Para el caso tanto de los modelos logit como los probit, el cambio en las probabilidades está dado por:

$$dP_i / dX_{ik} = f(\mathbf{X}'\boldsymbol{\beta}_k) \beta_k$$

Donde $f(\mathbf{X}'\boldsymbol{\beta}_k)$ es la función de densidad de probabilidad para cada caso.

Este modelo presenta las siguientes características:

- El logit va de menos infinito a infinito a medida que P va de cero a uno.
- L es lineal en X , sin embargo, las probabilidades por sí mismas no lo son.
- L puede estar en función de cualquier cantidad de variables explicativas.
- Si el valor de L es positivo, quiere decir que las probabilidades de que $Y = 1$, aumentan conforme las variables independientes se incrementan, y viceversa.
- La constante del modelo es el valor del logaritmo de las probabilidades a favor que $Y = 1$. El coeficiente de la variable independiente mide el cambio en L ocasionado por un cambio unitario en X , es decir, explica cómo el logaritmo natural de las probabilidades a favor de que $Y = 1$ cambia a medida que X cambia en una unidad.

Por otro lado, el modelo probit se basa en una función de distribución acumulativa (FDA) normal; por tal motivo a este modelo también se le conoce como normit. En este caso:

$$E(Y=1 | X_1) = P(I^* < I) = P(Z_i < \beta_1 + \beta_2 X_1) = F(\beta_1 + \beta_2 X_1)$$

Donde $E(Y=1 | X_1)$ significa la probabilidad de que Y suceda dado los valores de X ; I^* es un nivel crítico tal que si I excede a I^* , entonces ocurrirá $Y = 1$.

Así, para obtener información sobre I , lo mismo que para los parámetros, se toma la inversa de $F(\beta_1 + \beta_2 X_1)$ para obtener: $I_i = \beta_1 + \beta_2 X_1$

Si suponemos que al correr el modelo anterior encontramos que la constante es igual a -1.01, y que el coeficiente de la variable explicativa (ingreso, como ejemplo) es 0.04, y que

el nivel que se quiere medir en el ingreso es 6 mil pesos, para interpretar los resultados en términos de la probabilidad de que $Y_i = 1$ (sufra un accidente) ante cambios en una unidad del ingreso, requerimos analizar la tasa de cambio de la probabilidad de que $Y = 1$, ante cambios en x :

$$dP_i / dX_i = f(\beta_1 + \beta_2 X_i) \beta_2$$

Por lo que es necesario encontrar la función de densidad normal en $f[-1.01 + (0.04)(0.06)] = f(-0.72)$. Se requiere hacer referencia a las tablas de distribución normal, donde se encuentra que para $Z = -0.72$, la densidad normal es 0.30, valor que se multiplica por el coeficiente de la variable explicativa ($0.30 * 0.04 = 0.014$), con lo cual encontramos que para un nivel de ingresos de seis mil pesos, si el ingreso aumenta en mil pesos, la probabilidad de que el individuo choque aumenta en un 1.4%.

En suma, la única diferencia entre los modelos logit y probit es que utilizan diferentes funciones: probit utiliza la función de distribución normal, logit utiliza la función logística. Sin embargo, ante la similitud de ambas metodologías surge una pregunta de manera natural: ¿cuál de los dos es preferible? De acuerdo a Gujarati (2003), los modelos son muy semejantes para la mayoría de las aplicaciones, y la única diferencia radica en que la distribución logística tiene extremos más anchos, asemejándose a una distribución t con pocos grados de libertad, como se muestra en la gráfica 4.1. Lo mismo se puede observar en las distribuciones acumulativas de los modelos probit y logit que se muestran en la gráfica 4.2.

Gráfico 4.1. Funciones de densidad normal y logística

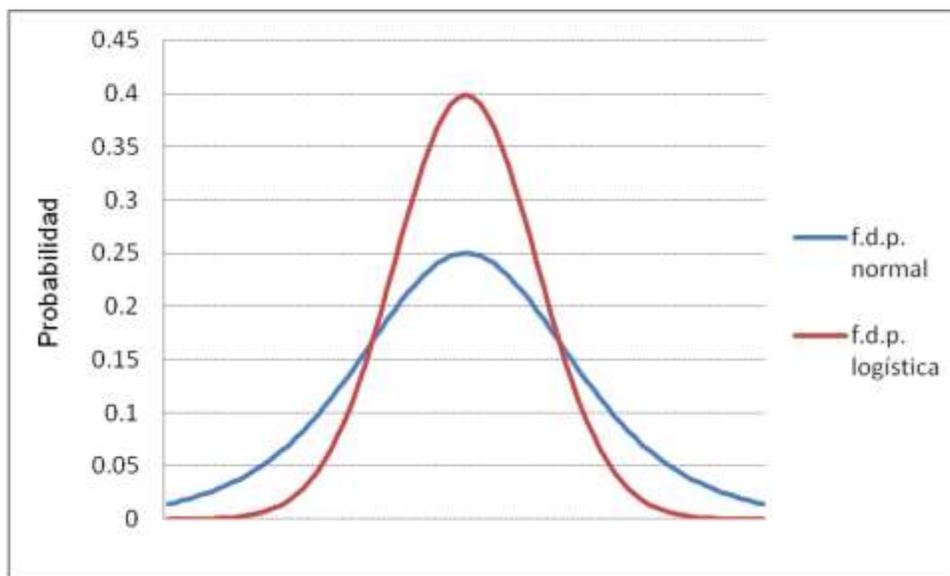
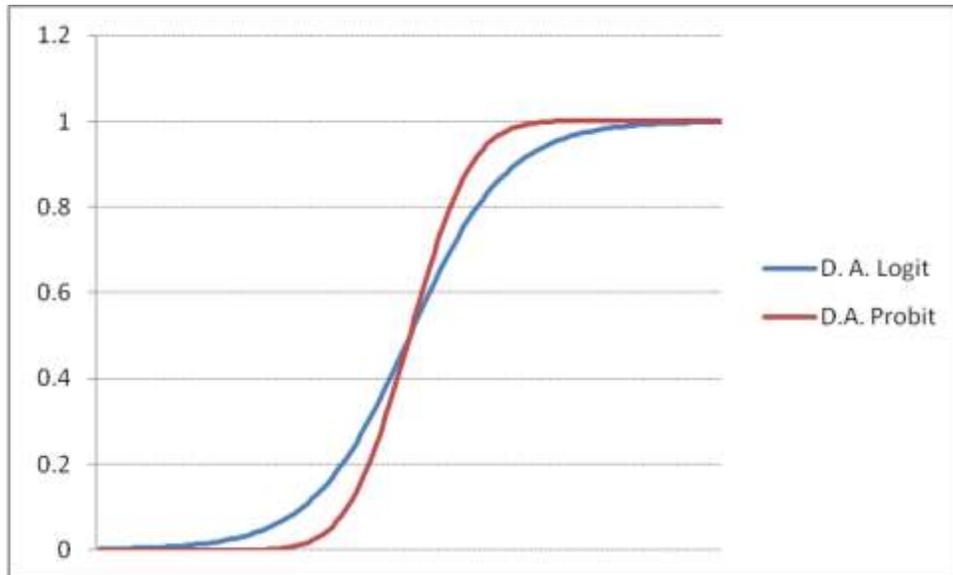


Gráfico 4.2. Distribuciones acumulativas logit y probit



IV.II. Estimaciones robustas.

Las regresiones en general pueden utilizar algunas metodologías complementarias para mejorar sus resultados. Las herramientas pueden estar concentradas en el trato de las observaciones o en la forma de ejercer la inferencia. Por ejemplo, se pueden ponderar las observaciones de la base de datos, lo que permite otorgarle pesos distintos a diferentes casos de acuerdo a una variable que sirve como criterio de ponderación⁵, modificando así los resultados de las estimaciones. Por otro lado, se pueden utilizar los comandos y metodologías como “*robusto*” o “*svy*” para lograr estimaciones más eficientes.

En la presente investigación se utiliza la metodología “*robusto*” con el objetivo de contrastar los resultados obtenidos. “Las estimaciones robustas surgen de un cálculo distinto de la varianza de los estimadores. Para su ilustración, supongamos la regresión por mínimos cuadrados ordinarios, cuyos coeficientes estimados son:

$$\beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Donde \mathbf{y} es un vector ($n \times 1$) y constituye la variable dependiente del modelo. \mathbf{X} a su vez constituye un vector de orden ($n \times k$), y está constituido por las variables independientes. Puesto que todo está considerado condicional de \mathbf{X} , entonces $(\mathbf{X}'\mathbf{X})^{-1}$ puede ser tratado como una matriz constante. Por lo tanto, la varianza de β es:

$$V(\beta) = (\mathbf{X}'\mathbf{X})^{-1}V(\mathbf{X}'\mathbf{y}) (\mathbf{X}'\mathbf{X})^{-1}$$

⁵ En el capítulo 6 se aborda el tema de las ponderaciones.

La varianza del primer elemento de $\mathbf{X}'\mathbf{y}$ es:

$$\mathbf{X}'_1\mathbf{y} = x_{1,1}y_1 + x_{2,1}y_2 + \dots + x_{n,1}y_n$$

Considerando de nuevo que \mathbf{X} se puede tratar como constante, y suponiendo independencia entre las y_i , se puede escribir la varianza de la siguiente forma:

$$V(\mathbf{X}'\mathbf{y}) = x_{1,1}^2 V(y_1) + x_{2,1}^2 V(y_2) + \dots + x_{n,1}^2 V(y_n)$$

Recordando que la estimación de $V(y_i)$ es el cuadrado del residual, e_j^2 , que se obtiene de la siguiente forma: $e_j = y_j - x_j\beta$; y estimando la diagonal de la matriz de covarianza para $\mathbf{X}'\mathbf{y}$, se encuentra que:

$$V(\mathbf{X}'\mathbf{y}) = \sum_{j=1} e_j^2 x'_j x_j$$

Y por lo tanto:

$$V(\beta) = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{j=1} e_j^2 x'_j x_j \right) (\mathbf{X}'\mathbf{X})^{-1}$$

Nótese que dado que el factor $(\sum_{j=1} e_j^2 x'_j x_j)$ se encuentra en medio de la expresión del lado derecho, al estimador de varianza robusta también se le llama estimador *sándwich*⁶.

IV.III. El método *bootstrap*.

De acuerdo con Mooney y Duval (1993) el *bootstrap* es una técnica de inferencia no paramétrica, intensiva computacionalmente. *Bootstrap* difiere del acercamiento de inferencia paramétrica tradicional en que emplea numerosos procesos computacionales para estimar la forma de la distribución de muestreo de un estadístico, en lugar de establecer fuertes supuestos acerca de la distribución y de fórmulas analíticas

“El *bootstrap* se basa en una analogía entre la muestra y la población de la cual fue extraída. La idea central es que algunas veces resulta mejor realizar conclusiones acerca de las características de la población estrictamente de la muestra que se tiene a la mano, en vez de establecer supuestos probablemente equivocados sobre la población”⁷.

La diferencia entre esta técnica estadística y las simulaciones de *Monte Carlo* es que en el caso de la última, se debe conocer la distribución que sigue el fenómeno que se analiza, lo cual en muchos casos no es posible; mientras que el *bootstrap* permite al

⁶ STATA Press (2005).

⁷ Ibid.

investigador realizar inferencia estadística sin la necesidad de conocer la distribución de la población.

