

UNIVERSIDAD DE SANTIAGO DE CHILE
FACULTAD DE CIENCIA
DEPARTAMENTO DE MATEMÁTICA Y CIENCIA DE LA COMPUTACIÓN



Evaluación del impacto de la ley Emilia en los accidentes de tránsito usando series de tiempo interrumpidas

Darling Navarrete Soto

Profesor Guía: Felipe Elorrieta López

Trabajo de Titulación presentado en conformidad a los requisitos
para obtener el grado de Ingeniero Estadístico

SANTIAGO - CHILE

2019

© 2019, Darling Navarrete Soto

Se autoriza la reproducción parcial de este escrito, con fines académicos, por cualquier forma, medio o procedimiento, siempre y cuando se incluya la cita bibliográfica del documento.

Evaluación del impacto de la ley Emilia en los accidentes de tránsito usando series de tiempo interrumpidas

Seminario de Tesis

Este trabajo de Titulación fue preparado bajo la supervisión del Profesor guía Felipe Elorrieta López del Departamento de Matemática y Ciencias de la Computación y ha sido aprobado por los miembros de la comisión, de la candidata señorita Darling Navarrete Soto, la señora Claudia Matus Correa y el señor Claudio Vargas.

Sr. Felipe Elorrieta L.
Profesor guía

Sra. Claudia Matus C.
Profesora comisión

Sr. Claudio Vargas R.
Profesor comisión

Sr. Rafael Labarca B.
Director del Departamento

Agradecimientos

En primer lugar, quiero expresar mi eterno agradecimiento a mi profesor guía Felipe Elorrieta, por su enorme paciencia, voluntad, apoyo y entrega en todo este proceso, porque de seguro fui la alumna que más veces fue a visitarlo a su oficina, la más preguntona y también la más divertida, sin su apoyo y tiempo invertido en mí, nada de esto sería posible, agradezco también todos estos años en los que fui su alumna, siempre sentí su amor por lo que hace, preocupación por sus alumnos, su dedicación y por sobre todo su humanidad. Nunca olvidare estos meses de sufrimiento, pero tampoco olvidare que con usted a mi lado todo fue muchísimo más fácil. Mi agradecimiento eterno es suyo.

Yerald Aguilar que sería de mi vida sin ti. La palabra incondicional la conocí contigo, cada día que vivo a tu lado me asombro aún más de lo hermosa persona que eres, eternamente daré gracias a la vida por haberte conocido, y también que me hayas acompañado en todos mis años de carrera, emocionalmente hiciste que este camino fuera mucho más fácil. Decidimos durante estos años tomar nuestro propio rumbo y formar nuestro hogar junto a Simba y han sido los mejores años de mi vida, tú mejor que nadie sabía lo importante que era para mí dar ese paso, porque las peores cosas también las vivimos juntos, estuviste para darme el aliento necesario para terminar este proceso y para seguir adelante en todos los sentidos, Simba y yo realmente somos muy afortunados, eres un grande y todo este trabajo está dedicado a ti. Se que no eres perfecto, pero en mi mundo tu eres la perfección. Te amo.

Padre mío, eres mi padre y madre a la vez, todo lo que he necesitado para ser quien soy me lo has dado tú. Mis años de carrera tuvieron más bajos que altos, pero tuve la fortuna de siempre poder contar contigo, las cosas fueron difícil para ambos y para ti lo siguen siendo, pero de corazón espero que la vida me dé la oportunidad de poder devolverte la mano y que al menos un par de años en tu vida sean un poco más fácil, sacarte un poco de carga y que no tengas tanta presión en tus hombros. Eres un hombre que tiene una bondad de la que yo tengo mucho que aprender, un papá del que mis hermanos y yo nos sentimos eternamente agradecidos y orgullosos y que sin ti el barco se hubiese hundido hace rato. Espero con esto poder darte un poco de satisfacción y que sientas que al menos por mi parte, valió la pena todo

el sufrimiento y sacrificio que has tenido. Gracias por ser como eres y por siempre estar ahí, TE AMO PAPÁ.

Quiero agradecer también a mis hermanos Ignacio, Matias y Bastian, que son parte fundamental en mi vida. Son una motivación para ser mejor persona y espero que esto los pueda motivar a pensar que todo lo que se propongan lo pueden lograr, solo tienen que tener determinación y esforzarse por sus sueños, que sean felices y siempre me tendrán aquí para lo que necesiten.

Como olvidar a Ignacio Calderón quien hizo de estos años universitario inolvidables, tenemos tantas historias que contar juntos que podría de seguro escribir un libro. Este tiempo fue de un apañe constante, juntos fuimos uno, lo que le pasaba a uno también le pasaba al otro y tuvimos la inmensa fortuna de terminar este largo camino juntos. Creo que el destino pone a la persona correcta en tu camino, para hacerlo más fácil y contigo no se puede definir de otra manera, estuviste en las buenas y malas, reímos eternamente, compartimos también todos los momentos de mayor estrés y aprendí tantas cosas como por ejemplo el amor por viajar, la determinación por lograr lo que te propongas, y que por sobre todo que puedo contar contigo así como tú puedes hacerlo conmigo, todo esto será inolvidable para mí y espero que lo sea también para ti. Deseo lo mejor en tu vida para este nuevo comienzo y que por sobre todo seas feliz. Te adoro baby!

Hay profesores que pueden marcar la vida y la profesora Gabriela Valdés justamente logro esto conmigo, es una profesora realmente admirable, con un amor único por la carrera, y por sobre todo una persona con un corazón gigante. Recuerdo su contención en momentos difíciles, su preocupación por cada uno de sus alumnos y la dedicación en cada actividad que tenía preparada. Como olvidar haber tenido una clase donde el tío Mario, a cada uno le compró algo para compartir, y de seguro esa fue la mejor catedra en todos estos años de universidad, ahí aprendí que un profesor no deja de ser humano y que la educación no se limita solo a la académico, que cuando algo se ama se transmite correctamente. Agradezco haberla conocido en este proceso y que este amor por la carrera se lo debo a cada una de sus clases.

Finalmente agradezco a Noelia y Tamara, ellas estuvieron en el proceso más difícil que me tocó vivir, si no hubiese sido por ellas probablemente no estaría ahora escribiendo esto, quizás las personas no dimensionan cuanto puede significar en otras su apoyo o sus palabras, sus familias estuvieron incondicionalmente para mí y esa red de apoyo que me brindaron jamás se olvida, estaré eternamente agradecida de haber conocido a estos dos ángeles.

Resumen

El 21 de enero del año 2013 Emilia Figueroa Silva de 9 meses murió, luego de que Nelson Fariña impactara el auto de sus padres con 1.9 gramos del alcohol por litro de sangre dándose a la fuga. Fue condenado a dos años de pena remitida y el pago de una multa de 8 UTM, en ese entonces la legislación no contemplaba penas efectivas para este tipo de delitos. Es por esto que se creó la ley Emilia que corresponde a una modificación a la ley de tránsito en lo que refiere al delito de manejo en estado de ebriedad causando lesiones graves, gravísimas o con resultado de muerte, obligando a detener la marcha del vehículo, prestar ayuda y dar cuenta a la autoridad en todo accidente que se produzcan lesiones.

Para analizar el impacto que ha tenido la entrada en vigencia de esta ley (16 de septiembre del año 2014), se utilizó series de tiempo interrumpidas (ITS), que consiste en interrumpir la serie en un tiempo definido y se espera como consecuencia que los datos reflejen un cambio de nivel o tendencia. Este estudio se enfocó en muertes y hospitalizaciones causadas por accidentes de tránsito, donde se obtuvo la frecuencia semanal de cada una de ellas para crear la serie temporal y modelos a utilizar. Dado que la variable respuesta es de tipo conteo se utilizó la regresión binomial negativa, ya que los datos poseen sobredispersión.

Del estudio se obtuvo que en las muertes analizadas a lo largo de todo el país luego de un año de la implementación de la ley, disminuyeron un 0.79%, mientras que en la región Metropolitana este decremento fue de un 3.3%. Por otra parte, las hospitalizaciones causadas por accidentes de tránsito a lo largo de todo el país tuvieron una disminución de un 1.8%, pero este análisis en la Región Metropolitana tuvo un aumento de un 1.4%, lo que puede deberse a la obligación de detener la marcha del vehículo y prestar ayuda.

Es importante mencionar que si bien, existió una disminución tanto en muertes como en hospitalizaciones (a excepción de hospitalizaciones en la Región Metropolitana) posterior a la implementación de la ley, no representa un impacto cuantioso en las condiciones que se producen los accidentes de tránsito en nuestro país.

Palabras claves: Ley Emilia, accidentes de tránsito, ITS, regresión binomial negativa.

Índice general

Índice general	5
1. Introducción	7
1.1. Motivación	7
1.2. Planteamiento del Problema	8
1.3. Objetivos	9
1.3.1. Objetivo General	9
1.3.2. Objetivos Específicos	9
1.4. Metodología de Trabajo	9
2. Marco Teórico	11
2.1. Modelos de Regresión	11
2.1.1. Regresión lineal simple	11
2.1.1.1. Supuestos del Modelo	12
2.1.1.2. Estimación de los parámetros	12
2.1.1.2.1. Estimación de mínimos cuadrados	12
2.1.1.3. Bondad de ajuste	13
2.1.1.4. Error estándar de la regresión	13
2.1.2. Regresión múltiple	13
2.1.2.1. Estimación de parámetros a través de mínimos cuadrados	14
2.1.2.2. Valores ajustados y validación cruzada	15
2.1.3. Regresión no lineal	15
2.1.4. Modelos lineales generalizados (GLM)	17
2.1.5. Familia Exponencial	18
2.1.5.1. Regresión Poisson	19
2.1.5.2. Regresión Binomial Negativa	20

2.2.	Conceptos previos de series de tiempo	20
2.2.1.	Series de Tiempo	20
2.2.2.	Proceso Estocástico	21
2.2.3.	Estacionaridad	21
2.2.4.	Función de autocovarianza y función de autocorrelación	22
2.2.4.1.	Función de autocovarianza	22
2.2.4.2.	Función de autocorrelación	22
2.2.5.	Componente Aleatorio	22
2.2.6.	Procesos Estacionarios	23
2.2.6.1.	Modelo Ar(p)	23
2.2.6.2.	Modelo MA(q)	24
2.2.7.	Procesos Lineales no estacionarios	24
2.2.7.1.	Proceso autoregresivo integrado de medias móviles <i>ARIMA</i> (p, d, q)	24
2.2.8.	Tendencia	25
2.2.9.	Estacionalidad	26
2.3.	Modelamiento y pruebas para un cambio estructural	28
2.3.1.	Prueba de Chow para parámetros diferentes	29
2.3.2.	Prueba de Chow para el coeficiente de pendiente	31
2.3.3.	Pruebas de Intercepto	32
2.4.	Modelos de series de tiempo interrumpidas	33
2.4.1.	Modelo de regresión	34
2.4.2.	Modelo de regresión segmentada	35
2.4.3.	Consideraciones metodológicas en el análisis de ITS	36
2.4.3.1.	Estacionalidad	36
2.4.3.2.	Variables de Confusión	36
2.4.3.3.	Sobredispersión	37
2.4.3.4.	Autocorrelación	37
3.	Análisis de muertes por accidentes de tránsito	38
3.1.	Serie de tiempo de muertes por accidentes de tránsito	42
3.1.1.	Tendencia de la serie	44
3.1.2.	Test de Chow	45
3.2.	Serie de Muertes por accidentes de tránsito en la Región Metropolitana	47

3.2.1.	Representación gráfica de la tendencia	48
3.2.2.	Test de Chow	49
3.3.	Serie de muertes en peatones y ocupante de vehículos causadas por accidentes de tránsito	51
3.3.1.	Serie de Peatones causadas por accidentes de tránsito	51
3.3.1.1.	Representación gráfica de la tendencia	52
3.3.1.2.	Test de Chow	52
3.3.2.	Serie de muertes de ocupante de vehículo causadas por accidentes de tránsito	54
3.3.2.1.	Representación gráfica de la tendencia	55
3.3.2.2.	Test de Chow	56
4.	Análisis de hospitalizaciones por accidentes de tránsito	59
4.1.	Serie de tiempo de hospitalizaciones causadas por accidentes de tránsito	61
4.1.1.	Tendencia de la serie	62
4.1.2.	Test de Chow	63
4.2.	Serie de tiempo de hospitalizaciones causadas por accidentes de tránsito en la Región Metropolitana	65
4.2.1.	Gráfico de la tendencia de la serie	65
4.2.2.	Test de Chow	66
5.	Modelos de ITS	68
5.1.	Modelos de ITS de muertes causadas por accidentes de tránsito	68
5.1.1.	Base de datos y análisis descriptivos	68
5.1.2.	Formulación modelos ITS	69
5.2.	Análisis de Supuestos del Modelo	73
5.2.1.	Supuestos	73
5.2.1.1.	Test de Box-Ljung	73
5.2.1.2.	Test de Breusch-Pagan	74
5.2.1.3.	Tabla resumen supuestos de los modelos	74
5.2.1.4.	Serie de tiempo muertes y su tendencia estimada	75
5.2.2.	Evaluación del impacto de la ley Emilia	76
5.3.	Modelos de ITS de muertes causadas por accidentes de tránsito en la Región Metropolitana	77
5.3.1.	Base de datos y análisis descriptivos	77
5.3.2.	Formulación modelos ITS	78

5.4.	Análisis de Supuestos del Modelo	81
5.4.1.	Supuestos	82
5.4.1.1.	Test de Box-Ljung	82
5.4.1.2.	Test de Breusch-Pagan	82
5.4.1.3.	Tabla resumen supuestos de los modelos	83
5.4.1.4.	Serie de tiempo muertes en la RM su tendencia estimada	83
5.4.2.	Evaluación del impacto de la ley Emilia	84
5.5.	Modelos de ITS de muertes en peatones y ocupante de vehículos causadas por accidentes de tránsito	86
5.5.1.	Modelos de ITS de muertes de peatones causadas por accidentes de tránsito	86
5.5.1.1.	Base de datos y análisis descriptivos	86
5.5.1.2.	Formulación modelos ITS	87
5.5.1.3.	Análisis de Supuestos del Modelo	90
5.5.1.4.	Supuestos	91
5.5.1.4.1.	Test de Box-Ljung	91
5.5.1.4.2.	Test de Breusch-Pagan	91
5.5.1.4.3.	Tabla resumen supuestos de los modelos	92
5.5.1.4.4.	Serie de tiempo muertes de peatones y su tendencia estimada	92
5.5.2.	Modelos de ITS de muertes de ocupante de vehículo causadas por accidentes de tránsito	93
5.5.2.1.	Base de datos y análisis descriptivos	93
5.5.2.2.	Formulación modelos ITS	94
5.5.3.	Análisis de Supuestos del Modelo	98
5.5.3.1.	Supuestos	98
5.5.3.1.1.	Test de Box-Ljung	98
5.5.3.1.2.	Test de Breusch-Pagan	99
5.5.3.1.3.	Tabla resumen supuestos de los modelos	99
5.5.3.1.4.	Serie de tiempo muertes ocupante de vehículo y su tendencia estimada	100
5.5.4.	Evaluación del impacto de la ley Emilia	101
6.	Modelos de ITS de hospitalizaciones causadas por accidentes de tránsito	103
6.1.	Base de datos y análisis descriptivo	103
6.2.	Modelos para series de tiempo interrumpidas	104

6.3. Análisis de Supuestos del Modelo	107
6.4. Supuestos	108
6.4.1. Test de Box-Ljung	108
6.4.2. Breusch-Pagan	108
6.4.2.1. Tabla resumen supuestos de los modelos	109
6.4.3. Serie de tiempo hospitalizaciones y su tendencia	109
6.4.4. Evaluación del impacto de la ley Emilia	110
6.5. Modelos de ITS de hospitalizaciones causadas por accidentes de tránsito en la Región Me- tropolitana	112
6.5.1. Base de datos y análisis descriptivo	112
6.5.2. Modelos para series de tiempo interrumpidas	113
6.6. Análisis de Supuestos del Modelo	116
6.6.1. Supuestos	117
6.6.1.1. Test de Box-Ljung	117
6.6.1.2. Breusch-Pagan	117
6.6.1.3. Tabla resumen supuestos de los modelos	118
6.6.2. Serie de tiempo hospitalizaciones y su tendencia	118
6.6.3. Evaluación del impacto de la ley Emilia	119
7. Conclusiones	121
Bibliografía	123
8. Anexo	125

Índice de Figuras

2.1. Cambio Estructural	29
2.2. Ejemplos de modelos de impacto usando ITS	34
3.1. Serie de muertes por accidentes de tránsito	43
3.2. Ventana de muertes por accidentes de tránsito	44
3.3. Ventana de la tendencia con sus quiebres	45
3.4. Serie muertes por accidentes de tránsito en la RM	48
3.5. Tendencia de la serie en la RM	49
3.6. Serie muertes de peatones causadas por accidentes de tránsito	51
3.7. Tendencia de la serie de muertes de peatones	52
3.8. Serie muertes de ocupante de vehículo causadas por accidentes de tránsito	55
3.9. Tendencia de la serie en muertes de ocupante de vehículo	56
4.1. Serie de tiempo hospitalizaciones causadas por accidentes de tránsito	61
4.2. Serie de tiempo hospitalizaciones causadas por accidentes de tránsito	62
4.3. Tendencia serie hospitalizaciones	63
4.4. Serie de tiempo hospitalizaciones en la Región Metropolitana	65
4.5. Tendencia serie de tiempo hospitalizaciones en la Región Metropolitana	66
5.1. Serie de tiempo muertes y tendencias estimada	75
5.2. Comparación estimaciones del mejor modelo sin intervención versus mejor modelo con intervención	76
5.3. Serie de tiempo muertes en la RM y su tendencia	84
5.4. Comparación estimaciones del mejor modelo sin intervención versus mejor modelo con intervención	85
5.5. Serie de tiempo muertes de peatones y su tendencia	93

5.6. Serie de tiempo muertes ocupante de vehículo y su tendencia	101
5.7. Comparación de las estimaciones del impacto de la ley en los peatones versus ocupantes de vehículo	102
6.1. Serie de tiempo hospitalizaciones y su tendencia	110
6.2. Comparación estimaciones del mejor modelo sin intervención versus con intervención . . .	111
6.3. Serie de tiempo hospitalizaciones y su tendencia	119
6.4. Comparación estimaciones del mejor modelo sin intervención versus con intervención . . .	120

Índice de Tablas

3.1. Tabla de resultados para calcular test de Chow en muertes	46
3.2. Tabla de resultados para calcular test de Chow en muertes, segunda intervención	47
3.3. Tabla de resultados para calcular test de Chow de muertes en la RM	49
3.4. Tabla de resultados para calcular test de Chow de muertes en la RM, segunda intervención	50
3.5. Tabla de resultados para calcular test de Chow de muertes de peatones	53
3.6. Tabla de resultados para calcular test de Chow de muerte de peatones, segunda intervención	53
3.7. Tabla de resultados para calcular test de Chow en muertes de ocupante de vehículo	56
3.8. Tabla de resultados para calcular test de Chow en muertes de ocupante de vehículo, segunda intervención	57
4.1. Tabla de resultados para calcular test de Chow en hospitalizaciones	63
4.2. Tabla de resultados para calcular test de Chow en hospitalizaciones segunda intervención .	64
4.3. Tabla de resultados para calcular test de Chow en hospitalizaciones de la RM	66
4.4. Tabla de resultados para calcular test de Chow en hospitalizaciones de la RM, segunda intervención	67
5.1. Estadística descriptiva	69
5.2. Resumen modelos propuestos	69
5.3. Resumen Modelo 1 muertes	69
5.4. Resumen Modelo 2 muertes	70
5.5. Resumen Modelo 3 muertes	70
5.6. Resumen Modelo 4 muertes	71
5.7. Resumen Modelo 5 muertes	71
5.8. Resumen Modelo 6 muertes	72
5.9. Resumen Modelo 7 muertes	72
5.10. Tabla resumen AIC y SCR	73

5.11. Tabla resumen Test de Box-Ljung	73
5.12. Tabla resumen Test de Breusch-Pagan	74
5.13. Tabla resumen supuestos de los modelos	74
5.14. Tabla resumen total de muertes para los modelos	77
5.15. Estadística descriptiva	78
5.16. Estadística descriptiva	78
5.17. Resumen Modelo 1 muertes RM	78
5.18. Resumen modelo glm.nb(muerte~tiempo)	79
5.19. Resumen Modelo 3 muertes RM	79
5.20. Resumen Modelo 4 muertes RM	79
5.21. Resumen Modelo 5 muertes RM	80
5.22. Resumen Modelo 6 muertes RM	80
5.23. Resumen Modelo 7 muertes RM	81
5.24. Tabla resumen AIC y SCR	81
5.25. Tabla resumen Test de Box-Ljung	82
5.26. Tabla resumen Test de Breusch-Pagan	82
5.27. Tabla resumen supuestos de los modelos	83
5.28. Tabla resumen total de muertes para los modelos	85
5.29. Estadística descriptiva	87
5.30. Tabla resumen modelos propuestos	87
5.31. Resumen Modelo 1 muertes peatones	87
5.32. Resumen Modelo 2 muertes peatones	88
5.33. Resumen Modelo 3 muertes peatones	88
5.34. Resumen Modelo 4 muertes peatones	89
5.35. Resumen Modelo 5 muertes peatones	89
5.36. Resumen Modelo 6 muertes peatones	90
5.37. Resumen Modelo 7 muertes peatones	90
5.38. Tabla resumen AIC y SCR	91
5.39. Tabla resumen Test de Box-Ljung	91
5.40. Tabla resumen Test de Breusch-Pagan	92
5.41. Tabla resumen supuestos de los modelos	92
5.42. Estadística descriptiva	94

5.43. Tabla resumen modelos propuestos	94
5.44. Resumen Modelo 1 muertes ocupante de vehículo	95
5.45. Resumen Modelo 2 muertes ocupante de vehículo	95
5.46. Resumen Modelo 3 muertes ocupante de vehículo	95
5.47. Resumen Modelo 4 muertes ocupante de vehículo	96
5.48. Resumen Modelo 5 muertes ocupante de vehículo	96
5.49. Resumen Modelo 6 muertes ocupante de vehículo	97
5.50. Resumen Modelo 7 muertes ocupante de vehículo	97
5.51. Tabla resumen AIC y SCR	98
5.52. Tabla resumen Test de Box-Ljung	99
5.53. Tabla resumen Test de Breusch-Pagan	99
5.54. Tabla resumen supuestos de los modelos	100
5.55. Tabla resumen total de muertes de peatones y ocupante de vehículo luego de un año de la implementación de la ley Emilia	102
6.1. Estadística descriptiva	104
6.2. Tabla resumen modelos propuestos	104
6.3. Resumen Modelo 1 hospitalizaciones	104
6.4. Resumen Modelo 2 hospitalizaciones	105
6.5. Resumen Modelo 3 hospitalizaciones	105
6.6. Resumen Modelo 4 hospitalizaciones	105
6.7. Resumen Modelo 5 hospitalizaciones	106
6.8. Resumen Modelo 6 hospitalizaciones	106
6.9. Resumen Modelo 7 hospitalizaciones	107
6.10. Tabla resumen AIC y SCR	107
6.11. Tabla resumen Test de Box-Ljung	108
6.12. Tabla resumen Test de Breusch-Pagan	108
6.13. Tabla resumen supuestos de los modelos	109
6.14. Tabla resumen total de hospitalizaciones un año después de la implementación de la ley Emilia	111
6.15. Estadística descriptiva	112
6.16. Tabla resumen modelos propuestos	113
6.17. Resumen Modelo 1 hospitalizaciones en la RM	113
6.18. Resumen Modelo 2 hospitalizaciones RM	113

6.19. Resumen Modelo 3 hospitalizaciones en la RM	114
6.20. Resumen Modelo 4 hospitalizaciones en la RM	114
6.21. Resumen Modelo 5 hospitalizaciones en la RM	115
6.22. Resumen Modelo 6 hospitalizaciones en la RM	115
6.23. Resumen Modelo 7 hospitalizaciones en la RM	116
6.24. Tabla resumen AIC y SCR	116
6.25. Tabla resumen Test de Box-Ljung	117
6.26. Tabla resumen Test de Breusch-Pagan	117
6.27. Tabla resumen supuestos de los modelos	118
6.28. Tabla resumen total de hospitalizaciones en la Región Metropolitana un año después de la implementación de la ley Emilia	120

Capítulo 1

Introducción

1.1. Motivación

El día 21 de enero del 2013 a raíz de un fatal accidente de tránsito, Emilia Silva Figueroa de 9 meses de vida, murió luego de que Nelson Fariña Jara impactara el auto de sus padres con 1.9 gramos de alcohol por litro de sangre y posteriormente se diera a la fuga en la comuna de Vitacura. Luego del fallecimiento de la pequeña Emilia, Fariña fue condenado a dos años de pena remitida y el pago de una multa de 8 UTM por conducción en estado de ebriedad, debido a esto los padres decidieron iniciar una campaña en redes sociales llamada “Ley Emilia”, ya que, en ese entonces la legislación no contemplaba penas efectivas para este tipo de asesinatos.

El 16 de septiembre del 2014 se implementó una modificación a la ley de tránsito (Ley N 18.290) llamada “Ley Emilia” (Ley N 20.770), en lo que se refiere al delito de manejo en estado de ebriedad, causando lesiones graves, gravísimas o, con resultado de muerte. Dicha ley obliga entre otras cosas a: “Detener la marcha del vehículo, prestar la ayuda posible y dar cuenta a la autoridad de todo accidente en que se produzcan lesiones”(1); lo que establece sanciones más duras a quienes manejen en estado de ebriedad y provoquen un accidente, además de tipificar como “delito” huir del lugar y negarse a la realización del alcotest.

Para analizar el impacto que ha tenido en los accidentes de tránsito la entrada en vigencia de la Ley Emilia, se usará el enfoque de serie temporal interrumpida (ITS). { *ITS es un diseño que registra gran cantidad de observaciones antes y después de aplicar un tratamiento. La serie se interrumpe en un punto claramente definido en el tiempo y se espera como consecuencia, que los datos reflejen un cambio de nivel,*

tendencia, o ambos (2). Esto permite que se pueda evaluar la presencia o eficacia del tratamiento. En este caso el tratamiento es la entrada en vigencia de la ley, donde se tiene bien definido en el tiempo de pre y post intervención.

1.2. Planteamiento del Problema

En Chile 1 de cada 10 accidentes con resultado de muerte se debe al consumo de alcohol en los conductores. Durante los últimos años se han logrado importantes avances en materia de leyes por parte del Ministerio de Transporte y Telecomunicaciones, a través de CONASET (Comisión Nacional de Seguridad de Tránsito), con la implementación de la “Ley tolerancia cero en el año 2012 y su posterior modificación con la entrada en vigencia de la “Ley Emilia”.(3)

Desde el 2012 según CONASET se registraron bajas importantes en lo que se refiere a fallecidos en accidentes de tránsito ligados al consumo de alcohol en los conductores, llegando a 126 víctimas en el año 2017, la cifra más baja en los últimos 30 años.

Dentro de los métodos de estudio para analizar esta problemática se considerará el enfoque de las ITS. El cual es útil para detectar cambios en la variable dependiente que se relacione con el tratamiento.

Una ITS estándar utiliza el siguiente modelo de regresión segmentada (4):

$$Y_t = \beta_0 + \beta_1 T + \beta_2 X_t + \beta_3 T X_t$$

Donde Y_t es el resultado en el tiempo t , T es el tiempo transcurrido desde el inicio del estudio con la unidad que representa la frecuencia con la que se toman las observaciones (mes o año) y X_t es una variable ficticia que indica el periodo antes del tratamiento (codificado en 0) o el periodo posterior al tratamiento (codificado en 1). Por otra parte β_0 , representa el nivel de la línea base a $T = 0$, β_1 , se interpreta a medida que aumenta el cambio en el resultado asociado con una unidad de tiempo, β_2 es el cambio de nivel que ocurre después del tratamiento y β_3 indica el cambio de pendiente después de la intervención. Cabe destacar que la fórmula anterior es utilizada solo cuando los datos tienen una tendencia lineal, pero en

el desarrollo del estudio se tendrán datos con 3 enfoques distintos, lo que nos llevará analizar datos con distinta tendencia.

El fin de este estudio es que, a través de este tipo de modelo, se pueda determinar el impacto a nivel social que ha tenido la implementación de esta nueva ley y evaluar si han disminuidos los accidentes de tránsito provocados por conductores que están bajo la influencia del alcohol.

1.3. Objetivos

1.3.1. Objetivo General

- Determinar el efecto de la ley Emilia en la tasa de accidentes de tránsito usando series de tiempo interrumpidas.

1.3.2. Objetivos Específicos

- Discutir tipos de análisis de modelos de series de tiempo interrumpidas.
- Proponer un modelo de serie de tiempo interrumpidas basado en un análisis descriptivo.
- Realizar análisis de mortalidad y hospitalizaciones con datos recolectados.
- Ajustar el modelo propuesto.
- Determinar los factores que inciden en el cambio de comportamiento de la serie.
- Validar el modelo propuesto.

1.4. Metodología de Trabajo

En primera instancia, se realizará una revisión bibliográfica en la literatura pertinente sobre series de tiempo interrumpidas, lo que nos permitirá obtener el conocimiento teórico y poder entender cómo se construyen este tipo de modelos, para su posterior implementación.

Luego de la implementación y de haber elegido el diseño de la ITS se debe crear una hipótesis de cómo el tratamiento afectaría al resultado si fuera eficaz, y evaluar si el cambio será gradual en el gradiente de

la tendencia, o en el nivel o en ambos, y si se mostrará inmediatamente o habrá un periodo de retraso en el efecto. Esta decisión se debe realizar en base a la literatura existente y el conocimiento.

Por otro lado, se debe realizar un análisis estadístico descriptivo que debe incluir diagramas de resumen, gráficos de dispersión de la serie, con esto identificar la tendencia, los patrones estacionales, ver si existen valores atípicos y realizar comparaciones bivariadas.

Posteriormente se debe crear el modelo de regresión y, dado que se tienen datos de tipo discreto (N de accidentes por conducción ligada al alcohol en un tiempo determinado) se usará un modelo de regresión Binomial Negativo debido a la sobredispersión en los datos.

Existen una serie puntos metodológicos a evaluar en series de tiempo para mejorar la solidez del análisis, la estacionalidad por ejemplo puede causar problemas, si existe una distribución desigual antes y después del tratamiento, lo que lleva a la autocorrelación y dispersión excesiva. La estacionalidad podría solucionarse con un modelo estratificado por algún periodo de tiempo o utilizando funciones más complejas como los términos de Fourier o splines. La dispersión excesiva que provocaría una estimación incorrecta de los errores estándar se debe solucionar realizando un ajuste de escala para corregir el modelo. La autocorrelación se explica en gran medida por la estacionalidad, por lo tanto, después de controlar este factor esto deja de ser un problema. (4)

Este procedimiento se debe realizar de forma independiente para accidentes de tránsito bajo la influencia del alcohol con, resultado de muerte, lesiones graves y lesiones gravísimas (5). Y así en el estudio evaluar diferentes tipos de tendencias y que el modelo para los tres enfoques sea distinto.

Capítulo 2

Marco Teórico

2.1. Modelos de Regresión

En un modelo de regresión, el concepto básico es que pronosticamos las series de tiempo de interés asumiendo que tiene una relación lineal con otras series de tiempo. La variable de pronóstico también se denomina variable regresiva o dependiente, mientras que la variable de predicción se denomina variable independiente o explicativa.

2.1.1. Regresión lineal simple

El modelo de regresión simple permite una relación lineal entre la variable de pronóstico y y una variable predictiva única X : (11)

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t$$

Los coeficientes β_0 y β_1 denotan la intersección y la pendiente de la línea respectivamente. La intersección β_0 representa el valor predicho de y cuando $x=0$. La pendiente β_1 representa el cambio promedio predicho en y que resulta del incremento en una unidad de x .

Este modelo podemos expresarlo matricialmente de la siguiente manera:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

2.1.1.1. Supuestos del Modelo

Cuando usamos un modelo de regresión lineal, implícitamente estamos haciendo suposiciones con respecto a las variables del modelo, como por ejemplo que la relación entre la variable de pronóstico y las variables predictoras satisface la ecuación lineal.

- $E(\epsilon_t) = 0$
- Las perturbaciones son no correlacionadas, de lo contrario, los pronósticos serán ineficientes, ya que hay más información en los datos que pueden ser explotados.
- Las perturbaciones no están relacionadas con las variables predictoras; de lo contrario, habría más información que debería incluirse en la parte sistemática del modelo.
- $\epsilon_t \sim N(0, \sigma^2)$ esto permite producir fácilmente intervalos de predicción.
- Las variables explicativas son no estocásticas (fijas), es decir, cada predictor X no es una variable aleatoria.

2.1.1.2. Estimación de los parámetros

En la práctica, tenemos observaciones, pero no conocemos los valores de los coeficientes β_0 , β_1 . Estos deben ser estimados a partir de los datos.

2.1.1.2.1. Estimación de mínimos cuadrados El principio de mínimos cuadrados proporciona una manera de elegir los coeficientes de manera efectiva al minimizar la suma de los errores cuadrados. Es decir, elegimos los valores de β_0 , β_1 que minimizan.

$$\sum_{t=1}^T \epsilon_t^2 = \sum_{t=1}^T (y_t - \beta_0 - \beta_1 x_{1,t})^2$$

Esto proporciona el menor valor para la suma de los errores cuadrados. Los coeficientes estimados $(\hat{\beta}_0, \hat{\beta}_1)$ se obtienen a través de la ecuación:

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

2.1.1.3. Bondad de ajuste

Para determinar qué tan bien un modelo de regresión lineal se ajusta a los datos es a través del coeficiente de determinación (R^2). Esto es el cuadrado de la correlación entre los valores observados y los predichos. Alternativamente puede calcularse como:

$$R^2 = \frac{\sum (\hat{y}_t - \bar{y})^2}{\sum (y_t - \bar{y})^2}$$

Esto refleja la proporción de variación en la variable de pronóstico que se explica por el modelo de regresión.

El coeficiente de determinación nos indica que tan cerca están las predicciones con los valores reales, es decir, si R^2 está cercano a 1 las predicciones se acercan a los valores, mientras que si R^2 está cercano a 0 las predicciones no están relacionadas con los valores reales.

Es importante que señalemos que el valor R^2 nunca disminuirá al añadir un predictor adicional al modelo y esto puede llevar a un sobreajuste. Validar el rendimiento del pronóstico de un modelo en los datos de la prueba es mucho mejor que medir el valor R^2 en los datos de entrenamiento.

2.1.1.4. Error estándar de la regresión

Otra medida de lo bien que el modelo puede encajar con los datos es la desviación estándar de los residuos, que se calcula utilizando

$$\hat{\sigma}_\epsilon = \sqrt{\frac{1}{T - k - 1} \sum_{t=1}^T \epsilon_t^2}$$

Donde k es el número de predictores en el modelo. Note que dividimos por $T - k - 1$ porque hemos estimado los parámetros $k + 1$ la intercepción y un coeficiente para cada variable predictiva al calcular los residuos.

El error estándar está relacionado con el tamaño del error promedio que produce el modelo. Podemos comparar este error con la media de la muestra de y o con la desviación estándar de y para obtener alguna perspectiva sobre la precisión del modelo.

2.1.2. Regresión múltiple

Cuando hay dos o más variables predictoras, el modelo se llama modelo de regresión múltiple. Cada una de las variables predictoras debe ser numérica. Los coeficientes miden el efecto de cada predictor después

de tener en cuenta los efectos de todos los otros predictores en el modelo. Así, los coeficientes miden los *efectos marginales* (11).

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \dots + \beta_k x_{k,t} + \epsilon_t$$

Donde y es la variable a pronosticar y x_1, \dots, x_k son las variables predictoras. Cada una de las variables predictoras debe ser numérica. Los coeficientes β_1, \dots, β_k miden el efecto de cada predictor después de tener en cuenta los efectos de todos los demás predictores del modelo. Así, los coeficientes miden los efectos marginales de las variables predictoras.

Este modelo puede ser conveniente escribirlo de forma matricial donde todos los valores de la variable de pronóstico se dan en una sola ecuación. Donde $y = (y_1, \dots, y_t)'$, $\epsilon = (\epsilon_1, \dots, \epsilon_t)'$, $\beta = (\beta_0, \dots, \beta_k)'$ y

$$X = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} & \cdots & x_{k,1} \\ 1 & x_{1,2} & x_{2,2} & \cdots & x_{k,2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1,t} & x_{2,t} & \cdots & x_{k,t} \end{pmatrix}$$

Entonces

$$y = X\beta + \epsilon$$

Donde ϵ tiene media 0 y varianza $\sigma^2 I$. Nótese que la matriz X tiene filas t que reflejan el número de observaciones y columnas $k+1$ que reflejan el intercepto que representa la columna de unos más el número de predictores.

2.1.2.1. Estimación de parámetros a través de mínimos cuadrados

La estimación de los mínimos cuadrados se realiza minimizando la expresión:

$$\epsilon' \epsilon = (y - X\beta)'(y - X\beta)$$

Puede demostrarse que esto se minimiza cuando β toma el valor:

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

Esto se conoce a veces como la “ecuación normal”. Los coeficientes estimados requieren la inversión de la matriz $(X'X)^{-1}$. Si X no es de rango de columna completa, entonces la matriz $X'X$ es singular y el modelo no puede ser estimado.

La desviación residual se estima utilizando:

$$\widehat{\sigma}_\epsilon^2 = \frac{1}{T - k - 1} (y - X\widehat{\beta})'(y - X\widehat{\beta})$$

2.1.2.2. Valores ajustados y validación cruzada

La ecuación normal muestra que los valores ajustados pueden calcularse utilizando

$$\widehat{y} = X\widehat{\beta} = X(X'X)^{-1}x'y = Hy$$

Donde $H = X(X'X)^{-1}X'$ se conoce como la “matriz sombrero” porque se utiliza para calcular \widehat{y} . Si los valores diagonales de H se denotan como h_1, \dots, h_t , entonces se puede calcular la estadística de validación cruzada usando

$$CV = \frac{1}{T} \sum_{t=1}^T [e_t / (1 - h_t)]^2$$

Donde e_t es el residuo obtenido al ajustar el modelo a todas las observaciones T . Por lo tanto no es necesario ajustar los modelos separados T cuando se calcula la estadística CV .

2.1.3. Regresión no lineal

Existen muchos casos en los que una forma funcional no lineal es más adecuada. A modo de simplificar el análisis, asumiremos que solo tenemos un predictor de X .

La forma más simple de modelar la relación no lineal es transformar la variable de pronóstico y/o la variable de predicción antes de estimar un modelo de regresión. La transformación más utilizada es el logaritmo natural, si bien esto proporciona una forma funcional no lineal, el modelo sigue siendo lineal en los parámetros.