Con respecto a *jackknife*, otra metodología no paramétrica similar al *bootstrap*, esta herramienta también evalúa la variabilidad de un estadístico examinando la variación dentro de un grupo de información, en vez de utilizar los supuestos de la inferencia paramétrica. El *jackknife* es una técnica menos general que el *bootstrap*, y explora la variación de la muestra en una forma distinta...“si el lema del *bootstrap* es el “muestreo con reemplazo”, el lema del *jackknife* es “la exclusión”. Es decir, el estadístico de *jackknife* es desarrollado a través de la exclusión sistemática conjuntos de la información, de observación en observación, y evaluando la variación del estimador resultante”⁸

“Como el *bootstrap*, el *jackknife* ha sido utilizado principalmente como una herramienta inferencial alternativa útil cuando los supuestos de la inferencia paramétrica tradicional no aplican. De cualquier forma, se ha demostrado que, a diferencia del *bootstrap*, el *jackknife* falla para estadísticos no marcadamente no lineales tales como la mediana muestral (Efron, 1982, p.9). Sin embargo aunque el *bootstrap* supera al *jackknife* en su generalidad debido a la calidad, existe evidencia que muestra que el *jackknife* puede ser superior al *bootstrap* en el área de muestreo complejos... (Fay, 1985)”⁹

Los pasos básicos que se realizan en el *bootstrap* se explican a continuación:

1. “Se construye una distribución de probabilidad empíricamente, $F(x)$, de una muestra otorgándoles una probabilidad de $1/n$ a cada punto, x_1, x_2, \dots, x_n . Esta es la función de distribución empírica (FDE) de x , la cual es el estimador no paramétrico de máxima verosimilitud, $F(X)$ (Roghati, 1984, pp. 234-236).
2. De la FDE, $F(x)$, se define un tamaño de muestra aleatoria simple con reemplazos. Este es el re-muestreo, X_b^* .
3. Se calcula el estadístico de interés, θ_b^* .
4. Se repiten los pasos 2 y 3 B veces, donde B es un número grande. La magnitud práctica de B depende de la prueba que se llevará a cabo. Típicamente, B se considera que sea entre 50 a 200 para estimar el error estándar de θ , y menos de 1,000 para estimar los intervalos de confianza alrededor de θ (Efron y Tibshirani, 1986, sec.9).

⁸ Ibid.

⁹ Ibid.

5. Se construye una distribución probabilística de las $B\theta_b^*$, estableciendo una probabilidad de $1/B$ a cada punto, $\theta_1^*, \theta_2^*, \dots, \theta_B^*$. Esta distribución constituye el estimador "bootstrapped" de la distribución de muestreo de θ : $F^*(\theta^*)$... Esta distribución puede ser utilizada para realizar inferencia acerca de θ .

V. Información

La información se obtuvo de un proceso de aplicación de encuestas. El cuestionario se diseñó con el fin de que la entrevista fuera corta, fácil de contestar y no pidiera información que el entrevistador difícilmente accediera a dar. El cuestionario se muestra en el anexo 2.

El tamaño de la muestra se definió en 401 observaciones, considerando que el límite de error máximo es de $\pm 5\%$ con un nivel de confianza del 95%, para una población grande. Se realizó un muestreo simple sin remplazo donde el directorio telefónico de los hogares se definió como marco muestral.

La metodología para la aplicación de la encuesta se explica a continuación. (1) Se seleccionaron números telefónicos de forma aleatoria, a través del directorio telefónico. (2) Se realiza la llamada y se establece el "rapport" o primer acercamiento por parte del entrevistador, identificándose, mencionando el propósito de la llamada, y el uso académico y confidencial que se le da a la información recaudada. (3) Se aplica la encuesta a la persona que contesta el teléfono; en caso de que quien conteste no conduzca algún automóvil, se pide hablar con algún miembro del hogar que sí conduzca. En caso de que nadie conduzca en ese hogar, o se nieguen a contestar, concluye sin éxito la entrevista telefónica. (4) Se aplica el cuestionario. (5) Se despide el entrevistador agradeciendo la colaboración.

Además, con el objetivo de buscar una misma probabilidad de que cualquier miembro del hogar conteste el teléfono, la entrevista se llevó a cabo los fines de semana, en un horario de 9:00 a 19:00. El periodo de levantamiento de información comprendió los meses de marzo y abril de 2006.

Las variables explicativas que se utilizan son las características de las personas (sexo, edad, estado civil, menor nivel educativo terminado e ingreso); hábitos de manejo (horas manejadas a la semana, fumar mientras maneja, uso del cinturón de seguridad, cambiar constantemente de música, escuchar el estéreo a un volumen alto, conducir con niños en el automóvil y hablar por teléfono); hábitos de consumo de alcohol (medido en copas que consume a la semana); características del automóvil que normalmente conduce (si es automóvil propio, el color del coche que conduce, la marca del coche, el año del modelo

del automóvil y si cuenta con seguro); la hora en que sucedió el choque, la magnitud del mismo; así como si había llovido al momento del incidente¹⁰.

Es importante considerar el apunte que Dufour (1999) hace acerca de las encuestas para calcular el consumo de alcohol de las personas. Dado que se basan en información reportada por los mismos entrevistados, el estudio está sujeto a errores, tanto intencionales como no intencionales, por parte del entrevistado. Aunque esta situación se presenta en cualquier encuesta, el riesgo de contar con este tipo de sesgos o errores aumenta cuando se trata de temas de adicciones, como las drogas o el alcohol.

V.I Las variables

Las variables que finalmente se utilizaron para el análisis econométrico se describen a continuación:

<i>Accidente</i>	Variable dicotómica que toma el valor de uno si el conductor sufrió durante el último año un accidente vial y cero en caso contrario. Esta es la variable dependiente del modelo.
<i>Edad</i>	Es una de las tres variables no dicotómicas utilizadas. Captura la edad del conductor.
<i>Sexo</i>	Variable de engaño o “dummy”, que captura el género del individuo. Toma el valor de uno cuando el automovilista es hombre y cero cuando es mujer.
<i>Soltero</i>	Variable de engaño que toma el valor de uno cuando el conductor es soltero, y cero cuando no lo es. Como se puede apreciar en el cuadro 4.1 esta variable es utilizada en el modelo uno como referencia.
<i>Hasta Secundaria</i>	Variable dicotómica que agrupa a los individuos que a lo mucho cuenten con nivel educativo de secundaria.
<i>Técnica</i>	Esta variable agrupa a los individuos que su mayor nivel educativo terminado es la educación técnica.
<i>Preparatoria</i>	Variable dummy cuyo valor es de uno cuando el mayor nivel educativo del individuo es preparatoria.
<i>Profesional</i>	Variable dicotómica cuyo valor es de uno cuando el mayor nivel educativo es profesional. Esta variable sirve como referencia en los modelos dos, tres, cuatro y cinco.

¹⁰ Ver encuesta en anexo 1.

<i>Posgrado</i>	Esta variable agrupa a los individuos que cuentan con el mayor nivel educativo.
<i>Fuma</i>	Esta variable es una de las proxies que se utilizan para aproximar a los hábitos de manejo de los individuos. Fuma toma el valor de uno cuando los individuos suelen fumar mientras manejan.
<i>Cinturón</i>	Esta variable captura el efecto del hábito de utilizar siempre el cinturón de seguridad cuando se conduce un automóvil, sobre las probabilidades de sufrir un accidente. Toma el valor de uno cuando el conductor siempre utiliza el cinturón, y cero de cualquier otra forma.
<i>Discos</i>	Variable dummy que toma el valor de uno cuando el conductor suele cambiar de música o estaciones de radio mientras conduce, y cero de cualquier otra forma.
<i>Música Alta</i>	Representa otra de las variables proxies a los hábitos de manejo; toma el valor de uno cuando el conductor suele manejar utilizando el volumen del radio a un nivel alto.
<i>Niños</i>	Esta variable dicotómica representa otra de las posibles distracciones de un conductor. Toma el valor de uno cuando el conductor suele manejar con niños abordo y cero de cualquier otra forma.
<i>Celular</i>	Variable de engaño que toma el valor de uno cuando el conductor acostumbra a hablar por teléfono celular mientras conduce, independientemente de si él realiza la llamada o él recibe la llamada; toma el valor de cero cuando no acostumbra a utilizar el celular mientras maneja.
<i>B_abstemia</i>	Esta es una de las cuatro variables que sirven para aproximar los hábitos de consumo de alcohol de los conductores. Esta variable toma el valor de uno cuando el individuo no toma ni siquiera una cerveza o copa a la semana. Esta es la variable de referencia en todos los modelos utilizados para las variables proxies de hábitos de consumo de alcohol.
<i>B_moderado</i>	Variable dummy que toma el valor de uno cuando el individuo se define como un bebedor moderado, es decir que considera que toma de una a dos cervezas o copas semanalmente; toma el valor de cero de cualquier otra forma.
<i>B_medio</i>	Esta variable toma el valor de uno cuando el conductor se define como un bebedor medio, es decir que toma de tres a siete copas o cervezas semanalmente.
<i>B_fuerte</i>	Esta variable toma un valor de uno cuando el conductor admite que toma al menos ocho cervezas o copas a la semana, mientras que de cualquier otra forma toma el valor de cero.

<i>Propio</i>	Variable dicotómica que toma el valor de uno cuando el automóvil es propio y cero cuando no lo es.
<i>Seguro</i>	Variable dummy que toma el valor de uno cuando el automóvil que conduce el individuo cuenta con seguro, y cero de cualquier otra forma.
<i>Rojo</i>	Variable de engaño que toma el valor de uno cuando el color del automóvil que conduce el individuo es de una tonalidad clara, y cero cuando es de colores oscuros.
<i>Tonalidad</i>	Variable que toma el valor de uno cuando el color del automóvil que conduce el individuo es de color rojo, y toma el valor de cero cuando no es de ese color.
<i>Antigüedad</i>	variable que captura los años de antigüedad que tiene el automóvil que conduce cada individuo.
<i>Chevrolet</i>	Variable dicotómica que toma el valor de uno cuando el automóvil que maneja el individuo es de la marca Chevrolet (independientemente del modelo) y cero de cualquier otra forma.
<i>Nissan</i>	Variable dicotómica que toma el valor de uno cuando el automóvil que maneja el individuo es de la marca Nissan (independientemente del modelo) y cero de cualquier otra forma.
<i>Otros</i>	Variable dummy que toma el valor de uno cuando el automóvil que conduce el individuo no es de la marca Chevrolet, Nissan, VolkWagen, Ford o Chrysler; y cero de cualquier otra forma. Es utilizada como variable de referencia en los modelos, para las variables que capturan la marca de los automóviles.
<i>Ponderador</i>	Variable que fue utilizada para ponderar los datos basados en el número de horas manejadas por individuo.

V.II. Estadísticas descriptivas de la muestra

En el cuadro 5.1 se muestran las estadísticas descriptivas de las variables. Es interesante observar que la media de la edad de los conductores es de alrededor de 37 años, que manejan 16 horas a la semana y que cuentan con un coche de antigüedad de 8.46 años.

En el mismo cuadro se puede apreciar cómo cambian las variables al ponderar los casos¹¹. Así, se puede observar que salvo “Horas manejadas”, todas las variables se ajustan sin sufrir fuertes cambios, tanto en sus medias como en sus desviaciones

¹¹ La metodología que se utiliza para ponderar las observaciones es la ponderación analítica. En la siguiente sección se profundiza en este tema.

estándar. También se muestra que al aplicar el ponderador no se modifica el tamaño de muestra, manteniéndolo en 401 observaciones¹².