Al aplicar esto el modelo queda de la siguiente manera (11):

$$\log y = \beta_0 + \beta_1 \log x + \epsilon$$

En este modelo β_1 puede interpretarse como una elasticidad, es decir, es el cambio porcentual medio en y resultante de un incremento del 1% en x .

Otra forma útil de modelar la relación no lineal se especifica como *log-lineal* que consiste en transformar solamente la variable de pronóstico y la forma *lineal-log* se obtiene transformando el predictor.

Es importante mencionar que, para realizar una transformación logarítmica a una variable, todos sus valores observados deben ser mayores que cero. En el caso que la variable x contenga ceros, se utiliza el *log* de transformación $(x+1)$. Esto tiene un efecto similar al de tomar logaritmos, pero evita el problema de los ceros, también tiene el efecto de que los ceros en la escala original permanecen en la escala transformada.

Hay casos en los que la simple transformación de los datos no será adecuada y puede ser necesaria una especificación más general. Entonces el modelo que usaremos es:

$$y = f(x) + \epsilon$$

Donde f es una función no lineal. En la especificación de la regresión no lineal que sigue, f es una función más flexible de x , en comparación con una simple transformación logarítmica o de otro tipo.

Una de las especificaciones más simples es hacer que f sea lineal. Es decir, introducimos puntos donde la pendiente de f puede cambiar. Estos puntos se llaman nudos. Esto puede lograrse dejando $x_{1,t} = x$ e introduciendo una variable $x_{2,t}$ tal que

$$x_{2,t} = (x - c)_+ = \begin{cases} 0 & x < c \\ (x - c) & x \geq c \end{cases}$$

La notación $(x - c)_+$ significa el valor $(x - c)$ si es positivo y 0 en caso contrario. Esto fuerza a la pendiente a doblarse en el punto c . Se pueden incluir dobleces adicionales en la relación añadiendo más variables de la forma anterior.

Un ejemplo de esto es el siguiente: Considerando $x = t$ y ajustando una tendencia lineal por partes a una serie temporal. Las relaciones lineales interrumpidas construidas de esta manera son un caso especial de *splines de regresión*. En general, se obtiene un spline de regresión lineal utilizando

$$x_1 = x, x_2 = (x - c_1)_+, \dots, x_k = (x - c_{k-1})_+$$

Donde c_1, \dots, c_{k-1} son los nudos (puntos donde la línea puede doblarse). Seleccionar el número de nudos $(k-1)$ y la posición en la que deben colocarse puede resultar difícil y un tanto arbitrario.

Se puede obtener un resultado más suave usando cubos por partes en lugar de líneas por partes. Estos están limitados a ser continuos (se unen) y suaves (para que no haya cambios repentinos de dirección). En general, una spline de regresión cúbica se escribe como

$$x_1 = x \quad x_2 = x^2 \quad x_3 = x^3 \quad x_4 = (x - c_1)_+ \quad \dots \quad x_k = (x - c_{k-3})_+$$

Los splines cúbicos usualmente dan un mejor ajuste a los datos. Sin embargo, las predicciones de y se vuelven poco fiables cuando x está fuera del rango de los datos históricos.

2.1.4. Modelos lineales generalizados (GLM)

Un modelo lineal generalizado es un conjunto de variables independientes y_1, y_2, \dots, y_n con función de densidad, o función de probabilidad, que puede escribirse como (12)

$$p(y_i|\theta_i, \phi) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi) \right\}$$

Donde,

θ_i es el parámetro natural o canónico.

ϕ es un parámetro adicional de escala o dispersión.

$a_i(\cdot)$, $b(\cdot)$ y $c(\cdot)$ son funciones específicas.

- Si ϕ es conocido este es un modelo de la familia exponencial lineal.
- Si ϕ es desconocido es un modelo de dispersión exponencial.

Un modelo lineal generalizado tiene tres componentes básicos:

1. **Componente aleatoria:** La componente aleatoria de un modelo lineal generalizado (GLM) consiste en una variable aleatoria Y con observaciones independientes (y_1, y_2, \dots, y_N) .

En muchas aplicaciones las observaciones de Y son binarias y se identifican como *éxito* y *fracaso*. Aunque de manera más general, cada Y_i , indica el número de éxitos de entre un número fijo de ensayos, y se modeliza como una *distribución binomial*.

Otro caso, es cuando cada observación es un recuento, entonces se puede asignar a Y una *distribución Poisson* o *binomial negativa*. En el caso en que las observaciones sean continuas se puede asumir para Y una *distribución normal*.

2. **Componente Sistemática:** Esta componente especifica las variables explicativas, que entran en forma de efectos fijos en un modelo lineal, es decir, las variables x_j se relacionan mediante

$$\alpha + \beta_1 x_1 + \dots + \beta_k x_k$$

Dicha combinación lineal de variables explicativas se denomina *predictor lineal*. Otra forma de expresarlo es como un vector (η_1, \dots, η_N)

$$\eta_i = \sum_{j=1}^p \beta_j x_{ij}$$

Donde x_{ij} es el valor del j -ésimo predictor en el i -ésimo individuo, e $i=1, \dots, N$. El término independiente (α) se obtendría haciendo que todos los x_{ij} sean igual a 1 para todos los i en la ecuación anterior.

3. **Función de enlace:** Se denota la esperanza de Y como $\mu = E(Y)$, entonces la función enlace especifica una función $g(\cdot)$ que relaciona μ con el predictor lineal,

$$g(\mu) = \alpha + \beta_1 x_1 + \dots + \beta_k x_k$$

Así, la función link $g(\cdot)$ relaciona las componentes aleatoria y sistemática. De este modo, para $i=1, \dots, N$,

$$\begin{aligned} \mu_i &= E(Y_i) \\ \eta_i &= g(\mu_i) = \sum_{j=1}^p \beta_j x_{ij} \end{aligned}$$

La función g más simple es $g(\mu) = \mu$, esto es, la identidad que da lugar al modelo de regresión lineal.

2.1.5. Familia Exponencial

Sea Y_1, \dots, Y_n variables aleatorias independientes, si cada una tiene función de densidad o función de probabilidad de la forma (14)

$$f(y_i | \theta_i, \phi) = \exp[\phi \{y_i \theta_i - b(\theta_i)\} + c(y_i, \phi)]$$

Se dice entonces que pertenece a la familia exponencial. Donde el parámetro de interés θ depende del valor esperado de y , ϕ es un parámetro de escala positivo y b y c son funciones arbitrarias.

La distribución Poisson pertenece a la familia exponencial y es importante abordar este modelo en el estudio, ya que, los datos que se analizarán son discretos y de tipo conteo.

2.1.5.1. Regresión Poisson

La distribución Poisson describe la frecuencia esperada de un conjunto de probabilidades para una variable discreta, es decir, que toma únicamente valores enteros positivos. En cada punto de esta distribución, se representa la probabilidad de que un determinado número de eventos ocurra durante un periodo de tiempo en un espacio o población específica. De esta manera, una variable de tipo Poisson consiste en el número de veces que un acontecimiento ocurre en un tiempo y población definida. Estas variables se presentan como tasas de un evento de interés. Algunas características importantes que se destacan son las siguientes:

- No adopta valores negativos, pues el conteo de eventos siempre tendrá resultados mayores a cero.
- Las tasas suelen ser bajas y se asume como constante dentro de cada unidad de observación.
- Independencia de eventos, es decir, la ocurrencia de uno de ellos no hace más o menos probable la ocurrencia de otro.
- En una observación dada, el promedio de la variable Poisson es igual a su varianza.

En el modelo GLM se usa habitualmente el logaritmo de la media para la función de enlace, de modo que el modelo *log-lineal* con una variable explicativa X se puede expresar como (16):

$$\text{Ln}(y_t) = \beta_0 + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \dots + \beta_k X_{k,t} + \epsilon_t$$

Donde el $\text{Ln}(y_t)$ es el logaritmo natural de la tasa o del número de eventos por unidad de observación. Por otra parte se tienen variables independientes $(X_{1,t}, \dots, X_{1,k})$.

Una aplicación en este tipo de regresión es la estimación de tasas relativas (TR), ajustadas por diversas variables independientes. En este caso la TR puede estimarse como una razón de funciones teniendo en el numerador de la ecuación para la categoría de “exposición” en el denominador el grupo de “no expuesto” (o de referencia), para un modelo múltiple esto quedaría de la siguiente manera:

$$TR = \frac{(X_{1,t} = 1)}{(X_{1,t} = 0)} = \frac{\epsilon^{\beta_0} \epsilon^{\beta_1 (X_{1,t}=1)} \epsilon^{\beta_2 X_{2,t}}, \dots, \epsilon^{\beta_k X_{k,t}}}{\epsilon^{\beta_0} \epsilon^{\beta_1 (X_{1,t}=0)} \epsilon^{\beta_2 X_{2,t}}, \dots, \epsilon^{\beta_k X_{k,t}}}$$

De esta forma se calcula la medida de asociación que sería el factor por el cual se multiplicaría la tasa al cambiar de categoría, desde un estado de no expuesto a uno de expuesto. En el caso de variables

independientes cuantitativas, la TR indicaría el factor por el que se multiplica la tasa con cada aumento en una unidad en la escala de exposición.

2.1.5.2. Regresión Binomial Negativa

Si una v.a Y se distribuye como una binomial negativa, entonces la función de probabilidad es,

$$P(y|k, \mu) = \frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+1)} \left(\frac{k}{\mu+k}\right)^k \left(1 - \frac{k}{\mu+k}\right)^y$$

con $y=0,1,2,\dots$ donde k y μ son los parámetros de la distribución.

Se tiene que,

$$E(Y) = \mu$$

$$Var(Y) = \mu + \frac{\mu^2}{k}$$

El parámetro $\frac{1}{k}$ es un parámetro de dispersión, de modo que si $\frac{1}{k} \rightarrow 0$ entonces $Var(Y) \rightarrow \mu$ y la distribución binomial negativa converge a una distribución de Poisson.

Por otro lado, para un valor fijo de k esta distribución pertenece a la familia exponencial natural, de modo que se puede definir un modelo GLM binomial negativo. En general, se usa una función link de tipo logaritmico.

Esta regresión se puede utilizar para datos sobredispersos de recuentos, es decir, cuando la varianza condicional es mayor que la media condicional. Se puede considerar como una generalización de la regresión Poisson, ya que tiene su misma estructura de medias y además un parámetro adicional para el modelo de sobredispersión.(12)

2.2. Conceptos previos de series de tiempo

2.2.1. Series de Tiempo

En diversos fenómenos naturales o sociales se pueden encontrar variables (cuantitativas) medidas en periodos constantes de tiempo (días, mes, semestres, años, etc..). Estas variables se pueden representar a partir de una serie de observaciones (y_t). Esto se describe como una serie temporal o serie cronológica, donde existe un conjunto arbitrario de tiempo, denominado T , el cual no es aleatorio y es de la forma $T = 1, 2, \dots, n$. Luego una serie temporal se representa como cualquier conjunto de observaciones y_t , $t \in T$

ordenadas en el tiempo, por lo tanto, t es estrictamente creciente.

Una serie temporal se puede descomponer en el siguiente modelo, llamado modelo aditivo, debido a la no interacción entre sus componentes:

$$Y_t = N_t + T_t + S_t + \varepsilon_t, \quad \varepsilon_t \sim RB$$

Donde,

N_t : Nivel de la serie

T_t : Tendencia de la serie

S_t : Componente estacional de la serie

ε_t : Componente aleatoria de la serie

2.2.2. Proceso Estocástico

Un proceso estocástico es un conjunto de variables aleatorias $\{Y_t, t \in T\}$, donde T es un conjunto arbitrario. En series de tiempo, el conjunto T es ordenado y corresponde a periodos de tiempo (días, meses, años, etc.). Para cada valor t del conjunto T $\{Y_t\}$ es una variable aleatoria, y los valores observados de las variables aleatorias en distintos instantes forman una serie de tiempo.

2.2.3. Estacionaridad

Un proceso estocástico $\{Y_t\}$ se dice estrictamente estacionario, si todas las distribuciones de dimensión finita permanecen invariantes en el tiempo, es decir, $F(Y_1, \dots, Y_n) = F(Y_{k+1}, \dots, Y_{k+n}) \forall n, k \in \mathbb{Z}$. Esta es una condición muy fuerte, ya que, para contrastarla se debe disponer de todas las distribuciones conjuntas para cualquier selección de variables del proceso.

Por otra parte, un proceso estocástico débilmente estacionario, implica que la media, la varianza y la covarianza son constantes en el tiempo.

Un proceso estocástico $\{Y_t\}$ se dice estacionario de segundo orden si:

- $E(Y_t) = \mu$, es decir, la media es constante.
- $V(Y_t) = \sigma^2$
- $Cov(Y_t, Y_{t+k}) = \gamma(k)$ para algún t, k , es decir, el valor de la covarianza depende solamente de la distancia entre dos periodos y no del tiempo en el que se ha calculado la covarianza .

2.2.4. Función de autocovarianza y función de autocorrelación

2.2.4.1. Función de autocovarianza

Sea $\{Y_t\}$ una serie temporal estacionaria, entonces la función de autocovarianza de $\{Y_t\}$ se define como:

$$\gamma(k) = Cov(Y_t, Y_{t\pm k})$$

Donde k es la diferencia de tiempo entre las observaciones.

2.2.4.2. Función de autocorrelación

Sea $\{Y_t\}$ una serie temporal estacionaria, entonces la función de autocorrelación de dicha serie se define como:

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}, \quad \text{donde } -1 \leq \rho(k) \leq 1$$

Donde, k es el periodo que separa dos observaciones.

2.2.5. Componente Aleatorio

ϵ_t o también llamado *ruido blanco* es la parte impredecible de la serie temporal, es decir, tiene un comportamiento permanentemente aleatorio y se caracteriza por que sus observaciones en dos tiempos diferentes no guardan correlación.

Es necesario establecer algunos supuestos sobre los errores que deben validarse antes de utilizar los resultados obtenidos de la estimación del modelo:

- $E[\epsilon_j]=0$
- $Var [\epsilon_j] = \sigma_\epsilon^2$
- $Cov (\epsilon_t, \epsilon_{t+k}) = 0, \quad \forall k \neq 0$

Se trata de un proceso en el que todas sus variables son independientes.

La media cero indica que la función de autocorrelación coincide con su autocovarianza y considerar que la varianza del proceso es constante, indica un proceso estacionario. Adicionalmente, el ruido blanco en sentido estricto presenta un valor en cada instante del tiempo que no depende de cual haya sido su valor en los instantes anteriores y que no ejerce ninguna influencia en sus valores futuros, es decir, que no es posible

obtener más información de la serie de tiempo que aporte a la estimación de valores futuros a partir de un modelo de regresión.

2.2.6. Procesos Estacionarios

2.2.6.1. Modelo Ar(p)

Los modelos autoregresivos se basan en la idea que el valor actual de la serie Y_t puede explicarse en función de p valores pasados $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ donde p determina el número necesario de rezagos para pronosticar un valor actual.

El modelo autoregresivo de orden p está dado por:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} \quad t \in$$

Esta ecuación de manera más compacta puede ser escrita como:

$$\Phi(B)Y_t = \epsilon_t$$

Donde B es definido como el operador de rezago y $\Phi(\cdot)$ es el polinomio característico definido como:

$$\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$$

Las propiedades de este modelo se pueden apreciar bajo un Ar(1), es decir,

$$Y_t = \phi Y_{t-1} + \epsilon_t$$

Bajo estacionariedad se tiene que:

- $E(Y_t) = 0$
- $\text{Var}(Y_t) = \frac{\sigma_t^2}{1 - \phi^2}$
- $\gamma(k) = \phi^k \gamma(0)$, con $\gamma(0) = \frac{\sigma_t^2}{1 - \phi}$
- $\rho(k) = \phi^k$

2.2.6.2. Modelo MA(q)

Sea ε_t una sucesión de ruido blanco, $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma^2$. Consideremos el proceso Y_t definido por:

$$Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q}$$

Este modelo se conoce como modelo de medias móviles (MA(q)). Otra forma de denotar este modelo es la siguiente:

$$Y_t = \Theta_q(B)\varepsilon_t$$

Donde $\Theta_q(B)$ es definido como el polinomio de medias móviles.

Las propiedades de este modelo se pueden apreciar bajo un MA(1), es decir,

$$Y_t = \varepsilon_t - \theta\varepsilon_{t-1}$$

Bajo estacionariedad se tiene que:

- $E(Y_t)=0$
- $Var(Y_t)=\sigma^2(1+\theta^2) < \infty$
- $\rho(k) = \begin{cases} 1 & \text{si } |k| = 0 \\ \frac{-\theta}{1+\theta^2} & \text{si } |k| = 1 \\ 0 & \text{si } |k| \geq 1 \end{cases}$

2.2.7. Procesos Lineales no estacionarios

2.2.7.1. Proceso autoregresivo integrado de medias móviles *ARIMA* (p,d,q)

Los modelos antes mencionados se basan en el supuesto de estacionariedad, es decir, que su media y varianza son constantes en el tiempo y la covarianza es invariante en el tiempo. Pero muchas series de tiempo no son estacionarias, porque pueden ir cambiando de nivel en el tiempo o la varianza no es constante, a este tipo de proceso se le considera procesos integrados. Por lo tanto, se debe diferenciar una serie de tiempo d veces para hacerla estacionaria y luego aplicar a esta serie diferenciada un modelo ARMA(p,q), entonces la serie original es ARIMA (p,d,q), es decir, una serie de tiempo autoregresiva integrada de media móvil, donde p denota el número de términos autoregresivos, d el número de veces que la serie debe ser diferenciada para hacerla estacionaria y q el número de términos de la media móvil invertible (20).

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d Y_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \varepsilon_t$$

Este modelo expresado en forma del polinomio operador de rezagos es:

$$\Phi_P(B)(1 - B)^d Y_t = \Theta_q(B) \varepsilon_t$$

Donde ε_t es un proceso de ruido blanco, y $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ son los parámetros del modelo.

La construcción de un modelo ARIMA se lleva a cabo de manera iterativa donde se pueden distinguir las siguientes etapas:

- **Identificación:** Se intentará sugerir un modelo ARIMA (p,d,q) que sea apropiado para reproducir la serie de tiempo. Es posible identificar más de un modelo candidato que pueda describir la serie.
- **Estimación:** Teniendo el modelo apropiado para la serie, se realiza inferencia sobre los parámetros.
- **Validación:** Se realizan diagnosticos para validar si el modelo seleccionado se ajusta a los datos, de no ser así, se escoge el próximo modelo candidato y se repiten los pasos anteriores.
- **Predicción:** Una vez seleccionado el mejor modelo se pueden realizar pronósticos en términos probabilísticos de los valores futuros.

2.2.8. Tendencia

La tendencia representa el comportamiento predominante de la serie que permanece constante en el tiempo, donde se pueden clasificar dos tipos de tendencia, tendencia lineal y tendencia no lineal. La tendencia lineal muestra que las observaciones de la variable cuantitativa se mueven a un ritmo constante. Por otra parte, la tendencia no lineal es cuando la serie presenta un comportamiento curvilíneo a largo plazo, dentro de las cuales se pueden distinguir, la logarítmica, exponencial, polinomial, entre otras.

Dentro del enfoque de este estudio es importante exponer lo siguiente:

1. **Regresión Polinomial:** Cuando se gráfica una serie de tiempo y las observaciones trazadas cambian de dirección más de una vez, entonces la tendencia óptima para este tipo de series es la polinomial, en donde su modelo matemático tiene como argumento el tiempo t (13):

$$T(t) = \beta_0 + \beta_1 t + \dots + \beta_p t^p$$

El orden o grado del polinomio se puede determinar por la cantidad de fluctuaciones de las observaciones o por la cantidad de máximos o mínimos que se observan a lo largo de la serie. Un polinomio de grado 2 suele tener solo un máximo o un mínimo, mientras que uno de grado 3 tiene uno o dos máximos y mínimos y el de grado 4 tiene hasta 3.

Cuando se ajusta un polinomio de grado p y existen observaciones que no se ajustan y no son capturadas por dicho polinomio, el recurso podría ser aumentar el grado en una unidad y obtener una curva que se ajustará mejor, esto generalmente provoca que una de las ramas de la función ajustada altere su dirección. Desde el punto de vista de la predicción, realizando este procedimiento no se obtienen resultados totalmente satisfactorios.

2.2.9. Estacionalidad

La estacionalidad se representa por un patrón de repetición dentro de un periodo de tiempo, por lo que es una función periódica. El periodo depende del origen de los datos que puede ser, anual, trimestral, mensual, semanal o diario. Por ejemplo, si se analiza el número de accidentes de tránsito diarios en Santiago, sería lógico pensar que la cantidad de accidentes aumentará los días viernes, sábado y domingo en relación al resto de los días de la semana y que ese comportamiento será similar el resto de las semanas. Si se obtuvieran estos datos durante un mes, es muy probable que las 4 semanas tengan un comportamiento muy similar o cíclico, es decir, que tengan un comportamiento estacional.

Cuando se corrige esta fuente de variación de los datos (desestacionalización), se pueden realizar comparaciones de un periodo con el periodo anterior o con cualquier otro y obtener resultados, que ya no se verán afectados por variaciones cíclicas que puedan provocar predicciones o conclusiones fallidas.

Una regresión armónica y dummy nos ayuda a estimar el efecto estacional de la serie:

- **Regresión Armónica:** El uso de esta regresión es apropiada cuando la serie tiene amplitudes constantes, para determinar el efecto estacional de la serie.

Sea el modelo:

$$Y_t = A \text{ sen}(w t + \varphi) \quad t = 1, \dots, N$$

Donde,

A = Amplitud de la serie.

w = Número de ciclos por unidad de tiempo ($\frac{2\pi}{P} = 2\pi f$).

P = Periodo de la serie.

f = Frecuencia de la serie ($f = \frac{1}{P}$)

φ = Desfase de la serie.

Si consideramos la siguiente transformación,

$$\begin{aligned} Y_t &= A \operatorname{sen}(2\pi f t + \varphi) \\ &= A[\operatorname{sen}(2\pi f t) \cos(\varphi) + \cos(2\pi f t) \operatorname{sen}(\varphi)] \\ &= \underbrace{A \cos(\varphi)}_{\beta_1} \operatorname{sen}(2\pi f t) + \underbrace{A \operatorname{sen}(\varphi)}_{\beta_2} \cos(2\pi f t) \\ &= \beta_1 \operatorname{sen}(2\pi f t) + \beta_2 \cos(2\pi f t) \end{aligned}$$

Luego, esto puede ser expresado como un modelo de regresión lineal simple

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_t \end{pmatrix} = \begin{pmatrix} \operatorname{sen}(2\pi f) & \cos(2\pi f) \\ \operatorname{sen}(4\pi f) & \cos(4\pi f) \\ \vdots & \vdots \\ \operatorname{sen}(2\pi f t) & \cos(2\pi f t) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

Cuando la serie temporal tiene más de un periodo podemos definir el modelo general como:

$$Y_t = \beta_{11} \operatorname{sen}(2\pi f_1 t) + \beta_{12} \cos(2\pi f_1 t) + \beta_{21} \operatorname{sen}(2\pi f_2 t) + \beta_{22} \cos(2\pi f_2 t) + \dots + \beta_{q1} \operatorname{sen}(2\pi f_q t) + \beta_{q2} \cos(2\pi f_q t)$$

$$\sum_{j=1}^q \beta_{j1} \operatorname{sen}(2\pi f_j t) + \beta_{j2} \cos(2\pi f_j t)$$

Donde, f_j corresponde a las frecuencias de Fourier y se calculan en base a los espectros de la serie.

■ **Regresión con variables dummy:**

En general, se supone que las variables explicativas en cualquier analisis de regresión son de naturaleza cuantitativa, es decir, que se registran en una escala bien definida, como temperatura, distancia, edad, etc. En muchas otras aplicaciones las cualitativas, por ejemplo, sexo (masculino, femenino),

situación laboral (empleado, desempleado), etc., se definen en una escala nominal, es decir, que no tienen una escala natural de medición. Tales variables suelen indicar la presencia o ausencia de un atributo, como empleado o desempleado, graduado o no graduado, fumador o no fumador, sí o no, aceptación o rechazo. Estas pueden cuantificarse de manera que cuando se toma el valor 1 indica generalmente la presencia del atributo y 0 la ausencia. Tales variables son muy útiles en las aplicaciones de series temporales. Puesto que la variable de observación es el tiempo, una variable ficticia representa si, en cada periodo de tiempo, ha ocurrido un determinado evento. Estas variables a menudo se utilizan para aislar ciertos periodos que pueden ser sistemáticamente diferentes de otros periodos cubiertos por un conjunto de datos. Este tipo de variables son claves en el estudio de eventos, es decir, el objetivo es ver si un evento en particular influye en algún resultado.

El modelo propuesto a continuación contiene la variable respuesta otra explicativa y una dummy. Por ejemplo si se quisiera evaluar la rentabilidad de una empresa de aerolínea durante el periodo t , donde R_t es la rentabilidad de la empresa durante el periodo t , x_t es la rentabilidad del mercado y d_t es la variable dummy que indica si ocurrió o no un determinado suceso (previamente establecido que pueda afectar la rentabilidad de la empresa). (15)

$$R_t = \beta_0 + \beta_1 x_t + \beta_2 d_t + \epsilon_t$$

2.3. Modelamiento y pruebas para un cambio estructural

A continuación, se muestran tres enfoques utilizados para determinar la presencia de un cambio estructural. Cada enfoque implica el cálculo de la suma de cuadrados residual (SCR) a partir de un modelo restringido y no restringido, que se engloban en una estadística F . El modelo restringido asume que las regresiones del subperíodo no son diferentes y que la intercepción y la pendiente permanecen iguales durante todo el período. Por el contrario, el modelo sin restricciones asume que los coeficientes son diferentes para los subperíodos. A partir de ahí, la inferencia se hace comparando la estadística F con el valor crítico F . Cada enfoque se discute a continuación junto con su respectiva prueba de Chow, que es un enfoque alternativo. [(21)]

2.3.1. Prueba de Chow para parámetros diferentes

Cuando los datos muestran un cambio y se tiene un antes y un después, se pueden reconocer dos periodos muy diferentes dentro de la misma serie. Un modelo que permite que los coeficientes sean diferentes en los dos periodos es el siguiente [(17)]:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_k \\ y_{k+1} \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ k & 0 \\ 0 & (k+1) \\ \vdots & \vdots \\ 0 & N \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_k \\ \epsilon_{k+1} \\ \vdots \\ \epsilon_N \end{pmatrix}$$

Donde de 1 a k se tienen las observaciones del primer periodo y de k+1 a N se tienen las observaciones del segundo periodo. Gráficamente esto se puede representar de la siguiente manera:

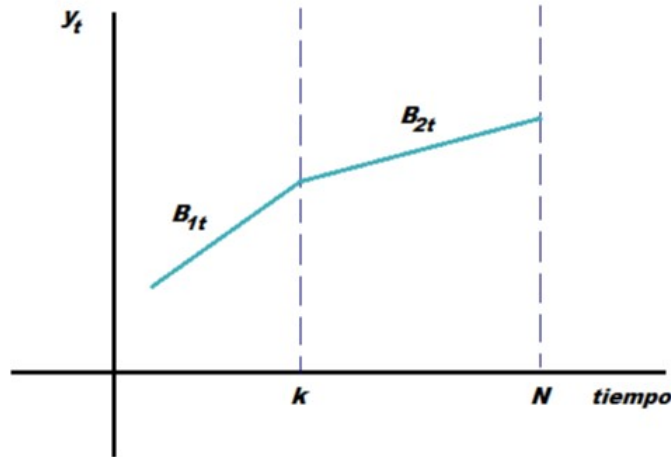


Figura 2.1: Cambio Estructural

A modo de simplificar el planteamiento del sistema de ecuación antes mencionado, su notación quedará de la siguiente manera (1):

$$\begin{pmatrix} y_{pre} \\ y_{post} \end{pmatrix} = \begin{pmatrix} i_{pre} & x_{pre}^* \\ i_{post} & x_{post}^* \end{pmatrix} \begin{pmatrix} \alpha \\ \beta^* \end{pmatrix} + u \quad u \sim N(0, \sigma^2 I)$$

Este modelo es restringido, donde el periodo 1 es antes del “break” y el periodo 2 después del cambio, dejamos entonces y_i, x_i ($i = 1, 2$) para indicar la ruptura adecuada de los datos. Del mismo modo $n = n_1 + n_2$, donde n_1 son el número de observaciones para el primer periodo y n_2 para el segundo periodo.

El modelo sin restricciones es de la forma (2):

$$\begin{pmatrix} y_{pre} \\ y_{post} \end{pmatrix} = \begin{pmatrix} i_{pre} & 0 & x_{pre}^* & 0 \\ 0 & i_{post} & 0 & x_{post}^* \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1^* \\ \beta_2^* \end{pmatrix} + u \quad u \sim N(0, \sigma^2 I)$$

El modelo restringido en (1) asume que el intercepto y la pendiente permanecen iguales durante todo el periodo en que $\alpha = \alpha_1 = \alpha_2$ y $\beta^* = \beta_1^* = \beta_2^*$. Sus grados de libertad son $n_1 + n_2 - k$, o $n - k$ donde k es el número de parámetros estimados.

El modelo sin restricciones en (2) asume que el intercepto y la pendiente son distintos. Cada periodo se revierte de manera diferente para obtener los siguientes grados de libertad: $n_1 - k$ para el primer período y $n_2 - k$ para el segundo período. La suma de los grados de libertad genera los grados de libertad para el modelo sin restricciones: $n_1 - k + n_2 - k = n_1 + n_2 - 2k = n - 2k$

Luego, se debe obtener la suma de cuadrados residual para el modelo restringido, es decir, SCR_R y el modelo no restringido $SCR_{NR} = SCR_1 + SCR_2$

Por lo tanto, la estadística F se obtiene de la siguiente manera:

$$F = \frac{(SCR_R - SCR_{NR})/k}{(SCR_{NR})/(n - 2k)} \sim F_{[k, (n-2k)]}$$

Los grados de libertad para el numerador se establecen de la siguiente manera:

$$SCR_R - SCR_{NR} = n - k - (n - 2k) = k$$

u en (1) y (2) representa el término de error.

La hipótesis nulas y alternativas son las siguientes:

$$H_0 : \alpha_1 = \alpha_2, \beta_1^* = \beta_2^* \text{ v/s } H_1 : \alpha_1 \neq \alpha_2, \beta_1^* \neq \beta_2^*$$

En la hipótesis nula no hay ruptura en el intercepto y la pendiente versus la hipótesis alternativa que existe una ruptura en el intercepto y la pendiente. Por lo tanto, el valor calculado de F se compara con el

valor crítico de F de la tabla. La hipótesis nula se rechaza si el valor calculado de F es mayor que el valor crítico y viceversa.

2.3.2. Prueba de Chow para el coeficiente de pendiente

En este caso, el interés está en establecer si la ruptura afecta la pendiente. Al igual que en el punto anterior, dejamos y_i , x_i ($i=1,2$) para indicar la partición apropiada de los datos en los dos periodos. De igual forma $n=n_1 + n_2$, por la cantidad de observaciones de los dos periodos. (17)

El modelo restringido resultante es (3):

$$\begin{pmatrix} y_{pre} \\ y_{post} \end{pmatrix} = \begin{pmatrix} i_{pre} & 0 & x_{pre}^* \\ 0 & i_{post} & x_{post}^* \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta^* \end{pmatrix} + u \quad u \sim N(0, \sigma^2 I)$$

Donde sus grados de libertad son: $n-k-1$

El modelo correspondiente sin restricciones es de la forma (4):

$$\begin{pmatrix} y_{pre} \\ y_{post} \end{pmatrix} = \begin{pmatrix} i_{pre} & 0 & x_{pre}^* & 0 \\ 0 & i_{post} & 0 & x_{post}^* \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1^* \\ \beta_2^* \end{pmatrix} + u \quad u \sim N(0, \sigma^2 I)$$

Alternativamente se puede escribir de la siguiente manera (5):

$$\begin{pmatrix} y_{pre} \\ y_{post} \end{pmatrix} = \begin{pmatrix} i_{pre} & 0 & x_{pre}^* & 0 \\ 0 & i_{post} & 0 & x_{post}^* \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 - \alpha_1 \\ \beta_1^* \\ \beta_2^* - \beta_1^* \end{pmatrix} + u \quad u \sim N(0, \sigma^2 I)$$

Para este modelo sus grados de libertad son: $n-2k$

La estadística F resultante se calcula:

$$F = \frac{(SCR_R - SCR_{NR}) (k - 1)}{(SCR_{NR}) / (n - 2k)} \sim F_{[(k-1), (n-2k)]}$$

Donde los grados de libertad para el numerador se establece:

$$SCR_R - SCR_{NR} = n - k - 1 - (n - 2k) = k - 1$$

En los tres modelos planteados anteriormente u representa el término de error. Esto se compara con el valor F de la tabla con un nivel de significancia del 5%.

La hipótesis nula y alternativa son:

$$H_0 : \beta_1^* = \beta_2^* \text{ v/s } H_1 : \beta_1^* \neq \beta_2^*$$

En la hipótesis nula no hay “ruptura” en la pendiente versus la hipótesis alternativa que si existe una “ruptura” en la pendiente.

2.3.3. Pruebas de Intercepto

En este caso, es de interés determinar si la ruptura estructural afecta las intercepciones, ya que las intercepciones de las ecuaciones de regresión interceptan el eje y en el mismo punto. Además, dado que se supone que la pendiente es común, el objetivo principal es determinar entre el modelo restringido y el no restringido, si es mayor o menor que el otro en la dirección de la variable dependiente.

De las pruebas anteriores, dejamos $y_i, x_i(i=1,2)$ para indicar la partición apropiada de los datos para el periodo 1 y el periodo 2. Lo mismo con la cantidad de observaciones para los periodos n_1 y n_2 , para el periodo 1 y 2 respectivamente.(17)

El modelo restringido resultante es (6):

$$\begin{pmatrix} y_{pre} \\ y_{post} \end{pmatrix} = \begin{pmatrix} i_{pre} & x_{pre}^* \\ i_{post} & x_{post}^* \end{pmatrix} \begin{pmatrix} \alpha \\ \beta^* \end{pmatrix} + u \quad u \sim N(0, \sigma^2 I)$$

Con $n-k$ grados de libertad.

El modelo no restringido es de la forma (7):

$$\begin{pmatrix} y_{pre} \\ y_{post} \end{pmatrix} = \begin{pmatrix} i_{pre} & 0 & x_{pre}^* \\ i_{post} & i_{post} & x_{post}^* \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 - \alpha_1 \\ \beta^* \end{pmatrix} + u \quad u \sim N(0, \sigma^2 I)$$

Donde los grados de libertad son $n-k-1$

La estadística F resultante se calcula:

$$F = \frac{(SCR_R - SCR_{NR})/(1)}{(SCR_{NR}/(n - k - 1))} \sim F[1, (n - k - 1)]$$

Donde los grados de libertad para el numerador se establecen de la siguiente manera:

$$(SCR_R - SCR_{NR}) = n - k - (n - k - 1) = 1$$

En los modelos (6) y (7) u representa el término de error. Esto se compara con el valor F de la tabla con un nivel de significancia del 5%.

La hipótesis nula y alternativa son:

$$H_0 : \alpha_1 = \alpha_2 \text{ v/s } H_1 : \alpha_1 \neq \alpha_2$$

Es decir, en H_0 no existe interrupción en el intercepto versus H_1 , donde existe una ruptura en el intercepto.

2.4. Modelos de series de tiempo interrumpidas

Una serie de tiempo interrumpida (ITS) es un modelo, que se caracteriza porque la serie se interrumpe, debido a la aplicación de un “tratamiento”. Los datos reflejan esta interrupción mostrando un cambio de nivel, tendencia o ambos. Gráficamente podemos ver estos patrones de cambio de la siguiente manera: (2)

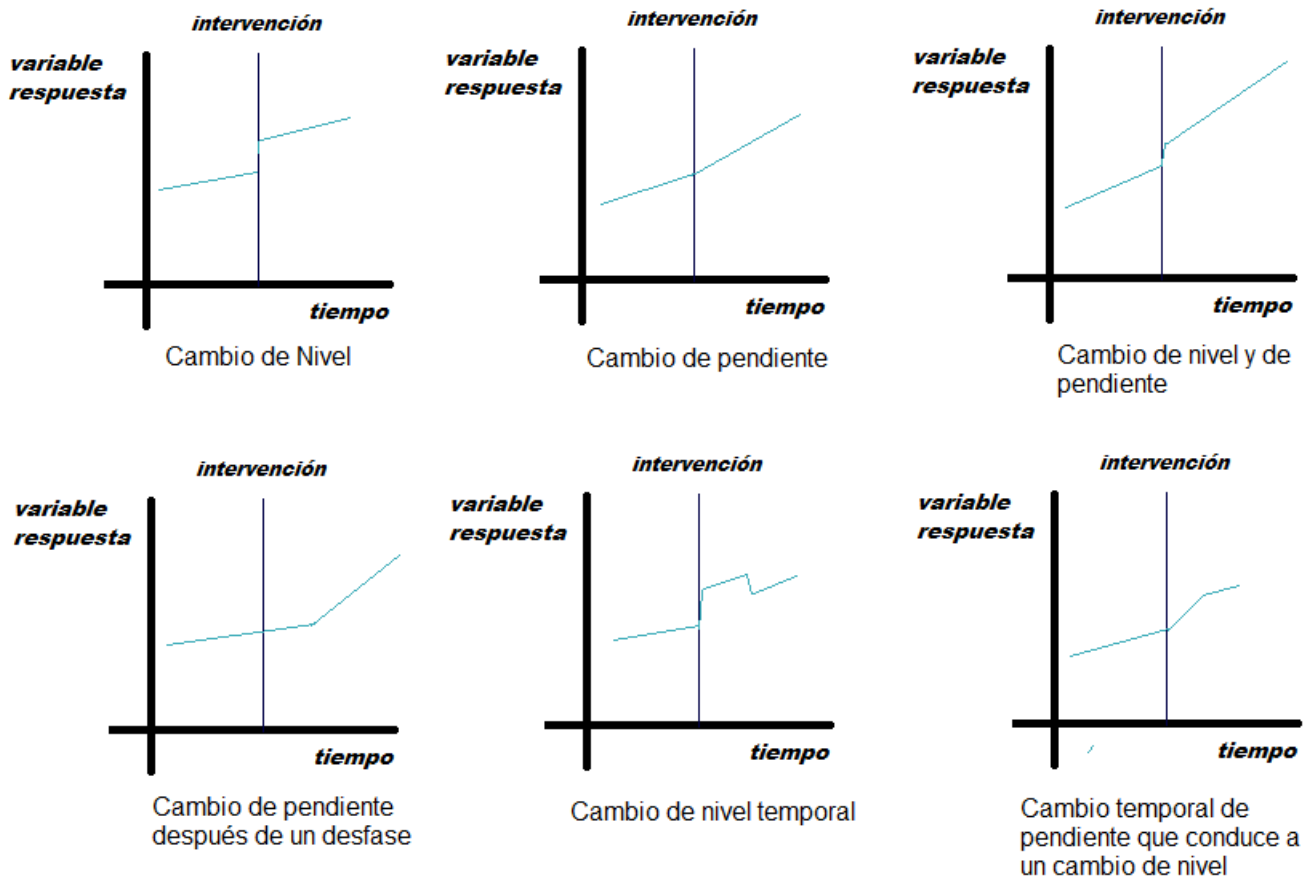


Figura 2.2: Ejemplos de modelos de impacto usando ITS

El cambio de nivel puede tomar tres formas que dependen del efecto que tenga la intervención: El primer tipo corresponde al cambio constante en el nivel debido al efecto que produce la intervención, el segundo es el efecto causado sobre las observaciones después de aplicada la intervención y el tercero es el efecto de la intervención sobre las observaciones que tiende a regresar hacia la línea de base con el transcurrir del tiempo.

La tendencia o pendiente puede tener un aumento o una disminución debido a la intervención y que se puede observar de manera persistente en el tiempo; se distinguen dos tipos: la primera presenta el resultado de la intervención que se produce inmediatamente después y permanece durante un periodo prolongado de tiempo y la segunda indica cuando el resultado de la intervención no persiste en el tiempo por lo que se refleja la finalización de la misma.

2.4.1. Modelo de regresión

Para determinar el efecto de la intervención en ITS, podemos recurrir al siguiente modelo: (19)

$$Y_t = \beta_0 + \beta_1 * Tiempo_t + \beta_2 * Intervencion_t + \beta_3 * Tiempo\ después\ de\ la\ intervencion_t + \epsilon_t$$

Donde,

- Y_t : Es el valor de la variable respuesta en cada unidad de tiempo t.
- β_0 : Nivel basal de los resultados de interés en el tiempo cero.
- β_1 : Nivel basal.
- $Tiempo_t$: Tiempo en unidad de medida especificada, desde el inicio del periodo de observación.
- β_2 : Cambio en el nivel de sensibilidad inmediatamente después de la intervención.
- $Intervencion_t$: Dicotomía antes y después de la intervención, es decir, 0 antes de la intervención y 1 después de la intervención.
- $\beta_1 + \beta_3$: Pendiente tras la intervención.
- $Tiempo\ después\ de\ la\ intervencion_t$: Tiempo en una unidad de medida después de la intervención.
- ϵ_t : Es la variabilidad aleatoria no explicada por el modelo.

2.4.2. Modelo de regresión segmentada

En el análisis estándar de STI, se utiliza el siguiente modelo, donde se requiere un mínimo de tres variables para un análisis de ITS: (2)

- T: Tiempo transcurrido desde el inicio del estudio con la unidad que representa la frecuencia con que se toman las observaciones (por ejemplo: mes, semanas, año, etc.).
- X_t : Una variable ficticia que indica el periodo de preintervención (codificado en 0) o el periodo postintervención (codificado 1).
- Y_t : El resultado en el tiempo t.
- ϵ_t : Error aleatorio.

$$Y_t = \beta_0 + \beta_1 T + \beta_2 X_t + \beta_3 T X_t + \epsilon_t$$

Donde,

- β_0 : Representa el nivel de la línea base en $T=0$.
- β_1 : Representa el cambio en la variable respuesta asociado con un aumento de la unidad de tiempo (que representa la tendencia antes de la intervención).
- β_2 : Cambio de nivel en el periodo posterior a la intervención ($H_0: \beta_2=0$).
- β_3 : Cambio de tendencia tras la intervención ($H_0: \beta_3=0$).

El modelo de regresión anterior representa el modelo de impacto “Cambio de nivel y pendiente” de la figura 4.4.

El modelo de regresión para representar el modelo de impacto “Cambio de nivel” de la figura 4.4 se puede representar de la siguiente manera debido a que no posee un cambio en la tendencia:

$$Y_t = \beta_0 + \beta_1 T + \beta_2 X_t + \epsilon_t$$

Por otra parte, para representar el modelo de impacto “Cambio de pendiente” de la figura 4.4, se excluye el término $\beta_2 X_t$ del modelo de regresión segmentada:

$$Y_t = \beta_0 + \beta_1 T + \beta_3 T X_t + \epsilon_t$$

2.4.3. Consideraciones metodológicas en el análisis de ITS

2.4.3.1. Estacionalidad

Es importante mencionar que la estacionalidad en este tipo de series puede causar dos problemas: Primero, si existe una distribución desigual en la frecuencia con las que se toman las observaciones antes y después de la intervención, esto podría sesgar los resultados, especialmente en el análisis de series cortas. En segundo lugar, se pueden provocar problemas de autocorrelación y sobre dispersión, ya que, por ejemplo los resultados de un mes tienden a ser más similares a meses vecinos dentro de una misma época del año.

2.4.3.2. Variables de Confusión

Otro punto importante en este tipo de modelos es que, en general no se ven afectados por variables de confusión (variables que pueden alterar las estimaciones de las medidas de asociación si no se reparten de forma homogénea entre los grupos de estudio) típicas que permanecen bastante constantes, como por ejemplo la edad de la población, debido a que cambian con lentitud a lo largo del tiempo. Sin embargo, este modelo puede verse afectado por factores de confusión que cambian más rápidamente con el tiempo,

por ejemplo una enfermedad infecciosa que en particular sea propensa a brotes. Cuando estos factores han sido medidos se pueden controlar incluyendo variables que los representen en el modelo de regresión.

Cuando los factores de confusión no se miden o se desconocen, se pueden utilizar una serie de adaptaciones de diseño para controlar posibles eventos simultáneos entre los que se tienen: agregar un grupo de control que no se hubiese visto afectado por la intervención (conocido como serie de tiempo interrumpido controlado); utilizar un diseño en el que la intervención se introduce en diferentes momentos o agregar fases adicionales de modo que la intervención se introduzca primero y luego retirarla para establecer si esto conduce a una inversión del efecto.

2.4.3.3. Sobredispersión

Por otra parte, en este tipo de diseños se pueden presentar exceso de dispersión. Por ejemplo, una suposición de la distribución Poisson es que la varianza es igual al conteo esperado, pero en los análisis de datos reales la varianza suele ser mayor, este fenómeno se conoce como sobre dispersión. Esto provoca una estimación errada de los errores estándar, por lo que es necesario aplicar un ajuste simple para obtener los errores estándar apropiados en el modelo.

2.4.3.4. Autocorrelación

En los modelos de regresión estándar se supone que las observaciones son independientes, pero en datos de series temporales se viola este supuesto debido a que las observaciones consecutivas tienden a ser más similares entre si que las que están más separadas, este enfoque se conoce como autocorrelación. En estudios epidemiológicos principalmente, la autocorrelación se explica en gran medida por otras variables como por ejemplo la estacionalidad, por lo tanto, después de controlar estos factores la autocorrelación residual generalmente deja de ser un problema. Sin embargo, la autocorrelación siempre debe evaluarse examinando la función de autocorrelación parcial. Cuando se mantiene la correlación en los residuos se debe ajustar para el uso de métodos como el modelo autoregresivo integrado de medias móviles (ARIMA).

Finalmente, se han descrito una serie de técnicas de comprobación de modelos, como los residuos de trazado y las funciones de autocorrelación parcial. Además, se puede realizar análisis de sensibilidad para probar el impacto de variar una gama de supuestos del modelo, como diferentes retrasos, tipos de modelos de impacto o enfoque de ajuste para la estacionalidad.

Capítulo 3

Análisis de muertes por accidentes de tránsito

Para la creación de la base de datos de muertes por accidentes de tránsito, se descargaron desde la página web del DEIS (Departamento de Estadísticas e Información de Salud) los registros de mortalidad desde el año 1997 al 2016. Cada uno de los años cuenta con los siguientes registros de las personas fallecidas:

- **EST_CIVIL:** Código del Estado Civil
 - 1= Soltero
 - 2= Casado
 - 3= Viudo
 - 4= Divorciado
 - 9= Ignorado
- **EDAD_TIPO:** Tipo de edad
 - 1= Años
 - 2= Meses
 - 3= Días
 - 4= Horas y Minutos
- **EDAD_CANT:** Edad cantidad en años
- **CURSO_INS:** Último curso escolar aprobado

- **NIVEL_INS:** Código del nivel educativo
 - 1= Superior
 - 2= Medio
 - 3= Secundaria
 - 4= Básico o Primario
 - 0-5=Ninguno
 - 9= Ignorado

- **ACTIV:** Código para la condición de actividad
 - 0= Inactivo
 - 1= Activo
 - 2= Cesante o desocupado
 - 9= Ignorado

- **OCUPA:** Código de la ocupación. El dato está condicionado al dato de la actividad
 - **Actividad= 0**
 - 2= Labores de casa
 - 3= Estudiante
 - 4= Rentista
 - 5= Jubilado
 - 6= Inválido ó recluso
 - 7= Otros
 - 8= Ninguno

 - **Actividad= 1 (Gran Grupo de Actividad)**
 - 0= Fuerza Armadas y del Orden
 - 1= Miembros del Poder Judicial

- 2= Profesionales, Científicos o Intelectuales
- 3= Técnicos y Profesionales medios
- 4= Empleados de oficina
- 5= Trabajadores de servicios y vendedores de comercio y mercado
- 6= Agricultores y trabajadores calificados agropecuarios y pesqueros
- 7= Oficiales, operarios y artesanos de artes mecánicas y otros oficios
- 8= Operadores de instalación, máquinas y montadores
- 9= Trabajadores no calificados

- Actividad= 2

- X= Cesante o desocupado

- Actividad= 9

- X= Ignorado

- **CATEG:** Código de la categoría ocupacional. El valor está condicionado al dato de la actividad
 - Actividad= 0
 - 0= Inactivo

 - Actividad= 1
 - 1= Patrón
 - 2= Empleado
 - 3= Obrero
 - 4= Trabajador por cuenta propia
 - 9= Ignorado

 - Actividad= 2

- 9= Ignorado

- Actividad= 9

- 9= Ignorado

- **DIA_DEF:** Día de la fecha de defunción

- **MES_DEF:** Mes de la fecha de defunción

- **AÑO_DEF:** Año de la fecha de defunción

- **LUGAR_DEF:** Código del lugar ocurrencia de la defunción
 - 1= Hospital o Clínica
 - 2= Casa Habitación
 - 3= Otro

- **REG_RES:** Región de residencia

- **SER_RES:** Servicio de Salud de residencia

- **COMUNA:** Código de la comuna de residencia

- **URBA_RURAL:** Código para definir área Urbano o Rural
 - 1= Urbano
 - 2= Rural

- **DIAG1:** Causa básica de defunción

- **DIAG2:** Causa externa de defunción

- **AT_MEDICA:** Código para identificar si tuvo atención médica
 - 1= Si
 - 2= No
 - 9= Ignorado

- **CAL_MEDICO:** Código para calidad de quién certifica la defunción

- 1= Médico tratante
 - 2= Médico Legista - Patólogo
 - 3= Otro
 - 4= Información de matronas
 - 5= Testigo
 - 9= Ignorado
- **COD_MENOR:** Código para identificar si el fallecido es menor de 1 año
- 0= Mayor de 1 año
 - 1= Menor de 1 año

Para obtener todas las muertes provocadas por accidentes de tránsito, la variable “*DIAG2*” se debe filtrar por todas causas que comienzan con V.

3.1. Serie de tiempo de muertes por accidentes de tránsito

La serie se creó obteniendo las frecuencias de muertes semanales desde el año 2000 al año 2017.

La serie de muertes por accidentes de tránsito a lo largo de todo el país es la siguiente:

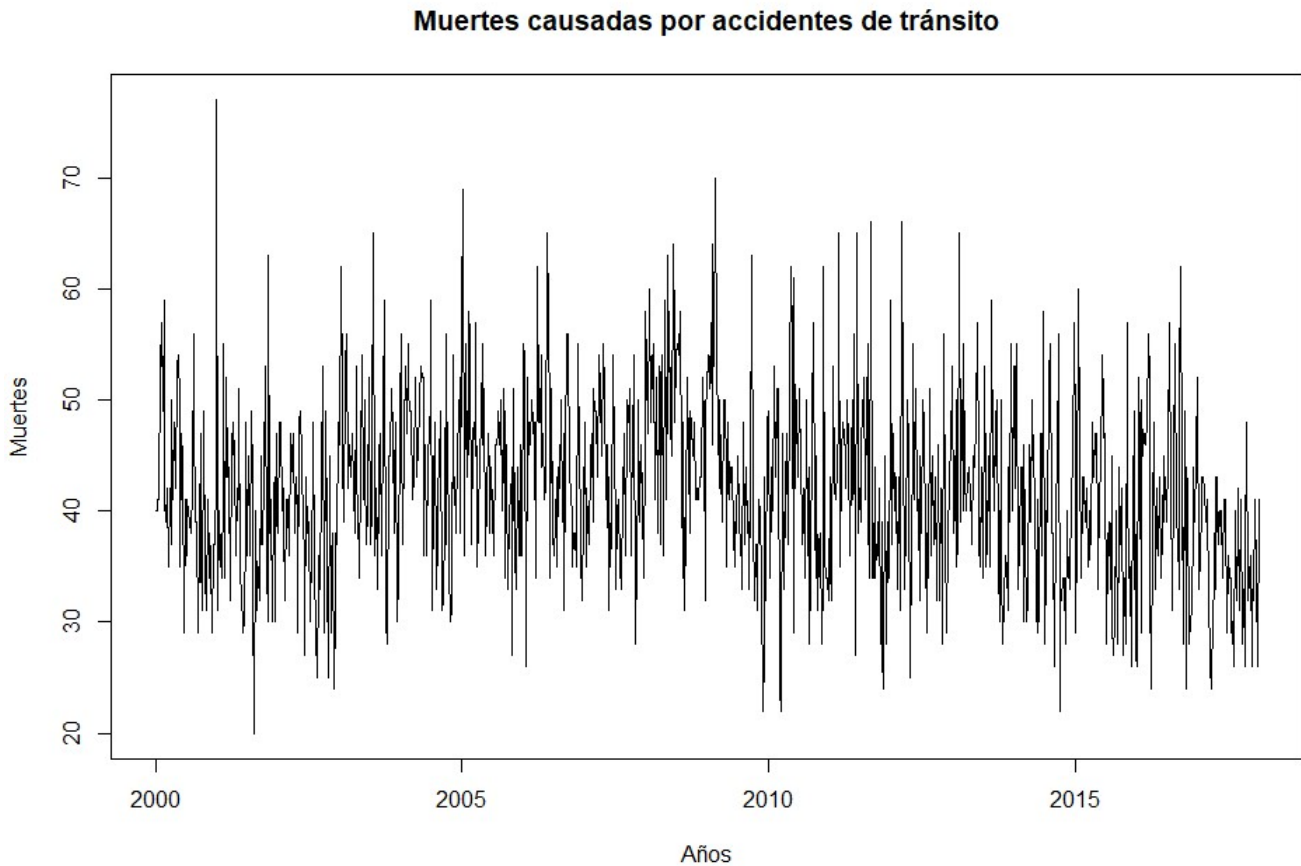


Figura 3.1: Serie de muertes por accidentes de tránsito

Es fácil notar que la serie posee mucha variabilidad, esto se debe a que la serie es semanal y a la gran cantidad de registros que se tienen.

Para lograr visualizar de mejor manera el comportamiento de los datos antes y después de la implementación de la ley, realizaremos un “zoom” a través del programa R-Studio desde el año 2013 hasta el fin de la serie, tal como se muestra a continuación:

Ventana de muertes causadas por accidentes de tránsito

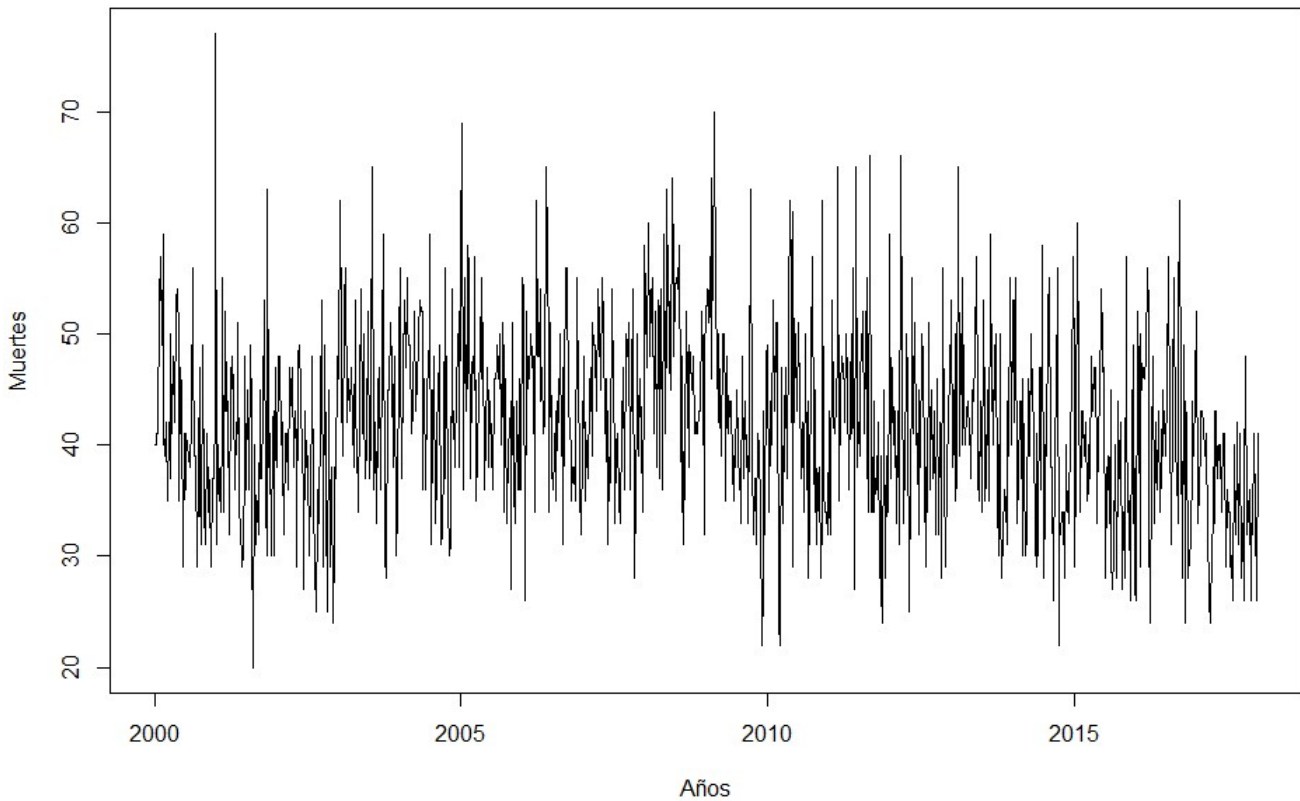


Figura 3.2: Ventana de muertes por accidentes de tránsito

Luego de la implementación de la ley Emilia que fue el 16 de septiembre del 2014, la serie presenta una leve disminución, esto se puede apreciar observando a continuación, la gráfica de la tendencia.

3.1.1. Tendencia de la serie

La tendencia de la serie se estima en el programa R-Studio a través del método de medias móviles. Este método se basa en sustituir el valor observado de la variable en cada periodo por la media de dicho valor y algunos anteriores y posteriores a él. El primer valor de una media móvil de orden k es:

$$MM_1(k) = \frac{X_1 + X_2 + \dots + X_K}{k};$$

El segundo valor es:

$$MM_2(k) = \frac{X_2 + X_3 + \dots + X_{K+1}}{k}$$

De esta misma manera se calculan los valores sucesivos y se obtiene una serie alisada, de la cual se han eliminado o suavizado las fluctuaciones con periodicidad igual o inferior al orden de la media móvil. Esta

serie alisada puede considerarse como una serie de estimaciones de la tendencia en cada periodo. (18)

En la siguiente figura se observa la tendencia de la serie, donde la línea roja representa la implementación de la ley y la línea azul el quiebre de tendencia más importante que existió luego de la implementación.

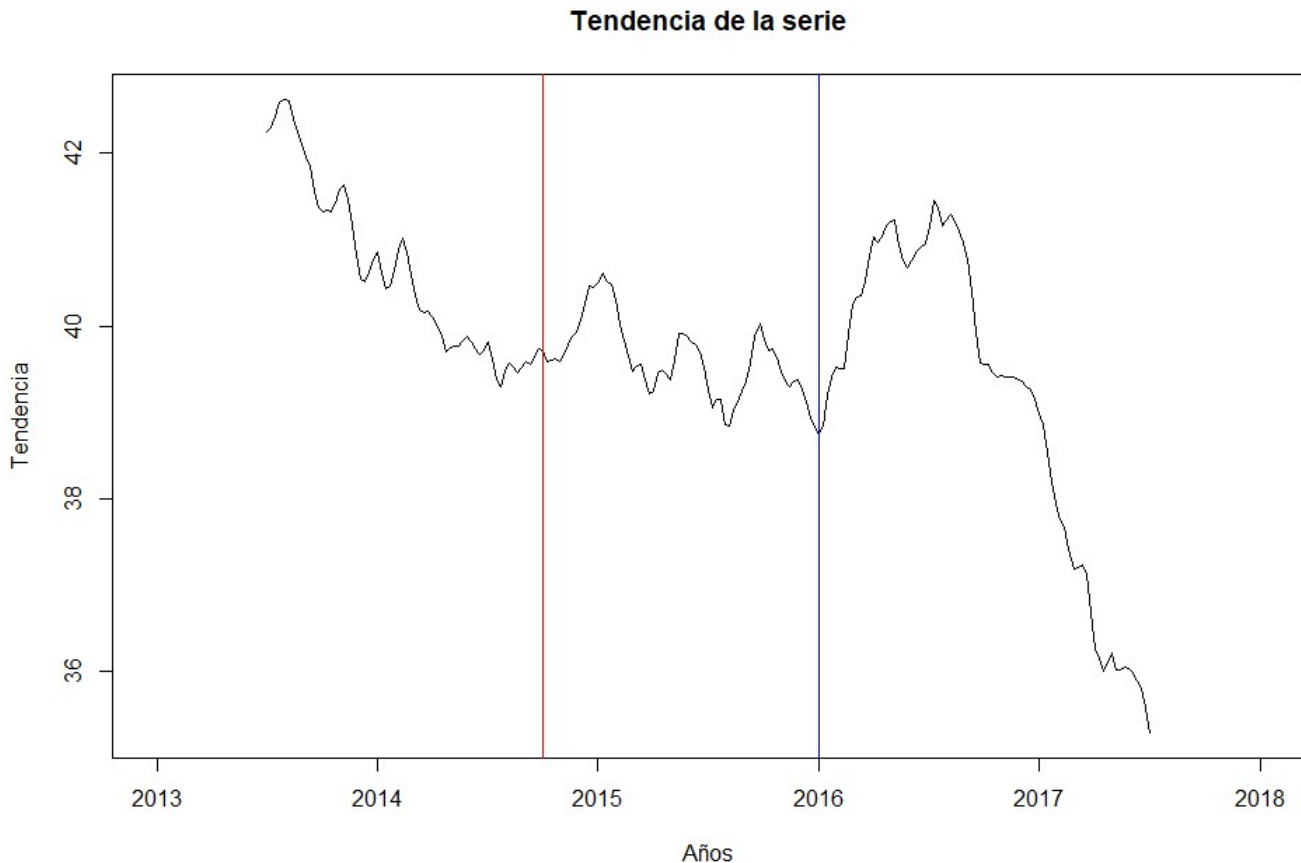


Figura 3.3: Ventana de la tendencia con sus quiebres

En la figura anterior podemos apreciar que desde el año 2013 existe una tendencia decreciente y que luego de la implementación de la ley existe un comportamiento cíclico que tiene una leve disminución de tendencia. En el año 2016 donde se marca el segundo cambio de tendencia se tiene un aumento que dura hasta el mes de Agosto y luego comienza un decrecimiento que llega a su mínimo al término del año 2017.

3.1.2. Test de Chow

Dado que conocemos con exactitud cuando fue aplicada la intervención, a través del test de Chow podemos comprobar si realmente existe un cambio de tendencia en los datos. Donde la hipótesis es:

H_0 : No existe un cambio en el intercepto y la tendencia v/s H_1 : Existe un cambio en el intercepto y la tendencia

Los valores obtenidos en los modelos para el cálculo del estadístico se muestran a continuación:

Variable	Resultado
SCR_R	66060.1
SCR_1	53825.23
SCR_2	10687.13
n	937
k	2

Cuadro 3.1: Tabla de resultados para calcular test de Chow en muertes

La fórmula para calcular el test de Chow es la siguiente:

$$F = \frac{(SCR_R - SCR_{NR})/k}{(SCR_{NR})/(n - 2k)} \sim F_{[k, (n-2k)]}$$

La SCR_{NR} está representada por $SCR_1 + SCR_2$ y reemplazando los valores obtenidos en la formula anterior nos queda:

$$F = \frac{(66060.1 - (53825.23 + 10687.13))/2}{(53825.23 + 10687.13)/(937 - 2 * 2)} \sim F_{[2, 933]}$$

$$F_{obs} = 11.14402$$

$$F_{cri} = F[(0, 05; 2; 933)] = 3.0053$$

Dado que el $F_{obs} > F_{cri}$ se rechaza la hipótesis nula, es decir, existe un cambio de nivel y tendencia en los datos luego de la implementación de la ley.

De forma análoga aplicaremos este test para probar si el cambio de tendencia que se ve reflejado en la línea azul de la figura 3.3 se valida con el test:

Los valores obtenidos en los modelos para el cálculo del estadístico se muestran a continuación:

Variable	Resultado
SCR_R	66060.1
SCR_1	62784.4
SCR_2	1783.962
n	884
k	2

Cuadro 3.2: Tabla de resultados para calcular test de Chow en muertes, segunda intervención

$$F = \frac{(66060.1 - (62784.4 + 1783.962))/2}{(62784.4 + 1783.962)/(937 - 2 * 2)} \sim F_{[2,933]}$$

$$F_{obs} = 10.7315$$

$$F_{cri} = F[(0, 05; 2; 937)] = 3.006$$

Dado que el $F_{obs} > F_{cri}$ se rechaza la hipótesis nula, es decir, existe un cambio en el intercepto y tendencia en los datos después del año 2016.

3.2. Serie de Muertes por accidentes de tránsito en la Región Metropolitana

Dado que la ley Emilia tiene una mayor importancia en áreas urbanas que rurales es importante analizar la ley en la Región Metropolitana. Los datos corresponden al conteo de muertes por semana desde el año 2007 al 2017.

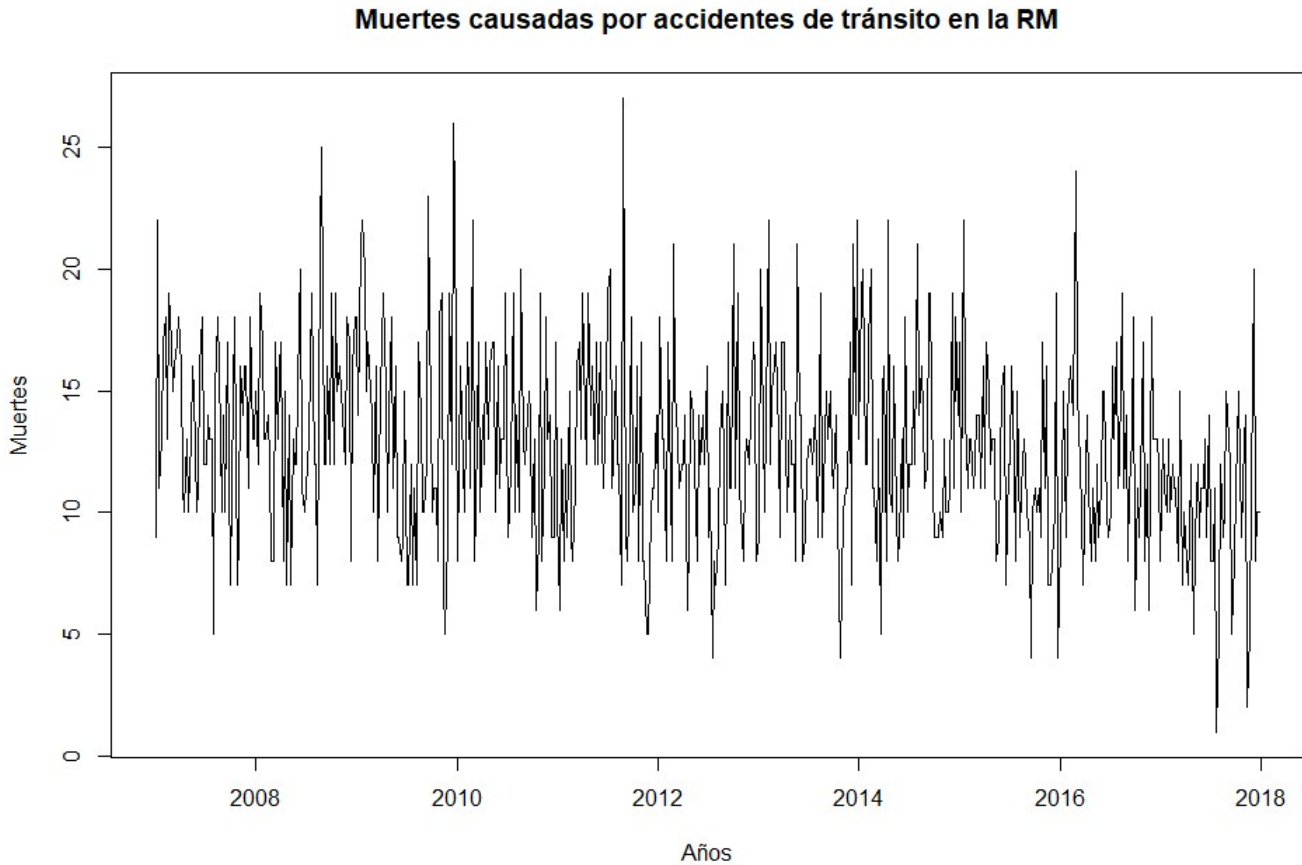


Figura 3.4: Serie muertes por accidentes de tránsito en la RM

3.2.1. Representación gráfica de la tendencia

La siguiente figura muestra la tendencia de la serie desde el año 2013 para poder apreciar de mejor manera el comportamiento de los datos luego de la implementación de la ley. Es importante notar que el decaimiento de tendencia se hace notorio luego de la implementación, y posteriormente en Julio del año 2016 se presenta un descenso que llega hasta su mínimo al terminar el año 2017.

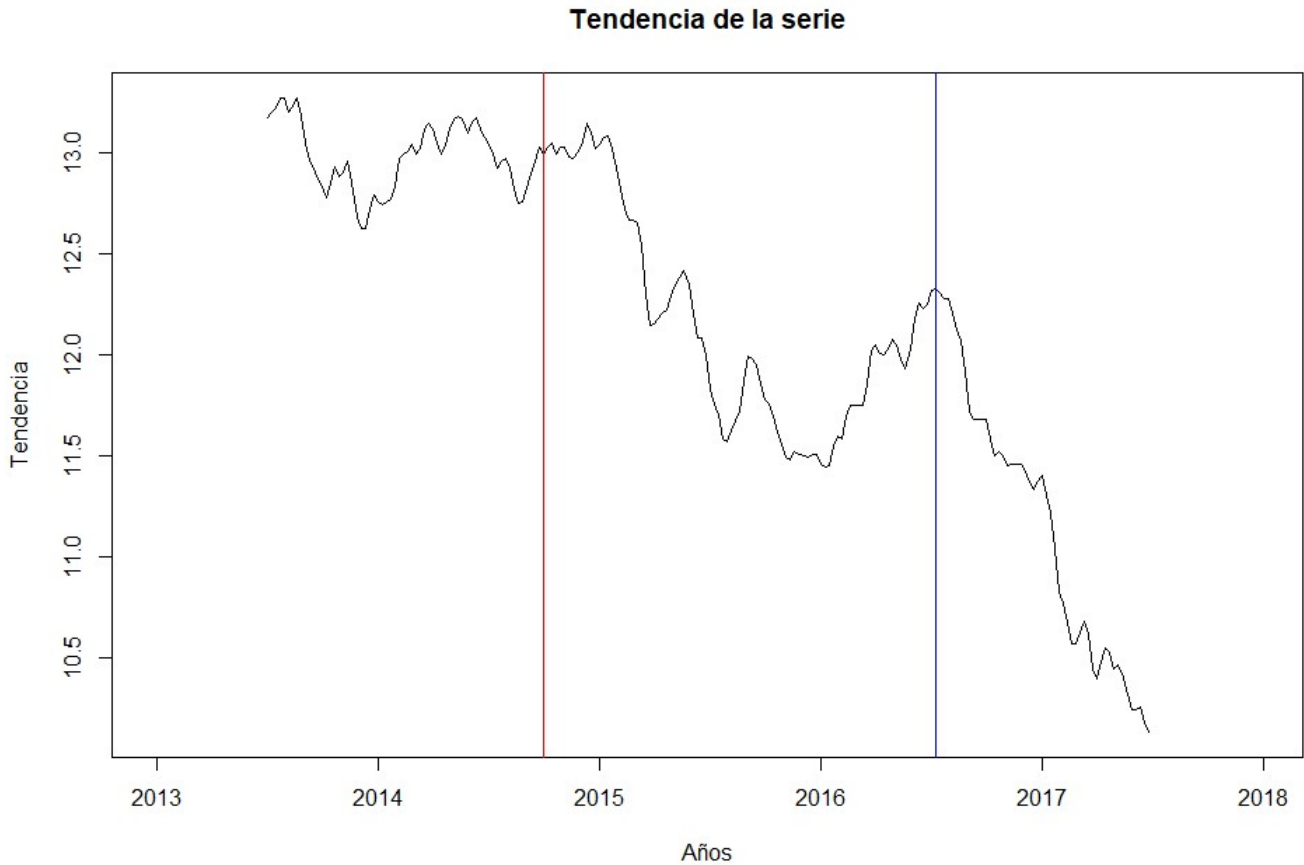


Figura 3.5: Tendencia de la serie en la RM

En la figura la línea roja representa la intervención y la línea azul el cambio más importante de tendencia.

3.2.2. Test de Chow

Para el cálculo del test, debemos obtener una serie de valores que se muestran en la siguiente tabla:

Variable	Resultado
SCR_R	8685.35
SCR_1	6520.84
SCR_2	2092.03
n	572
k	2

Cuadro 3.3: Tabla de resultados para calcular test de Chow de muertes en la RM

La SCR_{NR} está representada por $SCR_1 + SCR_2$ y reemplazando los valores obtenidos en la formula anterior nos queda:

$$F = \frac{(8685.35 - (6520.84 + 2092.03))/2}{(6520.84 + 2092.03)/(572 - 2 * 2)} \sim F_{[2,568]}$$

$$F_{obs} = 4.7795$$

$$F_{cri} = F[(0, 05; 2; 568)] = 3.011$$

Dado que el $F_{obs} > F_{cri}$ se rechaza la hipótesis nula, es decir, existe un cambio de nivel y tendencia en los datos luego de la implementación de la ley.

De manera análoga aplicaremos este test para probar si el cambio de tendencia que se ve reflejado en la línea azul de la figura 3.5 se valida con el test:

Los valores obtenidos en los modelos para el cálculo del estadístico se muestran a continuación:

Variable	Resultado
SCR _R	66060.1
SCR ₁	8167.468
SCR ₂	463.382
n	572
k	2

Cuadro 3.4: Tabla de resultados para calcular test de Chow de muertes en la RM, segunda intervención

$$F = \frac{(66060.1 - (8167.468 + 463.382))/2}{(8167.468 + 463.382)/(572 - 2 * 2)} \sim F_{[2,568]}$$

$$F_{obs} = 3.5866$$

$$F_{cri} = F[(0, 05; 2; 937)] = 3.011$$

Dado que el $F_{obs} > F_{cri}$ se rechaza la hipótesis nula, es decir, existe un cambio en el intercepto y tendencia en los datos después de mayo del año 2016.

3.3. Serie de muertes en peatones y ocupante de vehículos causadas por accidentes de tránsito

Es importante analizar en cuál de estas dos áreas la ley Emilia tuvo un mayor impacto, para esto replicaremos el análisis anterior a las muertes de peatones y ocupantes de vehículo.

3.3.1. Serie de Peatones causadas por accidentes de tránsito

Según la clasificación de ICD(6) las muertes de peatones están clasificadas desde la V011 a V099 para el “DIAG2”, para formar la serie se calculó la frecuencia semanal de muertes de peatones en accidentes de tránsito.

Graficamente podemos ver el comportamiento de la serie desde el año 2000 al 2017.

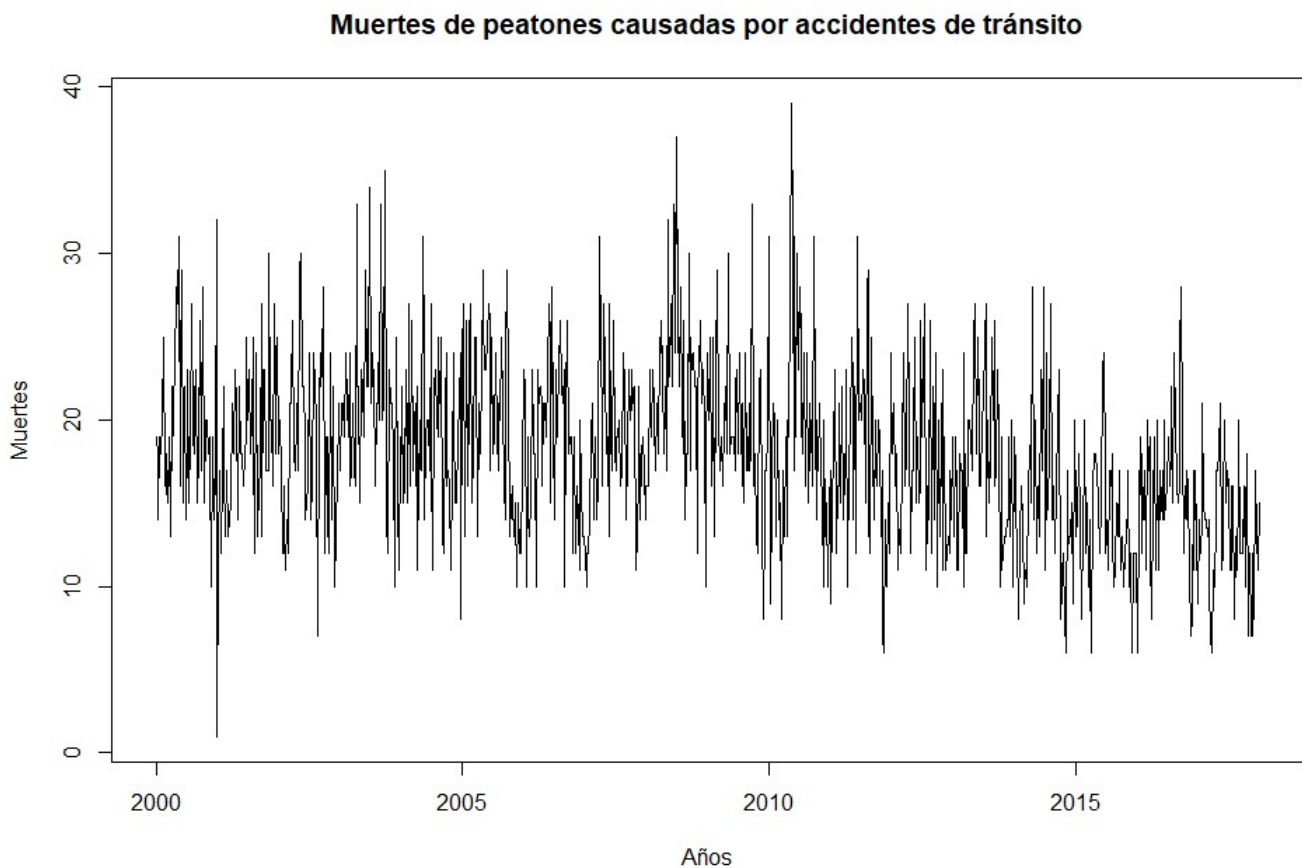


Figura 3.6: Serie muertes de peatones causadas por accidentes de tránsito

De la figura 3.8 podemos apreciar que tiene un comportamiento bastante homogéneo con respecto a la media, pero alrededor del año 2010 tiene una leve disminución en las muertes.