Con respecto a la frecuencia de las variables dicotómicas, se aprecia el cuadro 5.2 que en todas las variables se alcanza al menos un 10% con respecto al total de la muestra como, siendo “bebedores fuertes” y “Chrysler” las variables con menor respuesta, alcanzando apenas el 10% (10.22% y 10.97% respectivamente).

Cuadro 5.1. Estadísticas descriptivas de la muestra

Variable	Sin Ponderar					Ponderando				
	Obs	Media	Desv. Est.	Mín	Máx	Obs	Media	Desv. Est.	Mín	Máx
edad	401	37.30	13.24	14	72	401	36.98	12.31	14	72
hasta secundaria	401	0.20	0.40	0	1	401	0.20	0.40	0	1
técnica	401	0.15	0.36	0	1	401	0.16	0.37	0	1
preparatoria	401	0.24	0.43	0	1	401	0.24	0.43	0	1
profesional	401	0.36	0.48	0	1	401	0.36	0.48	0	1
posgrado	401	0.05	0.21	0	1	401	0.04	0.19	0	1
sexo	401	0.55	0.50	0	1	401	0.61	0.49	0	1
soltero	401	0.30	0.46	0	1	401	0.24	0.43	0	1
fuma	401	0.14	0.35	0	1	401	0.16	0.37	0	1
cinturón	401	0.81	0.39	0	1	401	0.80	0.40	0	1
discos	401	0.33	0.47	0	1	401	0.31	0.46	0	1
música alta	401	0.21	0.41	0	1	401	0.20	0.40	0	1
ninos	401	0.36	0.48	0	1	401	0.37	0.48	0	1
musica_alta	401	0.21	0.41	0	1	401	0.20	0.40	0	1
celular	401	0.33	0.47	0	1	401	0.38	0.49	0	1
b_moderado	401	0.21	0.41	0	1	401	0.21	0.41	0	1
b_medio	401	0.15	0.36	0	1	401	0.17	0.37	0	1
b_fuerte	401	0.10	0.30	0	1	401	0.12	0.33	0	1
hrs__manejo	401	19.91	16.39	1	94					
accidente	401	0.17	0.37	0	1	401	0.18	0.38	0	1
auto_propio	401	0.87	0.34	0	1	401	0.83	0.37	0	1
seguro	401	0.83	0.37	0	1	401	0.85	0.36	0	1
rojo	401	0.12	0.33	0	1	401	0.13	0.33	0	1
tonalidad	401	0.63	0.48	0	1	401	0.62	0.49	0	1
antiguedad	401	8.46	6.69	0	35	401	7.78	6.45	0	35
chevrolet	401	0.22	0.41	0	1	401	0.20	0.40	0	1
nissan	401	0.16	0.36	0	1	401	0.16	0.37	0	1
vw	401	0.13	0.34	0	1	401	0.14	0.34	0	1

¹² Se utiliza la ponderación analítica que se explica en el capítulo VI..

Cuadro 5.2. Frecuencia de variables dicotómicas

Variable	Valor		Proporción del total	
	0	1	0	1
hasta_secu	319	82	0.796	0.204
técnica	339	62	0.845	0.155
prepa	306	95	0.763	0.237
posgrado	382	19	0.953	0.047
hombre	179	222	0.446	0.554
solteros	279	122	0.696	0.304
fuma	345	56	0.860	0.140
cinturón	77	324	0.192	0.808
discos	269	132	0.671	0.329
niños	257	144	0.641	0.359
música_alta	317	84	0.791	0.209
celular	267	134	0.666	0.334
b_moderado	315	86	0.786	0.214
b_medio	340	61	0.848	0.152
b_fuerte	360	41	0.898	0.102
accidente	334	67	0.833	0.167
auto_propio	53	348	0.132	0.868
seguro	67	334	0.167	0.833
rojo	352	49	0.878	0.122
tonalidad	149	252	0.372	0.628
chevrolet	314	87	0.783	0.217
nissan	338	63	0.843	0.157
vw	348	53	0.868	0.132
ford	309	92	0.771	0.229
chrysler	357	44	0.890	0.110

VI. Resultados

El modelo de regresión Logit multinomial se utiliza para analizar la relación estadística que guardan las variables proxys de los hábitos de manejo y de consumo de alcohol, así como el género de los conductores, con las probabilidades de sufrir un accidente vial en el área metropolitana de Monterrey. Para ello, se ponderan los casos de acuerdo con las horas manejadas por el entrevistado; es decir, que se controla de acuerdo con las horas que maneja el conductor a la semana, con el objetivo de que esta variable no influya sobre las demás sesgando los resultados.

Existen varias metodologías para ponderar los casos, como la ponderación por frecuencia, la ponderación por muestreo (o también llamada como ponderación por probabilidad), la ponderación analítica y la ponderación por importancia¹³.

Con el fin de esclarecer la ponderación por frecuencia se plantea el siguiente ejemplo. Suponga que tiene los siguientes datos:

Cuadro 6.1. Ejemplo de nueve observaciones sin ponderar

Y	X1	X2
18	1	1
19	0	1
20	1	0
18	0	1
18	1	1
21	1	1
20	1	1
19	0	1
18	0	1

El utilizar Mínimos Cuadrados Ordinarios (MCO) por ejemplo, sobre esa base, regresando la variable Y_t contra X_1 y X_2 , equivaldría a regresar por la misma metodología Y_t contra X_1 , X_2 y un ponderador de frecuencia, de tal forma que los datos se resumirían de la siguiente forma:

¹³ STATA Press (2005).

Cuadro 6.2. Ejemplo de nueve observaciones ponderando por frecuencia

Y	X1	X2	Pond
18	1	1	2
18	0	1	2
19	0	1	2
20	1	0	1
20	1	1	1
21	1	1	1

Es decir que el ponderador por frecuencia se utiliza con el fin de redefinir cuántas observaciones representa cada uno de los casos y es equivalente a duplicar cada una de las observaciones de acuerdo a su ponderador, y utilizar la base de datos sin ponderar.

Por otro lado, la ponderación por muestreo es el peso que se le dan a las observaciones denotando la inversa de la probabilidad de que una observación sea incluida debido al diseño muestral. Al definir una variable como el ponderador por muestreo, lo que en realidad se está haciendo es especificar el número de casos de la población total que la observación incluida en la muestra está representando. Por ejemplo, si una observación tiene una probabilidad de ser incluida de 1/4, entonces el peso asignado a esta observación, de acuerdo con la ponderación por muestreo, será de 4.

El tercer caso, la ponderación analítica, se trata de otorgarle el peso a las observaciones de una forma inversamente proporcional a su varianza. Es decir que la observación “i” cuenta con una varianza σ^2 / w_i , donde w_i representa el peso asignado de dicha observación.

Por último, la ponderación por importancia no tiene una clara definición estadística; sin embargo, como su nombre lo indica, lo que busca es que cada una de las observaciones tenga un peso de acuerdo a su importancia global. Este ponderador es utilizado comúnmente como la ponderación analítica, sin embargo bajo este esquema los ponderadores no son normalizados. Esta metodología es la que utiliza en automático el programa SPSS, y debe de utilizarse con ciertas precauciones. Por ejemplo si se utiliza directamente una variable como ponderador, la metodología provoca que la muestra “explote”: se repiten las observaciones, disminuyen las varianzas, y por lo tanto las variables se vuelven alta y estadísticamente significativas, aunque en la realidad no lo sean.

Para evitar este problema se debe construir una nueva variable que se genere a partir de la variable bajo la cual se quiere ponderar, y se multiplique por el tamaño de la muestra y se divida por la suma total de la variable para toda la muestra. En este caso, dado que se busca ponderar de acuerdo con las horas manejadas, sería:

$$\text{Ponderador} = \frac{(\text{Horas manejadas}) \cdot (401)}{7985.5}$$

Donde:

Horas manejadas = total de horas manejadas en una semana del individuo i .

401 = tamaño de la muestra.

7985.5 = suma de las horas manejadas de todas las observaciones de la muestra.

Para el presente análisis se utiliza la ponderación analítica. Sin embargo, los resultados utilizando esa metodología y la ponderación por importancia resultan idénticos.

Así, una vez ponderados los casos, se estiman dos modelos alternativos: el inicial incluye todas las variables explicativas que se consideraron posibles de guardar una relación significativa con la variable dependiente, además de la factibilidad de que el entrevistado respondiera con honestidad. En este sentido se excluyeron variables que aunque análisis previos y la intuición indicarían que deberían ser incluidos, la viabilidad de que el entrevistado conteste esas preguntas por teléfono es baja, como es el caso de variables relacionadas al ingreso o la responsabilidad del choque; de igual forma se evita preguntar frontalmente si el entrevistado suele manejar cuando consume alcohol, y por tal motivo se utiliza como variable proxy sus hábitos de consumo de alcohol independientemente si maneja cuando toma o no.

La forma para depurar el modelo, siguiendo las observaciones realizadas por Johnson (1998), comienza con un modelo general del cual se van eliminando, por pasos, las variables que resultan no significativas. Esto debido a que pueden existir relaciones entre las variables no significativas que provoquen que aunque algunas individualmente sí sean importantes, en conjunto no sean significativas. Los resultados de los modelos se muestran en el cuadro 6.3.

En la regresión (1) las variables estadísticamente significativas son: los bebedores fuertes, los conductores que cuentan con automóvil propio, la tonalidad del auto y los

coches de marca Nissan; todas con signo positivo, lo que implica que contar con el atributo aumenta las probabilidades de chocar. Como se puede observar en el cuadro, algunas variables se acercan a la significancia estadística a un nivel del 10% como es el caso de la edad (12%), fumar cuando conduce (13%) y utilizar el teléfono celular mientras conduce (11%).

En el modelo (2) se excluyen todas aquellas variables que no estuvieron cerca de ser estadísticamente significativas. En las estimaciones se observan algunos cambios: la edad se vuelve significativa al 10% de manera negativa, lo que implica que al aumentar la edad, las probabilidades de chocar disminuyen; esto se puede explicar por el aprendizaje del conductor y por cambios en su comportamiento y personalidad y, por lo tanto, en su postura frente al riesgo. Asimismo, el uso del celular mientras conduce resulta estadísticamente significativo al 10%, de forma positiva; esto quiere decir que los conductores que utilizan el celular mientras manejan enfrentan mayores probabilidades de chocar que aquellos que no lo utilizan. Además, la variable tonalidad del coche deja de ser significativa.