3.3.1.1. Representación gráfica de la tendencia

De la siguiente figura se observa que la tendencia de la serie tenía un comportamiento descendiente desde el año 2013 y que luego de la implementación de la ley tenemos un segundo quiebre en Agosto del año 2016, donde la serie decrece considerablemente.

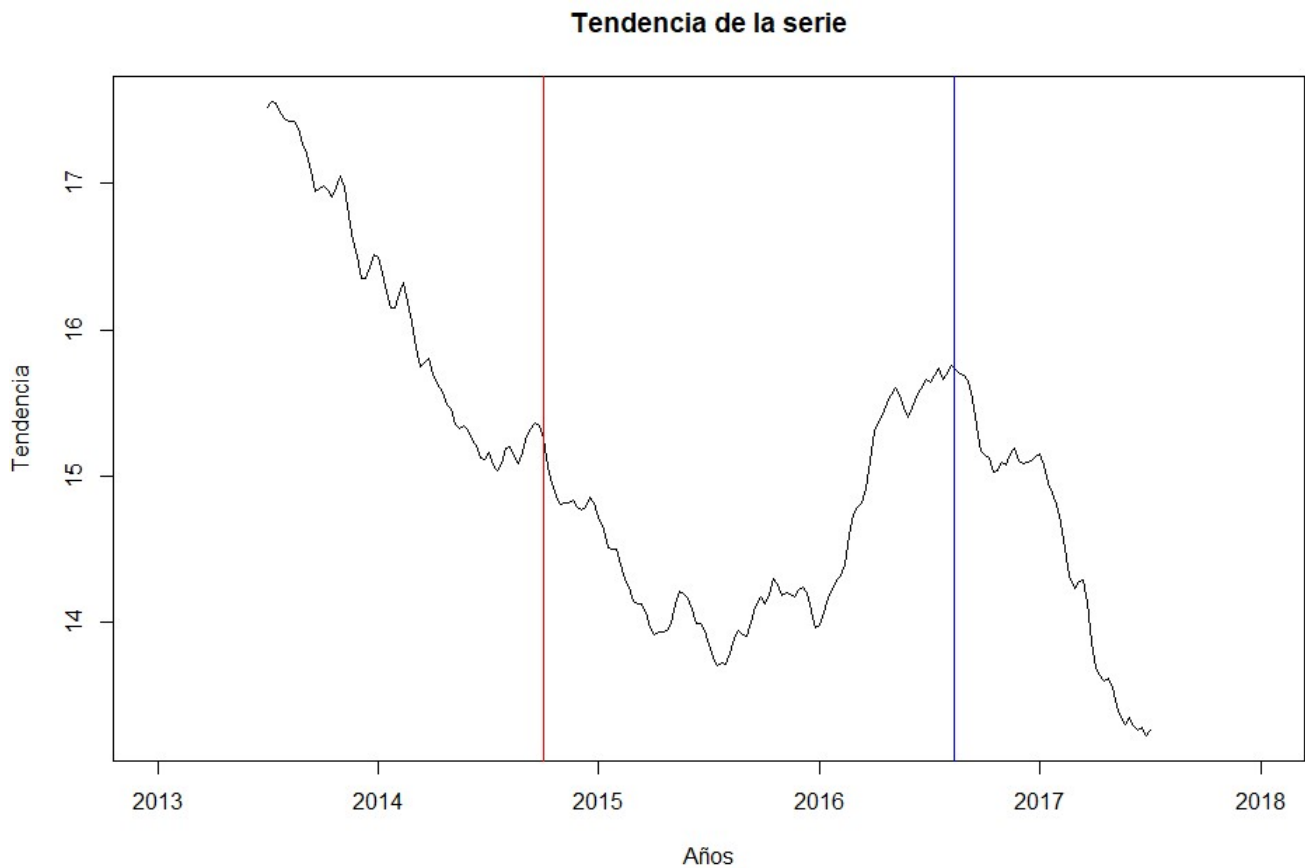


Figura 3.7: Tendencia de la serie de muertes de peatones

De la figura 3.7 tenemos que la línea roja representa la implementación de la ley Emilia y la línea azul el cambio de tendencia que se puede apreciar en el mes de Agosto del año 2016. Para comprobar esto realizaremos a continuación el test de Chow.

3.3.1.2. Test de Chow

Para comprobar si existe un cambio de nivel y tendencia luego de la implementación de la ley debemos obtener una serie de valores que se muestran en la siguiente tabla:

Variable	Resultado
SCR_R	23771.49
SCR_1	20115.85
SCR_2	2854.446
n	937
k	2

Cuadro 3.5: Tabla de resultados para calcular test de Chow de muertes de peatones

La SCR_{NR} está representada por $SCR_1 + SCR_2$ y reemplazando los valores obtenidos en la formula anterior nos queda:

$$F = \frac{(23771.49 - (20115.85 + 2854.446))/2}{(20115.85 + 2854.446)/(937 - 2 * 2)} \sim F_{[2,933]}$$

$$F_{obs} = 16.20151$$

$$F_{cri} = F[(0, 05; 2; 933)] = 3.005$$

Dado que el $F_{obs} > F_{cri}$ se rechaza la hipótesis nula, es decir, existe un cambio de nivel y tendencia en los datos luego de la implementación de la ley.

De manera análoga aplicaremos este test para probar si el cambio de tendencia que se ve reflejado en la línea azul de la figura 3.7.

Los valores obtenidos en los modelos para el cálculo del estadístico se muestran a continuación:

Variable	Resultado
SCR_R	23771.49
SCR_1	23296.5
SCR_2	279.7775
n	937
k	2

Cuadro 3.6: Tabla de resultados para calcular test de Chow de muerte de peatones, segunda intervención

$$F = \frac{(23771.49 - (23296.5 + 279.7775))/2}{(23296.5 + 279.7775)/(937 - 2 * 2)} \sim F_{[2,933]}$$

$$F_{obs} = 3.846006$$

$$F_{cri} = F[(0, 05; 2; 933)] = 3.005$$

Dado que el $F_{obs} > F_{cri}$ se rechaza la hipótesis nula, es decir, existe un cambio en el intercepto y tendencia en los datos en agosto del año 2016.

3.3.2. Serie de muertes de ocupante de vehículo causadas por accidentes de tránsito

Según la clasificación de ICD(6) las muertes de ocupantes de vehículo están clasificadas para “DIAG2” desde el V401 al V799, pero según el análisis previo que se realizó en los datos, el 70 % de los casos estaban clasificados como “Otros”, por esta razón también se incluyó esta variable al estudio. La serie se formó calculando la frecuencia semanal de muertes de ocupante de vehículo.

Graficamente podemos ver el comportamiento de la serie desde el año 2000 al 2017.

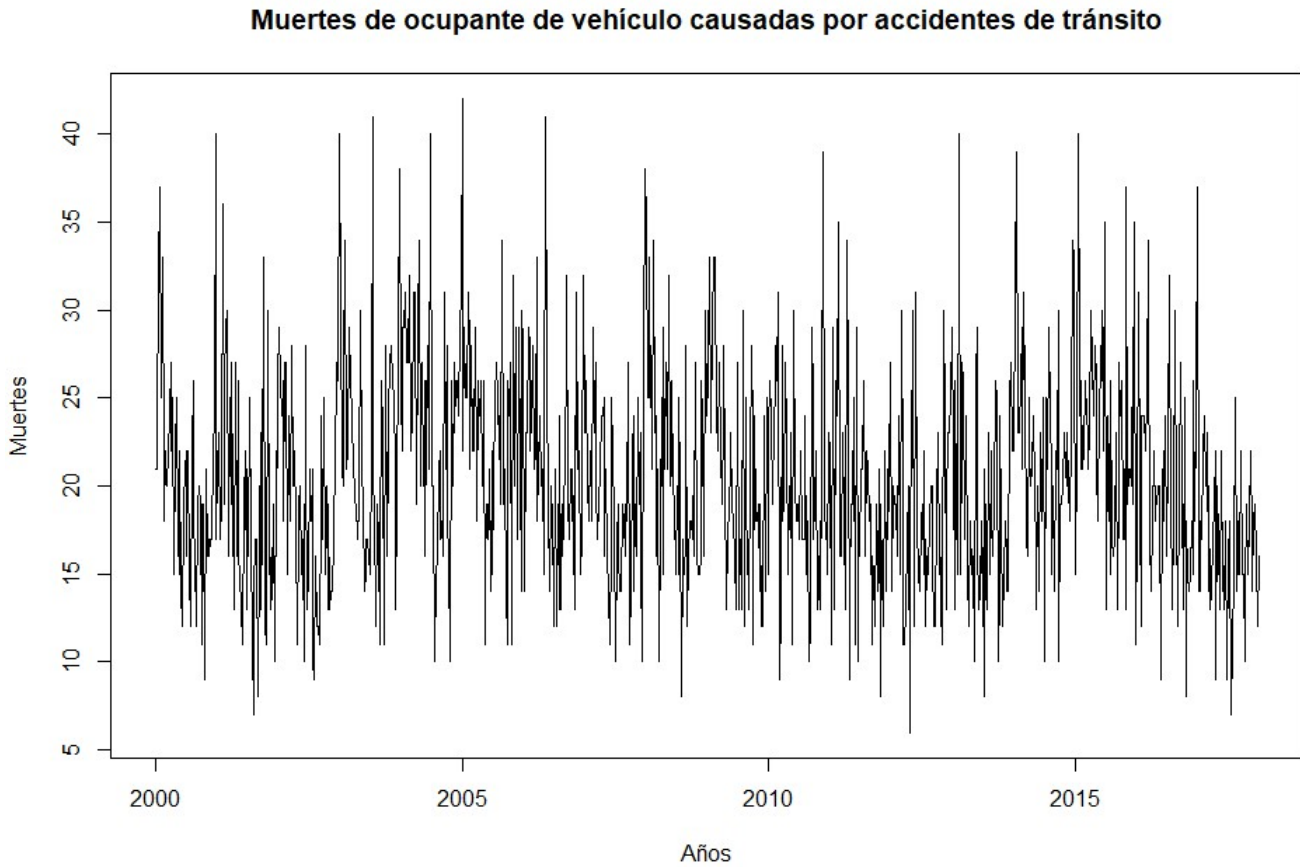


Figura 3.8: Serie muertes de ocupante de vehículo causadas por accidentes de tránsito

De la figura 3.8 se observa que la serie tiene un comportamiento estacional a lo largo de la serie.

3.3.2.1. Representación gráfica de la tendencia

De la siguiente figura se observa que la tendencia es ascendente desde el año 2013 hasta septiembre del año 2015, es decir, tuvo un aumento inclusive después de la implementación de la ley, pero luego se observa un comportamiento descendente muy marcado al llegar al final de la serie.

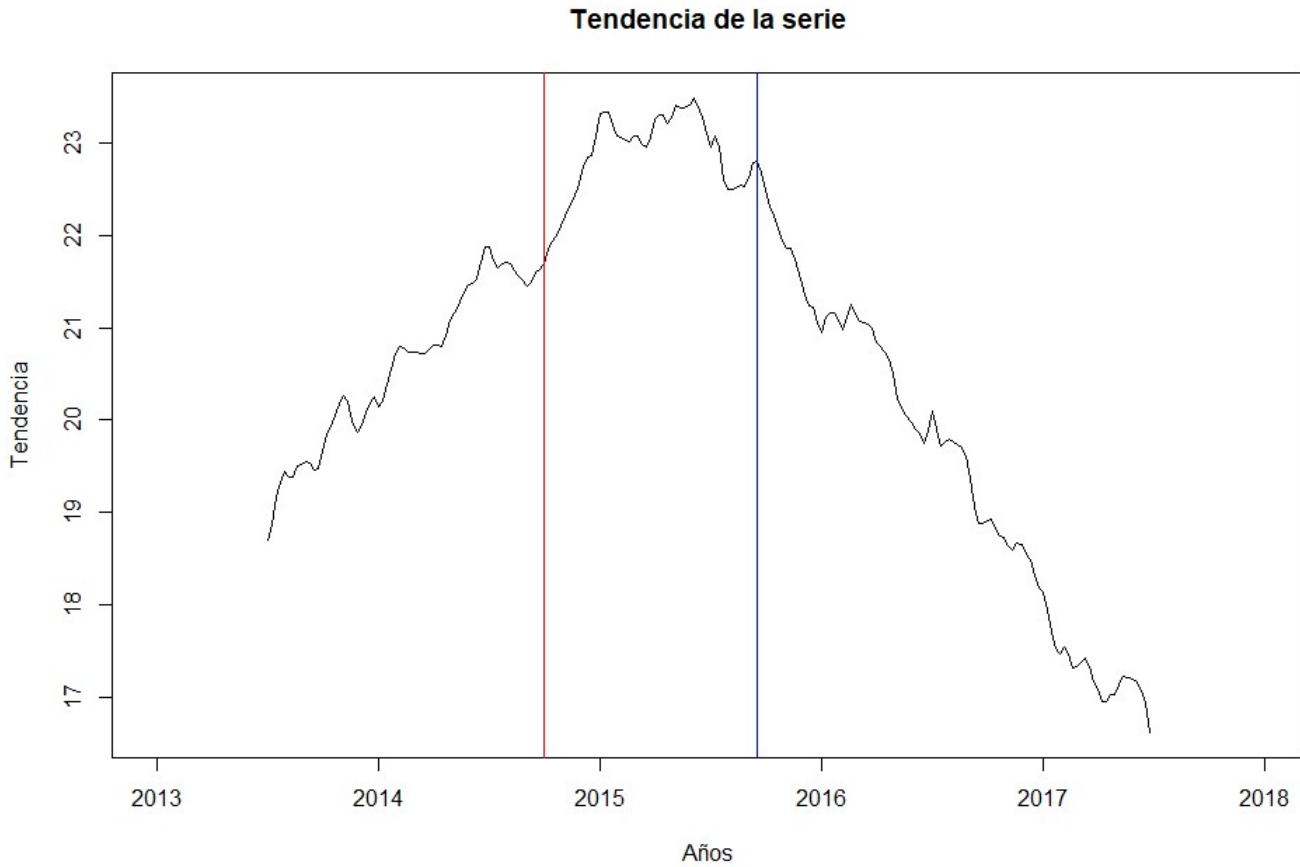


Figura 3.9: Tendencia de la serie en muertes de ocupante de vehículo

De la figura 3.9 tenemos que la línea roja representa la implementación de la ley Emilia y la línea azul cuando la tendencia comenzó a descender a principios de septiembre del año 2015.

3.3.2.2. Test de Chow

Para comprobar si existe un cambio de nivel y tendencia luego de la implementación de la ley debemos obtener una serie de valores que se muestran en la siguiente tabla:

Variable	Resultado
SCR_R	35211.15
SCR_1	28849.66
SCR_2	5148.102
n	936
k	2

Cuadro 3.7: Tabla de resultados para calcular test de Chow en muertes de ocupante de vehículo

La SCR_{NR} está representada por $SCR_1 + SCR_2$ y reemplazando los valores obtenidos en la formula anterior nos queda:

$$F = \frac{(35211.15 - (28849.66 + 5148.102))/2}{(28849.66 + 5148.102)/(936 - 2 * 2)} \sim F_{[2,932]}$$

$$F_{obs} = 16.56022$$

$$F_{cri} = F[(0, 05; 2; 933)] = 3.005$$

Dado que el $F_{obs} > F_{cri}$ se rechaza la hipótesis nula, es decir, existe un cambio de nivel y tendencia en los datos luego de la implementación de la ley.

De manera análoga aplicaremos este test para probar si el cambio de tendencia que se ve reflejado en la línea azul de la figura 3.9.

Los valores obtenidos en los modelos para el cálculo del estadístico se muestran a continuación:

Variable	Resultado
SCR_R	35211.15
SCR_1	32933.69
SCR_2	1615.779
n	936
k	2

Cuadro 3.8: Tabla de resultados para calcular test de Chow en muertes de ocupante de vehículo, segunda intervención

$$F = \frac{(35211.15 - (32933.69 + 1615.779))/2}{(32933.69 + 1615.779)/(936 - 2 * 2)} \sim F_{[2,932]}$$

$$F_{obs} = 8.886323$$

$$F_{cri} = F[(0, 05; 2; 932)] = 3.005$$

Dado que el $F_{obs} > F_{cri}$ se rechaza la hipótesis nula, es decir, existe un cambio en el intercepto y tendencia en los datos en agosto del año 2016.

Capítulo 4

Análisis de hospitalizaciones por accidentes de tránsito

La creación de la base de datos de hospitalizaciones se realizó obteniendo la información del DEIS, en donde se descargaron las bases de “egresos” desde el año 2006 hasta el año 2017. Luego esto cada uno de los años se filtraron por la variable “DIAG2” en donde se seleccionaron solo los elementos que empiezan con la letra “V”. La variable “DIAG2” representa la causa externa por la que fue hospitalizado el paciente, y las observaciones que empiezan con “V” representan los accidentes de tránsito. La base cuenta con las siguientes variables:

- **SERV_SALUD:** Servicio de Salud, corresponde al servicio de Salud de ocurrencia.
- **ESTAB:** Código de Establecimiento, corresponde al Establecimiento de ocurrencia.
- **SEXO:** Los valores aceptados son:
 - 1=Hombre
 - 2=Mujeres
- **EDAD:** Edad en cantidad, expresadas en años.
- **PREVISIÓN:** Previsión en Salud.
 - 1=FONASA
 - 2=ISAPRE
 - 3=Sin Previsión (Cancela la Atención)

- 4=Otra
- **BENEF:** Clase de beneficiario, corresponde a los Tramos de Fonasa.
 - 1=A
 - 2=B
 - 3=C
 - 4=D
- **COMUNA:** Corresponde a la división político administrativo vigente desde el 2008.
- **FECHA_EGRE:** Fecha de alta del paciente, formato fecha dd/mm/aaaa.
- **SERV_CL_EG:** Servicio Clínico de Egreso, corresponde a los códigos de los Servicios Clínicos del REM-20.
- **DIAS_ESTAD:** Días de estadía total.
- **DIAG1:** Diagnostico principal.
- **DIAG2:** Causa externa.
- **COND_EGR:** Condición al egreso, códigos válidos.
- **INTERV_Q:** Intervención quirúrgica, corresponde:
 - 1=Si
 - 2=No
- **REGIÓN:** Región de residencia, campo creado a partir de la comuna de residencia del paciente, de acuerdo a la División Político Administrativo vigente desde el año 2008.
- **SERV_RES:** Servicio de Salud de referencia, campo creado a partir de la comuna de residencia del paciente, de acuerdo a la División Político Administrativo vigente desde el 2008.

4.1. Serie de tiempo de hospitalizaciones causadas por accidentes de tránsito

La serie de tiempo de hospitalizaciones causadas por accidentes de tránsito se obtuvo tomando la frecuencia semanal desde el año 2006 hasta el 2017.

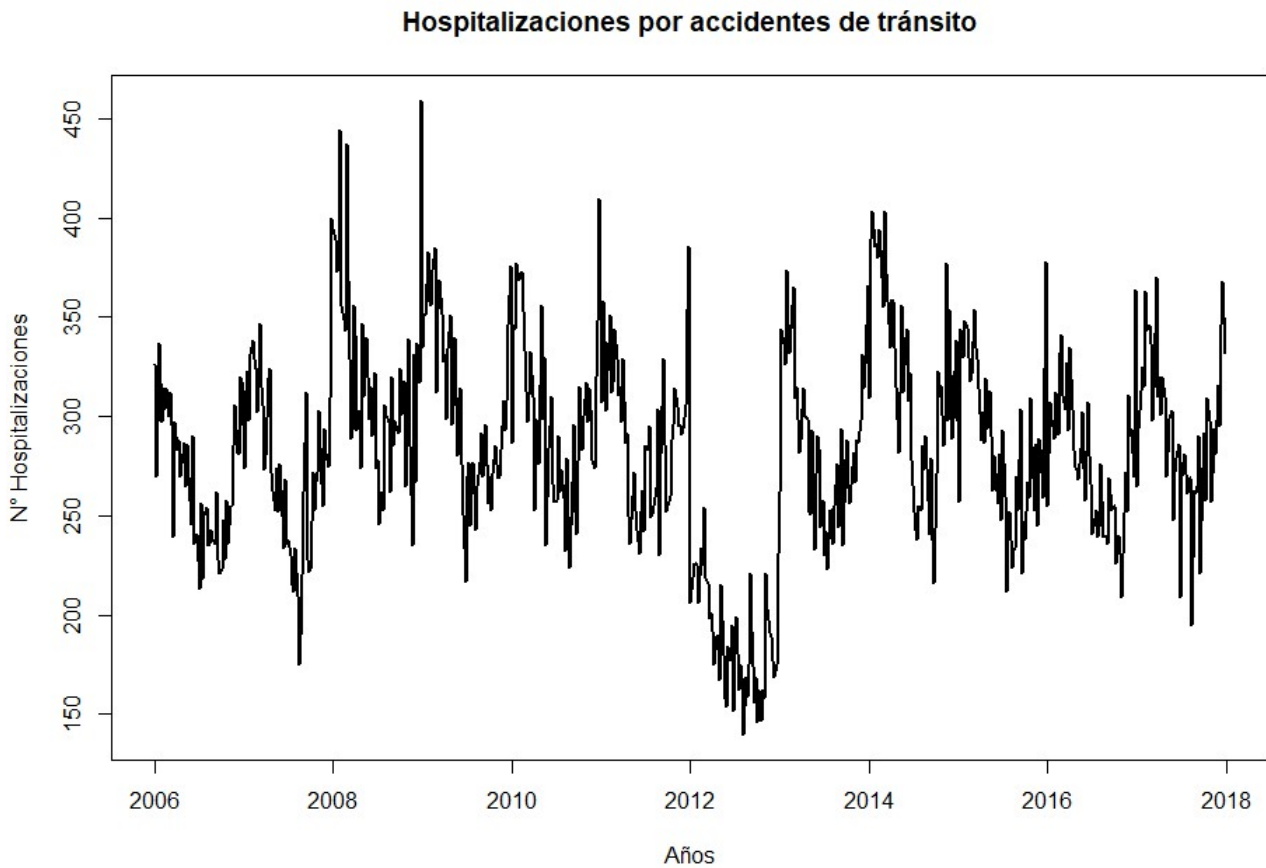


Figura 4.1: Serie de tiempo hospitalizaciones causadas por accidentes de tránsito

Dada la figura 4.2 se puede apreciar una caída muy pronunciada en el número de hospitalizaciones que duró a lo largo de un año. Esto puede deberse a un error en los datos que tiene el DEIS, por esta razón solo se tomarán los datos desde septiembre del 2013 para completar un ciclo completo antes de la implementación de la ley. A continuación, se presenta la serie con la que se realizará el resto del análisis.

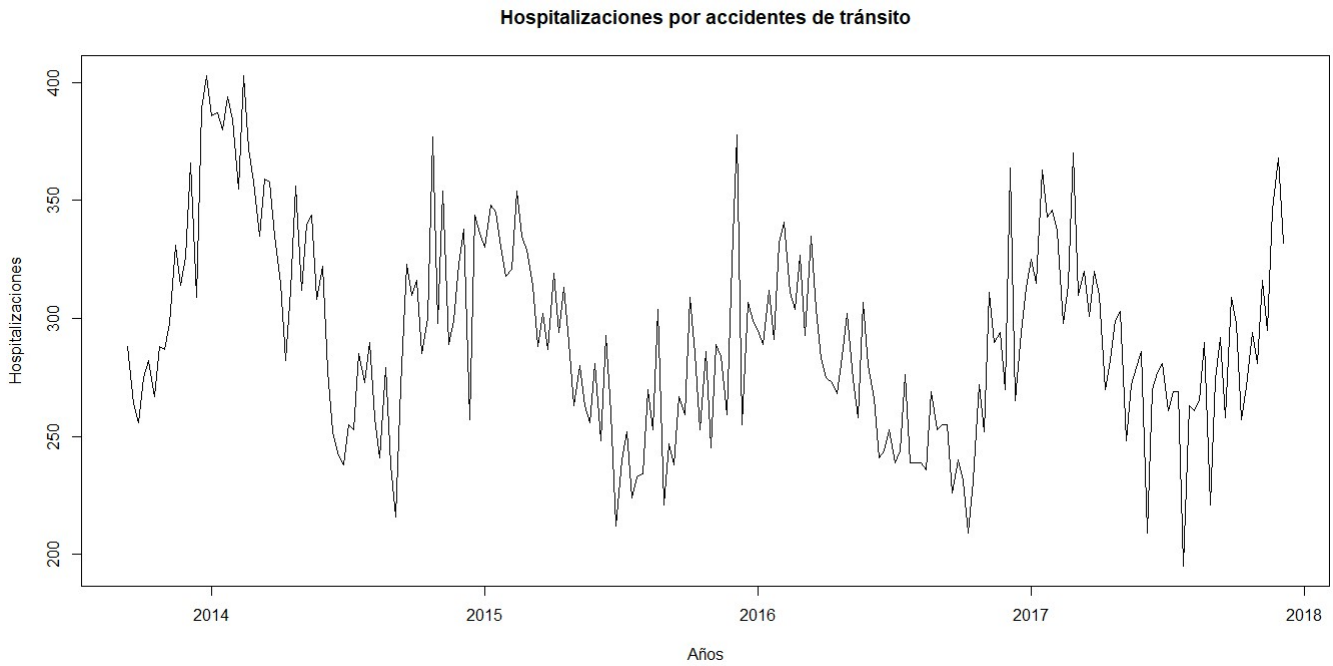


Figura 4.2: Serie de tiempo hospitalizaciones causadas por accidentes de tránsito

De la figura anterior tenemos que la serie presenta un comportamiento estacional.

4.1.1. Tendencia de la serie

En la siguiente figura se puede observar la tendencia de la serie, en donde la línea roja representa el momento de implementación de la ley, mientras que la línea azul un nuevo cambio de tendencia.

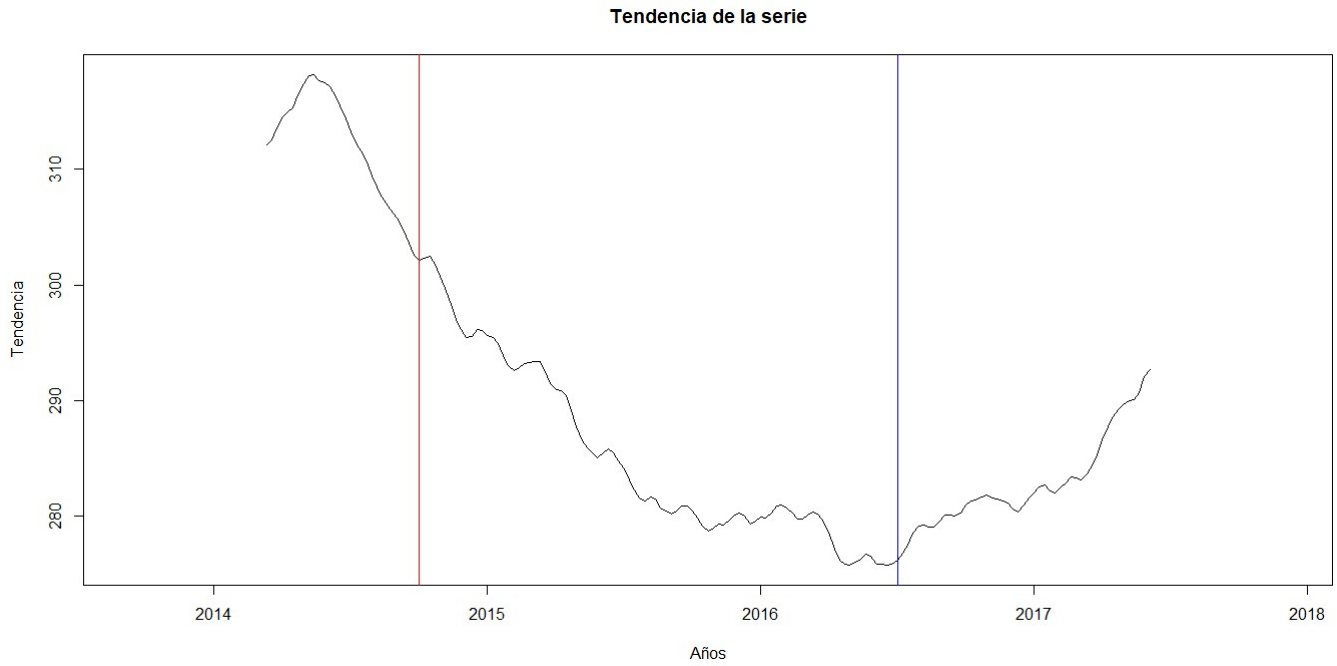


Figura 4.3: Tendencia serie hospitalizaciones

De la figura 4.3, se observa que existe un comportamiento descendiente antes y después de la implementación y que solo existe un aumento en la tendencia en la mitad del 2016 y hasta el fin de la serie.

4.1.2. Test de Chow

Para el cálculo del test de Chow, es necesario obtener los siguientes valores, a través del modelo completo y el modelo restringido para probar si existe o no un cambio estructural en la serie.

Variable	Resultado
SCR_R	358431.4
SCR_1	114830
SCR_2	228639.7
n	221
k	2

Cuadro 4.1: Tabla de resultados para calcular test de Chow en hospitalizaciones

La SCR_{NR} está representada por $SCR_1 + SCR_2$ y reemplazando los valores obtenidos en la formula anterior nos queda:

$$F = \frac{(358431.4 - (114830 + 228639.7))/2}{(114830 + 228639.7)/(221 - 2 * 2)} \sim F_{[2,217]}$$

$$F_{obs} = 4.73$$

$$F_{cri} = F[(0, 05; 2; 217)] = 3.037$$

Dado que el $F_{obs} > F_{cri}$ se rechaza la hipótesis nula, es decir, existe un cambio de nivel y tendencia en los datos luego de la implementación de la ley.

A través de este mismo método comprobaremos si el segundo cambio de tendencia que está marcado con la línea azul en la figura 4.3 se corrobora.

Variable	Resultado
SCR _R	358431.4
SCR ₁	227813.3
SCR ₂	105585.4
n	221
k	2

Cuadro 4.2: Tabla de resultados para calcular test de Chow en hospitalizaciones segunda intervención

La SCR_{NR} está representada por $SCR_1 + SCR_2$ y reemplazando los valores obtenidos en la formula anterior nos queda:

$$F = \frac{(358431.4 - (227813.3 + 105585.4))/2}{(227813.3 + 105585.4)/(221 - 2 * 2)} \sim F_{[2,217]}$$

$$F_{obs} = 7.99$$

$$F_{cri} = F[(0, 05; 2; 217)] = 3.037$$

Dado que el $F_{obs} > F_{cri}$ se rechaza la hipótesis nula, es decir, existe un cambio de nivel y tendencia a partir de junio del año 2016.

4.2. Serie de tiempo de hospitalizaciones causadas por accidentes de tránsito en la Región Metropolitana

Es importante realizar el estudio centrándose en la Región Metropolitana, debido a que es la región más urbana del país y evaluar si la implementación de la ley tiene un real impacto en las hospitalizaciones causadas por accidentes de tránsito.

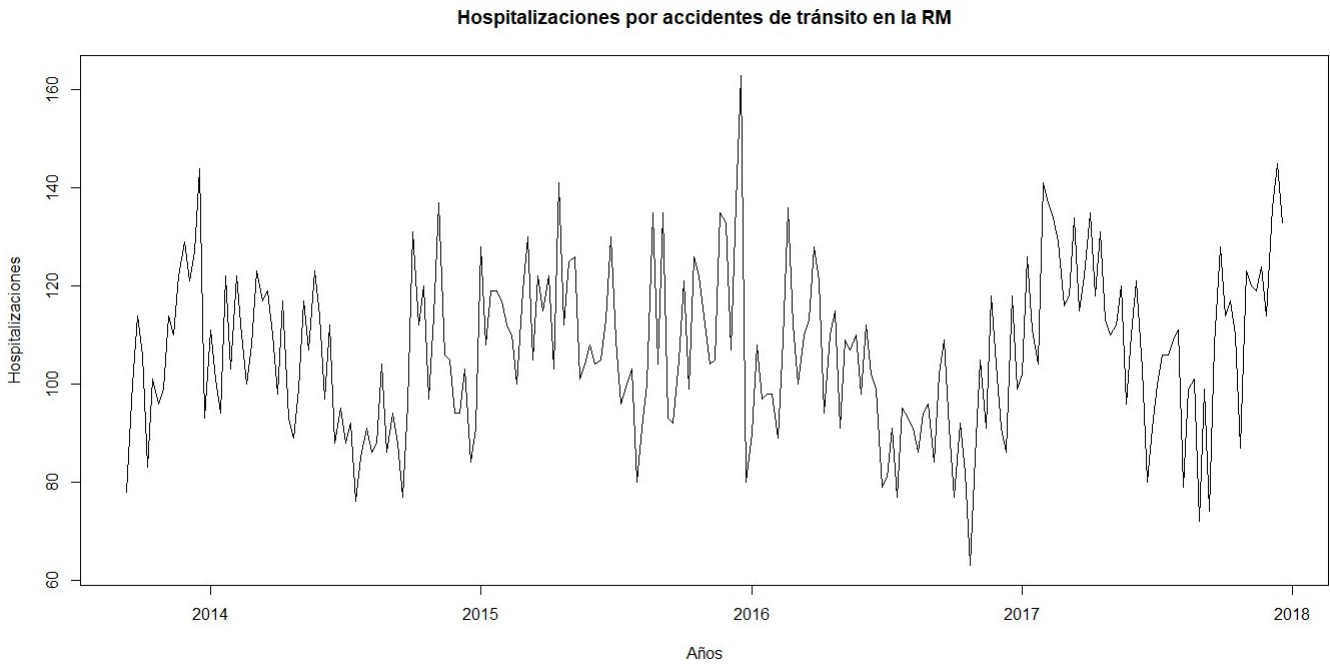


Figura 4.4: Serie de tiempo hospitalizaciones en la Región Metropolitana

4.2.1. Gráfico de la tendencia de la serie

El comportamiento de la tendencia en la Región Metropolitana al contrario de todo el país alcanza su “peak” luego de la implementación de la ley (marcado con la línea roja), esto puede ser explicado por la obligación que se tiene de detener la marcha del vehículo y prestar ayuda en caso de accidente y con esto lograr una disminución en las muertes, pero aumentar las hospitalizaciones.

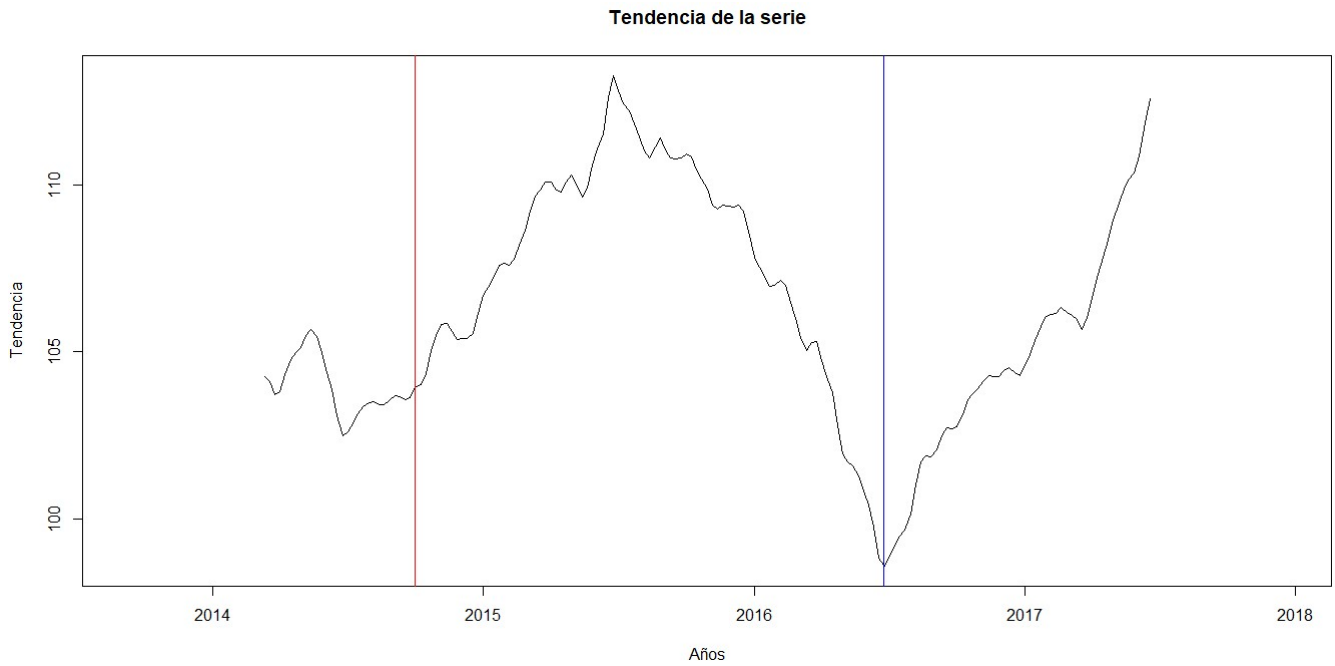


Figura 4.5: Tendencia serie de tiempo hospitalizaciones en la Región Metropolitana

4.2.2. Test de Chow

Al igual que en las series anteriores, se deben obtener las respectivas sumas de cuadrados de los modelos para determinar si se acepta o rechaza la hipótesis nula. A continuación se muestra la tabla de valores:

Variable	Resultado
SCR_R	60082.12
SCR_1	9824.82
SCR_2	48310.43
n	223
k	2

Cuadro 4.3: Tabla de resultados para calcular test de Chow en hospitalizaciones de la RM

La SCR_{NR} está representada por $SCR_1 + SCR_2$ y reemplazando los valores obtenidos en la formula anterior nos queda:

$$F = \frac{(60082.12 - (9824.82 + 48310.43))/2}{(9824.82 + 48310.43)/(223 - 2 * 2)} \sim F_{[2,219]}$$

$$F_{obs} = 3.600027$$

$$F_{cri} = F[(0, 05; 2; 219)] = 3.03709$$

Dado que el $F_{obs} > F_{cri}$ se rechaza la hipótesis nula, es decir, existe un cambio de nivel y tendencia en los datos luego de la implementación de la ley.

Ahora debemos replicar el mismo análisis para el cambio de tendencia que está marcado con una línea azul en la figura 4.5.

Variable	Resultado
SCR _R	60082.12
SCR ₁	37271.22
SCR ₂	21931.07
n	223
k	2

Cuadro 4.4: Tabla de resultados para calcular test de Chow en hospitalizaciones de la RM, segunda intervención

La SCR_{NR} está representada por $SCR_1 + SCR_2$ y reemplazando los valores obtenidos en la formula anterior nos queda:

$$F = \frac{(60082.12 - (37271.22 + 21931.07))/2}{(37271.22 + 21931.07)/(223 - 2 * 2)} \sim F_{[2,219]}$$

$$F_{obs} = 9.503$$

$$F_{cri} = F[(0, 05; 2; 219)] = 3.03709$$

Dado que el $F_{obs} > F_{cri}$ se rechaza la hipótesis nula, es decir, existe un cambio de nivel y tendencia desde el mes de junio del año 2016. Este aumento puede deberse a diversos factores, uno de ellos puede ser una disminución en la propaganda de esta ley.

Capítulo 5

Modelos de ITS

5.1. Modelos de ITS de muertes causadas por accidentes de tránsito

5.1.1. Base de datos y análisis descriptivos

Antes de la formulación de los modelos, es necesario realizar un pequeño análisis descriptivo de las variables, para obtener una mayor precisión en los resultados

La base de datos consta de 5 variables que se detallan a continuación:

- **Muerte:** Variable de tipo conteo, se tienen las frecuencias de muertes causadas por accidentes de tránsito desde el año 2000 al 2016.
- **Tiempo:** Variable discreta que es utilizada para representar la tendencia de la serie.
- **Tiempo²:** Variable discreta que representa la tendencia cuadrática de la serie.
- **Intervención:** Variable dicotómica que sirve para representar la intervención de la ley. El periodo pre-intervención = 0, mientras que el periodo pos-intervención = Tendencia de la serie.
- **Intervención2:** Variable dicotómica que representa el nuevo cambio de tendencia ocurrido en mayo del año 2016. El periodo pre-intervención = 0, mientras que el periodo pos-intervención = Tendencia de la serie. .

A continuación, se presenta un pequeño análisis descriptivo de las variables en estudio:

Variable	Minímo	Máximo	Media	Varianza
Muerte	20	77	41.93	71.9309
Tiempo	1	937	469	73242.17
Tiempo ²	1	877969	293125	68728618146
Intervención	0	937	155.5	108860
Intervención2	0	937	98.28	77553.46

Cuadro 5.1: Estadística descriptiva

5.1.2. Formulación modelos ITS

A continuación se presenta una tabla resumen de 7 modelos modelos posibles, donde hasta el Modelo 5 se fueron agregando 1 a 1 las variables en estudio, y en los últimos 2 se probarón interacciones que pudiesen capturar el cambio que existe en los datos, para luego poder cuantificar el impacto de la ley. Dado que la variable respuesta es de tipo conteo, se prueban modelos poisson y binomial negativa.

Nombre	Regresión	Modelo
Modelo 1	Poisson	(Muerte \sim Tiempo+ Tiempo ²)
Modelo 2	Binomial Negativa	(Muerte \sim Tiempo+ Tiempo ²)
Modelo 3	Binomial Negativa	(Muerte \sim Tiempo + Tiempo ² + harmonic(Tiempo,1,52))
Modelo 4	Binomial Negativa	(Muerte \sim Tiempo+ Tiempo ² + Intervención + harmonic (Tiempo,1,52))
Modelo 5	Binomial Negativa	(Muerte \sim Tiempo + Tiempo ² + Intervención +Intervención2 + harmonic (Tiempo,1,52))
Modelo 6	Binomial Negativa	(Muerte \sim Tiempo * Intervención + Tiempo ² + Intervención2+ harmonic (Tiempo,1,52))
Modelo 7	Binomial Negativa	(Muerte \sim Tiempo + Tiempo ² + Intervención2 + Intervención * harmonic (Tiempo,1,52))

Cuadro 5.2: Resumen modelos propuestos

■ **Modelo 1:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.6886	0.0153	241.11	0.0000
Tiempo	0.0005	0.0001	6.74	0.0000
Tiempo2	-0.0000	0.0000	-8.37	0.0000

Cuadro 5.3: Resumen Modelo 1 muertes

En este modelo tenemos que la Deviance(D)=1500.34 y los grados de libertad(gl)=934, para evaluar si existe o no sobredispersión en los datos se calcula $D/gl=1.60$, dado que este valor es muy superior a 1 indica que existe sobredispersión en los datos.

Para solucionar esta problemática se probará un modelo binomial negativo, que también tiene la variable respuesta de tipo conteo y ajusta la sobredispersión.

■ **Modelo 2:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.6887	0.0193	191.18	0.0000
Tiempo	0.0005	0.0001	5.33	0.0000
Tiempo2	-0.0000	0.0000	-6.64	0.0000

Cuadro 5.4: Resumen Modelo 2 muertes

Al evaluar el ajuste del modelo podemos observar que $D/gl=1.00$, lo que indica que existe un buen ajuste. Por esta razón los siguientes modelos a probar se realizarán con la distribución binomial negativa.

Dado que la serie presenta estacionalidad es importante ajustarla a través de la función *harmonic* que básicamente especifica el número de pares de seno y coseno a incluir junto con el periodo, en este caso es 52 por el número de semanas durante un año.

■ **Modelo 3:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.6844	0.0188	195.88	0.0000
Tiempo	0.0005	0.0001	5.55	0.0000
Tiempo2	-0.0000	0.0000	-6.81	0.0000
harmonic(Tiempo, 1, 52)1	0.0617	0.0088	7.00	0.0000
harmonic(Tiempo, 1, 52)2	-0.0040	0.0088	-0.45	0.6503

Cuadro 5.5: Resumen Modelo 3 muertes

En el siguiente modelo se incluye la intervención, es decir, este modelo considera un antes y después de la aplicación de la ley, el resto de las variables se mantienen con respecto al modelo anterior.

■ **Modelo4:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.6768	0.0202	181.64	0.0000
Tiempo	0.0006	0.0001	5.00	0.0000
Tiempo2	-0.0000	0.0000	-5.22	0.0000
Intervención	0.0000	0.0000	1.03	0.3024
harmonic(Tiempo, 1, 52)1	0.0618	0.0088	7.02	0.0000
harmonic(Tiempo, 1, 52)2	-0.0045	0.0088	-0.51	0.6087

Cuadro 5.6: Resumen Modelo 4 muertes

De la tabla 5.6 podemos apreciar que la intervención no es significativa para el modelo, es decir, la implementación de la ley Emilia no estaría ayudando a explicar la variabilidad.

El siguiente modelo propuesto contiene una segunda intervención que representa el segundo cambio de tendencia apreciado en la figura 3.3 con la línea azul.

■ **Modelo 5:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.6726	0.0206	178.14	0.0000
Tiempo	0.0006	0.0001	5.10	0.0000
Tiempo2	-0.0000	0.0000	-5.30	0.0000
Tnervención	0.0000	0.0000	0.57	0.5702
Intervención2	0.0000	0.0000	1.08	0.2813
harmonic(Tiempo, 1, 52)1	0.0615	0.0088	6.98	0.0000
harmonic(Tiempo, 1, 52)2	-0.0042	0.0088	-0.48	0.6325

Cuadro 5.7: Resumen Modelo 5 muertes

De la tabla 5.7 podemos apreciar que la segunda intervención ingresada al modelo que representa el cambio de tendencia del año 2016 donde existe un aumento en las muertes tampoco es significativa para el modelo .

■ **Modelo 6:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.6785	0.0208	177.18	0.0000
Tiempo	0.0006	0.0001	4.62	0.0000
Intervención	0.0011	0.0005	2.08	0.0374
Tiempo2	-0.0000	0.0000	-4.78	0.0000
Intervención2	0.0002	0.0001	2.29	0.0219
harmonic(Tiempo, 1, 52)1	0.0590	0.0089	6.65	0.0000
harmonic(Tiempo, 1, 52)2	-0.0040	0.0088	-0.46	0.6452
Tiempo:Intervención	-0.0000	0.0000	-2.04	0.0412

Cuadro 5.8: Resumen Modelo 6 muertes

En la tabla 5.8 podemos apreciar que el modelo propuesto, tiene una interacción entre la intervención y el tiempo que es significativa, además esta interacción aumenta la significancia del resto de las variables.

■ **Modelo 7:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.6711	0.0206	177.80	0.0000
Tiempo	0.0006	0.0001	5.17	0.0000
Tiempo2	-0.0000	0.0000	-5.37	0.0000
Intervención	0.0000	0.0000	0.56	0.5778
harmonic(Tiempo, 1, 52)1	0.0661	0.0097	6.83	0.0000
harmonic(Tiempo, 1, 52)2	-0.0039	0.0097	-0.41	0.6833
Intervención2	0.0000	0.0000	1.16	0.2470
Intervención:harmonic(Tiempo, 1, 52)1	-0.0000	0.0000	-1.16	0.2468
Intervención:harmonic(Tiempo, 1, 52)2	-0.0000	0.0000	-0.11	0.9086

Cuadro 5.9: Resumen Modelo 7 muertes

En la tabla 5.9 podemos apreciar que el modelo propuesto, tiene una interacción entre la intervención y el armónico que no es significativa.

5.2. Análisis de Supuestos del Modelo

Es importante evaluar el cumplimiento de algunos supuestos básicos de los modelos y en base a eso poder elegir el más adecuado. Previo a eso se evalúa el AIC y la suma cuadrática de los residuos.

Modelo	AIC	SCR
Modelo 2	6592.838	938.69
Modelo 3	6549.027	939.05
Modelo 4	6549.961	939.13
Modelo 5	6550.798	939.05
Modelo 6	6548.643	938.64
Modelo 7	6553.453	939.02

Cuadro 5.10: Tabla resumen AIC y SCR

El mejor modelo es el que tiene el menor AIC, junto con la menor suma cuadrática de los residuos. La decisión que se tome frente al mejor modelo depende de otras variables además de las que se presentan en la tabla anterior, es por esto que a continuación se realiza un análisis de supuestos.

5.2.1. Supuestos

5.2.1.1. Test de Box-Ljung

Este test tiene como función probar la aleatoriedad e independencia en forma conjunta.

H_0 : Los errores están no correlacionados *v/s* H_1 : Los errores están correlacionados

Modelo	P-valor	Resultado
Modelo 2	0.0002	Se rechaza
Modelo 3	0.122	No se rechaza
Modelo 4	0.150	No se rechaza
Modelo 5	0.182	No se rechaza
Modelo 6	0.381	No se rechaza
Modelo 7	0.144	No se rechaza

Cuadro 5.11: Tabla resumen Test de Box-Ljung

En la tabla 5.11 se hizo una tabla resumen del resultado del test de Box-Ljung para todos los modelos planteados anteriormente con un lag=50, podemos apreciar que a excepción del primer modelo, los residuos no están correlacionados, es decir, que no existe una dependencia temporal.

5.2.1.2. Test de Breusch-Pagan

Esta prueba se utiliza para probar la homocedasticidad de los residuos.

H_0 : Residuos Homocedásticos *v/s* H_1 : Residuos Heterocedásticos

Modelo	P-valor	Resultado
Modelo 2	0.511	No se rechaza
Modelo 3	0.245	No se rechaza
Modelo 4	0.325	No se rechaza
Modelo 5	0.454	No se rechaza
Modelo 6	0.161	No se rechaza
Modelo 7	0.144	No se rechaza

Cuadro 5.12: Tabla resumen Test de Breusch-Pagan

En la tabla 5.12 en los 6 modelos propuestos no se rechaza la hipótesis nula, es decir, los residuos son homocedásticos, por lo tanto, se cumple el supuesto de homocedasticidad

5.2.1.3. Tabla resumen supuestos de los modelos

A continuación, se muestra una tabla resumen de todos los supuestos del modelo, y en base a esto elegir el mejor para evaluar finalmente el impacto de la ley:

Modelo	AIC	SCR	Box-Ljung	Breusch-Pagan
Modelo 2	6592.838	938.69	Se rechaza	No se rechaza
Modelo 3	6549.027	939.05	No se rechaza	No se rechaza
Modelo 4	6549.961	939.13	No se rechaza	No se rechaza
Modelo 5	6550.798	939.05	No se rechaza	No se rechaza
Modelo 6	6548.643	938.64	No se rechaza	No se rechaza
Modelo 7	6553.453	939.02	No se rechaza	No se rechaza

Cuadro 5.13: Tabla resumen supuestos de los modelos

De la tabla 5.13, el modelo elegido dada la significancia de las variables, AIC, SCR, el test de box-Ljung y Breusch-Pagan es el Modelo 6. Este modelo se ocupará para evaluar el impacto que tuvo la ley Emilia en las muertes de todo el país causadas por accidentes de tránsito.

5.2.1.4. Serie de tiempo muertes y su tendencia estimada

En la siguiente figura se muestra la serie de muertes con un “zoom” desde el año 2013 al 2017, para visualizar de mejor manera la serie y su tendencia estimada durante la implementación de la ley. Se observa que la línea celeste más cercana al eje y representa la intervención y que luego de esto la tendencia estimada del Modelo 6 tiene un decaimiento hasta la segunda intervención que representa un cambio de tendencia registrado en los datos. En el año 2016 existe un aumento en la tendencia estimada de la serie, que luego llega a su mínimo al finalizar el año 2017, tal como se ve en la serie que presenta una disminución significativa.

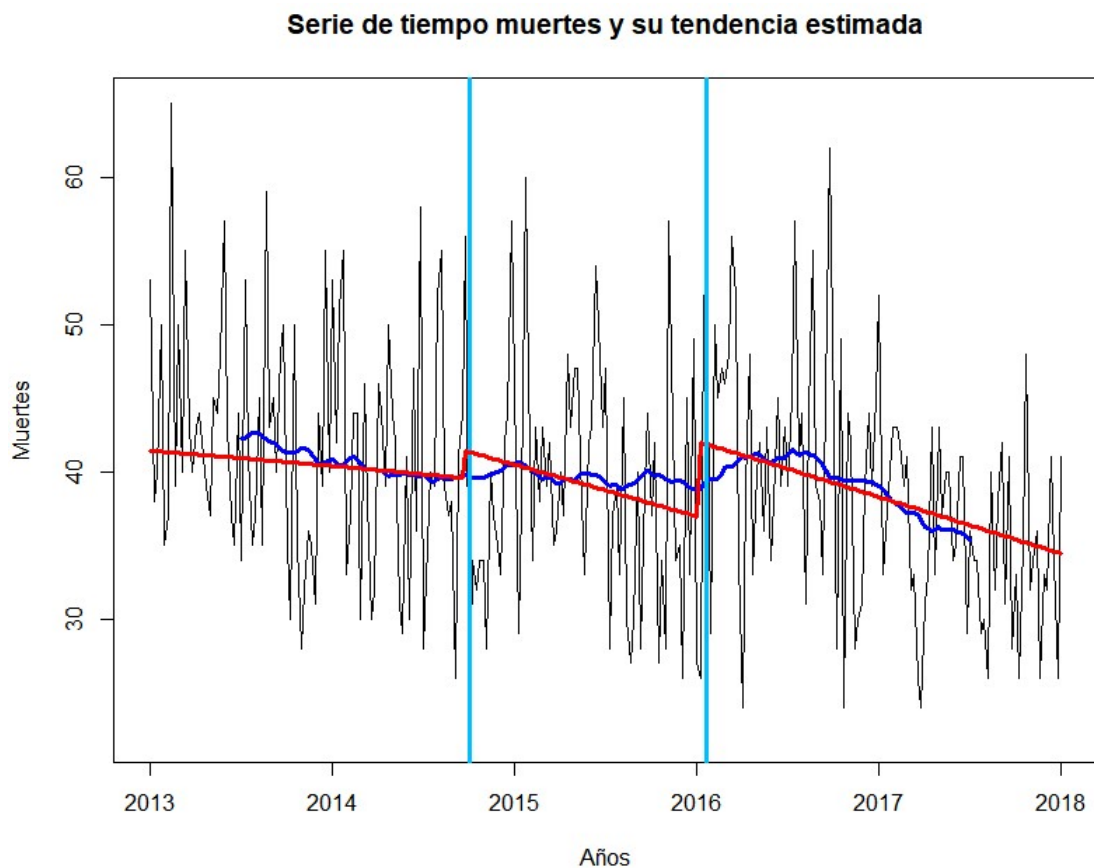


Figura 5.1: Serie de tiempo muertes y tendencias estimada

5.2.2. Evaluación del impacto de la ley Emilia

Con el modelo seleccionado en el punto anterior, se calcularon las estimaciones del modelo sin intervención ($\text{muerte} \sim \text{tiempo} + \text{tiempo}^2 + \text{harmonic}$) y las estimaciones del modelo incluyendo la intervención ($\text{muerte} \sim \text{tiempo} * \text{intervención} + \text{tiempo}^2 + \text{intervención}^2 + \text{harmonic}$). Estos datos fueron graficados para ver cuantas muertes predecía el modelo sin intervención y cuantas muertes hubo realmente y evaluar así el impacto de la ley.

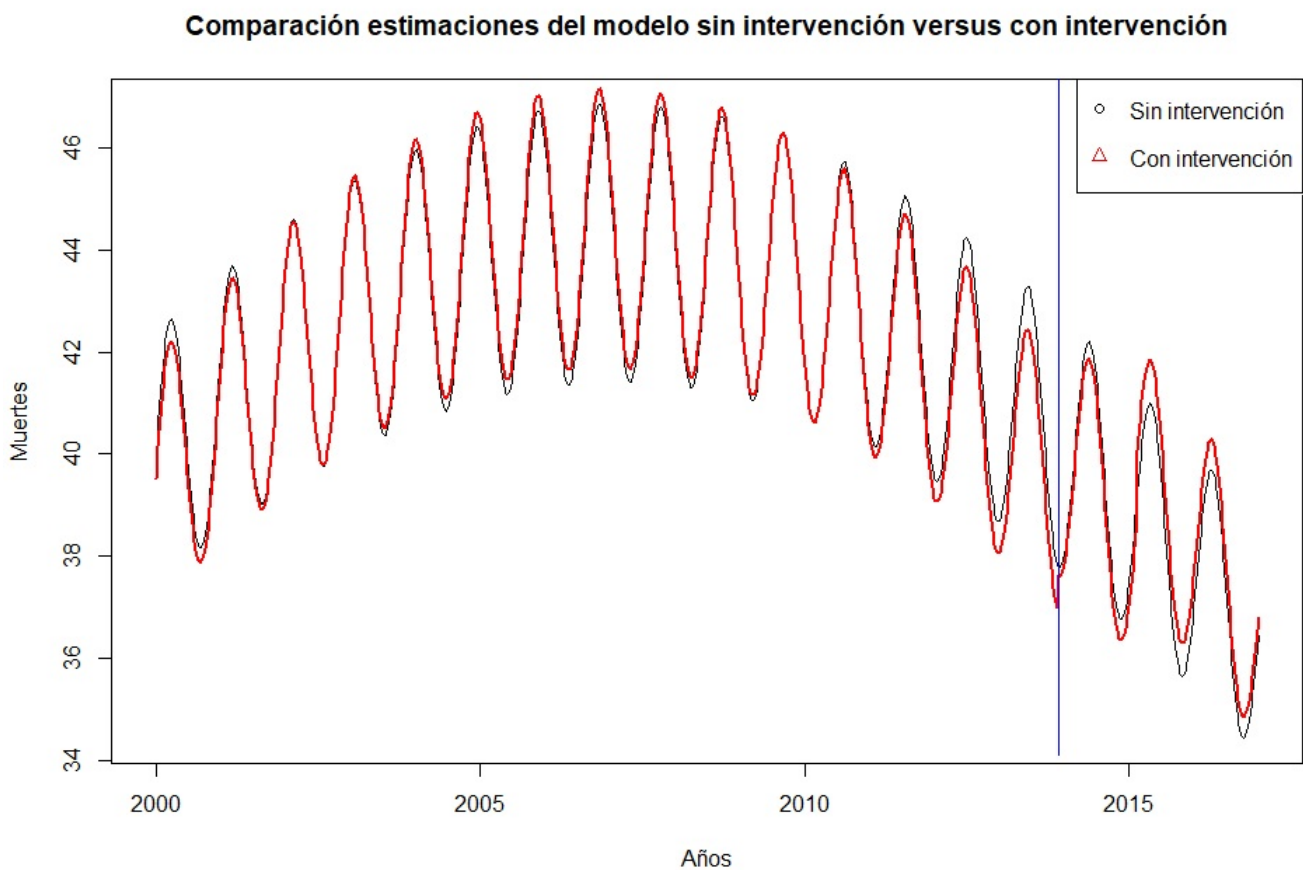


Figura 5.2: Comparación estimaciones del mejor modelo sin intervención versus mejor modelo con intervención

De la figura 5.2 tenemos que alrededor del año 2011, el modelo sin intervención estima mayor cantidad de muertes previo a la implementación de la ley y luego de la implementación marcado con la línea azul, hubo una disminución en la cantidad de muertes, que no fue capturada por el modelo con la intervención. Analizando esta gráfica y el estimador de la intervención 5.8 que tiene un signo positivo, podemos inclinarnos a que al menos en la muertes causadas por accidentes de tránsito a lo largo de todo el país la ley Emilia no tuvo un impacto significativo.

	Moodelo sin intervención	Modelo con intervención	Valor Real
Total	2103.2305	2086.48153	2089

Cuadro 5.14: Tabla resumen total de muertes para los modelos

De la tabla 5.14 tenemos que el total de muertes estimadas para el modelo sin intervención fue de 2103.2305, mientras que para el modelo con intervención fue de 2086.48153, donde el porcentaje de variación es de un 0.79 %, es decir, luego de la implementación de la ley el modelo estimó una disminución en la cantidad de muertes del 0.79 %. Es lógico que la cantidad de muertes estimadas del modelo con intervención sea muy cercano al valor real.

5.3. Modelos de ITS de muertes causadas por accidentes de tránsito en la Región Metropolitana

Del total de muertes analizadas a lo largo de todo el país desde 2007 hasta el año 2017, el 30 % corresponde a la Región Metropolitana, esto debido a que es la región más urbana del país, es por esto que se realiza el análisis de los modelos.

5.3.1. Base de datos y análisis descriptivos

A continuación, se presentan las variables que son parte de los 6 modelos propuestos y un análisis descriptivo.

La base de datos consta de 4 variables que se detallan a continuación:

- **Muerte:** Variable de tipo conteo, se tienen las frecuencias de muertes causadas por accidentes de tránsito desde 2007 a 2017 en la Región Metropolitana.
- **Tiempo:** Variable discreta que es utilizada para representar la tendencia de la serie.
- **Intervención:** Variable dicotómica que sirve para representar la intervención de la ley. El periodo pre-intervención = 0, mientras que el periodo pos-intervención = Tendencia de la serie.
- **Intervención2:** Variable dicotómica que sirve para representar el cambio de tendencia del año 2016, que se puede apreciar en la figura 3.5 representado por la línea azul. El periodo pre-intervención = 0, mientras que el periodo pos-intervención = Tendencia de la serie.

A continuación, se presenta un pequeño análisis descriptivo de las variables en estudio:

Variable	Minímo	Máximo	Media	Varianza
Muerte	1	27	12.71	16.005
Tiempo	1	572	286.5	27313
Intervención	0	572	107.5	43871.13
Intervención2	0	572	31.13	16389.61

Cuadro 5.15: Estadística descriptiva

5.3.2. Formulación modelos ITS

A continuación se presentan 7 modelos, dependiendo del cumplimiento de los supuestos se elegirá el mejor para luego evaluar la implementación de la ley.

Nombre	Regresión	Modelo
Modelo 1	Poisson	(Muerte \sim Tiempo)
Modelo 2	Binomial Negativa	(Muerte \sim Tiempo)
Modelo 3	Binomial Negativa	(Muerte \sim Tiempo + harmonic(Tiempo,1,52))
Modelo 4	Binomial Negativa	(Muerte \sim Tiempo+ Intervención + harmonic (Tiempo,1,52))
Modelo 5	Binomial Negativa	(Muerte \sim Tiempo + Intervención +Intervención2 + harmonic (Tiempo,1,52))
Modelo 6	Binomial Negativa	(Muerte \sim Tiempo * Intervención + harmonic (Tiempo,1,52))
Modelo 7	Binomial Negativa	(Muerte \sim Tiempo + Intervención * harmonic (Tiempo,1,52))

Cuadro 5.16: Estadística descriptiva

■ **Modelo 1:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.6618	0.0228	116.74	0.0000
Tiempo	-0.0004	0.0001	-5.97	0.0000

Cuadro 5.17: Resumen Modelo 1 muertes RM

Al igual que en el análisis de muertes a lo largo de todo el país, es necesario evaluar si existe sobredispersión en los datos. Tenemos que $D=693.95$ y que los $gl=570$, por lo tanto, $D/gl=1.217$. Dado esto usaremos un modelo binomial negativo para solucionar la sobredispersión.

■ **Modelo 2:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.6622	0.0250	106.29	0.0000
Tiempo	-0.0004	0.0001	-5.48	0.0000

Cuadro 5.18: Resumen modelo glm.nb(muerte~tiempo)

Al evaluar el ajuste del modelo podemos observar que $D/gl=1.02$, por lo que ya no existe sobredispersión. Los próximos modelos propuestos se realizarán con la misma distribución.

■ **Modelo 3:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.6589	0.0249	106.60	0.0000
Tiempo	-0.0004	0.0001	-5.38	0.0000
harmonic(Tiempo, 1, 52)1	0.0322	0.0181	1.78	0.0750
harmonic(Tiempo, 1, 52)2	0.0360	0.0180	2.00	0.0459

Cuadro 5.19: Resumen Modelo 3 muertes RM

En la figura 5.19 podemos apreciar que el la función harmonic es significativa para el modelo. Ahora se presenta el cuarto modelo que incluye la intervención al estudio:

■ **Modelo 4:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.6313	0.0278	94.58	0.0000
Tiempo	-0.0002	0.0001	-2.29	0.0218
Intervención	-0.0002	0.0001	-2.25	0.0246
harmonic(Tiempo, 1, 52)1	0.0311	0.0180	1.73	0.0836
harmonic(Tiempo, 1, 52)2	0.0384	0.0180	2.14	0.0326

Cuadro 5.20: Resumen Modelo 4 muertes RM

De la tabla 5.20 tenemos que la intervención de la ley en este modelo es significativa, es decir, la implementación de la ley estaría explicando la variabilidad del modelo.

El siguiente modelo propuesto contiene una la segunda intervención, que más que una intervención representa el cambio de tendencia de la serie.

■ **Modelo 5:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.6303	0.0278	94.52	0.0000
Tiempo	-0.0002	0.0001	-2.26	0.0238
Intervención	-0.0002	0.0001	-1.71	0.0876
Intervención2	-0.0001	0.0001	-1.00	0.3196
harmonic(Tiempo, 1, 52)1	0.0287	0.0181	1.58	0.1141
harmonic(Tiempo, 1, 52)2	0.0371	0.0180	2.06	0.0395

Cuadro 5.21: Resumen Modelo 5 muertes RM

De la tabla 5.21, se observa la segunda intervención no es significativa para el modelo y disminuye levemente la significancia del resto de las variables.

■ **Modelo 6:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.6296	0.0279	94.38	0.0000
Tiempo	-0.0002	0.0001	-2.24	0.0251
Intervención	0.0007	0.0009	0.80	0.4248
harmonic(Tiempo, 1, 52)1	0.0303	0.0180	1.69	0.0918
harmonic(Tiempo, 1, 52)2	0.0375	0.0180	2.09	0.0369
Tiempo:Intervención	-0.0000	0.0000	-1.03	0.3018

Cuadro 5.22: Resumen Modelo 6 muertes RM

En la tabla 5.22 podemos apreciar que el modelo propuesto, tiene una interacción entre la intervención y el tiempo que no es significativa, por lo que este último no es el mejor modelo a considerar dado el resto de las posibilidades que han sido presentadas, es importante mencionar también que al agregar esta interacción al modelo ninguna de las 2 intervenciones planteadas estarían explicando el modelo.

■ **Modelo 7:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.6310	0.0278	94.58	0.0000
Tiempo	-0.0002	0.0001	-2.29	0.0220
Intervención	-0.0002	0.0001	-2.28	0.0225
harmonic(Tiempo, 1, 52)1	0.0364	0.0199	1.82	0.0683
harmonic(Tiempo, 1, 52)2	0.0383	0.0199	1.93	0.0535
Intervención:harmonic(Tiempo, 1, 52)1	-0.0001	0.0001	-0.61	0.5427
Intervención:harmonic(Tiempo, 1, 52)2	-0.0000	0.0001	-0.02	0.9805

Cuadro 5.23: Resumen Modelo 7 muertes RM

En la tabla 5.23 podemos apreciar que el modelo propuesto, tiene una interacción entre la intervención y el armónico que no es significativa.

5.4. Análisis de Supuestos del Modelo

Uno de los indicadores importantes para la elección del mejor modelo es el AIC y la SCR, es por esto que para los 6 modelos propuestos analizaremos estos indicadores.

Modelo	AIC	SCR
Modelo 2	3173.186	583.21
Modelo 3	3170.042	583.49
Modelo 4	3167.003	583.21
Modelo 5	3168.005	583.02
Modelo 6	3167.941	583.03
Modelo 7	3170.632	583.14

Cuadro 5.24: Tabla resumen AIC y SCR

Dados los resultados de la tabla 5.24, el Modelo 4 es el que tiene menor AIC y el Modelo 5 menor SCR.

5.4.1. Supuestos

5.4.1.1. Test de Box-Ljung

Este test tiene como función probar la aleatoriedad e independencia en forma conjunta.

H_0 : Los errores están no correlacionados *v/s* H_1 : Los errores están correlacionados

Modelo	P-valor	Resultado
Modelo 2	0.161	No se rechaza
Modelo 3	0.274	No se rechaza
Modelo 4	0.324	No se rechaza
Modelo 5	0.323	No se rechaza
Modelo 6	0.307	No se rechaza
Modelo 7	0.335	No se rechaza

Cuadro 5.25: Tabla resumen Test de Box-Ljung

En la tabla 5.25 se hizo una tabla resumen del resultado del test de Box-Ljung donde para todos los modelos planteados, los errores no están correlacionados, es decir, no existe una dependencia temporal.

5.4.1.2. Test de Breusch-Pagan

Este test se utiliza para probar la homocedasticidad de los residuos.

H_0 : Residuos Homocedásticos *v/s* H_1 : Residuos Heterocedásticos

Modelo	P-valor	Resultado
Modelo 2	0.297	No se rechaza
Modelo 3	0.0143	No se rechaza al 1 %
Modelo 4	0.017	No se rechaza al 1 %
Modelo 5	0.075	No se rechaza
Modelo 6	0.119	No se rechaza
Modelo 7	0.138	No se rechaza

Cuadro 5.26: Tabla resumen Test de Breusch-Pagan

En la tabla 5.26, en el global mayoritariamente no se rechaza la hipótesis nula, es decir, para esos modelos los residuos son homocedásticos.

5.4.1.3. Tabla resumen supuestos de los modelos

A continuación se muestra una tabla resumen de todos los supuestos del modelo, y en base a esto elegir el mejor para evaluar finalmente el impacto de la ley:

Modelo	AIC	SCR	Box-Ljung	Breusch-Pagan
Modelo 2	3173.186	583.21	No se rechaza	No se rechaza
Modelo 3	3170.042	583.49	No se rechaza	No se rechaza al 1 %
Modelo 4	3167.003	583.21	No se rechaza	No se rechaza al 1 %
Modelo 5	3168.005	583.02	No se rechaza	No se rechaza
Modelo 6	3167.941	583.03	No se rechaza	No se rechaza
Modelo 7	3170.632	583.14	No se rechaza	No se rechaza

Cuadro 5.27: Tabla resumen supuestos de los modelos

Dada la significancia de las variables en estudio, y de los demás criterios de evaluación, el modelo seleccionado para evaluar el impacto de la ley es el Modelo 4.

5.4.1.4. Serie de tiempo muertes en la RM su tendencia estimada

En la siguiente figura se tiene un “zoom” de la serie de tiempo de muertes en la RM, donde la línea azul representa la tendencia y la roja representa la tendencia estimada del Modelo 4, luego de la intervención que viene dada por la línea celeste existe una pequeña disminución en la tendencia estimada que se mantiene constante hasta el final de la serie.

Serie de tiempo muertes en la RM y su tendencia estimada

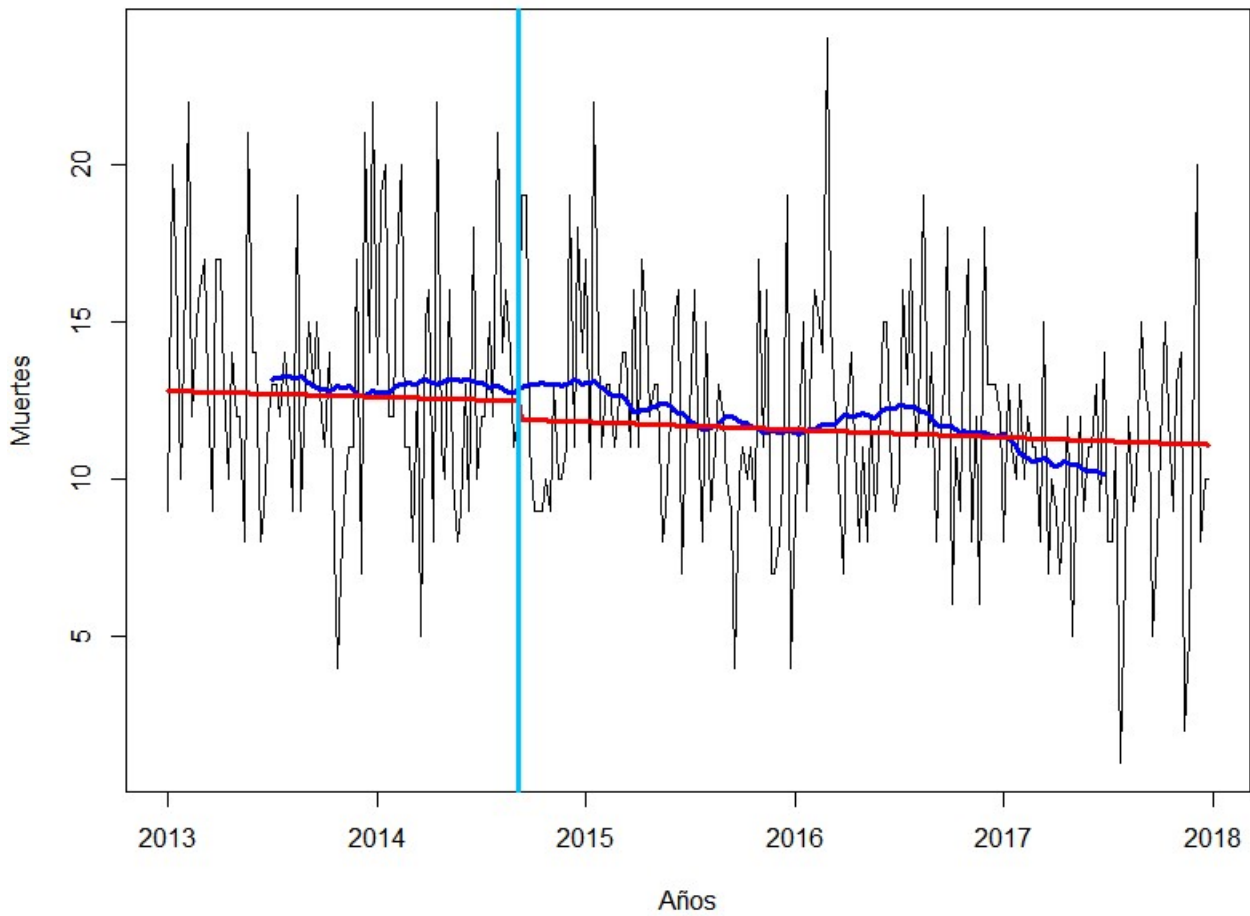


Figura 5.3: Serie de tiempo muertes en la RM y su tendencia

5.4.2. Evaluación del impacto de la ley Emilia

Dado el mejor modelo del punto anterior, se analizan las estimaciones de muertes en la Región Metropolitana, sin la intervención de la ley versus el modelo con la intervención. A continuación, se presenta la gráfica de las estimaciones obtenidas.

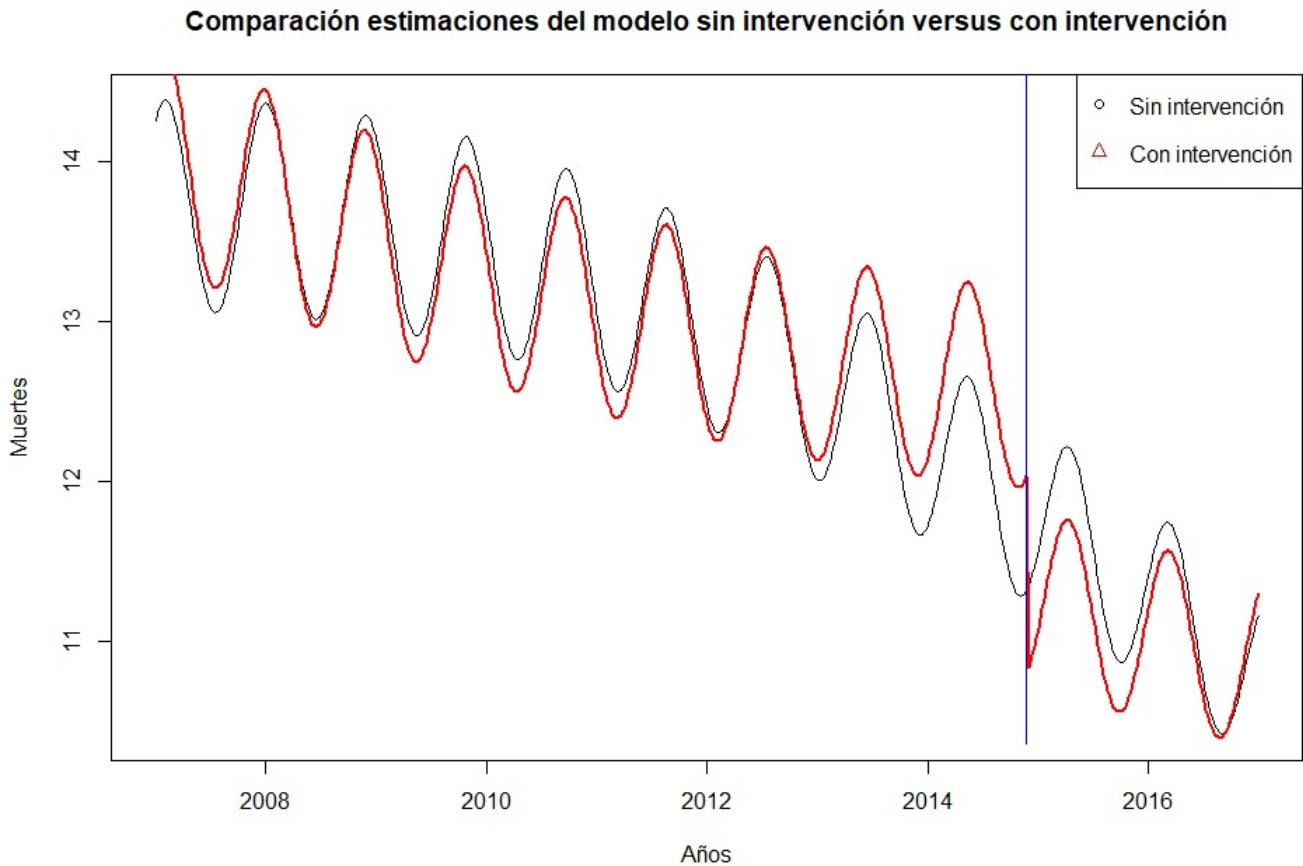


Figura 5.4: Comparación estimaciones del mejor modelo sin intervención versus mejor modelo con intervención

De la figura 6.3 tenemos que alrededor del año 2013, el modelo sin intervención estima menor cantidad de muertes previo a la implementación de la ley y luego de la implementación marcado con la línea azul, las estimaciones del modelo con intervención son notoriamente menores. Del gráfico 3.5 tenemos que luego de la implementación de la ley la tendencia disminuyó, que es lo mismo que se puede observar en esta figura.