Cuadro 6.3. Resultados de los modelos Logit

Variables	(1)			(2)		
	B	Desv. Std.	P-Valor	B	Desv. Std.	P-Valor
edad	-0.0237	0.0153	0.1230	-0.0209	0.0121	0.0840
hasta_secu	-0.3273	0.4770	0.4930			
prepa	0.2558	0.3799	0.5010			
tecnica	-0.5689	0.5167	0.2710			
posgrado	0.6476	0.6980	0.3530			
hombre	0.0368	0.3422	0.9140			
solteros	0.0018	0.3992	0.9960			
fuma	0.6059	0.4041	0.1340	0.5247	0.3658	0.1510
cinturon	0.2307	0.3883	0.5520			
discos	0.0334	0.3375	0.9210			
musica_alta	-0.1735	0.3968	0.6620			
ninios	-0.1958	0.3491	0.5750			
celular	0.4922	0.3160	0.1190	0.5365	0.2789	0.0540
b_moderado	-0.0191	0.3846	0.9600			
b_medio	-0.3778	0.4616	0.4130			
b_fuerte	0.9364	0.4713	0.0470	0.8118	0.3715	0.0290
auto_propio	1.0601	0.4907	0.0310	1.1156	0.4557	0.0140
seguro	0.1338	0.4971	0.7880			
rojo	-0.4160	0.4727	0.3790			
tonalidad	0.5642	0.3283	0.0860	0.4378	0.2994	0.1440
antiguedad	0.0057	0.0257	0.8260			
chevrolet	-0.4391	0.5603	0.4330			
nissan	1.1925	0.4879	0.0150	1.0646	0.3330	0.0010
vw	0.3747	0.5594	0.5030			
ford	0.4508	0.5009	0.3680			
chrysler	0.7569	0.5699	0.1840			
_cons	-2.9521	1.0429	0.0050	-2.6823	0.6517	0.0000
Log-Likelihood		-166.2442			-173.0510	
LR chi-2		40.6600			27.2900	
Prob > chi-2		0.0444			0.0003	
Pseudo R2		0.1095			0.0731	

Al depurar el modelo (1), y eliminar los sesgos generados por la inclusión de variables no relevantes se logran resultados más eficientes, en donde permanecen “Nissan”, “Auto propio” y “Bebedor fuerte” como las únicas variables significativas al 5%; mientras que se confirma que el tono del auto y fumar mientras manejan no son variables estadísticamente significativas, a diferencia del uso del celular y de la edad, como se mencionó anteriormente.

En el modelo (2) se pierde valor en la Pseudos R^2 , sin embargo es un efecto natural al depurar un modelo de 27 variables a un modelo de siete variables explicativas. El rol que juega la constante no es muy claro e importante en el análisis por lo que no se profundiza en ella.

Asimismo, al comparar los dos modelos también se puede apreciar que los signos de los parámetros no cambian, y los coeficientes no experimentan fuertes alteraciones. Sin

embargo, sí se generan cambios en lo relacionado a la significancia estadística, como ya se mencionó.

Además, es importante notar que los valores de verosimilitud (40.66 y 27.29), con sus respectivos p-valores (0.0444 y 0.0003), para ambos casos muestran que las variables en su conjunto son distintas de cero.

De acuerdo con Sribney (2005), a pesar de que algunos investigadores prefieren utilizar el comando *svy* cuando se está utilizando información proveniente de encuestas, cuando se utilizan metodologías de muestreo sencillas, como en la presente investigación, se obtienen los mismos resultados independientemente de si se utiliza el comando *svy* o no; por ello se decidió no utilizar el comando *svy*, sin embargo, sí se utilizó el comando *robusto*, cuyos resultados se discuten más adelante. Asimismo, más adelante se utiliza la herramienta estadística de bootstrap para analizar la consistencia o sensibilidad del modelo y de la información generada.

VI.I. Interpretación de resultados

Es importante mencionar que siempre es complicado interpretar los resultados de los modelos de variable dependiente binaria. De acuerdo con Wooldridge (2001), desde el punto de vista práctico, y dadas las computadoras y programas modernos, lo más difícil de los modelos logit y probit es presentar e interpretar los resultados. En otras palabras, el costo principal de estos modelos es que son difíciles de interpretar.

Asimismo, antes de interpretar cualquier resultado es necesario subrayar que no se toma en cuenta en la presente investigación la magnitud del accidente, únicamente si éste sucede o no. Por lo tanto, una variable puede influir marginalmente en las probabilidades de sufrir un accidente, sin embargo su efecto puede ser sobresaliente si se trata únicamente de accidentes fatales. Se debe de considerar esta situación para no llegar a conclusiones que, aunque la evidencia empírica pareciera respaldarlas, son erróneas.

Así, al aplicar un antilogaritmo a los resultados del modelo logit, se transforman los coeficientes a momios de probabilidad. El cuadro 6.4 muestra los resultados del modelo (2) a niveles, utilizando el antilogaritmo, y mostrando además el impacto de las variables en el cambio de las probabilidades de sufrir un accidente vial.

Cuadro 6.4. Resultados del modelo (2) a niveles, momios e impacto de las variables dependientes en el cambio de probabilidades de sufrir un accidente vial

Variables	B	Momio de probabilidad	Desv. Est.	P-Valor	(dp/dxi) * Bi
edad	-0.0209	-0.9793	0.0118	0.084	-0.0027
fuma	0.5247	1.6899	0.6181	0.151	0.0907
celular	0.5365	1.7101	0.4768	0.054	0.0872
b_fuerte	0.8118	2.2521	0.8366	0.029	0.1613
propio	1.1156	3.0514	1.3904	0.014	0.1658
tonalidad	0.4378	1.5492	0.4639	0.144	0.0642
nissan	1.0646	2.8995	0.9656	0.001	0.2277
Constante	-2.6823				
Log-likelihood			-173.0509		
LR chi-2			27.29		
Prob > chi-2			0.0003		
Pseudo R2			0.0731		

Como se puede apreciar, la variable “edad” es la única variable con signo negativo. Así se encuentra que, manteniendo las demás variables constantes, al aumentar un año la edad del conductor, el logaritmo natural de las probabilidades a favor de sufrir un accidente vial disminuye en un factor del 0.0206 (que se obtiene de restar: $1 - 0.9794$). En otras palabras, se estima que por cada año adicional de un conductor, el momio de sufrir un accidente automovilístico disminuye en un 2.01%. Se puede observar en la sexta columna del cuadro 6.4 el efecto de un cambio unitario de las variables explicativas en las probabilidades de sufrir un accidente vial. Para la variable edad se concluye que un año adicional en la edad del conductor disminuye en un 2.7% la probabilidad de sufrir un accidente vial.

De la misma forma, controlando las demás variables, podemos decir que utilizar el teléfono celular mientras se conduce incrementa el momio de probabilidad de sufrir un accidente vial en un factor de 1.71; o en otras palabras, se estima que utilizar el teléfono celular mientras se conduce aumenta el momio de chocar en un 171%. Además, los conductores que utilizan el celular enfrentan un incremento en la probabilidad de sufrir un accidente vial de 8.72%.

Asimismo, los hallazgos muestran que los conductores que se definen como bebedores fuertes enfrentan un incremento en el momio de probabilidad de sufrir un accidente

automovilístico en un factor del 2.25. Es decir, las personas cuyos hábitos de consumo de alcohol los ubican como bebedores fuertes enfrentan un incremento en el momio de sufrir un accidente vial de 225%. Es decir que, manteniendo las demás variables constantes, los conductores que se autodefinen como bebedores fuertes experimentan un crecimiento en las probabilidades de sufrir un accidente vial del 16.13% respecto a los conductores que no son bebedores fuertes.

Por otro lado, se entiende auto propio aquel que pertenece a la familia o al hogar, no necesariamente al conductor. Por ejemplo, un conductor de 18 años puede no ser el propietario del automóvil, sin embargo el auto es propio dado que pertenece al jefe del hogar. En este sentido, manteniendo las demás variables constantes, los conductores que manejan autos propios enfrentan un incremento en el momio de probabilidad en un factor 3.05 mayor que aquellos que no lo hacen. Es decir que las probabilidades de chocar aumentan en 16.58% cuando el coche es propiedad del conductor. Este resultado sugiere que al conducir un auto ajeno, el conductor suele ser más cuidadoso de lo que normalmente es.

Por último, controlando las demás variables, los conductores con autos marca Nissan enfrentan un incremento en el momio de probabilidad de sufrir un accidente automovilístico en un factor del 2.89. En otras palabras, los conductores de autos Nissan enfrentan un aumento en el momio de sufrir un accidente vial de 289% respecto a los conductores de autos de otras marcas. Así, los conductores de automóviles marca Nissan, enfrentan una probabilidad de sufrir un accidente vial 22.77% mayor al resto de los conductores, manteniendo las demás variables constantes. La explicación que puede respaldar este hallazgo es que los taxistas en su mayoría utilizan esta marca de automóviles, lo cual aumenta su proporción del parque vehicular, y lo que parece más importante, los hábitos de manejo de los taxistas resultan peligrosos. En este sentido es común observar taxis que se paran en media avenida sin utilizar las luces intermitentes; pararse en el segundo o tercer carril; dar vueltas prohibidas; no respetar las señales, etc. Esta situación claramente aumenta el riesgo de los conductores, situación que se refleja subyacentemente en las probabilidades de chocar un automóvil Nissan, lo cual se manifiesta en las cifras oficiales mostradas en el cuadro 1.2.

Cabe señalar que tanto fumar mientras se conduce, como la tonalidad del color del automóvil no resultaron ser variables estadísticamente significativas ni a un nivel de confianza del 90%.

VI.II. Comparación de resultados utilizando los modelos logit y probit.

Como ya se mencionó, los resultados que se obtienen tanto por logit como por probit son muy similares, dado que lo que diferencia a estas metodologías es la distribución de probabilidad que se utiliza. El cuadro 6.5 muestra los resultados que arrojan ambos modelos.

Cuadro 6.5 resultados del modelo (2) por logit y probit

Variables	Logit		Probit		Diferencia
	B	P-Valor	B	P-Valor	
edad	-0.0209	0.084	-0.011159	0.093	-0.0097
fuma	0.5247	0.151	0.289515	0.1670	0.2352
celular	0.5365	0.054	0.301555	0.0550	0.2350
b_fuerte	0.8118	0.029	0.476042	0.0280	0.3358
propio	1.1156	0.014	0.597393	0.015	0.5182
tonalidad	0.4378	0.144	0.226627	0.1740	0.2111
nissan	1.0646	0.001	0.617134	0.0010	0.4474
Constante	-2.6823	0	-1.544428	0	-1.1379
Log-likelihood	-173.0509		-173.0937		0.0428
LR chi-2	27.29		27.20		0.09
Prob > chi-2	0.0003		0.0003		0
Pseudo R2	0.0731		0.0729		0.0002

Como se puede observar, tanto los signos como la significancia de las variables son iguales en ambos casos, sin embargo, la magnitud de los coeficientes cambian. Las variables más sensibles a ello son manejar un auto propio, los coches marca Nissan y los bebedores fuertes. De esta forma, puede surgir una vez más la duda de cuál estimación es mejor. De acuerdo con Gujarati (2003), no existe una razón de peso para elegir entre un modelo logit y un probit, sin embargo, en la práctica muchos investigadores utilizan el modelo logit por su comparativa simplicidad matemática. En este caso, de acuerdo con los valores de Log-likelihood y la pseudo R², los estimadores de logit son mejores, aunque la diferencia es muy pequeña.

VI.III. Propiedades de las estimaciones

VI.III. I. Multicolinealidad

Con respecto al problema de multicolinealidad, en los modelos de variables discretas normalmente se analiza a través de la matriz de correlación o a través del valor VIF (variante inflation factor) aunque únicamente se utiliza en el caso de modelos no binarios.

Desde una perspectiva intuitiva, el problema de la multicolinealidad impide interpretar los coeficientes de las salidas, dado que si, por ejemplo, en un modelo: $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + e_t$, y además $X_1 = Z + \rho X_2$, la interpretación de $\beta_1=0.5$ sería: manteniendo los demás regresores constantantes, por cada movimiento en una unidad de X_1 , Y varía en 0.5. Sin embargo, dada la relación de X_1 con X_2 , no se puede mantener el segundo regresor constante, y por lo tanto se pierde la interpretación de los resultados.