	Modelo sin intervención	Modelo con intervención	Valor Real
Total	615.0291485	594.5215861	596

Cuadro 5.28: Tabla resumen total de muertes para los modelos

De la tabla 5.28 tenemos que el total de muertes estimadas para el modelo sin intervención fue de 615.0291, mientras que para el modelo con intervención fue de 594.5215, obteniendo un porcentaje de variación del 3.3 %, es decir, luego de la implementación de la ley Emilia las muertes causadas por accidentes de tránsito en la Región Metropolitana disminuyeron un 3.3 %.

5.5. Modelos de ITS de muertes en peatones y ocupante de vehículos causadas por accidentes de tránsito

5.5.1. Modelos de ITS de muertes de peatones causadas por accidentes de tránsito

Es importante analizar el impacto de la ley en los peatones y ver que tan sensibles pueden ser a la ley o si se han visto beneficiados por una conducción más prudente por parte de los conductores.

5.5.1.1. Base de datos y análisis descriptivos

A continuación, se presentan las variables que son parte de los 7 modelos propuestos y un análisis descriptivo.

La base de datos consta de 4 variables que se detallan a continuación:

- **Muerte:** Variable de tipo conteo, se tienen las frecuencias de muertes causadas por accidentes de tránsito desde 1997 a 2016 en la Región Metropolitana.
- **Tiempo:** Variable discreta que es utilizada para representar la tendencia de la serie.
- **Intervención:** Variable dicotómica que sirve para representar la intervención de la ley. El periodo pre-intervención = 0, mientras que el periodo pos-intervención = Tendencia de la serie.
- **Intervención2:** Variable dicotómica que sirve para representar el cambio de tendencia del año 2016, que se puede apreciar en la figura 3.5 representado por la línea azul. El periodo pre-intervención = 0, mientras que el periodo pos-intervención = Tendencia de la serie.
- **Autoregresiva:** Variable autoregresiva de orden 1 (AR(1))

A continuación, se presenta un pequeño análisis descriptivo de las variables en estudio:

Variable	Minímo	Máximo	Media	Varianza
Muerte	1	39	18.03	28.35
Tiempo	2	937	469.5	73086
Tiempo ²	4	877969	293438	6871013
Intervención	0	937	157.3	109691.5
Intervención2	0	937	22.75	16389.61
Autoregresiva	1	39	18.04	28.34

Cuadro 5.29: Estadística descriptiva

5.5.1.2. Formulación modelos ITS

A continuación se presenta una tabla con los 7 modelos propuestos para luego realizar el respectivo análisis de supuestos.

Nombre	Regresión	Modelo
Modelo 1	Poisson	(Muerte ~ Tiempo + Tiempo ²)
Modelo 2	Binomial Negativa	(Muerte ~ Tiempo+ Tiempo ²)
Modelo 3	Binomial Negativa	(Muerte ~ Tiempo + Tiempo ² + harmonic(Tiempo,1,52))
Modelo 4	Binomial Negativa	(Muerte ~ Tiempo+ Tiempo ² + Intervención + harmonic (Tiempo,1,52))
Modelo 5	Binomial Negativa	(Muerte ~ Tiempo + Tiempo ² + Intervención +Intervención2+ Autoregresiva + harmonic (Tiempo,1,52))
Modelo 6	Binomial Negativa	(Muerte ~ Tiempo * Intervención +Tiempo ² +harmonic (Tiempo,1,52))
Modelo 7	Binomial Negativa	(Muerte ~ Tiempo + Tiempo ² + Intervención * harmonic (Tiempo,1,52))

Cuadro 5.30: Tabla resumen modelos propuestos

■ **Modelo 1:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.9162	0.0228	127.74	0.0000
Tiempo	0.0005	0.0001	4.77	0.0000
Tiempo2	-0.0000	0.0000	-8.11	0.0000

Cuadro 5.31: Resumen Modelo 1 muertes peatonales

Al igual que en el análisis de muertes a lo largo de todo el país, es necesario evaluar si existe sobre-dispersión en los datos. Tenemos que $D=1266.49$ y que los $gl=933$, por lo tanto $D/gl=1.3574$. Dado esto

usaremos un modelo binomial negativo para solucionar la sobredispersión.

■ **Modelo 2:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.9163	0.0266	109.78	0.0000
Tiempo	0.0005	0.0001	4.11	0.0000
Tiempo2	-0.0000	0.0000	-7.04	0.0000

Cuadro 5.32: Resumen Modelo 2 muertes peatones

Al evaluar el ajuste del modelo podemos observar que $D/gl=1.01$, por lo que ya no existe sobredispersión. Los próximos modelos propuestos se realizarán con la misma distribución.

■ **Modelo 3:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.9111	0.0249	116.85	0.0000
Tiempo	0.0005	0.0001	4.37	0.0000
Tiempo2	-0.0000	0.0000	-7.44	0.0000
harmonic(Tiempo, 1, 52)1	0.0102	0.0119	0.86	0.3910
harmonic(Tiempo, 1, 52)2	-0.1332	0.0119	-11.21	0.0000

Cuadro 5.33: Resumen Modelo 3 muertes peatones

En la figura 5.33 podemos apreciar que la función harmonic es significativa para el seno de la función. Ahora se presenta el cuarto modelo que incluye la intervención al estudio:

■ **Modelo 4:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.9213	0.0268	109.09	0.0000
Tiempo	0.0004	0.0002	2.83	0.0046
Tiempo2	-0.0000	0.0000	-4.13	0.0000
Intervención	-0.0001	0.0001	-1.02	0.3085
harmonic(Tiempo, 1, 52)1	0.0098	0.0119	0.83	0.4074
harmonic(Tiempo, 1, 52)2	-0.1325	0.0119	-11.14	0.0000

Cuadro 5.34: Resumen Modelo 4 muertes peatones

De la tabla 5.34 tenemos que la intervención de la ley en este modelo no es significativa, es decir, la implementación de la ley no estaría explicando la variabilidad del modelo, pero si se observa el parámetro estimado de la intervención, tenemos que existe un cambio de signo con respecto al estimador del tiempo, que indicaría un cambio de tendencia.

El siguiente modelo propuesto contiene la segunda intervención, que más que una intervención representa el cambio de tendencia de la serie, además de la variable autoregresiva.

■ **Modelo 5:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.8686	0.0430	66.72	0.0000
Tiempo	0.0004	0.0002	2.56	0.0103
Tiempo2	-0.0000	0.0000	-3.74	0.0002
Autoregresiva	0.0029	0.0018	1.61	0.1070
Intervención2	-0.0000	0.0001	-0.32	0.7496
Intervención	-0.0001	0.0001	-0.98	0.3287
harmonic(Tiempo, 1, 52)1	0.0096	0.0120	0.80	0.4240
harmonic(Tiempo, 1, 52)2	-0.1253	0.0127	-9.91	0.0000

Cuadro 5.35: Resumen Modelo 5 muertes peatones

De la tabla 5.35, se observa que esta segunda intervención, dado el signo del estimador también representa un cambio de tendencia.

■ **Modelo 6:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.9125	0.0275	106.01	0.0000
Tiempo	0.0005	0.0002	3.15	0.0017
Intervención	-0.0007	0.0005	-1.57	0.1167
Tiempo2	-0.0000	0.0000	-4.38	0.0000
harmonic(Tiempo, 1, 52)1	0.0105	0.0119	0.88	0.3782
harmonic(Tiempo, 1, 52)2	-0.1322	0.0119	-11.12	0.0000
Tiempo:Intervención	0.0000	0.0000	1.46	0.1451

Cuadro 5.36: Resumen Modelo 6 muertes peatones

En la tabla 5.36 podemos apreciar que el modelo propuesto, tiene una interacción entre la intervención y el tiempo que no es significativa, por lo que este último no es el mejor modelo a considerar dado el resto de las posibilidades que han sido presentadas.

■ **Modelo 7:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.9205	0.0268	109.03	0.0000
Tiempo	0.0005	0.0002	2.86	0.0042
Tiempo2	-0.0000	0.0000	-4.16	0.0000
Intervención	-0.0001	0.0001	-1.02	0.3068
harmonic(Tiempo, 1, 52)1	0.0135	0.0129	1.05	0.2951
harmonic(Tiempo, 1, 52)2	-0.1334	0.0129	-10.34	0.0000
Intervención:harmonic(Tiempo, 1, 52)1	-0.0000	0.0000	-0.72	0.4694
Intervención:harmonic(Tiempo, 1, 52)2	0.0000	0.0000	0.15	0.8838

Cuadro 5.37: Resumen Modelo 7 muertes peatones

En la tabla 5.37 podemos apreciar que el modelo propuesto, tiene una interacción entre la intervención y el armónico que no es significativa.

5.5.1.3. Análisis de Supuestos del Modelo

Uno de los indicadores importantes para la elección del mejor modelo es el AIC y la SCR, por lo que para los 6 modelos propuestos analizaremos estos indicadores.

Modelo	AIC	SCR
Modelo 2	5622.29	948.82
Modelo 3	5508.47	947.35
Modelo 4	5508.43	947.43
Modelo 5	5509.75	947.00
Modelo 6	5508.52	947.07
Modelo 7	5511.88	947.45

Cuadro 5.38: Tabla resumen AIC y SCR

Dados los resultados de la tabla 5.38, el Modelo 4 es el que tiene menor AIC y el Modelo 5 menor SCR.

5.5.1.4. Supuestos

5.5.1.4.1. Test de Box-Ljung Este test tiene como función probar la aleatoriedad e independencia en forma conjunta.

H_0 : Los errores están no correlacionados *v/s* H_1 : Los errores están correlacionados

Modelo	P-valor	Resultado
Modelo 2	0.000	Se rechaza
Modelo 3	0.006	Se rechaza
Modelo 4	0.000	Se rechaza
Modelo 5	0.010	Se rechaza
Modelo 6	0.000	Se rechaza
Modelo 7	0.000	Se rechaza

Cuadro 5.39: Tabla resumen Test de Box-Ljung

En la tabla 5.39 se hizo una tabla resumen del resultado del test de Box-Ljung donde para todos los modelos planteados, los errores están correlacionados, es decir, existe una dependencia temporal.

5.5.1.4.2. Test de Breusch-Pagan Este test se utiliza para probar la homocedasticidad de los residuos.

H_0 : Residuos Homocedásticos *v/s* H_1 : Residuos Heterocedásticos

Modelo	P-valor	Resultado
Modelo 2	0.019	No se rechaza al 1 %
Modelo 3	0.071	No se rechaza
Modelo 4	0.113	No se rechaza
Modelo 5	0.021	No se rechaza al 1 %
Modelo 6	0.142	No se rechaza
Modelo 7	0.194	No se rechaza

Cuadro 5.40: Tabla resumen Test de Breusch-Pagan

En la tabla 5.40, para los modelos propuestos no se rechaza la hipótesis nula, es decir, los residuos son homocedásticos.

5.5.1.4.3. Tabla resumen supuestos de los modelos A continuación, se muestra una tabla resumen de todos los supuestos del modelo, y en base a esto elegir el mejor para evaluar finalmente el impacto de la ley:

Modelo	AIC	SCR	Box-Ljung	Breusch-Pagan
Modelo 2	5622.29	948.82	Se rechaza	No se rechaza al 1 %
Modelo 3	5507.47	947.35	Se rechaza	No se rechaza
Modelo 4	5508.43	947.43	Se rechaza	No se rechaza
Modelo 5	5509.75	947.00	Se rechaza	No se rechaza al 1 %
Modelo 6	5508.52	947.07	Se rechaza	No se rechaza
Modelo 7	5511.88	947.45	Se rechaza	No se rechaza

Cuadro 5.41: Tabla resumen supuestos de los modelos

De los modelos planteados anteriormente, tenemos que la intervención no estaría explicando el comportamiento de los datos, también tenemos que el supuesto de no correlación de los residuos no se cumple. Esto implica que las estimaciones dejan de ser eficientes y que la inferencia estadística pueda verse afectada. De igual manera analizaremos las estimaciones con el Modelo 4.

5.5.1.4.4. Serie de tiempo muertes de peatones y su tendencia estimada En la siguiente figura se tiene la serie de tiempo de muertes de peatones desde el año 2013 al 2017 donde se adhiere la tendencia estimada del modelo seleccionado en el punto anterior, se observa que luego de la implementación de la ley marcado con la línea celeste, existe una disminución en la tendencia y que mantiene este comportamiento hasta finalizar la serie.

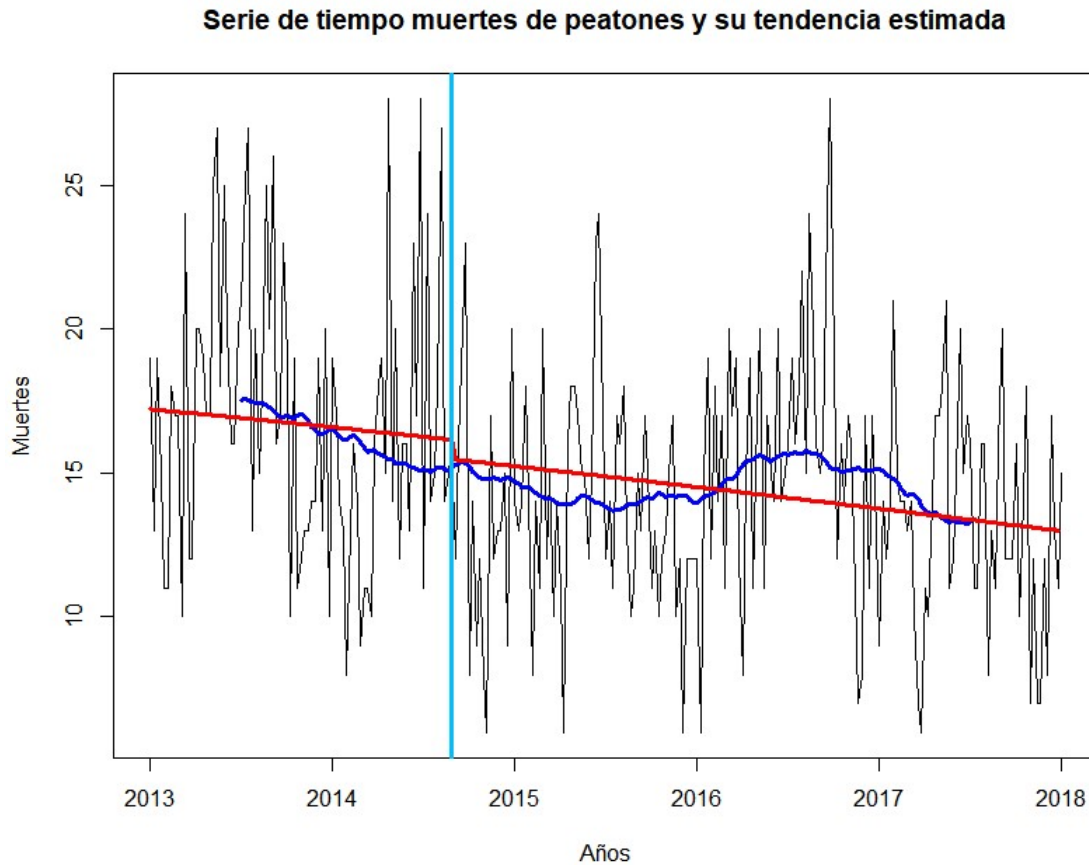


Figura 5.5: Serie de tiempo muertes de peatones y su tendencia

5.5.2. Modelos de ITS de muertes de ocupante de vehículo causadas por accidentes de tránsito

A continuación, se muestran 7 posibles modelos para evaluar el impacto de la ley que en este caso debería ser más sensible que en los peatones.

5.5.2.1. Base de datos y análisis descriptivos

A continuación, se presentan las variables que son parte de los 7 modelos propuestos y un análisis descriptivo.

La base de datos consta de 5 variables que se detallan a continuación:

- **Muerte:** Variable de tipo conteo, se tienen las frecuencias semanal de muertes de ocupantes de vehículo causadas por accidentes de tránsito desde el 2000 al 2017.
- **Tiempo:** Variable discreta que es utilizada para representar la tendencia de la serie.

- **Tiempo²**: Variable discreta que es utilizada para representar la tendencia cuadratica de la serie.
- **Intervención**: Variable dicotómica que sirve para representar la intervención de la ley. El periodo pre-intervención = 0, mientras que el periodo pos-intervención = Tendencia de la serie.
- **Intervención2**: Variable dicotómica que sirve para representar el cambio de tendencia de septiembre del 2015, que se puede apreciar en la figura 3.9 representado por la línea azul. El periodo pre-intervención = 0, mientras que el periodo pos-intervención = Tendencia de la serie.

A continuación, se presenta un pequeño análisis descriptivo de las variables en estudio:

Variable	Minímo	Máximo	Media	Varianza
Muerte	6	42	20.39	37.92727
Tiempo	1	937	470	73086
Tiempo ²	1	877969	293438	68710131667
Intervención	0	937	157.3	109691.5
Intervención2	0	937	68.4	57127.16

Cuadro 5.42: Estadística descriptiva

5.5.2.2. Formulación modelos ITS

A continuación se presenta una tabla resumen con los modelos propuestos.

Nombre	Regresión	Modelo
Modelo 1	Poisson	(Muerte ~ Tiempo + Tiempo ²)
Modelo 2	Binomial Negativa	(Muerte ~ Tiempo+ Tiempo ²)
Modelo 3	Binomial Negativa	(Muerte ~ Tiempo + Tiempo ² + harmonic(Tiempo,1,52))
Modelo 4	Binomial Negativa	(Muerte ~ Tiempo+ Tiempo ² + Intervención +Intervención2+ harmonic (Tiempo,1,52))
Modelo 5	Binomial Negativa	(Muerte ~ Tiempo + Tiempo ² + Intervención +Intervención2+ Autoregresiva + harmonic (Tiempo,1,52))
Modelo 6	Binomial Negativa	(Muerte ~ Tiempo * Intervención +Tiempo ² + Intervención2 +harmonic (Tiempo,1,52))
Modelo 7	Binomial Negativa	(Muerte ~ Tiempo + Tiempo ² + Intervención2 +Intervención * harmonic (Tiempo,1,52))

Cuadro 5.43: Tabla resumen modelos propuestos

- **Modelo 1:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.0099	0.0218	137.99	0.0000
Tiempo	0.0002	0.0001	2.03	0.0422
Tiempo2	-0.0000	0.0000	-3.00	0.0027

Cuadro 5.44: Resumen Modelo 1 muertes ocupante de vehículo

Al igual que en el análisis de muertes a lo largo de todo el país, es necesario evaluar si existe sobre-dispersión en los datos. Tenemos que $D=1699.992$ y que los $gl=933$, por lo tanto, $D/gl=1.822$. Dado esto usaremos un modelo binomial negativo para solucionar la sobredispersión.

▪ **Modelo 2:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.0094	0.0294	102.26	0.0000
Tiempo	0.0002	0.0001	1.53	0.1267
Tiempo2	-0.0000	0.0000	-2.25	0.0244

Cuadro 5.45: Resumen Modelo 2 muertes ocupante de vehículo

Al evaluar el ajuste del modelo podemos observar que $D/gl=1.00$, por lo que ya no existe sobredispersión. Los próximos modelos propuestos se realizarán con la misma distribución.

▪ **Modelo 3:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.9973	0.0277	108.10	0.0000
Tiempo	0.0002	0.0001	1.76	0.0782
Tiempo2	-0.0000	0.0000	-2.43	0.0150
harmonic(Tiempo, 1, 52)1	0.1044	0.0130	8.02	0.0000
harmonic(Tiempo, 1, 52)2	0.0948	0.0130	7.28	0.0000

Cuadro 5.46: Resumen Modelo 3 muertes ocupante de vehículo

En la figura 5.46 podemos apreciar que el armónico es significativa para el modelo, es decir, está función estaría ajustando la estacionalidad de los datos.

Ahora se presenta el cuarto modelo que incluye la intervención al estudio:

■ **Modelo 4:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.9510	0.0297	99.30	0.0000
Tiempo	0.0007	0.0002	4.00	0.0001
Tiempo2	-0.0000	0.0000	-4.80	0.0000
Intervención	0.0002	0.0001	4.22	0.0000
harmonic(Tiempo, 1, 52)1	0.1063	0.0129	8.23	0.0000
harmonic(Tiempo, 1, 52)2	0.0918	0.0129	7.11	0.0000

Cuadro 5.47: Resumen Modelo 4 muertes ocupante de vehículo

De la tabla 5.47 tenemos que la intervención de la ley en este modelo es significativa, es decir, la implementación de la ley estaría explicando la variabilidad del modelo, además el resto de las variables son significativas.

El siguiente modelo propuesto contiene una la segunda intervención, que más que una intervención representa el cambio de tendencia de la serie.

■ **Modelo 5:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.9791	0.0300	99.36	0.0000
Tiempo	0.0005	0.0002	2.54	0.0112
Tiempo2	-0.0000	0.0000	-3.19	0.0014
Intervención	0.0003	0.0001	4.81	0.0000
Intervención2	-0.0002	0.0001	-4.45	0.0000
harmonic(Tiempo, 1, 52)1	0.1024	0.0128	8.00	0.0000
harmonic(Tiempo, 1, 52)2	0.0948	0.0128	7.41	0.0000

Cuadro 5.48: Resumen Modelo 5 muertes ocupante de vehículo

De la tabla 5.48, tenemos que el estimador de la Intervención2 es negativo, es decir, existe una disminución de la tendencia en este punto.

A continuación, se muestran dos modelos con interacciones entre sus variables.

■ **Modelo 6:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.9829	0.0302	98.77	0.0000
Tiempo	0.0004	0.0002	2.32	0.0206
Intervención	0.0011	0.0008	1.33	0.1826
Tiempo2	-0.0000	0.0000	-2.94	0.0033
Intervención2	-0.0002	0.0001	-1.62	0.1055
harmonic(Tiempo, 1, 52)1	0.1028	0.0128	8.02	0.0000
harmonic(Tiempo, 1, 52)2	0.0934	0.0129	7.26	0.0000
Tiempo:Intervención	-0.0000	0.0000	-0.99	0.3244

Cuadro 5.49: Resumen Modelo 6 muertes ocupante de vehículo

En la tabla 5.49 podemos apreciar que el modelo propuesto, tiene una interacción entre la intervención y el tiempo que no es significativa, por lo que este último no es el mejor modelo a considerar dado el resto de las posibilidades que han sido presentadas, es importante mencionar también que al agregar esta interacción al modelo ninguna de las 2 intervenciones planteadas estarían explicando el modelo.

■ **Modelo 7:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.9769	0.0299	99.46	0.0000
Tiempo	0.0005	0.0002	2.60	0.0094
Tiempo2	-0.0000	0.0000	-3.25	0.0012
Intervención2	-0.0002	0.0001	-4.56	0.0000
Intervención	0.0003	0.0001	4.91	0.0000
harmonic(Tiempo, 1, 52)1	0.1139	0.0141	8.07	0.0000
harmonic(Tiempo, 1, 52)2	0.0999	0.0141	7.09	0.0000
Intervención:harmonic(Tiempo, 1, 52)1	-0.0001	0.0000	-1.94	0.0524
Intervención:harmonic(Tiempo, 1, 52)2	-0.0000	0.0000	-0.91	0.3642

Cuadro 5.50: Resumen Modelo 7 muertes ocupante de vehículo

En la tabla 5.50 podemos apreciar que el modelo propuesto, tiene una interacción entre la intervención y el armónico que no es significativa.

5.5.3. Análisis de Supuestos del Modelo

Uno de los indicadores importantes para la elección del mejor modelo es el AIC y la SCR, es por esto que para los 6 modelos propuestos analizaremos estos indicadores.

Modelo	AIC	SCR
Modelo 2	6007.564	939.6392
Modelo 3	5900.326	940.7358
Modelo 4	5884.237	940.8945
Modelo 5	5866.583	939.6254
Modelo 6	5867.613	939.741
Modelo 7	5866.96	939.802

Cuadro 5.51: Tabla resumen AIC y SCR

Dados los resultados de la tabla 5.51, el Modelo 5 es el que tiene menor AIC y SCR. A continuación, se realizará el análisis de supuestos para elegir el mejor modelo.

5.5.3.1. Supuestos

5.5.3.1.1. Test de Box-Ljung Esta prueba tiene como función probar la aleatoriedad e independencia en forma conjunta.

H_0 : Los errores están no correlacionados *v/s* H_1 : Los errores están correlacionados

Modelo	P-valor	Resultado
Modelo 2	0.000	Se rechaza
Modelo 3	0.000	Se rechaza
Modelo 4	0.000	Se rechaza
Modelo 5	0.178	No se rechaza
Modelo 6	0.230	No se rechaza
Modelo 7	0.118	No se rechaza

Cuadro 5.52: Tabla resumen Test de Box-Ljung

En la tabla 5.52 se hizo una tabla resumen del resultado del test de Box-Ljung donde, cuando se ingresa al modelo la segunda intervención, se cumple el supuesto para esos modelos.

5.5.3.1.2. Test de Breusch-Pagan Este test se utiliza para probar la homocedasticidad de los residuos.

$$H_0 : \text{Residuos Homocedásticos } v/s \ H_1 : \text{Residuos Heterocedásticos}$$

Modelo	P-valor	Resultado
Modelo 2	0.617	No se rechaza
Modelo 3	0.346	No se rechaza
Modelo 4	0.488	No se rechaza
Modelo 5	0.243	No se rechaza
Modelo 6	0.337	No se rechaza
Modelo 7	0.382	No se rechaza

Cuadro 5.53: Tabla resumen Test de Breusch-Pagan

En la tabla 5.53, se observa que para los 6 modelos propuestos se cumple el supuesto de normalidad de los residuos.

5.5.3.1.3. Tabla resumen supuestos de los modelos A continuación, se muestra una tabla resumen de todos los supuestos del modelo, y en base a esto elegir el mejor para evaluar finalmente el impacto de la ley:

Modelo	AIC	SCR	Box-Ljung	Breusch-Pagan
Modelo 2	6007.564	939.6392	Se rechaza	No se rechaza
Modelo 3	5900.326	940.7358	Se rechaza	No se rechaza
Modelo 4	5884.237	940.8945	Se rechaza	No se rechaza
Modelo 5	5866.583	939.6254	No se rechaza	No se rechaza
Modelo 6	5867.613	939.741	No se rechaza	No se rechaza
Modelo 7	5866.96	939.802	No se rechaza	No se rechaza

Cuadro 5.54: Tabla resumen supuestos de los modelos

Dado los modelos planteados anteriormente tenemos que el Modelo 5, es el modelo seleccionado para evaluar el impacto de la ley. El modelo tiene la mayor significancia de variables, junto con el cumplimiento de los supuestos.

5.5.3.1.4. Serie de tiempo muertes ocupante de vehículo y su tendencia estimada En la siguiente figura se observa la serie de tiempo, que tiene representado por la línea azul la tendencia y por la línea roja la tendencia estimada del modelo 5, donde la implementación de la ley está marcado por la línea celeste que se observa al finalizar el año 2014, luego de la implementación de la ley existe un aumento significativo en la tendencia hasta fines del año 2016, en este punto se visualiza otro cambio de tendencia donde existe una disminución que se mantiene constante al terminar el año 2017.

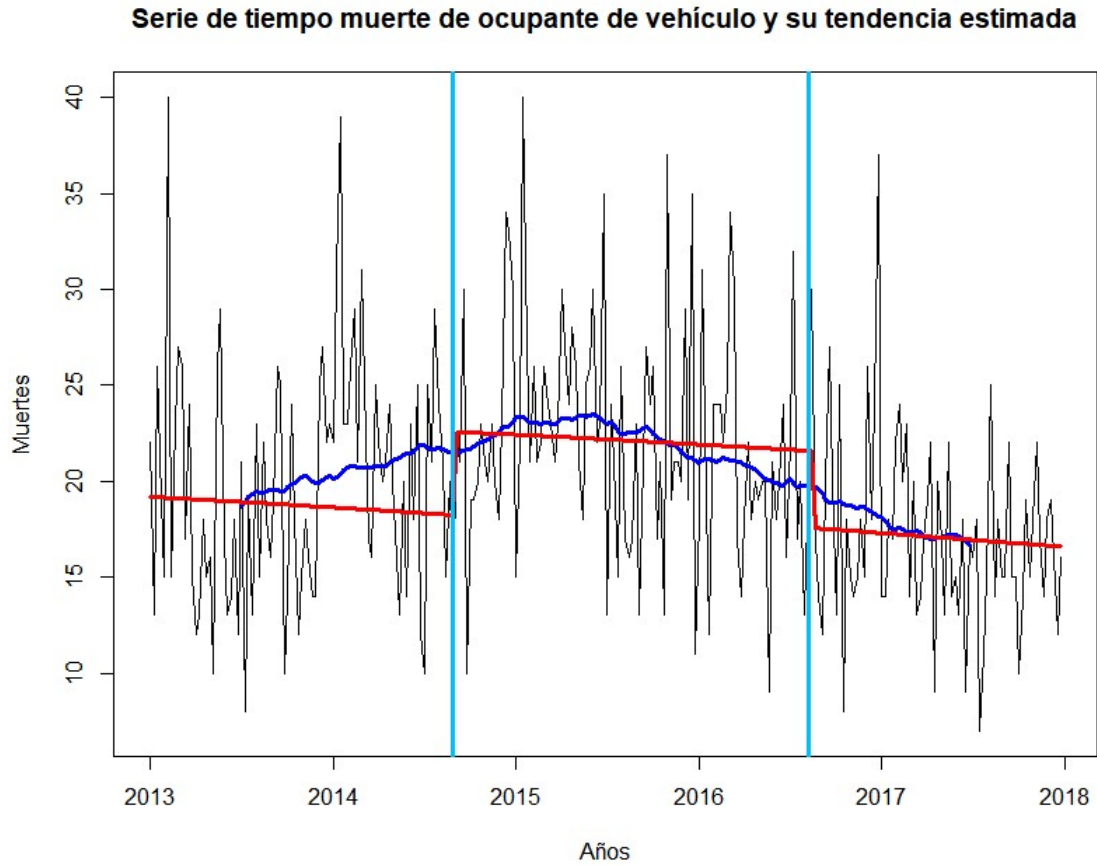


Figura 5.6: Serie de tiempo muertes ocupante de vehículo y su tendencia

5.5.4. Evaluación del impacto de la ley Emilia

A continuación, se realizará un análisis comparativo del impacto de la ley, es decir, evaluar quienes se vieron más beneficiados con la implementación de la ley Emilia.

Comparación estimaciones del modelo de peatones v/s ocupantes de vehículo con intervención

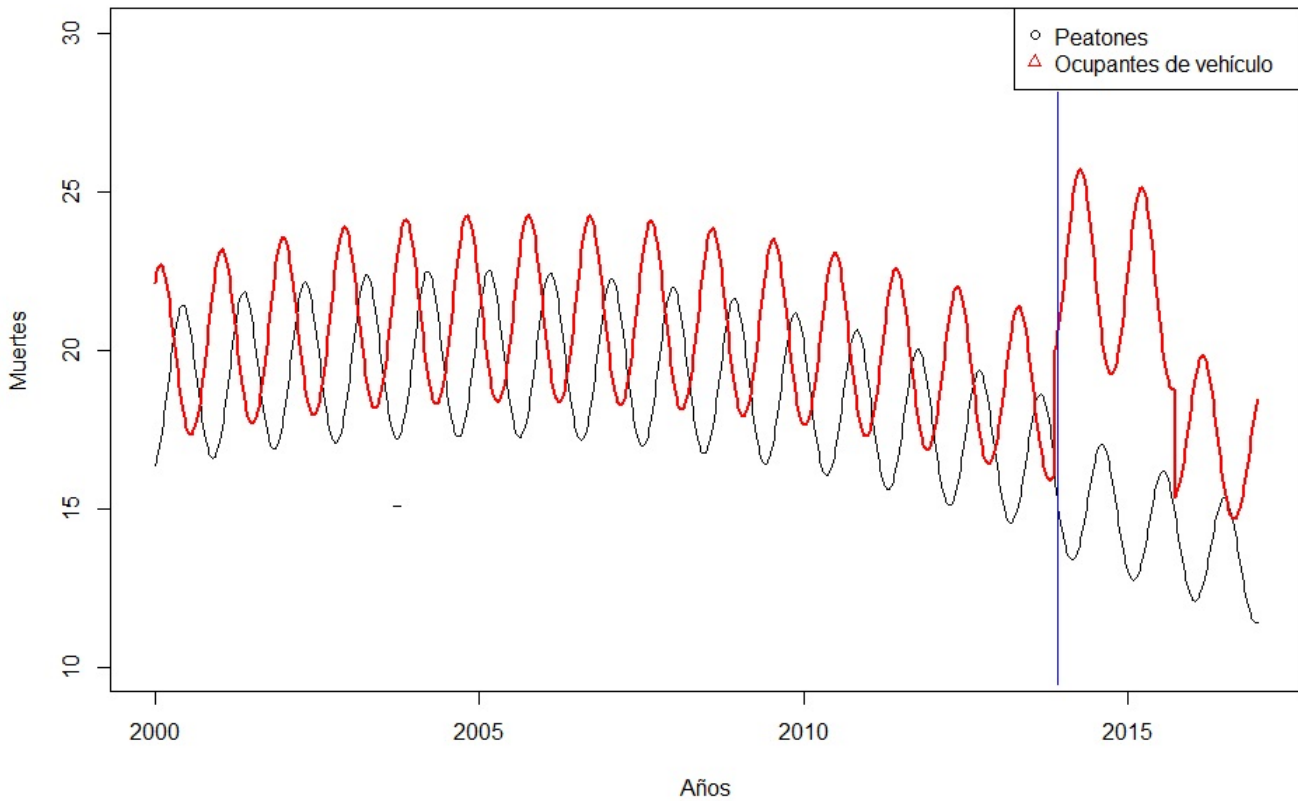


Figura 5.7: Comparación de las estimaciones del impacto de la ley en los peatones versus ocupantes de vehículo

De la figura 5.7 tenemos que el modelo de peatones estimó una caída considerable de muertes luego de la implementación de la ley, pero un aumento en las muertes de ocupantes de vehículo luego de la implementación. Hay que decir que, los peatones se vieron beneficiados por una conducción más prudente por parte de los conductores luego de la implementación.

	Peatones	Ocupante de vehículo
Total	802.2573385	1186.828606
Real	789	1217

Cuadro 5.55: Tabla resumen total de muertes de peatones y ocupante de vehículo luego de un año de la implementación de la ley Emilia

De la tabla 5.55 tenemos que el total de muertes estimadas para peatones es 32% menos que las de ocupantes de vehículo, luego de la implementación de la ley los peatones se vieron más beneficiados, con una conducción más prudente por parte de los conductores.

Capítulo 6

Modelos de ITS de hospitalizaciones causadas por accidentes de tránsito

En este capítulo se presentan 7 posibles modelos para el análisis de series de tiempo interrumpidas, previo a eso se debe realizar un pequeño análisis descriptivo para determinar el modelo más apropiado.

6.1. Base de datos y análisis descriptivo

La base de datos cuenta con 4 variables que se explican a continuación:

- **Hospitalizaciones:** Variable de tipo conteo, donde se tienen las frecuencias de hospitalizaciones causadas por accidentes de tránsito a lo largo de todo el país desde el año 2013 a 2017.
- **Tiempo:** Variable discreta, que representa la tendencia de la serie.
- **Intervención:** Variable dicotómica que representa el periodo pre y post implementación de la ley Emilia, donde 0=pre-intervención y Tiempo=post-intervención.
- **Autoregresiva:** Variable autoregresiva de orden 2

A continuación, se presenta un pequeño análisis descriptivo de las variables en estudio:

Variable	Minímo	Máximo	Media	Varianza
Hospitalizaciones	195	403	292.9	1773.7
Tiempo	4	224	114	4088.5
Intervención	0	224	107.6	5291.62
Autoregresiva	195	403	292.6	1767.7

Cuadro 6.1: Estadística descriptiva

6.2. Modelos para series de tiempo interrumpidas

A continuación, se presentan 7 posibles modelos con distintas variables e interacciones para decidir cuál es el que mejor se ajusta a los datos de hospitalizaciones que se tienen en base a la implementación de la ley Emilia.

Nombre	Regresión	Modelo
Modelo 1	Poisson	(Hospitalizaciones ~ Tiempo)
Modelo 2	Binomial Negativa	(Hospitalizaciones ~ Tiempo)
Modelo 3	Binomial Negativa	(Hospitalizaciones ~ Tiempo + harmonic(Tiempo,1,52))
Modelo 4	Binomial Negativa	(Hospitalizaciones ~ Tiempo+ Intervención + harmonic(tiempo,1,52))
Modelo 5	Binomial Negativa	(Hospitalizaciones ~ Tiempo+ Intervención+ Autoregresiva + harmonic(tiempo,1,52))
Modelo 6	Binomial Negativa	(Hospitalizaciones ~ Tiempo * Intervención+ Autoregresiva + harmonic(tiempo,1,52))
Modelo 7	Binomial Negativa	(Hospitalizaciones ~ Tiempo + Intervención *harmonic(Tiempo,1,52)+ Autoregresiva)

Cuadro 6.2: Tabla resumen modelos propuestos

■ Modelo 1:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	5.7521	0.0079	727.05	0.0000
Tiempo	-0.0006	0.0001	-10.42	0.0000

Cuadro 6.3: Resumen Modelo 1 hospitalizaciones

Del modelo anterior es importante mencionar que la Deviance=1212.2 y que los grados de libertad son 219, para ver el ajuste que tienen los datos al modelo se debe dividir la deviance en los grados de libertad

y si este valor es muy superior a 1 indica que existe sobredispersión. En este caso $D/df=5.535$, por lo tanto, existe una notoria sobredispersión, dado que la binomial negativa tiene la característica de solucionar esta problemática, el resto de los modelos se prueban con esta regresión.