Con respecto a la multicolinealidad de los modelos, en el anexo 3 se muestra la matriz de correlación. En ella se puede observar que el nivel más alto de correlación se presenta entre las variables “soltero” y “edad”, con un coeficiente de correlación de -0.4871. El segundo par de variables más correlacionadas son “antigüedad” y “seguro” con un coeficiente de -0.3815. Así, aunque ésta no es un *test* formal para probar la presencia o ausencia de multicolinealidad, las variables parecen no estar fuertemente correlacionadas, y por lo tanto se puede presumir que no existe este problema en ellas.

VI.III.II. Heteroscedasticidad.

Es importante analizar la presencia o ausencia de la heteroscedasticidad dado que en caso de que el término de error muestre la presencia de este problema, genera que los parámetros sean inconsistentes. Sin embargo, como lo expresan Davidson y MacKinnon (1982), dado que en los modelos basados en los métodos estimados por máxima versosimilitud (como los logits y probits) analizar la heteroscedasticidad “usualmente es más caro que, por ejemplo, en el caso de los estimadores de mínimos cuadrados para un modelo de regresión lineal, los investigadores suelen mostrarse reticentes a probar la especificación” y la heteroscedasticidad del modelo como se haría normalmente en el análisis de regresiones.

Así, en los modelos binarios recientemente se utilizan novedosas técnicas, como es el caso del procedimiento propuesto por Davidson y MacKinnon (1982), en el cual suponen que la heteroscedasticidad está en función de variables Z, las cuales son elegidas de las variables explicativas del modelo probit o logit. La prueba está basada en el principio del multiplicador de Lagrange. Los resultados estimados son utilizados para construir una regresión artificial y auxiliar diseñada para realizar la prueba de heteroscedasticidad.

Sin embargo, esta prueba es intensiva en programación y pocas veces es utilizada empíricamente. Una metodología que se utiliza con el fin de eludir las pruebas de heteroscedasticidad para modelos con variable dependiente binaria, es la denominada *robusto*, que aunque es raramente utilizado, ayuda a mejorar la estimación.

Robusto es un comando de programación que permite calcular un estimador robusto de la varianza; al cual “se le suele referir por distintos nombres: Hubert/White/sándwich son normalmente utilizados en el contexto de robustez contra heteroscedasticidad. Los estadísticos especializados en encuestas a menudo se refieren al cálculo de esta varianza como el método de linealización de las serie de Taylor de primer orden”¹⁴.

Los resultados del modelo estimado *robusto* se muestran en el cuadro 6.6. Como se puede apreciar, los resultados utilizando este comando son distintos al modelo (2): los parámetros estimados permanecen idénticos, sin embargo las desviaciones estándar aumentan, y por lo tanto los valores del p-valor aumentan también, manteniéndose como variables estadísticamente significativas únicamente “propio” y “Nissan”.

¹⁴ STATA Press (2005).

Cuadro 6.6. Resultados del modelo (2) utilizando el estimaciones robustas

Variable	Coeficiente	Desv. Std. Robusto	P-Valor
edad	-0.0208809	0.0146121	0.153
fuma	0.5246728	0.4771073	0.271
celular	0.5365135	0.3437183	0.119
b_fuerte	0.8118369	0.5240943	0.121
tonalidad	0.4377722	0.3778568	0.247
propio	1.115614	0.6518671	0.087
nissan	1.064571	0.4114508	0.01
_cons	-2.682292	0.8450974	0.002
Log-Likelihood		-173.05096	
LR chi-2		17.67	
Prob > chi-2		0.0136	

Estos resultados muestran la sensibilidad de las desviaciones estándar de los estimadores utilizando únicamente el logit ponderando los casos. Aunque no es evidencia suficiente, se puede considerar la existencia del problema de heteroscedasticidad, lo que perjudica el contraste de hipótesis. Situación que puede cuestionar los resultados.

Así, el cuadro 6.6 muestra que a un nivel de confianza del 5%, únicamente Nissan permanece como una variable estadísticamente significativa; mientras que a un nivel del 10% auto propio también resulta ser significativa. De estos resultados sobresale el hecho de que bebedor fuerte deja de ser una variable estadísticamente significativa, al aumentar su p-valor de 0.029 (en el modelo 2) a 0.271 (con estimaciones robustas). La variable tonalidad se alejó aun más de la significancia estadística al pasar de 0.144 a 0.247.

Asimismo, se estimó de manera análoga al modelo (1) utilizando el comando *robusto*, comenzando del modelo más general, es decir, incluyendo todas las variables, y a partir de sus resultados se depuró el modelo. El cuadro 6.7 muestra el resultado inicial y el final. Cabe resaltar que utilizar el comando *robusto* proporciona los mismos estimadores que ponderar por muestreo.

Se puede apreciar en el cuadro 6.7 que la variable Nissan es la única estadísticamente significativa a un nivel de confianza del 5%, mientras que el uso del teléfono celular resulta significativa a un nivel de confianza del 10%. La variable bebedor fuerte resulta ser no significativa, aunque su p-valor (0.14) no se aleja demasiado del nivel del 10%.

Cuadro 6.7. Resultados de los modelos logit inicial y final, utilizando estimaciones robustas

Variables	(1) b			(2) b		
	B	Std. Err.	P-valor	B	Std. Err.	P-valor
edad	-0.024	0.018	0.194			
tecnica	-0.242	0.717	0.736			
preparatoria	0.583	0.619	0.346			
carrera	0.327	0.582	0.574			
posgrado	0.975	0.772	0.207			
hombre	0.037	0.403	0.927			
solteros	0.002	0.461	0.997			
fuma	0.606	0.512	0.237			
cinturon	0.231	0.442	0.602			
discos	0.033	0.407	0.935			
musica_alta	-0.174	0.497	0.727			
nios	-0.196	0.444	0.659			
celular	0.492	0.327	0.132	0.619	0.353	0.079
b_moderado	-0.019	0.510	0.970			
b_medio	-0.378	0.608	0.534			
b_fuerte	0.936	0.618	0.130	0.780	0.528	0.140
propio	1.060	0.668	0.113			
seguro	0.134	0.646	0.836			
rojo	-0.416	0.532	0.434			
tonalidad	0.564	0.371	0.128			
antiguedad	0.006	0.033	0.863			
chevrolet	-0.439	0.542	0.418			
nissan	1.193	0.481	0.013	0.899	0.407	0.027
vw	0.375	0.637	0.556			
ford	0.451	0.564	0.424			
chrysler	0.757	0.664	0.254			
_cons	-3.279	1.254	0.009	-2.108	0.236	0.000

VI. III. IV. Sensibilidad a cambios en el tamaño de la muestra

A continuación se analiza la sensibilidad del modelo a cambios en el tamaño de la muestra. Para ello se decidió utilizar la metodología de *bootstrap*:

- I. Utilizando una muestra de 300 observaciones.
 - a. Realizando 500 iteraciones.
 - b. Realizando 1,000 iteraciones.
- II. Utilizando una muestra de 200 observaciones.
 - a. Realizando 500 iteraciones.
 - b. Realizando 1,000 iteraciones.

El *bootstrap* es una metodología bajo la cual se reduce la muestra original de tamaño “n” (en este caso $n = 401$), y se reproduce de manera aleatoria “q” veces; es decir que se eligen “q” submuestras de tamaño “m” (donde $n > m$) a partir de la información contenida en muestra original, seleccionando las observaciones con reemplazo, es decir que una observación puede estar incluida más de una vez en cada una de las “q” submuestras. Así, para cada una de las “q” submuestras se aplica un proceso estadístico, en este caso un modelo logit. Como señala Simon (1969), esta metodología representa la línea más avanzada de las técnicas de remuestreo, tanto desde una perspectiva teórica como práctica.

El uso del *bootstrap* comúnmente se presenta de una forma no paramétrica, y por lo tanto puede ser sustituto de los mismos; sin embargo, su uso se puede combinar con métodos paramétricos, como es el caso de la presente investigación.

La idea detrás del uso de este procedimiento, para la actual investigación, es que si los cambios en los coeficientes de las iteraciones no experimentan grandes cambios con respecto a sus valores originales (con la muestra total de 401 observaciones), entonces se puede afirmar que las estimaciones no son muy sensibles a cambios en el tamaño de la muestra. Sin embargo, siempre que se utiliza esta metodología surgen preguntas difíciles de contestar, como: ¿cómo y por qué se elige el tamaño de la submuestra?, o ¿cuántas repeticiones se requieren para obtener resultados confiables?

Con respecto al uso de modelos logit en el *bootstrap*, no se puede hacer uso de ponderaciones de las observaciones, por lo que en este caso no se le dan pesos a las observaciones. Sin embargo, los coeficientes en sí no son lo que interesa, sino la igualdad o diferencia estadística entre los parámetros calculados por *bootstrap* y los calculados con la muestra original (en este caso sin ponderar).

En el cuadro 6.8 se muestran los resultados del proceso de *bootstrap*, con tamaño de muestra de 300 casos, mientras que el cuadro 6.8 muestra lo mismo para el caso de tamaños de muestra de 200 casos. En la primera columna se muestran las variables. De la segunda a la cuarta columna se muestran los resultados del *bootstrap* utilizando 500 repeticiones; y de la quinta a la séptima columna se muestran los resultados de las 1,000 repeticiones.

Cuadro 6.8. Resultados del *bootstrap* del modelo (2), sin ponderar, utilizando 500 y 1,000 repeticiones, y un tamaño de muestra de n = 300

Variable	500 Repeticiones			1,000 Repeticiones		
	Observado	Sesgo	Error est.	Observado	Sesgo	Error est.
b_edad	-0.031	-0.002	0.013	-0.031	-0.002	0.015
b_fuma	0.325	-0.005	0.502	0.325	-0.005	0.476
b_celular	0.496	0.009	0.340	0.496	0.017	0.345
b_b_fuerte	0.400	0.023	0.596	0.400	-0.045	0.576
b_auto_propio	1.013	0.139	0.723	1.013	0.119	0.719
b_tonalidad	0.271	0.026	0.366	0.271	0.019	0.366
b_nissan	1.057	0.077	0.419	1.057	0.050	0.419
b_cons	-2.065	-0.143	0.805	-2.065	-0.125	0.856
se_edad	0.012	0.002	0.001	0.012	0.002	0.001
se_fuma	0.391	0.091	0.062	0.391	0.089	0.061
se_celular	0.283	0.059	0.021	0.283	0.058	0.020
se_b_fuerte	0.425	0.107	0.089	0.425	0.108	0.089
se_auto_propio	0.490	0.132	0.137	0.490	0.132	0.143
se_tonalidad	0.296	0.062	0.027	0.296	0.061	0.027
se_nissan	0.332	0.074	0.034	0.332	0.072	0.033
se_cons	0.638	0.156	0.125	0.638	0.157	0.129

Donde, la columna “observado” muestra los resultados de la regresión original con la muestra completa; en la primera parte muestra la estimación del parámetro y en la segunda, su error estándar. La columna de “sesgo” muestra la diferencia entre el valor estimado con la muestra completa y la media de las estimaciones con el *bootstrap*. La columna de “error estándar” muestra la desviación estándar del parámetro para cada una de las iteraciones.

Efron (1982)¹⁵ sugiere que si el resultado de dividir el sesgo entre la desviación estándar es menor a 0.25, entonces el sesgo en los estimadores no representa un problema importante. Así, para el caso de los coeficientes, nunca se presenta un resultado mayor a 0.25 en lo referente a los coeficientes usando tamaños de muestra de 300 y 200, y utilizando 500 y 1,000 iteraciones; por lo que el sesgo no resulta lo suficientemente grande para presentar un problema.

¹⁵ Citado por Mooney y Duval (1993).

Cuadro 6.9. Resultados del *bootstrap* del modelo (2), utilizando 500 y 1,000 repeticiones, y un tamaño de muestra de n = 200

Variable	500 Repeticiones			1,000 Repeticiones		
	Observado	Sesgo	Desv. Est.	Observado	Sesgo	Desv. Est.
b_edad	-0.031	-0.003	0.017	-0.031	-0.003	0.018
b_fuma	0.325	-0.021	0.643	0.325	-0.032	0.646
b_celular	0.496	0.001	0.435	0.496	0.024	0.440
b_fuerte	0.400	-0.012	0.712	0.400	0.007	0.722
b_tonalidad	0.271	0.028	0.478	0.271	0.024	0.468
b_propio	1.013	0.113	0.797	1.013	0.086	0.803
b_nissan	1.057	0.120	0.536	1.057	0.052	0.547
b_cons	-2.065	-0.066	1.001	-2.065	-0.042	1.000
se_edad	0.012	0.006	0.002	0.012	0.006	0.002
se_fuma	0.391	0.222	0.106	0.391	0.226	0.113
se_celular	0.283	0.146	0.034	0.283	0.145	0.033
se_b_fuerte	0.425	0.252	0.126	0.425	0.249	0.131
se_tonalidad	0.296	0.154	0.043	0.296	0.154	0.044
se_propio	0.490	0.280	0.169	0.490	0.284	0.174
se_nissan	0.332	0.178	0.052	0.332	0.181	0.056
se_cons	0.638	0.338	0.161	0.638	0.344	0.163

Sin embargo, el interés y el objetivo que se persigue al utilizar el *bootstrap* en la presente investigación son los niveles de la desviación estándar de los coeficientes. Es decir, comprobar si los coeficientes de los “n” b_i son estadísticamente distintos de los correspondientes a los coeficientes del modelo original. Más formalmente:

$$H_0: \beta_{i,j} = \beta_{i,401}$$

$$H_i: \beta_{i,j} \neq \beta_{i,401}$$

Donde:

$\beta_{i,i}$: Es el coeficiente i de la iteración j.

$\beta_{i,401}$: Es el coeficiente i de la regresión original con 401 observaciones.

Es importante notar que no todas las variables llegaron a cumplir con las 500 o 1,000 iteraciones; el caso en particular es el de la variable “propio” cuando se reduce la muestra a 300 observaciones, mientras que cuando se reduce la muestra a 200 casos las variables que cumplen con el número de repeticiones son edad (para el caso de las 1,000 repeticiones), fuma (en ambos casos), b_fuerte (ambos casos), propio (ambos casos) y tonalidad (únicamente en las 1,000 iteraciones). En el caso de las variables dicotómicas, esta situación se presenta debido a que la frecuencia de una de las dos respuestas

cuenta con un porcentaje pequeño de menciones (alrededor del 10%). Esto implica que al reducir la muestra, en ocasiones la variable no cuenta con suficientes casos para poder tomarse dentro de la regresión logística.

Cuadro 6.10. Resumen de los resultados de las pruebas “t” para comparar los coeficientes de las iteraciones y del modelo original

Bootstrap: logit 200 observaciones										
Variable	500 Repeticiones					1000 Repeticiones				
	Número de iteraciones	Hipótesis no rechazadas		Hipótesis rechazadas		Número de iteraciones	Hipótesis no rechazadas		Hipótesis rechazadas	
		Número	Porcentaje	Número	Porcentaje		Número	Porcentaje	Número	Porcentaje
edad	500	479	95.8	21	4.2	1000	957	95.7	43	4.3
fuma	499	480	96.2	19	3.8	996	958	96.2	38	3.8
celular	500	476	95.2	24	4.8	1000	949	94.9	51	5.1
b_fuerte	492	474	96.3	18	3.7	987	932	94.4	55	5.6
propio	475	451	94.9	24	5.1	952	906	95.2	46	4.8
tonalidad	500	472	94.4	28	5.6	952	897	94.2	55	5.8
nissan	500	465	93.0	35	7.0	1000	960	96.0	40	4.0
cons	500	489	97.8	11	2.2	1000	947	94.7	53	5.3

Bootstrap: logit 300 observaciones										
Variable	500 Repeticiones					1,000 Repeticiones				
	Número de iteraciones	Hipótesis no rechazadas		Hipótesis rechazadas		Número de iteraciones	Hipótesis no rechazadas		Hipótesis rechazadas	
		Número	Porcentaje	Número	Porcentaje		Número	Porcentaje	Número	Porcentaje
edad	500	489	97.8	11	2.2	1000	962	96.2	38	3.8
fuma	500	476	95.2	24	4.8	1000	955	95.5	45	4.5
celular	500	474	94.8	26	5.2	1000	949	94.9	51	5.1
b_fuerte	500	464	92.8	36	7.2	1000	943	94.3	57	5.7
propio	493	460	93.3	33	6.7	988	930	94.1	58	5.9
tonalidad	500	477	95.4	23	4.6	1000	952	95.2	48	4.8
nissan	500	470	94.0	30	6.0	1000	945	94.5	55	5.5
cons	500	467	93.4	33	6.6	1000	939	93.9	61	6.1

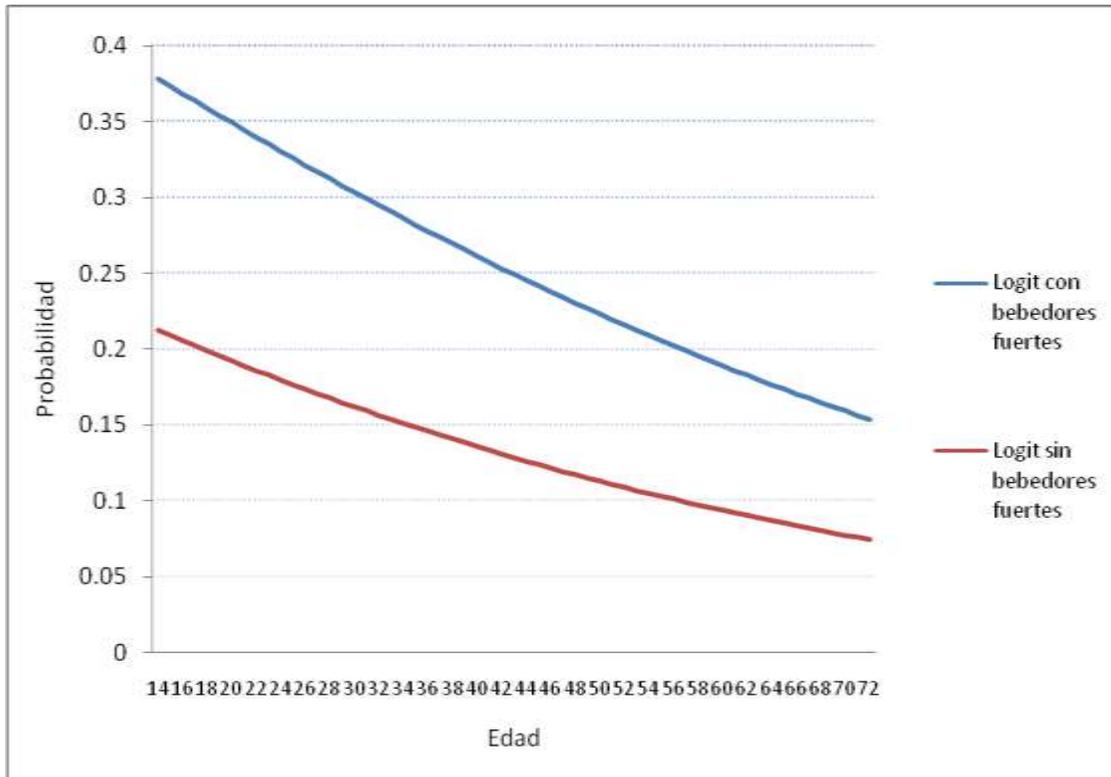
Como se puede apreciar en el cuadro 6.10 tanto al construir muestras de 300 casos, como de 200 observaciones, el porcentaje de parámetros derivados del *bootstrap* que no son estadísticamente distintos del estimador generado con la muestra completa, nunca descendió de 92%; es decir que incluso al trabajar con una submuestra de la mitad del tamaño de la muestra original, los parámetros permanecen al menos en el 92% de los casos, siendo estadísticamente iguales que en la regresión original con 401 observaciones.

VI.IV. Distribución de probabilidades, de acuerdo a la edad de los conductores

El gráfico 6.1 muestra la distribución de las probabilidades de sufrir un accidente vial en el AMM, por edad, para los bebedores fuertes y aquellos que no lo son; para ello se utiliza el

modelo 2, y se evalúan las variables con sus valores medios, con excepción de la variable b_fuerte, que se evalúa con un valor de 1 y 0.

Gráfico 6.1. Distribución logística de probabilidades de sufrir un accidente, por consumo de alcohol, de acuerdo a la edad



Se puede apreciar en el gráfico 6.1 que las probabilidades de sufrir un accidente vial se modifican también en función de si el conductor es un bebedor fuerte o no. En este sentido, los bebedores fuertes enfrentan mayores probabilidades de chocar que el resto de los conductores. En promedio la distancia entre las dos curvas es de 12.12%, sin embargo, se puede observar que la distancia entre las curvas va disminuyendo conforme aumenta la edad de los conductores.

Además, se puede observar en el gráfico 6.1 que la probabilidad de chocar va disminuyendo conforme la edad aumenta y la pendiente va disminuyendo conforme aumenta la edad. Esto muestra una relación no lineal entre estas dos variables. Por ello, se desarrollaron dos modelos incluyendo la variable edad al cuadrado, como se puede observar en el cuadro 6.11. El modelo (3) incluye tanto la edad, como la edad al cuadrado; mientras que el modelo (4) únicamente incluye edad al cuadrado. Así, se puede

apreciar que al incluir las dos variables, ambas se vuelven no significativas, además de que existe el problema de multicolinealidad como se puede observar en el cuadro 6.12, por lo que la prueba de significancia estadística se vuelve no confiable. De cualquier forma, los modelos (2) y (4) arrojan prácticamente los mismos resultados.

Cuadro 6.11. Resultados de los modelos utilizando edad al cuadrado

Variables	(3)			(4)		
	B	Desv. Std.	P-Valor	B	Desv.	P-Valor
Edad	0.026	0.072	0.715			
Edad^2	-0.001	0.001	0.507	0.000	0.000	0.075
Fuma	0.518	0.366	0.157	0.520	0.366	0.155
Celular	0.529	0.279	0.058	0.532	0.279	0.057
B_fuerte	0.833	0.373	0.025	0.821	0.371	0.027
Auto_propio	1.095	0.456	0.016	1.110	0.455	0.015
Tonalidad	0.448	0.300	0.136	0.445	0.300	0.138
Nissan	1.054	0.333	0.002	1.060	0.333	0.001
Constante	-3.491	1.384	0.012	-3.032	0.570	0.000
Log likelihood = -172.82132			Log likelihood = -172.88904			
Pseudo R2 = 0.0743			Pseudo R2 = 0.0739			

Al tomar en cuenta únicamente la edad al cuadrado, las variables mantienen su signo y su relevancia estadística de acuerdo a los resultados del modelo (2). Al comparar entre los modelos (2) y (4), el último presenta una pequeña mejoría con respecto al pseudo R^2 y al *Log likelihood*. Sin embargo, dado el ligero incremento en estos estimadores, se decide mantener la variable edad por su fácil interpretación.