■ **Modelo 2:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	5.7505	0.0188	306.09	0.0000
Tiempo	-0.0006	0.0001	-4.35	0.0000

Cuadro 6.4: Resumen Modelo 2 hospitalizaciones

Para este modelo tenemos que la Deviance=221.35 y los grados de libertad=219, al analizar la sobredispersión tenemos $D/df=1.01$, por lo que esta continuaremos el análisis con la distribución binomial negativa.

■ **Modelo 3:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	5.7236	0.0130	440.15	0.0000
Tiempo	-0.0004	0.0001	-4.29	0.0000
harmonic(Tiempo, 1, 52)1	0.0414	0.0089	4.67	0.0000
harmonic(Tiempo, 1, 52)2	-0.1383	0.0092	-15.10	0.0000

Cuadro 6.5: Resumen Modelo 3 hospitalizaciones

Dada la figura 4.3, se puede apreciar que existe estacionalidad en los datos (2.2.9), para ajustarla se ingresó la función Harmonic que además es significativa para el modelo.

■ **Modelo 4:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	5.6765	0.0177	320.70	0.0000
Tiempo	0.0020	0.0007	3.10	0.0019
Intervención	-0.0022	0.0006	-3.79	0.0001
harmonic(Tiempo, 1, 52)1	0.0524	0.0091	5.78	0.0000
harmonic(Tiempo, 1, 52)2	-0.1422	0.0089	-15.92	0.0000

Cuadro 6.6: Resumen Modelo 4 hospitalizaciones

Del cuadro 6.6 se puede apreciar que el ingreso de la intervención al modelo es significativa, si observamos también el estimador de la intervención, se puede apreciar que tiene un cambio de signo con respecto al tiempo del modelo, esto representa que la intervención provocó una disminución en la tendencia de los datos.

Para ajustar la autocorrelación se ingresó un AR(2) al modelo,

■ **Modelo 5:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	5.4631	0.0677	80.72	0.0000
Tiempo	0.0016	0.0007	2.46	0.0138
Intervención	-0.0017	0.0006	-2.96	0.0031
harmonic(Tiempo, 1, 52)1	0.0450	0.0091	4.93	0.0000
harmonic(Tiempo, 1, 52)2	-0.1113	0.0129	-8.64	0.0000
Autoregresiva	0.0007	0.0002	3.26	0.0011

Cuadro 6.7: Resumen Modelo 5 hospitalizaciones

Del cuadro 6.7 tenemos que la variable Autoregresiva es significativa para el modelo, al igual que el resto de las variables.

■ **Modelo 6:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	5.6197	0.0775	72.49	0.0000
Tiempo	-0.0006	0.0009	-0.73	0.4669
Intervención	-0.0012	0.0006	-2.01	0.0449
harmonic(Tiempo, 1, 52)1	0.0358	0.0092	3.90	0.0001
harmonic(Tiempo, 1, 52)2	-0.1223	0.0128	-9.52	0.0000
Autoregresiva	0.0005	0.0002	2.08	0.0372
Tiempo:Intervención	0.0000	0.0000	3.75	0.0002

Cuadro 6.8: Resumen Modelo 6 hospitalizaciones

Del cuadro 6.8, se puede notar que la interacción entre el Tiempo y la Intervención es significativa para el modelo, pero esto genera que la variable Tiempo, no esté ayudando a explicar la variabilidad.

■ **Modelo 7:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	5.4925	0.0684	80.25	0.0000
Tiempo	0.0023	0.0008	2.95	0.0031
Intervención	-0.0023	0.0007	-3.42	0.0006
harmonic(Tiempo, 1, 52)1	0.0671	0.0183	3.66	0.0003
harmonic(Tiempo, 1, 52)2	-0.1481	0.0200	-7.42	0.0000
Autoregresiva	0.0006	0.0002	2.56	0.0104
Intervención:harmonic(Tiempo, 1, 52)1	-0.0002	0.0001	-1.30	0.1939
Intervención:harmonic(Tiempo, 1, 52)2	0.0003	0.0001	2.29	0.0218

Cuadro 6.9: Resumen Modelo 7 hospitalizaciones

Del cuadro 6.9, podemos notar que la interacción es solo significativa para el primer harmonico.

A continuación, se presenta un análisis de supuestos para escoger el mejor modelo de los 7 propuestos anteriormente.

6.3. Análisis de Supuestos del Modelo

Previo al análisis de supuestos, ilustraremos el criterio de información de AKAIKE y la suma de cuadrados residual en post de seleccionar el mejor modelo.

Modelo	AIC	SCR
Modelo 2	2261.5	221.35
Modelo 3	2100.4	221.49
Modelo 4	2088.6	221.95
Modelo 5	2068.3	222.08
Modelo 6	2068.8	222.45
Modelo 7	2077.7	222.29

Cuadro 6.10: Tabla resumen AIC y SCR

De la tabla 6.10, tenemos que el menor AIC, es para el Modelo 5 y la menor SCR es para el primer modelo.

6.4. Supuestos

6.4.1. Test de Box-Ljung

Este test tiene como función probar la aleatoriedad e independencia en forma conjunta.

H_0 : Los errores están no correlacionados *v/s* H_1 : Los errores están correlacionados

Modelo	P-valor	Resultado
Modelo 2	0.0000	Se rechaza
Modelo 3	0.0000	Se rechaza
Modelo 4	0.0556	No se rechaza
Modelo 5	0.4008	No se rechaza
Modelo 6	0.4482	No se rechaza
Modelo 7	0.4529	No se rechaza

Cuadro 6.11: Tabla resumen Test de Box-Ljung

En la tabla anterior se hizo una tabla resumen del resultado del test de Box-Ljung para todos los modelos planteados anteriormente, donde en los últimos 4 modelos, los residuos no están correlacionados para un lag=50.

6.4.2. Breusch-Pagan

Este test se utiliza para probar la homocedasticidad de los residuos.

H_0 : Residuos Homocedásticos *v/s* H_1 : Residuos Heterocedásticos

Modelo	P-valor	Resultado
Modelo 2	0.0555	No se rechaza
Modelo 3	0.0129	No se rechaza al 1 %
Modelo 4	0.0000	Se rechaza
Modelo 5	0.0142	No se rechaza al 1 %
Modelo 6	0.0398	No se rechaza al 1 %
Modelo 7	0.0323	No se rechaza al 1 %

Cuadro 6.12: Tabla resumen Test de Breusch-Pagan

En el cuadro 6.12, tenemos que al 1 % no se rechaza la hipótesis nula en 5 de los modelos planteados.

6.4.2.1. Tabla resumen supuestos de los modelos

A continuación, se muestra una tabla resumen de todos los supuestos del modelo, y en base a esto elegir el mejor para evaluar finalmente el impacto de la ley:

Modelo	AIC	SCR	Box-Ljung	Breusch-Pagan
Modelo 2	2261.5	221.35	Se rechaza	No se rechaza
Modelo 3	2100.4	221.49	Se rechaza	No se rechaza al 1 %
Modelo 4	2088.6	221.95	No se rechaza	Se rechaza
Modelo 5	2068.3	222.08	No se rechaza	No se rechaza al 1 %
Modelo 6	2068.8	222.45	No se rechaza	No se rechaza al 1 %
Modelo 7	2077.7	222.29	No se rechaza	No se rechaza al 1 %

Cuadro 6.13: Tabla resumen supuestos de los modelos

Dada la significancia de los parámetros en cada modelo planteado junto con el cumplimiento de los supuestos y la evaluación del AIC Y la SCR, el Modelo 5, es el adecuado para evaluar el impacto de la ley, ya que, todas sus variables son significativas, tiene el menor AIC y una SCR que es solo un poco mayor al de menor valor, y se cumplen los supuestos.

6.4.3. Serie de tiempo hospitalizaciones y su tendencia

En la siguiente figura se observa la serie de tiempo de hospitalizaciones junto con su tendencia (marcado con la línea azul) y la tendencia estimada del Modelo 5 (representado por la línea roja), es importante notar que el comportamiento de la tendencia va en un aumento significativo antes de la intervención, pero luego de esto disminuye considerablemente.

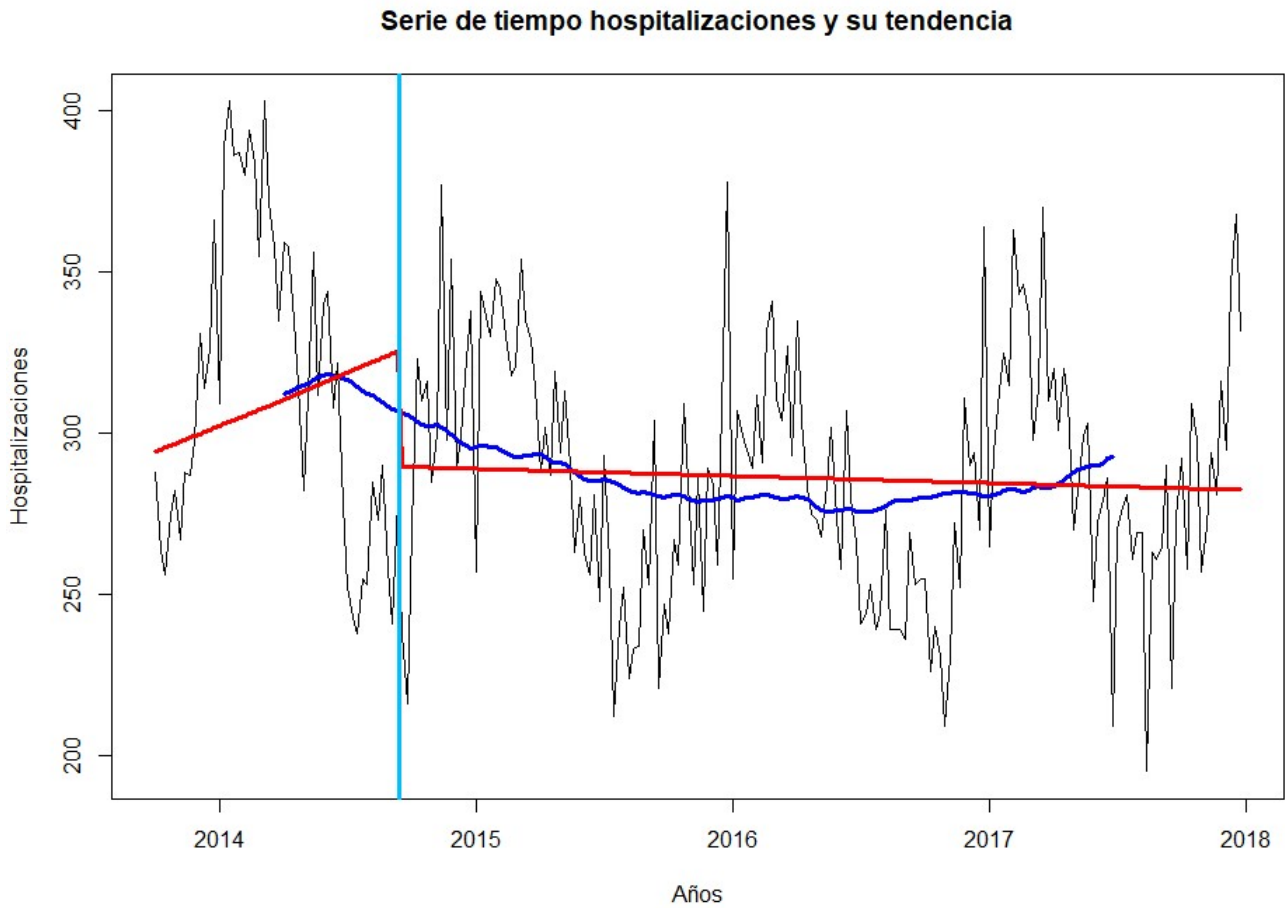


Figura 6.1: Serie de tiempo hospitalizaciones y su tendencia

6.4.4. Evaluación del impacto de la ley Emilia

Dado el mejor modelo del punto anterior, se grafican a continuación las estimaciones de dicho modelo sin intervención y con intervención para evaluar el impacto de la ley en las hospitalizaciones causadas por accidentes de tránsito a lo largo de todo el país.

Comparación estimaciones del modelo sin intervención y con intervención

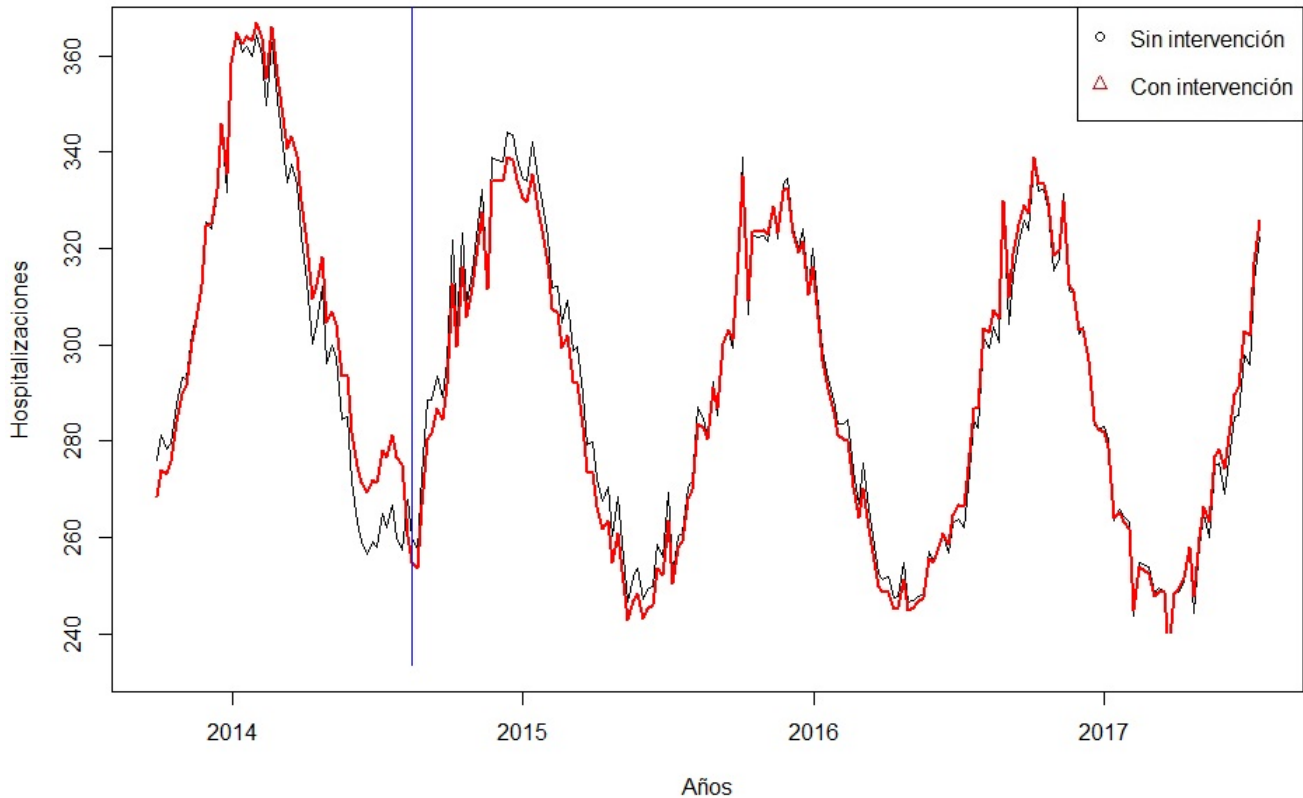


Figura 6.2: Comparación estimaciones del mejor modelo sin intervención versus con intervención

De la figura 6.2 tenemos que previo al impacto de la ley Emilia el modelo con intervención estima ligeramente más hospitalizaciones que sin intervención, pero luego de la implementación de la ley el comportamiento de ambos modelos es bastante similar. A continuación veremos si existe realmente una diferencia significativa entre estos modelos.

	Moodelo sin intervención	Modelo con intervención	Valor Real
Total	15661.79071	15377.02789	

15232 height

Cuadro 6.14: Tabla resumen total de hospitalizaciones un año después de la implementación de la ley Emilia

De la tabla 6.14 tenemos que desde la semana del 16 de septiembre del 2014 al 16 de septiembre del 2015 la estimación de hospitalizaciones con el modelo sin intervención fue de 15661.79, mientras que para

esas mismas semanas el modelo con intervención estimó 15377.03 hospitalizaciones, es decir, se estima que el impacto de la ley disminuyó un 1.8 % las hospitalizaciones causadas por accidentes de tránsito.

6.5. Modelos de ITS de hospitalizaciones causadas por accidentes de tránsito en la Región Metropolitana

Del total de hospitalizaciones causadas por accidente de tránsito en todo el país el 34 % ocurrió en la Región Metropolitana y como además es la región más urbana es importante replicar el mismo estudio para analizar el impacto de la ley.

6.5.1. Base de datos y análisis descriptivo

La base de datos cuenta con 4 variables que se explican a continuación:

- **Hospitalizaciones:** Variable de tipo conteo, donde se tienen las frecuencias de hospitalizaciones causadas por accidentes de tránsito a lo largo de todo el país desde el año 2013 a 2017.
- **Tiempo:** Variable discreta, que representa la tendencia de la serie, en este caso es una tendencia cuadrática que es la que mejor se ajusta a los datos.
- **Intervención:** Variable dicotómica que representa el periodo pre y post implementación de la ley Emilia, donde 0=pre-intervención y Tiempo=post-intervención.
- **Autoregresiva:** Variable autoregresiva de orden 2

A continuación, se presenta un pequeño análisis descriptivo de las variables en estudio:

Variable	Minímo	Máximo	Media	Varianza
Hospitalizaciones	63	163	107.1	269.8
Tiempo	9	49729	16839	198454
Intervención	0	223	107	5204.67
Autoregresiva	63	163	107	266.96

Cuadro 6.15: Estadística descriptiva

6.5.2. Modelos para series de tiempo interrumpidas

A continuación, se presentan 7 posibles modelos con distintas variables e interacciones. Luego de elegir el que mejor se ajuste a los datos se podrá evaluar el impacto de la ley.

Nombre	Regresión	Modelo
Modelo 1	Poisson	(Hospitalizaciones ~ Tiempo)
Modelo 2	Binomial Negativa	(Hospitalizaciones ~ Tiempo)
Modelo 3	Binomial Negativa	(Hospitalizaciones ~ Tiempo + harmonic(Tiempo,1,52))
Modelo 4	Binomial Negativa	(Hospitalizaciones ~ Tiempo+ Tiempo ² Intervención + harmonic(tiempo,1,52))
Modelo 5	Binomial Negativa	(Hospitalizaciones ~ Tiempo+ Intervención+ Autoregresiva + harmonic(tiempo,1,52))
Modelo 6	Binomial Negativa	(Hospitalizaciones ~ Tiempo * Intervención+ Tiempo ² + Autoregresiva + harmonic(tiempo,1,52))
Modelo 7	Binomial Negativa	(Hospitalizaciones ~ Tiempo+ Tiempo ² + Intervención *harmonic(Tiempo,1,52)+ Autoregresiva)

Cuadro 6.16: Tabla resumen modelos propuestos

■ **Modelo 1:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.6533	0.0131	354.00	0.0000
Tiempo	0.0002	0.0001	2.41	0.0162

Cuadro 6.17: Resumen Modelo 1 hospitalizaciones en la RM

Del modelo anterior es importante mencionar que la Deviance=553.01 y que los grados de libertad son 219, para evaluar la sobredispersión tenemos que D/gl=2.53, por lo tanto, existe una notoria sobredispersión, dado que la binomial negativa tiene la característica de solucionar esta problemática, ocuparemos esta regresión.

■ **Modelo 2:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.6536	0.0207	224.77	0.0000
Tiempo	0.0002	0.0002	1.51	0.1321

Cuadro 6.18: Resumen Modelo 2 hospitalizaciones RM

Para este modelo tenemos que la Deviance=221.77 y los grados de libertad=219, al analizar la sobre-dispersión tenemos D/gl=1.01, por lo que continuaremos el análisis con la distribución binomial negativa.

■ **Modelo 3:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.6415	0.0192	241.60	0.0000
Tiempo	0.0003	0.0001	2.30	0.0214
harmonic(Tiempo, 1, 52)1	0.0290	0.0133	2.19	0.0285
harmonic(Tiempo, 1, 52)2	-0.0814	0.0135	-6.02	0.0000

Cuadro 6.19: Resumen Modelo 3 hospitalizaciones en la RM

Del cuadro 6.19 podemos apreciar que la función harmonic ingresada, es la encargada de ajustar la estacionalidad de los datos, además estaría ayudando a explicar la variabilidad del modelo.

■ **Modelo 4:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.7753	0.0398	119.89	0.0000
Tiempo	-0.0048	0.0014	-3.32	0.0009
Tiempo2	0.0000	0.0000	2.69	0.0071
intervención	0.0030	0.0010	3.02	0.0026
harmonic(Tiempo2, 1, 52)1	0.0069	0.0141	0.49	0.6254
harmonic(Tiempo2, 1, 52)2	-0.0264	0.0143	-1.84	0.0655

Cuadro 6.20: Resumen Modelo 4 hospitalizaciones en la RM

Del cuadro 6.20 se puede apreciar que el ingreso de la intervención es significativa, es decir, la intervención de la ley estaría ayudando a explicar la variabilidad del modelo.

Se puede notar por el signo del estimador de la tendencia tenía un comportamiento decreciente antes de la intervención, pero que el ingreso de esta variable cambio su sentido.

Para ajustar la correlación se ingresó un AR(1) al modelo,

■ **Modelo 5:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.4575	0.0785	56.77	0.0000
Tiempo	-0.0038	0.0015	-2.62	0.0087
Tiempo2	0.0000	0.0000	2.56	0.0104
intervención	0.0022	0.0010	2.30	0.0215
autoregresiva	0.0028	0.0006	4.69	0.0000
harmonic(Tiempo, 1, 52)1	0.0065	0.0137	0.48	0.6325
harmonic(Tiempo, 1, 52)2	-0.0595	0.0138	-4.32	0.0000

Cuadro 6.21: Resumen Modelo 5 hospitalizaciones en la RM

Del cuadro 6.21 tenemos que la variable Autoregresiva es significativa para el modelo, y con esto se espera corregir la autocorrelación que se verá en el análisis de supuestos.

■ **Modelo 6:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.4888	0.0890	50.44	0.0000
Tiempo	-0.0075	0.0052	-1.45	0.1458
intervención	0.0055	0.0044	1.23	0.2190
autoregresiva	0.0027	0.0006	4.67	0.0000
Tiempo2	0.0001	0.0001	0.81	0.4161
harmonic(Tiempo, 1, 52)1	0.0062	0.0137	0.45	0.6494
harmonic(Tiempo, 1, 52)2	-0.0644	0.0153	-4.21	0.0000
Tiempo:intervención	-0.0001	0.0001	-0.74	0.4574

Cuadro 6.22: Resumen Modelo 6 hospitalizaciones en la RM

Del cuadro 6.22, se puede notar que la interacción entre el Tiempo y la Intervención no es significativa además de disminuir la significancia del resto de las variables.

■ **Modelo 7:**

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	4.4809	0.0825	54.30	0.0000
Tiempo	-0.0045	0.0018	-2.52	0.0117
Tiempo2	0.0000	0.0000	2.75	0.0060
intervención	0.0026	0.0012	2.20	0.0277
harmonic(Tiempo, 1, 52)1	-0.0092	0.0306	-0.30	0.7632
harmonic(Tiempo, 1, 52)2	-0.0379	0.0229	-1.65	0.0980
autoregresiva	0.0027	0.0006	4.61	0.0000
intervención:harmonic(Tiempo, 1, 52)1	0.0001	0.0002	0.57	0.5714
intervención:harmonic(Tiempo, 1, 52)2	-0.0002	0.0002	-1.17	0.2408

Cuadro 6.23: Resumen Modelo 7 hospitalizaciones en la RM

La interacción planteada en este modelo, no es significativa.

A continuación, se presenta un análisis de supuestos para escoger el mejor modelo de los 7 propuestos anteriormente.

6.6. Análisis de Supuestos del Modelo

Previo al análisis de supuestos, ilustraremos el criterio de información de AKAIKE y la suma de cuadrados residual en post de seleccionar el mejor modelo.

Modelo	AIC	SCR
Modelo 2	1876.79	223.09
Modelo 3	1844.10	223.11
Modelo 4	1869.22	222.99
Modelo 5	1815.96	222.74
Modelo 6	1817.41	222.75
Modelo 7	1818.34	222.79

Cuadro 6.24: Tabla resumen AIC y SCR

De la tabla 6.24, tenemos que el menor AIC y SCR es para el Modelo 5, donde además se tiene la mayor significancia de variables.

6.6.1. Supuestos

6.6.1.1. Test de Box-Ljung

Este test tiene como función probar la aleatoriedad e independencia en forma conjunta.

H_0 : Los errores están no correlacionados *v/s* H_1 : Los errores están correlacionados

Modelo	P-valor	Resultado
Modelo 2	0.0000	Se rechaza
Modelo 3	0.0000	Se rechaza
Modelo 4	0.0000	Se rechaza
Modelo 5	0.081	No se rechaza
Modelo 6	0.0713	No se rechaza
Modelo 7	0.0813	No se rechaza

Cuadro 6.25: Tabla resumen Test de Box-Ljung

En la tabla anterior se hizo una tabla resumen del resultado del test de Box-Ljung para todos los modelos planteados anteriormente, donde para un lag=50, el supuesto se cumple solo cuando se ingresa la variable autoregresiva al modelo.

6.6.1.2. Breusch-Pagan

Este test se utiliza para probar la homocedasticidad de los residuos.

H_0 : Residuos Homocedásticos *v/s* H_1 : Residuos Heterocedásticos

Modelo	P-valor	Resultado
Modelo 2	0.083	No se rechaza
Modelo 3	0.044	No se rechaza al 1%
Modelo 4	0.485	No se rechaza
Modelo 5	0.2177	No se rechaza
Modelo 6	0.1799	No se rechaza
Modelo 7	0.174	No se rechaza

Cuadro 6.26: Tabla resumen Test de Breusch-Pagan

En el cuadro 6.26, se cumple el modelo de homocedasticidad todos los modelos planteados.

6.6.1.3. Tabla resumen supuestos de los modelos

A continuación, se muestra una tabla resumen de todos los supuestos del modelo, y en base a esto elegir el mejor para evaluar finalmente el impacto de la ley:

Modelo	AIC	SCR	Box-Ljung	Breusch-Pagan
Modelo 2	1876.79	223.09	Se rechaza	No se rechaza
Modelo 3	1844.10	223.11	Se rechaza	Se rechaza
Modelo 4	1869.22	222.99	Se rechaza	No se rechaza
Modelo 5	1815.96	222.74	No se rechaza	No se rechaza
Modelo 6	1817.41	222.75	No se rechaza	No se rechaza
Modelo 7	1818.34	222.79	No se rechaza	No se rechaza

Cuadro 6.27: Tabla resumen supuestos de los modelos

Dado todo el análisis de supuestos de los modelos planteados, la significancia de las variables y el AIC y la SCR, el mejor modelo para evaluar el impacto de la ley es el Modelo 5, dado que tiene el mejor ajuste en todos los puntos mencionados anteriormente.

6.6.2. Serie de tiempo hospitalizaciones y su tendencia

En la siguiente figura se puede observar la serie de tiempo con su respectiva tendencia marcada por la línea azul, sobre esto está la tendencia estimada del mejor modelo seleccionado en el punto anterior, que es la línea roja, la intervención (línea celeste) nos permite visualizar que luego de la implementación de la ley existe un aumento de tendencia, es decir, aumentaron la cantidad de hospitalizaciones en la Región Metropolitana causadas por accidentes de tránsito. Finalmente el comportamiento de la tendencia va al alza hasta el final de la serie.

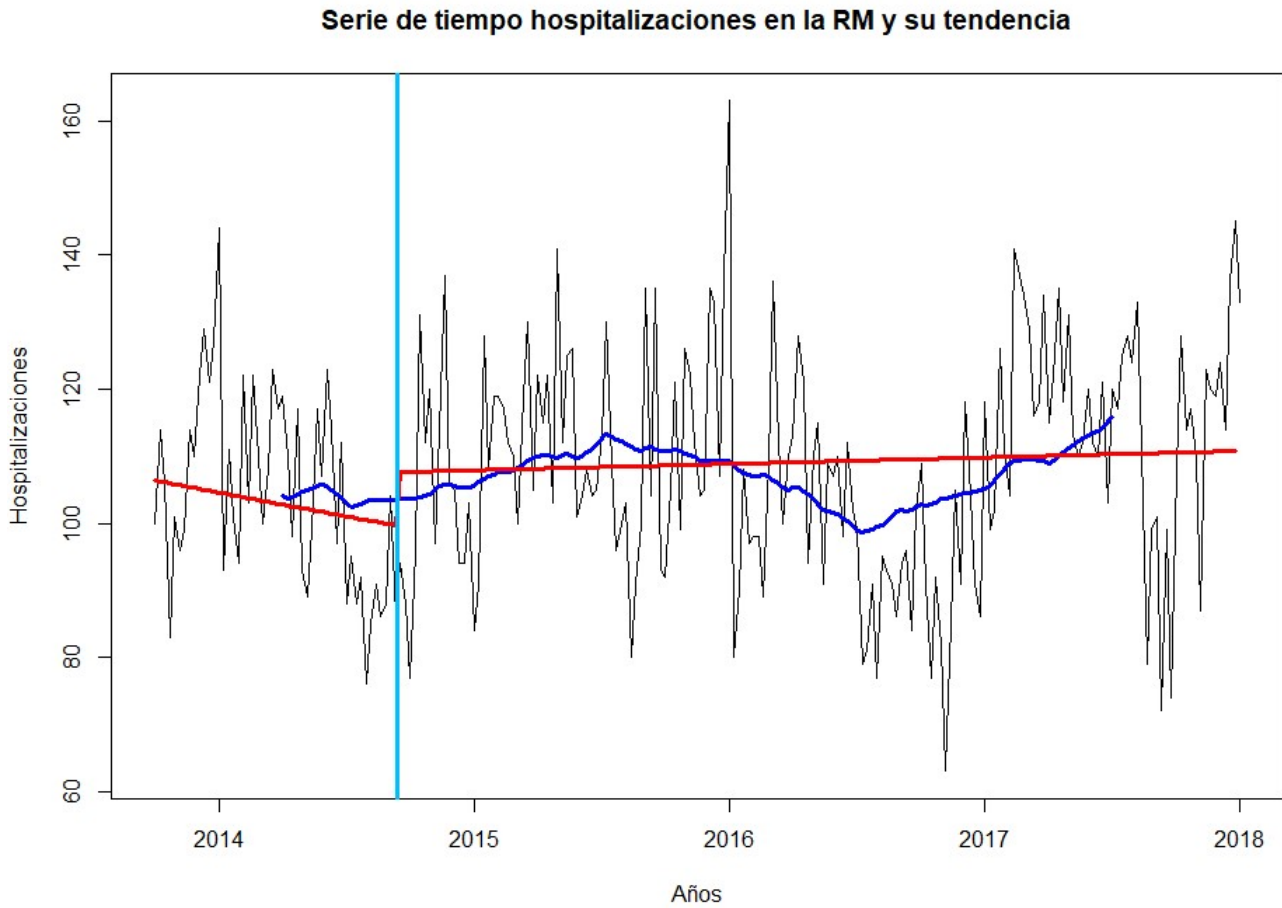


Figura 6.3: Serie de tiempo hospitalizaciones y su tendencia

6.6.3. Evaluación del impacto de la ley Emilia

Dado el mejor modelo del punto anterior, se grafican a continuación las estimaciones de dicho modelo sin intervención y con intervención para evaluar el impacto de la ley en las hospitalizaciones en la Región Metropolitana causadas por accidentes de tránsito.

Comparación estimaciones del modelo sin intervención y con intervención

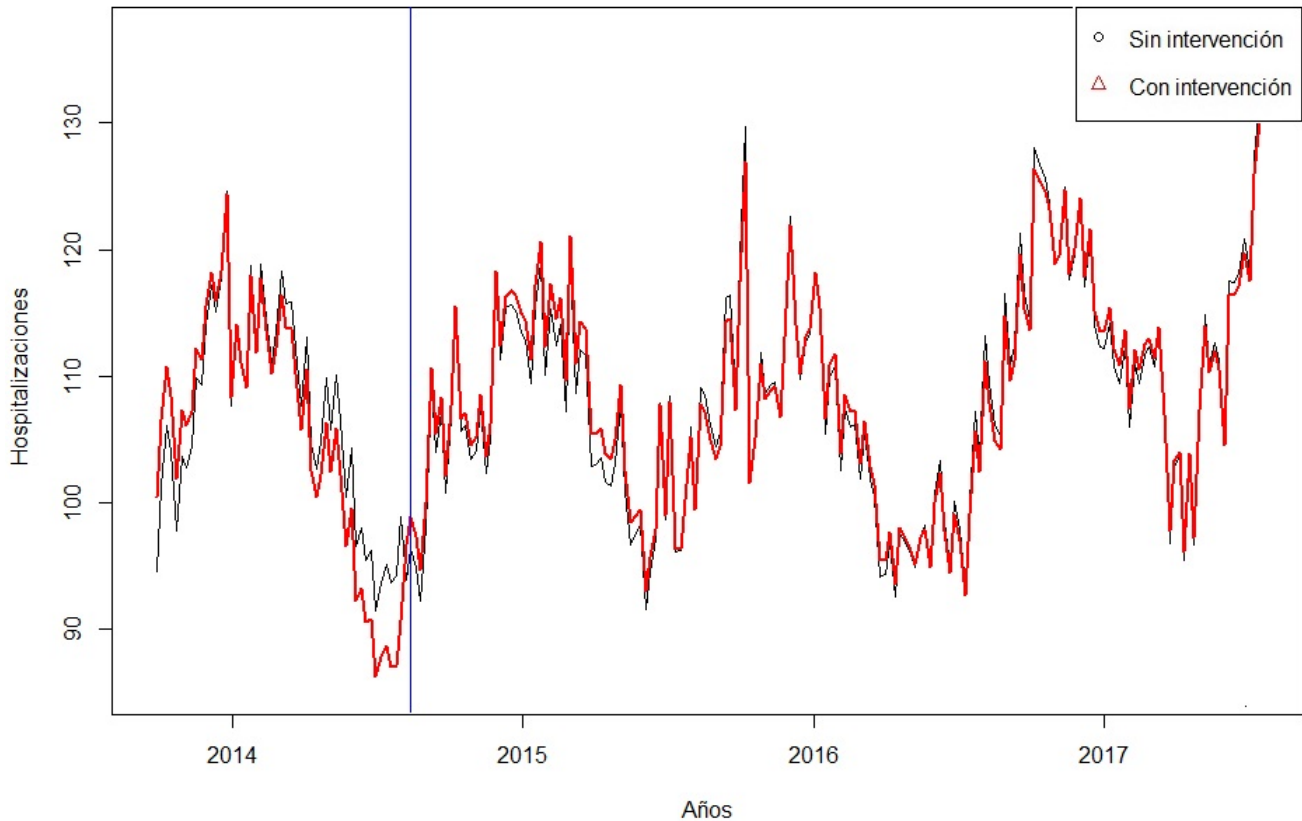


Figura 6.4: Comparación estimaciones del mejor modelo sin intervención versus con intervención

De la figura 6.4 tenemos que previo al impacto de la ley Emilia el modelo con intervención estima menos hospitalizaciones que sin intervención, pero luego de la implementación de la ley el comportamiento de ambos modelos es bastante similar, a continuación veremos si existe realmente una diferencia significativa entre estos.

	Modelo sin intervención	Modelo con intervención	Valor Real
Total	5616.222545	5700.023526	5734

Cuadro 6.28: Tabla resumen total de hospitalizaciones en la Región Metropolitana un año después de la implementación de la ley Emilia

De la tabla 6.28 tenemos que el total de hospitalizaciones durante un año luego de la implementación de la ley Emilia sin considerar esta intervención, estimó un total de 5616.22, es decir, un 1.4% menos que con la implementación de la ley. El modelo considerando la intervención estima ligeramente más hospitalizaciones en la Región Metropolitana

Capítulo 7

Conclusiones

La evaluación del impacto de la ley Emilia se realizó en dos grandes áreas, que son muertes y hospitalizaciones.

Las muertes causadas por accidentes de tránsito a lo largo de todo el país luego de la implementación de la ley tuvieron un decremento de un 0.79 %, aunque este valor no es significativo para asegurar que la implementación de la ley tuvo un impacto positivo considerando está pequeña reducción.

Al realizar el estudio en la Región Metropolitana dado que es la región más urbana del país y que corresponde al 30 % del total de muertes de los años en estudio, se obtuvo que las muertes disminuyeron un 3.3 %. Es lógico pensar que el impacto fuera mucho mayor en esta región. De igual manera este valor tampoco es significativo para aseverar que la ley Emilia impactó positivamente en la disminución de muertes.

Por otra parte, era de importancia realizar una comparativa entre las muertes de peatones y ocupantes de vehículo y analizar quienes fueron más sensibles a la implementación de la ley. De esto tenemos que los peatones se vieron beneficiados considerablemente por una conducción más prudente por parte de los conductores, pero por el contrario las muertes de ocupantes de vehículo tuvieron un aumento inclusive en el comportamiento previó. La diferencia que se estima en esta comparación es de un 32 % y es la más significativa de todo el estudio, de esta comparación se puede concluir que la ley Emilia impactó positivamente las muertes de peatones en comparación a los ocupantes de vehículo luego de su implementación.

Analizando las hospitalizaciones causadas por accidentes de tránsito, se tiene que luego de la implementación de la ley éstas disminuyeron un 1.8 % en todo el país. Esta diferencia no nos permite asegurar que existe un decrecimiento significativo y que la ley realmente impactó en la disminución de hospitalizaciones. Realizando el mismo análisis en la Región Metropolitana, tenemos que luego de la implementación de la ley hubo un aumento de un 1.4 % en las hospitalizaciones, esto se debe a que la ley obliga a detener la

marcha del vehículo y a prestar ayuda, provocando de cierta manera que se puede evitar una muerte, pero aumentar el número de hospitalizaciones.