Cuadro 6.12. Correlación entre edad y edad al cuadrado

	Edad	edad2
edad	1	
edad2	0.9848	1

También se utilizan variables de engaño para diferenciar los casos analizados de acuerdo con rangos de edad. El cuadro 6.13 muestra los distintos modelos y sus resultados. Se puede apreciar que la inclusión de tales variables de engaño no sólo no son significativas, sino que convierten a la edad y otras variables en no significativas.

Cuadro 6.13. Resultado de modelos utilizando variables de engaño por rangos de edad

Variables	(5)		(6)		(7)	
	Coef	P-Valor	Coef	P-Valor	Coef	P-Valor
edad	-0.006	0.776	-0.006	0.776	-0.006	0.776
fuma	0.505	0.171	0.505	0.171	0.505	0.171
celular	0.516	0.065	0.516	0.065	0.516	0.065
b_fuerte	0.824	0.029	0.824	0.029	0.824	0.029
auto_propio	1.141	0.012	1.141	0.012	1.141	0.012
tonalidad	0.417	0.167	0.417	0.167	0.417	0.167
nissan	1.056	0.002	1.056	0.002	1.056	0.002
dh29	0.188	0.693	2.006	0.209		
d30a59			1.818	0.201	-0.188	0.693
d60m	-1.818	0.201			-2.006	0.209
_cons	-3.228	0.002	-5.047	0.010	-3.040	0.000

En suma, aunque la relación entre la edad y las probabilidades de chocar no es lineal, no resultan significativas variables dicotómicas que se probaron para diferencia entre grupos de edad (hasta 29 años de edad, de 30 a 59 y 60 y más); y lo mismo sucedió al incluir la variable edad al cuadrado (edad²).

Por otro lado, se podría esperar una diferencia considerablemente mayor entre los bebedores fuertes y aquellos que no lo son; sin embargo, es necesario recordar que en el presente estudio no se diferencia entre los accidentes, por lo que no se toma en cuenta la magnitud de cada uno de los accidentes: un accidente fatal se cuantifica de la misma manera que un ligero accidente. Esta situación genera que la diferencia entre estos dos tipos de conductores no sea tan grande.

Un futuro estudio, con un tamaño de muestra tal que el número de casos de accidentes fueran representativos del total de accidentes en el AMM (es decir, alrededor de 400 observaciones en donde el conductor sí haya sufrido un accidente en el último año), podría enfocarse a la magnitud del accidente como variable dependiente, y analizar la incidencia del consumo de alcohol sobre las magnitudes de los accidentes. Sin embargo, de cualquier forma es imposible capturar accidentes en donde el conductor haya muerto, por lo que la única forma de considerar estos casos es utilizando la información oficial de los municipios del AMM.

VII. Conclusiones

Los accidentes viales representan un tema de interés para el análisis económico debido a su incidencia en el bienestar de la sociedad, así como en la eficiencia económica. En la presente investigación se realizó un proceso de encuestado, en donde se utilizó el directorio telefónico de Telmex como marco muestral, y se definió una muestra de 401 observaciones. Asimismo, se utilizó la metodología Logit para comprobar las tres hipótesis planteadas:

1. Los hábitos de manejo tienen un impacto significativo en las probabilidades de sufrir un accidente vial. Los hábitos negativos se resumen en distracciones que aumentan las probabilidades de sufrir un accidente vial en el Área Metropolitana de Monterrey.
2. Los hábitos de consumo de alcohol son determinantes de las probabilidades de sufrir un accidente vial. Por lo tanto, una persona abstemia se espera que tenga menores probabilidades de sufrir un accidente vial con respecto a los bebedores medios y fuertes.
3. Controlando las demás variables, existe una diferencia significativa en las probabilidades de sufrir un accidente vial de acuerdo con el género del conductor.

Con respecto a la primera hipótesis, la evidencia empírica demuestra que la única distracción que efectivamente impacta estadística y positivamente las probabilidades de sufrir un accidente es el uso del teléfono celular mientras se conduce. Por lo tanto, las demás variables Proxy utilizadas para medir los hábitos de manejo, es decir: cambiar de discos o estaciones constantemente; escuchar la música a un volumen alto; manejar con niños en el auto y utilizar el cinturón de seguridad, resultan no significativas. Así, la hipótesis uno se acepta, aunque sea únicamente para el caso del uso del teléfono celular.

De aquí se desprende que una acción que los gobiernos municipales del Área Metropolitana de Monterrey pueden tomar, es decir, prohibir a los conductores utilizar los teléfonos celulares, o restringirlos al uso únicamente de celulares con alta voz (hands free). El ayuntamiento de Saltillo por ejemplo implementó esta política pública: en el municipio, de acuerdo al Reglamento de Seguridad Pública, Tránsito y Vialidad del Municipio de Saltillo, Coahuila, Capítulo I, Artículo 82, "Al momento de conducir un

vehículo queda prohibido para su conductor el uso de teléfonos celulares, audífonos y cualquier otro elemento que pueda distraer su atención”.

Con respecto a la segunda hipótesis, los hábitos de consumo de alcohol, la evidencia empírica muestra que, controlando las demás variables, la probabilidad de que un individuo sufra un accidente vial aumenta únicamente cuando el conductor es un bebedor fuerte, es decir que toma siete o más copas a la semana. Así, no existe una diferencia estadística entre los abstemios, los bebedores ligeros y los medios.

Es importante considerar que los resultados que se derivan del estudio son conservadores, dado que los encuestados tienen incentivos a mentir. Además, es importante señalar que se mide únicamente la probabilidad de sufrir un accidente, y no se considera la magnitud del mismo. Un estudio con información derivada de los accidentes permitiría analizar cómo afecta el consumo de alcohol a las probabilidades de accidentarse y de la magnitud del mismo.

Por otro lado, la evidencia empírica muestra que no existen diferencias estadísticamente significativas entre las probabilidades de sufrir un accidente vial de un hombre y una mujer. Por lo tanto, la hipótesis número tres se rechaza.

Con ello, también se puede clasificar como mito la creencia de que existen diferencias en las probabilidades de sufrir un accidente de acuerdo al género de los automovilistas, dado que aun controlando las horas de manejo de cada grupo, la evidencia no muestra diferencias significativas.

VII.I. Limitantes y futuras mejoras

La limitante principal por la naturaleza misma del estudio es que la información utilizada genera resultados en un punto determinado del tiempo, y no toma en cuenta el resto del tiempo. Sin embargo, no se espera que cambie significativamente las conductas de los automovilistas, al menos en el mediano plazo.

Otra de las limitantes de la presente investigación es que se ha ignorado a un 10% de la población, debido a que no se han tomado en cuenta para la creación del directorio telefónico, debido a que una compañía distinta de Telmex les provee el servicio telefónico.

Sin embargo, no se espera que ese 10% de la población se comporte distinto del 90% de la población restante.

Asimismo, no se ha tomado en cuenta información relevante para las estimaciones de las probabilidades de sufrir un accidente vial, como puede ser el caso del ingreso de los individuos. Esto se debe a la gran dificultad que se tiene para que un individuo conteste la pregunta por teléfono a un desconocido, además de que también tiene incentivos a mentir. Ejemplo de ello es el de los modelos de los automóviles, en donde aunque se tomó en cuenta la marca de los autos, no se pudo extraer información sobre los modelos, debido a que los encuestados se negaban constantemente a contestar la pregunta.

Además, los resultados utilizando las estimaciones robustas sugieren que se tenga precaución con los hallazgos, dado que los cambios en los errores estándar afectan las pruebas de hipótesis.

Las futuras mejoras a la presente investigación son las implicaciones de los hallazgos en la industria aseguradora, así como la sugerencia de aplicación de políticas públicas diseñadas con el fin de minimizar los accidentes viales.

Referencias

Artís, Manuel; Ayuso, Mercedes; Guillén, Montserrat. (2002). "Detection of automobile insurance fraud with discrete choice models and misclassified claims". *The Journal of Risk and Insurance*, Vol. 69, No.3.

Chaloupka, Frank; Saffer, Henry; Grossman, Michael. (1993). "Alcohol control policies and motor vehicle fatalities". National Bureau of Economic Research. Investigación #3831. EE-UU.

Cohen, Alma; Einav, Liran. (2003). "The effects of mandatory seat belt laws on driving behavior and traffic fatalities". *The Review of Economics and Statistics*. EE-UU.

Cook, Philip; Ostermann, Jan; Sloan, Frank. (2005). "Are alcohol excise taxes good for us?, short and long-term effects on mortality rates". National Bureau of Economic Research (NBER). Workink paper #W11138. EE-UU.

Damaris, Alfred. (1992). "Logit Modeling, Practical Applications". SAGE University Paper. No.86.

Dionne; Vanasse. (1992). "Automobile insurance ratemaking in the presence of asymmetrical information". *Journal of Applied Econometrics, Series: Quantitative Applications in the Social Sciences*, vol.7, No.2.

Green, William. (1999). "Análisis Económico". Practice Hall, tercera edición. Madrid, España.

Golob, Thomas; Recker, Wilfred. (2002). "Relationships among urban freeways accidents, traffic flow, weather, and lighting conditions". *Journal of Transportation Engineering*. EE-UU.

Griffiths, William; Hill, Carter; Judge, George. (1997). "Learning and practicing econometrics". John Wiley & Sons. EE-UU.

Gujarati, Damodar. (2005). "Econometría". Mc Graw Hill, cuarta edición. México.

Haigney, Di; Taylor, Ray G. (1998). "Mobile phone use whilst driving: phone operation vs. Vehicle transmission. Road Safety Department. Birmingham, Inglaterra. Recuperado de: www.rosopa.com/roadsafety/info/mobile_phone_studies.pdf, el día 14 de agosto de 2007.

Harvey, Andrew. (1989). "Forecasting, structural time series models and the Kalman Filter". Cambridge University Press. Reino Unido.

Hedlund, James; McCartt, Anne. (2002). "Drunk Driving: Seeking additional solutions". Preusser Research Group, AAA Foundation for Traffic Safety. Washington, EE-UU. Recuperado de: <http://www.aaafoundation.org/pdf/DrunkDriving-SeekingAdditionalSolutions.pdf>, el día 10 de agosto de 2007.

Islam, Samantha; Mannering, Fred. (2007). "Age, Gender Major Factors In Severity Of Auto-accident Injuries". Purdue University. *ScienceDaily*., Recuperado de: <http://www.sciencedaily.com/releases/2007/01/070105074534.htm>, el día 10 de mayo de 2008.

Johnson, Dallas. (1998). "Métodos multivariados aplicados al análisis de datos". International Thomson Editores. Primera edición.

Levitt, Steven; Porter, Jack. (1999). "Estimating the effect of alcohol on drivers risk using only fatal accident statistics". National Bureau of Economic Research. Investigación #6944. EE-UU.

Lohr, Sharon. (1999). "Muestreo: diseño y análisis". International Thomson Editores. México.

Simon, J. L. (1969). "Basic research methods in social science". New, York: Random House. EE-UU.

Loureiro, Paulo; Sachsida, Adolfo; Moreira, Tito. (2004). "Traffic accidents: an econometric investigation. Economic Bulletin, Vol. 18, #3.

McKnight, James; McKnight Scout. (1991). "The effect of cellular phone use upon driver attention". Nacional Public Services Research Institute. Recuperado de:

<http://www.aaafoundation.org/resources/index.cfm?button=cellphone>, el día 10 de agosto de 2007.

Mooney; Duval (1993). "Bootstrapping. A non parametric approach to statistical inference". SAGE University Paper, series: Quantitative applications in the social sciences, #95. EE-UU.

Reaside, Robert. (2004). "Predicting and monitoring casualty numbers in Great Britain". Journal of Transportation and Statistics. Vol. 7, No. 1. Recuperado de: http://www.bts.gov/publications/journal_of_transportation_and_statistics/volume_07_number_01/, el día 13 de marzo de 2006.

Reglamento de Seguridad Pública, Tránsito y Vialidad del Municipio de Saltillo, Coahuila. Recuperado de: <http://www.ordenjuridico.gob.mx/Estatal/COAHUILA/Municipios/Saltillo/SALReg7.pdf>, el día 17 de octubre de 2007.

Rhum, Christopher. (1995). "Alcohol policies and highway vehicle fatalities". National Bureau of Economic Research. Investigación #5195.

Saffer, Henry. (1997). "Alcohol advertising and motor vehicle fatalities". The Review of Economics and Statistics, Vol. 79, No.3. EE-UU.

Spahr, Ronald; Esolas, Edmond. (1982). "1979 Automobile accident reports: Do driver characteristics support rate discrimination?" Journal of Risk and Insurance, Vol. 49, No. 1.

Sribney, Bill. "Do the svy commands handle zero weights different than non-svy commands do?". Stata Corp. Recuperado de: <http://www.stata.com/support/faqs/stat/zerowgt.html>, el día 30 de enero de 2008.

Suponcic, Susan; Tennyson, Sharon. (1995). "Rate regulation and the industrial organization of automobile insurance". National Bureau of Economic Research, Investigación #5275. EE-UU.

"STATA programming reference. Manual Release 9". (2005). STATA Press. EE-UU.

Stutts, Jane; Huang, Herman; Hunter, William. (2002). "Cell phone use while driving in North Carolina: 2002 Update Report". The University of North Carolina, Highway Safety Research Center.

Stutts, Jane; Feaganes, John; Rodgman, Eric; Hamlett, Charles; Meadows, Thomas; Reinfurt, Donald. (2003). "Distractions in everyday driving". Highway Safety Research Center, University of North Carolina at Chapel Hill. Recuperado de: <http://www.aaafoundation.org/pdf/DistractionsInEverydayDriving.pdf>, el día 10 de agosto de 2007.

Wooldridge, Jeffrey M. (2001). "Introducción a la econometría, un enfoque moderno". Thomson Learning, primera edición, México.

Anexo 1

El Norte.

Son un caos estadísticas municipales

Por Deyra Guerrero

(28-May-2007).-

En el área metropolitana de Monterrey, además de que cada municipio define qué criterios usar para establecer sanciones, también el control estadístico de accidentes, lesionados y muertos por alcohol es un descontrol.

No existe una metodología uniforme para presentar cifras y obtener conclusiones fiables de los operativos.

EL NORTE solicitó información sobre el número de multas y arrestos aplicados a conductores ebrios, así como de accidentes por alcohol, antes y después de las megamultas.

Sólo Monterrey, San Pedro y San Nicolás proporcionaron la información completa, aunque cada uno con sus criterios de organización y diferentes parámetros de comparación.

En Guadalupe, pese a que tienen contabilizados el total de accidentes viales, lesionados y muertos, se afirmó que no cuentan con la cifra de cuántos fueron relacionados con el alcohol.

En Escobedo sólo proporcionaron los accidentes, lesionados y muertos por alcohol de noviembre del 2006 al 24 de mayo del 2007, y argumentaron fallas técnicas para no entregar meses anteriores.

Fuente: Periódico El Norte. 28 de Mayo de 2007.

Anexo 2. Encuesta

1. ¿Usted es conductor de automóviles? Sí No → Termina la encuesta
Edad: _____ Mayor nivel educativo terminado: _____
Sexo: Masculino Femenino Estado civil: _____

2. Cuando usted conduce un automóvil...

- a) ¿**Suele** fumar? Sí No
- b) ¿**Siempre** utiliza el cinturón de seguridad? Sí No
- c) ¿**Suele** cambiar constantemente de discos o de estaciones? Sí No
- d) ¿**Suele** escuchar música a un **volumen alto**? Sí No
- e) ¿**La mayoría de las veces** conduce con niños en el automóvil? Sí No
- f) ¿**Suele** hablar por teléfono celular? Sí No

3. Con respecto a sus hábitos de consumo de alcohol, usted consume **a la semana**:

- i) Ni una cerveza o copa.
- ii) 1 o 2 cervezas o copas.
- iii) de 3 a 7 cervezas o copas.
- iv) Más de 7 cervezas o copas.

4. ¿Usted en promedio cuántas horas maneja **a la semana**? _____

5. ¿En el **último año** ha sufrido usted un accidente vial? Sí No
6. ¿Conduce usted un automóvil propio? Sí No
7. ¿Cuenta usted con seguro de automóvil? Sí No
8. ¿De qué **color** es el automóvil que conduce? _____
- 8.1 ¿Tonalidad? a) Claro b) Oscuro
9. ¿Cuál es la marca de su automóvil (Ford, Chrysler, Chevrolet, etc.)? _____
10. ¿Qué **año** es el modelo del automóvil que maneja? _____

En caso de que la persona **no haya chocado** termina la encuesta

11. La magnitud del accidente fue:

- a) Apenas un raspón b) Más o menos fuerte c) Considerablemente fuerte

12. ¿El accidente lo sufrió...?

- a) Entre semana b) En fin de semana
- I) Durante la mañana (7-11.59 AM) II) tarde (12-7.59 PM) III) Noche (8 PM- 6 AM)

13. ¿Recuerda aproximadamente la hora en que sucedió el accidente? _____

14. ¿Cuando chocó había llovido o estaba mojado el pavimento? Sí No

Anexo 3. Matriz de correlación

	edad	matr_jequi	servica	precio	prestado	numero	sistema	suma	californ	disco	musica	mas	celular	l_moderada	l_ruido	l_suerte	accidents	auto	precio	seguro	rojo	shakado	arrojadas	chevrolet	casan	vw	toy	chrysler					
edad	1																																
matr_jequi	0.0927	1																															
servica	0.0742	-0.2168	1																														
precio	-0.222	-0.2825	-0.2383	1																													
prestado	0.1736	-0.1131	-0.0814	-0.1243	1																												
numero	-0.536	0.107	-0.06	0.0756	0.0114	1																											
sistema	-0.4571	-0.0534	-0.1529	0.1457	0.1075	0.0485	1																										
suma	-0.0465	0.1347	-0.0528	0.0124	0.0465	0.1195	-0.005	1																									
californ	0.0185	-0.1037	0.0159	0.0126	-0.1443	-0.0939	-0.0965	-0.2237	1																								
disco	-0.3462	-0.1193	-0.1068	0.1588	-0.3054	0.2085	0.0857	-0.1186	-0.1186	1																							
musica	-0.1235	0.0071	0.0268	-0.0625	-0.0891	0.0001	0.259	-0.1909	0.3567	-0.0537	1																						
mas	-0.0126	-0.1363	-0.1421	0.2156	0.2441	0.0512	0.1061	-0.0654	-0.0775	0.1563	-0.0708	1																					
celular	-0.0413	-0.054	-0.0218	0.0375	0.0265	0.127	0.0242	-0.0902	0.0542	0.0094	0.0094	-0.0018	0.0087	1																			
l_moderada	-0.0307	0.0091	-0.0619	-0.04	-0.0510	0.2407	-0.0054	0.1015	-0.0403	0.0112	0.025	-0.0446	0.1246	-0.2213	1																		
l_ruido	-0.0531	0.0736	0.0378	0.0636	0.0022	0.2368	0.3462	0.1015	-0.2116	0.148	0.0682	-0.1325	-0.0297	-0.1763	-0.1429	1																	
accidents	-0.1182	-0.0775	-0.0426	0.0205	0.0029	-0.2147	0.1397	0.0317	0.0167	0.0246	0.0467	-0.1428	0.1079	-0.0223	-0.0222	0.5474	1																
auto	0.2032	-0.1124	0.0064	-0.077	0.0424	-0.0466	-0.1169	0.0715	-0.0243	-0.0102	0.1079	0.047	-0.0179	0.047	-0.015	-0.0424	0.0752	1															
precio	-0.0206	-0.2364	-0.0233	0.1238	0.037	-0.1083	-0.0089	-0.1088	0.1041	0.0718	0.017	0.0146	0.1047	-0.0153	-0.015	-0.0424	0.0752	0.0944	1														
seguro	0.0387	0.0263	-0.0332	-0.0199	0.0243	-0.0632	-0.0316	-0.0185	0.0486	0.0303	-0.0424	-0.0571	-0.066	-0.0094	-0.015	-0.0424	0.0752	0.0944	0.1609	1													
rojo	0.0718	0.07	-0.023	-0.0692	-0.0471	0.0001	0.0374	-0.0622	0.0706	0.0303	-0.0424	-0.0571	-0.066	-0.0094	-0.015	-0.0424	0.0752	0.0944	0.1609	-0.0574	1												
shakado	0.173	0.2975	-0.1193	-0.0523	0.0041	0.1844	-0.1052	0.0668	-0.016	-0.0068	-0.0068	-0.0068	-0.1929	-0.0239	-0.0139	0.034	0.0262	-0.2815	0.344	-0.039	1												
arrojadas	-0.0751	0.0461	-0.0075	0.034	-0.0624	0.059	0.0148	0.0415	0.0562	0.0413	-0.0157	-0.0297	-0.0453	-0.0039	-0.0179	-0.0058	0.0445	-0.2815	0.344	-0.039	-0.0332	1											
chevrolet	-0.0714	-0.048	-0.033	0.0485	0.005	0.0017	0.0273	-0.0033	-0.0167	0.0194	-0.1944	-0.0003	0.0574	0.0002	-0.0111	0.0728	0.1567	-0.1140	0.0464	-0.0395	0.0899	-0.0277	1										
casan	-0.0701	-0.018	-0.014	0.017	0.0516	-0.0199	0.04	0.0544	-0.0227	0.0143	-0.0002	-0.1048	0.0114	-0.0013	-0.0058	-0.0169	0.0089	0.0544	-0.0395	0.0899	-0.0277	-0.0254	-0.1685	1									
vw	-0.0714	0.0619	-0.039	-0.0379	-0.0285	-0.0643	0.0197	-0.0251	-0.0288	0.0386	0.1355	0.0284	-0.0095	0.0681	0.0703	-0.0058	0.0278	-0.0418	0.0157	-0.145	0.1757	-0.2872	-0.2156	-0.0126	1								
toy	0.0714	0.0619	-0.039	-0.0379	-0.0285	-0.0643	0.0197	-0.0251	-0.0288	0.0386	0.1355	0.0284	-0.0095	0.0681	0.0703	-0.0058	0.0278	-0.0418	0.0157	-0.145	0.1757	-0.2872	-0.2156	-0.0126	-0.1816	1							
chrysler	0.0059	0	0.0706	-0.008	-0.0407	-0.006	-0.0414	-0.054	-0.0114	-0.0252	-0.0043	0.0532	-0.2119	-0.0095	-0.0376	-0.0396	-0.0298	0.0059	-0.0069	0.0395	-0.0933	0.0296	-0.1548	-0.137	-0.1816	1							