Finalmente realizando un análisis global, tenemos que la ley Emilia no produjo ningún impacto significativo más que en la comparación de muertes de peatones y ocupantes de vehículo, esto puede deberse a diversos factores, pero uno de ellos es la poca propaganda que existe en los medios de comunicación con respecto a la ley o más bien a no conducir en estado de ebriedad, dichas propagandas se realizan solo en fiestas patrias o en fin de semana largo. Incitar una conducción responsable y sin alcohol en la sangre con una mayor frecuencia, podría ayudar en el impacto de la ley, aunque también hay que considerar que existe un tema cultural importante, es decir, es difícil pensar que una persona de 40 años que ha conducido en reiteradas ocasiones bajo la influencia del alcohol, tenga un cambio de pensamiento con la implementación de esta ley.

Para estudios futuros sería importante considerar variables como el tamaño del parque automotriz que existe en Chile; la población nacional para poder realizar tasas de accidentes por cantidad de personas y obtener con esto un mayor entendimiento del comportamiento de los datos; indagar por cuanto tiempo se propago en los medios de comunicación la ley Emilia después de su implementación, y comparar si el aumento de tendencia de las series obtenidas, coinciden con este punto.

Bibliografía

- [1] CHILE, BIBLIOTECA DEL CONGRESO NACIONAL DE CHILE /BCN, disponible en <https://www.bcn.cl/leychile/N?i=1066775f=2014-09-16>.
- [2] Lopez, Cummins y Gasparrini (2016). INTERRUPTED TIME SERIES REGRESSION FOR THE EVALUATION OF PUBLIC HEALTH INTERVENTIONS: A TUTORIAL [PDF file]. International Journal of Epidemiology. doi:10.1093/ije/dyw098.
- [3] CHILE, COMISIÓN NACIONAL DE SEGURIDAD DE TRÁNSITO, DISPONIBLE, disponible en <https://www.conaset.cl/alcohol-y-conduccion/>.
- [4] Lopez, Cummins y Gasparrini, (2016, p. 2-7).
- [5] CHILE, COMISIÓN NACIONAL DE SEGURIDAD DE TRÁNSITO, DISPONIBLE, disponible en <https://www.conaset.cl/ley-emilia/>.
- [6] CLASIFICACIÓN ESTADÍSTICA INTERNACIONAL DE ENFERMEDADES Y PROBLEMAS DE SALUD DÉCIMA REVISIÓN, disponible en <http://www.icd.who.int>.
- [7] Lopez, Cummins, Gasparrini (2018). THE USE OF CONTROLS IN INTERRUPTED TIME SERIES STUDIES OF PUBLIC HEALTH INTERVENTIONS [PDF file]. International Journal of Epidemiology. doi:10.1093/ije/dyy135.
- [8] Marco, Delgado, Vargas, Muñoz, Bhutta y Ferreccio (2018). TYPHOID FEVER IN CHILE 1969–2012: ANALYSIS OF AN EPIDEMIC AND ITS CONTROL [PDF file]. doi:10.4269/ajtmh.18-0125.
- [9] Box y Tiao (2005). INTERVENTION ANALYSIS WITH APPLICATIONS TO ECONOMIC AND ENVIRONMENTAL PROBLEMS [PDF file].
- [10] CHILE, COMISIÓN NACIONAL DE SEGURIDAD DE TRÁNSITO, disponible en <https://www.conaset.cl/programa/observatorio-datos-estadistica/>.

- [11] Hyndman, R.J., Athanasopoulos, G. (2018) FORECASTING: PRINCIPLES AND PRACTICE, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2.
- [12] MODELOS LINEALES GENERALIZADOS [PDF file.]. Recuperado de <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/Categor/Tema3Cate.pdf>
- [13] Pollock, DSG (2000). ESTIMACIÓN DE TENDENCIAS Y TENDENCIAS A TRAVÉS DE FILTROS RACIONALES DE ONDA CUADRADA. REVISTA DE ECONOMETRÍA, 99 (2), 317-334.
- [14] Pulzara, J.N (2018) FAMILIA EXPONENCIAL. Recuperado de http://rstudio-pubs-static.s3.amazonaws.com/432153_0247c7e75e584c4d941ec29bae9bbbd1.html.
- [15] BASIC REGRESSION ANALYSIS WITH TIME SERIES DATA[PDF file.]. Recuperado de <http://www.eco.uc3m.es/jgonzalo/teaching/TecnicasEconometricas/WooldridgeCh10-12.pdf>
- [16] Diaz-Quijano FA. REGRESIONES APLICADAS AL ESTUDIO DE EVENTOS DISCRETOS EN EPIDEMIOLOGÍA. Rev Univ Ind Santander Salud. 2016; 48(1): 9-15.
- [17] Majune, Socrates Kraido THEORY AND PRACTICE OF TESTING FOR A SINGLE STRUCTURAL BREAK IN STATA. Recuperado de <https://www.econstor.eu/bitstream/10419/172517/3/Single%20Structural%20Break.pdf>.
- [18] ANÁLISIS DE LA TENDENCIA, disponible en <http://www.ub.edu/aplicainfor/spss/cap8-3.htm>
- [19] León-Álvarez AL, Betancur-Gómez JI, Jaimes F, Grisales-Romero H. RONDA CLÍNICA Y EPIDEMIOLOGICA. SERIES DE TIEMPO INTERRUMPIDAS.Iatreia. 2017 Jul-Sept;30(3): 344-351.
doi: 10.17533/udea.iatreia.v30n3a11
- [20] De la Fuente, S. SERIES TEMPORALES: MODELO ARIMA. Recuperado de <http://www.estadistica.net/ECONOMETRIA/SERIES-TEMPORALES/modelo-arima.pdf>
- [21] Majune, Socrates Kraido (2018) : THEORY AND PRACTICE OF TESTING FOR A SINGLE STRUCTURAL BREAK IN STATA, ZBW

Capítulo 8

Anexo

- Creación de la serie temporal

```
#####  
##                               Obtención frecuencia semanal de muertes en STATA                               ##  
#####
```

```
keep DIA_DEF MES_DEF ANO_DEF  
gen fechareal=mdy( MES_DEF, DIA_DEF, ANO_DEF)  
format fechareal %td  
gen semanareal=week(fechareal)  
contract semanareal ANO_DEF, freq()
```

```
#####  
##                               Obtención frecuencia semanal de hospitalizaciones en STATA                               ##  
#####
```

```
keep DIA_EGRE MES_EGRE ANO_EGRE  
gen fechareal=mdy( MES_EGRE, DIA_EGRE, ANO_EGRE)  
format fechareal %td  
gen semanareal=week(fechareal)  
contract semanareal ANO_EGRE, freq()
```

```
#####
##                               Creación serie temporal muertes                               ##
#####

muertes
muerte1<-ts(muertes ,start = 2000, frequency = 52)
ts.plot(muerte1, col="black", main="Muertes causadas por accidentes de tránsito", lwd=1,
xlab="Años" , ylab="Muertes")
plot(decompose(muerte1))
muert<-window(muerte1,start=2013)
ts.plot(muerte1, col="black", main="Ventana de muertes causadas por accidentes de tránsito",
lwd=1, ylab="Muertes", xlab="Años")
x<-decompose(muerte1)
plot(decompose(muerte1))
real<-x$trend
plot(real, col="black", main="Tendencia de la serie", xlab="Años", ylab="Tendencia")
abline(v=2014.75,col="red")
abline(v=2015.71,col="blue")
#####
##                               Creación serie temporal muertes en la RM                               ##
#####

muertesRM
muerte1<-ts(muertesRM ,start = 2007, frequency = 52)
ts.plot(muerte1, col="black", main="Muertes causadas por accidentes de tránsito en la RM", lwd=1,
xlab="Años" , ylab="Muertes")
plot(decompose(muerte1))
muert<-window(muerte1,start=2013)
ts.plot(muerte1, col="black", main="Ventana de muertes causadas por accidentes de tránsito",
lwd=1, ylab="Muertes", xlab="Años")
x<-decompose(muerte1)
plot(decompose(muerte1))
```

```

real<-x$trend
plot(real, col="black", main="Tendencia de la serie", xlab="Años", ylab="Tendencia")
abline(v=2014.75,col="red")
abline(v=2015.71,col="blue")
#####
##                               Creación serie temporal muertes en la RM                               ##
#####

peatones
muerte1<-ts(peatones ,start = 2000, frequency = 52)
ts.plot(muerte1, col="black", main="Muertes de peatones causadas por accidentes de tránsito",
lwd=1, xlab="Años" , ylab="Muertes")
plot(decompose(muerte1))
muert<-window(muerte1,start=2013)
ts.plot(muerte1, col="black", main="Ventana de muertes causadas por accidentes de tránsito",
lwd=1, ylab="Muertes", xlab="Años")
x<-decompose(muerte1)
plot(decompose(muerte1))
real<-x$trend
plot(real, col="black", main="Tendencia de la serie", xlab="Años", ylab="Tendencia")
abline(v=2014.75,col="red")
abline(v=2015.71,col="blue")
#####
##                               Creación serie temporal muertes en la RM                               ##
#####

ocupante
muerte1<-ts(ocupante ,start = 2000, frequency = 52)
ts.plot(muerte1, col="black", main="Muertes de ocupante de vehículo causadas por accidentes de
tránsito", lwd=1, xlab="Años" , ylab="Muertes")
plot(decompose(muerte1))
muert<-window(muerte1,start=2013)

```

```
ts.plot(muerte1, col="black", main="Ventana de muertes causadas por accidentes de tránsito",
lwd=1, ylab="Muertes", xlab="Años")
x<-decompose(muerte1)
plot(decompose(muerte1))
real<-x$trend
plot(real, col="black", main="Tendencia de la serie", xlab="Años", ylab="Tendencia")
abline(v=2014.75,col="red")
abline(v=2015.71,col="blue")
```

```
#####
##                               Creación serie temporal hospitalizaciones                               ##
#####
```

```
hospitalizaciones
hospil<-ts(hospitalizaciones ,start = c(2013.75,2017), frequency = 52)
ts.plot(hospil, col="black", main="Hospitalizaciones causadas por accidentes de tránsito", lwd=1,
xlab="Años" , ylab="Hospitalizaciones")
x<-decompose(hospil)
plot(decompose(hospil))
real<-x$trend
plot(real, col="black", main="Tendencia de la serie", xlab="Años", ylab="Tendencia")
abline(v=2014.75,col="red")
abline(v=2016.5,col="blue")
```

```
#####
##                               Creación serie temporal hospitalizaciones                               ##
#####
```

```
hospitalizacionesRM
hospil<-ts(hospitalizacionesRM ,start = c(2013.75,2017), frequency = 52)
ts.plot(hospil, col="black", main="Hospitalizaciones causadas por accidentes de tránsito en la RM",
lwd=1, xlab="Años" , ylab="Hospitalizaciones")
```

```

x<-decompose(hospil)
plot(decompose(hospil))
real<-x$trend
plot(real, col="black", main="Tendencia de la serie", xlab="Años", ylab="Tendencia")
abline(v=2014.75,col="red")
abline(v=2016.5,col="blue")
#####
##                                     Test de Chow muertes                                     ##
#####

data<-muertes
attach(data)
modc<-lm(muerte~tiempo)
summary(modc)
gc<-modc$df.residual
rssc<-sum(residuals(modc)^2)
rssc
mod1<-lm(muerte[1:765]~tiempo[1:765])
summary(mod1)
g1<-mod1$df.residual
rss1<-sum(residuals(mod1)^2)
rss1
mod2<-lm(muerte[766:937]~tiempo[766:937])
summary(mod2)
g2<-mod2$df.residual
rss2<-sum(residuals(mod2)^2)
rss2
n=937
k=2
f=qf(.95,2,933)
f

```

```
chow=((rssc-(rss1+rss2))/(k-1))/((rss1+rss2)/(n-2*k))
chow
```

```
#####
##                               Test de Chow muertes en la RM          ##
#####
```

```
data<-muertesRM
attach(data)
modc<-lm(muerte~tiempo)
summary(modc)
gc<-modc$df.residual
rssc<-sum(residuals(modc)2)
rssc
mod1<-lm(muerte[1:401]~tiempo[1:401])
summary(mod1)
g1<-mod1$df.residual
rss1<-sum(residuals(mod1)2)
rss1
mod2<-lm(muerte[402:572]~tiempo[402:572])
98 summary(mod2)
g2<-mod2$df.residual
rss2<-sum(residuals(mod2)2)
rss2
n=572
k=2
f=qf(.95,2,568)
f
chow=((rssc-(rss1+rss2))/(k-1))/((rss1+rss2)/(n-2*k))
chow
```

```
#####
##                               Test de Chow muertes peatones          ##
#####
```

```
data<-peatones
attach(data)
modc<-lm(muerte~tiempo)
summary(modc)
gc<-modc$df.residual
rssc<-sum(residuals(modc)^2)
rssc
mod1<-lm(muerte[1:765]~tiempo[1:765])
summary(mod1)
g1<-mod1$df.residual
rss1<-sum(residuals(mod1)^2)
rss1
mod2<-lm(muerte[766:937]~tiempo[766:937])
summary(mod2)
g2<-mod2$df.residual
rss2<-sum(residuals(mod2)^2)
rss2
n=937
k=2
f=qf(.95,2,933)
f
chow=((rssc-(rss1+rss2))/(k-1))/((rss1+rss2)/(n-2*k))
chow
```

```
#####
##                               Test de Chow muertes ocupante de vehículo          ##
#####
```

```
data<-ocupante
attach(data)
modc<-lm(muerte~tiempo)
summary(modc)
gc<-modc$df.residual
rssc<-sum(residuals(modc)2)
rssc
mod1<-lm(muerte[1:765]~tiempo[1:765])
summary(mod1)
g1<-mod1$df.residual
rss1<-sum(residuals(mod1)2)
rss1
mod2<-lm(muerte[766:937]~tiempo[766:937])
summary(mod2)
g2<-mod2$df.residual
rss2<-sum(residuals(mod2)2)
rss2
n=937
k=2
f=qf(.95,2,933)
f
chow=((rssc-(rss1+rss2))/(k-1))/((rss1+rss2)/(n-2*k))
chow
```

```
#####
##                               Test de Chow hospitalizaciones                               ##
#####
```

```
data<-hospitalizaciones
attach(data)
modc<-lm(hospitalizaciones~tiempo)
summary(modc)
gc<-modc$df.residual
rssc<-sum(residuals(modc)2)
rssc
mod1<-lm(hospitalizaciones[1:52]~tiempo[1:52])
summary(mod1)
g1<-mod1$df.residual
rss1<-sum(residuals(mod1)2)
rss1
mod2<-lm(hospitalizaciones[53:223]~tiempo[53:223])
summary(mod2)
g2<-mod2$df.residual
rss2<-sum(residuals(mod2)2)
rss2
n=223
k=2
f=qf(.95,2,219)
f
chow=((rssc-(rss1+rss2))/(k-1))/((rss1+rss2)/(n-2*k))
chow
```

```
#####
##                               Test de Chow hospitalizaciones en la RM                               ##
#####
```

```
data<-hospitalizacionesRM
attach(data)
modc<-lm(hospitalizaciones~tiempo)
summary(modc)
gc<-modc$df.residual
rssc<-sum(residuals(modc)2)
rssc
mod1<-lm(hospitalizaciones[1:52]~tiempo[1:52])
summary(mod1)
g1<-mod1$df.residual
rss1<-sum(residuals(mod1)2)
rss1
mod2<-lm(hospitalizaciones[53:223]~tiempo[53:223])
summary(mod2)
g2<-mod2$df.residual
rss2<-sum(residuals(mod2)2)
rss2
n=223
k=2
f=qf(.95,2,219)
f
chow=((rssc-(rss1+rss2))/(k-1))/((rss1+rss2)/(n-2*k))
chow
```

```
#####
##                               ##
#####
```

```
data<-muertes
summary(data)
Análisis descriptivo3 mod<-glm(Muerte~Tiempo+Tiempo2, data, family = "poisson")
summary(mod)
xtable(summary(mod))
d<-mod$deviance
d/933
```

```
mod1<-glm.nb(Muerte Tiempo+ Tiempo2, data)
summary(mod1)
xtable(summary(mod1))
de<-mod1$deviance
de/933
Supuestos del Modelo AIC(mod1)
scr<-sum(resid(mod1)2)
scr
Box.test(mod1$residuals, type = "Ljung" , lag=50)
bptest(mod1)
```

```
mod2<-glm.nb(Muerte~Tiempo+Tiempo2+harmonic(Tiempo,1,52),data)
summary(mod2)
xtable(summary(mod2))
Supuestos del Modelo AIC(mod2)
scr2<-sum(resid(mod2)2)
scr2
Box.test(mod2$residuals, type = "Ljung" , lag=50)
bptest(mod2)
```

```
mod3<-glm.nb(Muerte~Tiempo+Tiempo2+ Intervención+harmonic(Tiempo,1,52),data)
summary(mod3)
xtable(summary(mod3))
Supuestos del Modelo AIC(mod3)
scr2<-sum(resid(mod3)2)
scr2
Box.test(mod3$residuals, type =“Ljung” , lag=50)
bptest(mod3)
```

```
mod4<-glm.nb(Muerte~Tiempo+Tiempo2+ Intervención+Intervención2+
harmonic(Tiempo,1,52),data)
summary(mod4)
xtable(summary(mod4))
Supuestos del Modelo AIC(mod4)
scr2<-sum(resid(mod4)2)
scr2
Box.test(mod4$residuals, type =“Ljung” , lag=50)
bptest(mod4)
```

```
mod5<-glm.nb(Muerte~Tiempo* Intervención+Tiempo2+Intervención2+
harmonic(Tiempo,1,52),data)
summary(mod5)
xtable(summary(mod5))
Supuestos del Modelo AIC(mod5)
scr2<-sum(resid(mod5)2)
scr2
Box.test(mod5$residuals, type =“Ljung” , lag=50)
bptest(mod5)
```

```
mod6<-glm.nb(Muerte~Tiempo+ Tiempo2+Intervención2+Intervención*
harmonic(Tiempo,1,52),data)
```

```

summary(mod6)
xtable(summary(mod5))
Supuestos del Modelo AIC(mod6)
scr2<-sum(resid(mod6)2)
scr2
Box.test(mod6$residuals, type =“Ljung” , lag=50)
bptest(mod6)

#####
##                               Modelos ITS muertes en la RM                               ##
#####

data<-muertesRM
summary(data)
Análisis descriptivo3 mod<-glm(Muerte~Tiempo, data, family = “poisson”)
summary(mod)
xtable(summary(mod))
d<-mod$deviance
d/568

mod1<-glm.nb(Muerte Tiempo, data)
summary(mod1)
xtable(summary(mod1))
de<-mod1$deviance
de/568
Supuestos del Modelo AIC(mod1)
scr<-sum(resid(mod1)2)
scr
Box.test(mod1$residuals, type =“Ljung” , lag=50)
bptest(mod1)

```

```
mod2<-glm.nb(Muerte~Tiempo+harmonic(Tiempo,1,52),data)
summary(mod2)
xtable(summary(mod2))
Supuestos del Modelo AIC(mod2)
scr2<-sum(resid(mod2)2)
scr2
Box.test(mod2$residuals, type =“Ljung” , lag=50)
bptest(mod2)
```

```
mod3<-glm.nb(Muerte~Tiempo+ Intervención+harmonic(Tiempo,1,52),data)
summary(mod3)
xtable(summary(mod3))
Supuestos del Modelo AIC(mod3)
scr2<-sum(resid(mod3)2)
scr2
Box.test(mod3$residuals, type =“Ljung” , lag=50)
bptest(mod3)
```

```
mod4<-glm.nb(Muerte~Tiempo+ Intervención+Intervención2+harmonic(Tiempo,1,52),data)
summary(mod4)
xtable(summary(mod4))
Supuestos del Modelo AIC(mod4)
scr2<-sum(resid(mod4)2)
scr2
Box.test(mod4$residuals, type =“Ljung” , lag=50)
bptest(mod4)
```

```
mod5<-glm.nb(Muerte~Tiempo* Intervención+harmonic(Tiempo,1,52),data)
summary(mod5)
xtable(summary(mod5))
Supuestos del Modelo AIC(mod5)
scr2<-sum(resid(mod5)2)
```

```

scr2
Box.test(mod5$residuals, type = "Ljung" , lag=50)
bptest(mod5)

mod6<-glm.nb(Muerte~Tiempo+ Intervención*harmonic(Tiempo,1,52),data)
summary(mod6)
xtable(summary(mod5))
Supuestos del Modelo AIC(mod6)
scr2<-sum(resid(mod6)2)
scr2
Box.test(mod6$residuals, type = "Ljung" , lag=50)
bptest(mod6)
#####
##                               Modelos ITS muertes peatones                               ##
#####

data<-peatones
summary(data)
Análisis descriptivo3 mod<-glm(Muerte~Tiempo+ Tiempo2, data, family = "poisson")
summary(mod)
xtable(summary(mod))
d<-mod$deviance
d/933

mod1<-glm.nb(Muerte Tiempo+ Tiempo2, data)
summary(mod1)
xtable(summary(mod1))
de<-mod1$deviance
de/933
Supuestos del Modelo AIC(mod1)
scr<-sum(resid(mod1)2)

```

```
scr
Box.test(mod1$residuals, type = "Ljung" , lag=50)
bptest(mod1)

mod2<-glm.nb(Muerte~Tiempo+ Tiempo2+harmonic(Tiempo,1,52),data)
summary(mod2)
xtable(summary(mod2))
Supuestos del Modelo AIC(mod2)
scr2<-sum(resid(mod2)2)
scr2
Box.test(mod2$residuals, type = "Ljung" , lag=50)
bptest(mod2)

mod3<-glm.nb(Muerte~Tiempo+ Tiempo2+ Intervención+harmonic(Tiempo,1,52),data)
summary(mod3)
xtable(summary(mod3))
Supuestos del Modelo AIC(mod3)
scr2<-sum(resid(mod3)2)
scr2
Box.test(mod3$residuals, type = "Ljung" , lag=50)
bptest(mod3)

mod4<-glm.nb(Muerte~Tiempo+ Tiempo2+ Intervención+Intervención2+
Autoregresiva+harmonic(Tiempo,1,52),data)
summary(mod4)
xtable(summary(mod4))
Supuestos del Modelo AIC(mod4)
scr2<-sum(resid(mod4)2)
scr2
Box.test(mod4$residuals, type = "Ljung" , lag=50)
bptest(mod4)
```

```

mod5<-glm.nb(Muerte~Tiempo* Intervención+ Tiempo2+harmonic(Tiempo,1,52),data)
summary(mod5)
xtable(summary(mod5))
Supuestos del Modelo AIC(mod5)
scr2<-sum(resid(mod5)2)
scr2
Box.test(mod5$residuals, type =“Ljung” , lag=50)
bptest(mod5)

```

```

mod6<-glm.nb(Muerte~Tiempo+ Tiempo2+ Intervención*harmonic(Tiempo,1,52),data)
summary(mod6)
xtable(summary(mod5))
Supuestos del Modelo AIC(mod6)
scr2<-sum(resid(mod6)2)
scr2
Box.test(mod6$residuals, type =“Ljung” , lag=50)
bptest(mod6)

```

```

#####
##                               Modelos ITS muertes ocupante de vehículo                               ##
#####

```

```

data<-ocupante
summary(data)
Análisis descriptivo3 mod<-glm(Muerte~Tiempo+ Tiempo2, data, family = “poisson”)
summary(mod)
xtable(summary(mod))
d<-mod$deviance
d/933

```

```

mod1<-glm.nb(Muerte Tiempo+ Tiempo2, data)

```

```
summary(mod1)
xtable(summary(mod1))
de<-mod1$deviance
de/933
Supuestos del Modelo AIC(mod1)
scr<-sum(resid(mod1)2)
scr
Box.test(mod1$residuals, type = "Ljung" , lag=50)
bptest(mod1)

mod2<-glm.nb(Muerte~Tiempo+ Tiempo2+harmonic(Tiempo,1,52),data)
summary(mod2)
xtable(summary(mod2))
Supuestos del Modelo AIC(mod2)
scr2<-sum(resid(mod2)2)
scr2
Box.test(mod2$residuals, type = "Ljung" , lag=50)
bptest(mod2)

mod3<-glm.nb(Muerte~Tiempo+ Tiempo2+ Intervención+Intervención2+
harmonic(Tiempo,1,52),data)
summary(mod3)
xtable(summary(mod3))
Supuestos del Modelo AIC(mod3)
scr2<-sum(resid(mod3)2)
scr2
Box.test(mod3$residuals, type = "Ljung" , lag=50)
bptest(mod3)

mod4<-glm.nb(Muerte~Tiempo+ Tiempo2+ Intervención+Intervención2+Autoregresiva+
harmonic(Tiempo,1,52),data)
summary(mod4)
```

```

xtable(summary(mod4))
Supuestos del Modelo AIC(mod4)
scr2<-sum(resid(mod4)^2)
scr2
Box.test(mod4$residuals, type =“Ljung” , lag=50)
bptest(mod4)

```

```

mod5<-glm.nb(Muerte~Tiempo* Intervención+ Tiempo^2+ Intervención2+
harmonic(Tiempo,1,52),data)
summary(mod5)
xtable(summary(mod5))
Supuestos del Modelo AIC(mod5)
scr2<-sum(resid(mod5)^2)
scr2
Box.test(mod5$residuals, type =“Ljung” , lag=50)
bptest(mod5)

```

```

mod6<-glm.nb(Muerte~Tiempo+ Tiempo^2+ Intervención2+ Intervención*
harmonic(Tiempo,1,52),data)
summary(mod6)
xtable(summary(mod6))
Supuestos del Modelo AIC(mod6)
scr2<-sum(resid(mod6)^2)
scr2
Box.test(mod6$residuals, type =“Ljung” , lag=50)
bptest(mod6)

```

```

#####
##                               ##
#####

```

```

data<-hospitalizaciones

```

```
summary(data)
```

```
Análisis descriptivo mod<-glm(Hospitalizaciones~Tiempo, data, family = "poisson")
```

```
summary(mod)
```

```
xtable(summary(mod))
```

```
d<-mod$deviance
```

```
d/219
```

```
mod1<-glm.nb(Hospitalizaciones ~Tiempo, data)
```

```
summary(mod1)
```

```
xtable(summary(mod1))
```

```
de<-mod1$deviance
```

```
de/219
```

```
Supuestos del Modelo AIC(mod1)
```

```
scr<-sum(resid(mod1)2)
```

```
scr
```

```
Box.test(mod1$residuals, type = "Ljung" , lag=50)
```

```
bptest(mod1)
```

```
mod2<-glm.nb(Hospitalizaciones~Tiempo+harmonic(Tiempo,1,52),data)
```

```
summary(mod2)
```

```
xtable(summary(mod2))
```

```
Supuestos del Modelo AIC(mod2)
```

```
scr2<-sum(resid(mod2)2)
```

```
scr2
```

```
Box.test(mod2$residuals, type = "Ljung" , lag=50)
```

```
bptest(mod2)
```

```
mod3<-glm.nb(Hospitalizaciones~Tiempo+ Intervención+harmonic(Tiempo,1,52),data)
```

```
summary(mod3)
```

```
xtable(summary(mod3))
```

```
Supuestos del Modelo AIC(mod3)
```

```
scr2<-sum(resid(mod3)2)
```

```
scr2
Box.test(mod3$residuals, type = "Ljung" , lag=50)
bptest(mod3)

mod4<-glm.nb(Hospitalizaciones~Tiempo+ Intervención+Autoregresiva+
harmonic(Tiempo,1,52),data)
summary(mod4)
xtable(summary(mod4))
Supuestos del Modelo AIC(mod4)
scr2<-sum(resid(mod4)^2)
scr2
Box.test(mod4$residuals, type = "Ljung" , lag=50)
bptest(mod4)

mod5<-glm.nb(Hospitalizaciones~Tiempo* Intervención+Autoregresiva+
harmonic(Tiempo,1,52),data)
summary(mod5)
xtable(summary(mod5))
Supuestos del Modelo AIC(mod5)
scr2<-sum(resid(mod5)^2)
scr2
Box.test(mod5$residuals, type = "Ljung" , lag=50)
bptest(mod5)

mod6<-glm.nb(Hospitalizaciones~Tiempo+ Autoregresiva+
Intervención*harmonic(Tiempo,1,52),data)
summary(mod6)
xtable(summary(mod6))
Supuestos del Modelo AIC(mod6)
scr2<-sum(resid(mod6)^2)
scr2
Box.test(mod6$residuals, type = "Ljung" , lag=50)
```

```
bptest(mod6)
```

```
#####
##                               Modelos ITS hospitalizaciones en la RM                               ##
#####
```

```
data<-hospitalizacionesRM
```

```
summary(data)
```

```
Análisis descriptivo mod<-glm(Hospitalizaciones~Tiempo, data, family = "poisson")
```

```
summary(mod)
```

```
xtable(summary(mod))
```

```
d<-mod$deviance
```

```
d/219
```

```
mod1<-glm.nb(Hospitalizaciones Tiempo, data)
```

```
summary(mod1)
```

```
xtable(summary(mod1))
```

```
de<-mod1$deviance
```

```
de/219
```

```
Supuestos del Modelo AIC(mod1)
```

```
scr<-sum(resid(mod1)2)
```

```
scr
```

```
Box.test(mod1$residuals, type = "Ljung" , lag=50)
```

```
bptest(mod1)
```

```
mod2<-glm.nb(Hospitalizaciones~Tiempo+harmonic(Tiempo,1,52),data)
```

```
summary(mod2)
```

```
xtable(summary(mod2))
```

```
Supuestos del Modelo AIC(mod2)
```

```
scr2<-sum(resid(mod2)2)
```

```
scr2
Box.test(mod2$residuals, type = "Ljung" , lag=50)
bptest(mod2)

mod3<-glm.nb(Hospitalizaciones~Tiempo+ Tiempo2+ Intervención+harmonic(Tiempo,1,52),data)
summary(mod3)
xtable(summary(mod3))
Supuestos del Modelo AIC(mod3)
scr2<-sum(resid(mod3)2)
scr2
Box.test(mod3$residuals, type = "Ljung" , lag=50)
bptest(mod3)

mod4<-glm.nb(Hospitalizaciones~Tiempo+ Intervención+Autoregresiva+
harmonic(Tiempo,1,52),data)
summary(mod4)
xtable(summary(mod4))
Supuestos del Modelo AIC(mod4)
scr2<-sum(resid(mod4)2)
scr2
Box.test(mod4$residuals, type = "Ljung" , lag=50)
bptest(mod4)

mod5<-glm.nb(Hospitalizaciones~Tiempo* Intervención+Autoregresiva+Tiempo2+
harmonic(Tiempo,1,52),data)
summary(mod5)
xtable(summary(mod5))
Supuestos del Modelo AIC(mod5)
scr2<-sum(resid(mod5)2)
scr2
Box.test(mod5$residuals, type = "Ljung" , lag=50)
bptest(mod5)
```

```

mod6<-glm.nb(Hospitalizaciones~Tiempo+ Tiempo2+ Autoregresiva+Intervención*
harmonic(Tiempo,1,52),data)
summary(mod6)
xtable(summary(mod6))
Supuestos del Modelo AIC(mod6)
scr2<-sum(resid(mod6)2)
scr2
Box.test(mod6$residuals, type =“Ljung” , lag=50)
bptest(mod6)

```

```

#####
##                               ##
#####

```

```

mod5<-glm.nb( Muerte~ Tiempo * Intervención + Tiempo2 + Intervención2 + harmonic (Tiem-
po,1,52))
x1=data.frame(Tiempo=data$Tiempo,Tiempo2=data $Tiempo2,
Intervención2=data$Intervención2,Intervención=data$Intervención,
harmonic=harmonic(data$Tiempo,1,52))
mod5<-glm.nb(Muerte Tiempo*Intervención+Tiempo2+Intervención2+ harmonic.1 + harmonic.2,
x1)
ts.plot(predict(mod5,type=“response”), ylab=“Predicciones”, xlab=“Años”, main=“Predicciones y
tendencia estimada”)
x2=data.frame(Tiempo,Intervención,Intervención2,Tiempo2,harmonic.1=0,harmonic.2=0)
lines(Tiempo,predict(mod5,x2,type=“response”),col=“red”)
abline(v=767, col=“blue”)
abline(v=834, col=“blue”)

```

```
#####
##                               ##
Predicciones y tendencia muertes en la RM
#####
```

```
mod3<-glm.nb( Muerte~ Tiempo + Intervención + harmonic (Tiempo,1,52))
x1=data.frame(Tiempo=data$Tiempo,Intervención=data$Intervención,
harmonic=harmonic(data$Tiempo,1,52))
mod3<-glm.nb(Muerte Tiempo+Intervención+ harmonic.1 + harmonic.2, x1)
ts.plot(predict(mod5,type="response"), ylab="Predicciones", xlab="Años", main="Predicciones y
tendencia estimada")
x2=data.frame(Tiempo,Intervención,Intervención2,Tiempo2,harmonic.1=0,harmonic.2=0)
lines(Tiempo,predict(mod5,x2,type="response"),col="red")
abline(v=453, col="blue")
```

```
#####
##                               ##
Predicciones y tendencia peatones
#####
```

```
mod4<-glm.nb( Muerte~ Tiempo+ Tiempo2+ Intervención + harmonic (Tiempo,1,52))
x1=data.frame(Tiempo=data$Tiempo,Tiempo2=data $Tiempo2,Intervención=data$Intervención,
harmonic=harmonic(data$Tiempo,1,52))
mod4<-glm.nb(Muerte Tiempo+Tiempo2+Intervención+ harmonic.1 + harmonic.2, x1)
ts.plot(predict(mod5,type="response"), ylab="Predicciones", xlab="Años", main="Predicciones y
tendencia estimada")
x2=data.frame(Tiempo,Intervención,Intervención2,Tiempo2,harmonic.1=0,harmonic.2=0)
lines(Tiempo,predict(mod5,x2,type="response"),col="red")
abline(v=767, col="blue")
```

```
#####
##                               ##
#####
```

```
mod4<-glm.nb( Muerte~ Tiempo+ Tiempo2+ Intervención+ Intervención2 + harmonic (Tiempo,1,52))
x1=data.frame(Tiempo=data$Tiempo,Tiempo2=data $Tiempo2,Intervención=data$Intervención,
Intervención2=data$Intervención2,harmonic=harmonic(data$Tiempo,1,52))
mod4<-glm.nb(Muerte~Tiempo+Tiempo2+Intervención+Intervención2+ harmonic.1 + harmonic.2,
x1)
ts.plot(predict(mod5,type="response"), ylab="Predicciones", xlab="Años", main="Predicciones y
tendencia estimada")
x2=data.frame(Tiempo,Intervención,Intervención2,Tiempo2,harmonic.1=0,harmonic.2=0)
lines(Tiempo,predict(mod5,x2,type="response"),col="red")
abline(v=767, col="blue")
abline(v=834, col="blue")
```

```
#####
##                               ##
#####
```

```
mod5<-glm.nb( Hospitalizaciones~ Tiempo+ Intervención+ Autoregresiva + harmonic (Tiempo,1,52))
x1=data.frame(Tiempo=data$Tiempo,Intervención=data$Intervención,
Autoregresiva=data$Autoregresiva,harmonic=harmonic(data$Tiempo,1,52))
mod5<-glm.nb(Hospitalizaciones~Tiempo+Intervención+Autoregresiva+harmonic.1 + harmonic.2,
x1)
ts.plot(predict(mod5,type="response"), ylab="Predicciones", xlab="Años", main="Predicciones y
tendencia estimada")
x2=data.frame(Tiempo,Intervención,Intervención2,Tiempo2,harmonic.1=0,harmonic.2=0)
lines(Tiempo,predict(mod5,x2,type="response"),col="red")
abline(v=767, col="blue")
```

```
#####
##                               ##
#####
```

```
mod5<-glm.nb( Hospitalizaciones~ Tiempo+ Intervención+ Autoregresiva + harmonic (Tiempo,1,52))
x1=data.frame(Tiempo=data$Tiempo,Intervención=data$Intervención,
Autoregresiva=data$Autoregresiva,harmonic=harmonic(data$Tiempo,1,52))
mod5<-glm.nb(Hospitalizaciones~Tiempo+Intervención+Autoregresiva+harmonic.1 + harmonic.2,
x1)
ts.plot(predict(mod5,type="response"), ylab="Predicciones", xlab="Años", main="Predicciones y
tendencia estimada")
x2=data.frame(Tiempo,Intervención,Intervención2,Tiempo2,harmonic.1=0,harmonic.2=0)
lines(Tiempo,predict(mod5,x2,type="response"),col="red")
abline(v=52, col="blue")
```

```
#####
##                               ##
#####
```

```
mod1<-glm.nb(Muertes~Tiempo + Tiempo2+ harmonic(Tiempo,1,52))
mod2<-glm.nb(Muertes~Tiempo*Intervención + Tiempo2+ Intervención2+ harmonic(Tiempo,1,52))
fitsin=mod1$fitted.value
fitcon=mod2$fitted.value
write.csv(fitsin ,file = C:"/Users/darli/Desktop/fitsinmuertes.csv")
write.csv(fitcon ,file = C:"/Users/darli/Desktop/fitconmuertes.csv")
plot(muertes, type = "l" , pch=1, col="black", xlim= c(2000,2017),xlab = "Años", main = "Compara-
ción estimaciones modelo sin intervención versus con intervención ")
lines(muertes, type="l", col="red", pch=2, lwd=2)
abline(v=767, col="blue")
```

```
legend("topright", legend =c("Sin intervención", "Con intervención") , pch=c(1,2), col=c("black",
"red"), adj = c(0,0.5))
```

```
#####
##                               Evaluación del Impacto de la ley en muertes en la RM                               ##
#####
```

```
mod1<-glm.nb(Muertes~Tiempo + harmonic(Tiempo,1,52))
mod2<-glm.nb(Muertes~Tiempo+Intervención + harmonic(Tiempo,1,52))
fitsin=mod1$fitted.value
fitcon=mod2$fitted.value
write.csv(fitsin ,file = C:="/Users/darli/Desktop/fitsinmuertesRM.csv")
write.csv(fitcon ,file = C:="/Users/darli/Desktop/fitconmuertesRM.csv")
plot(muertesRM, type = "l" , pch=1, col="black", xlim= c(2000,2017),xlab = "Años", main = "Com-
paración estimaciones modelo sin intervención versus con intervención ")
lines(muertesRM, type="l", col="red", pch=2, lwd=2)
abline(v=767, col="blue")
legend("topright", legend =c("Sin intervención", "Con intervención") , pch=c(1,2), col=c("black",
"red"), adj = c(0,0.5))
```

```
#####
##                               Evaluación del Impacto de la ley en muertes peatones versus ocupante de vehículo                               ##
#####
```

```
mod1<-glm.nb(Muertes~Tiempo +Tiempo2+ Intervención harmonic(Tiempo,1,52))
mod2<-glm.nb(Muertes~Tiempo+ Tiempo2+Intervención + Intervención2+ harmonic(Tiempo,1,52))
fitpea=mod1$fitted.value
fitocu=mod2$fitted.value
write.csv(fitpea ,file = C:="/Users/darli/Desktop/fitpeatones.csv")
```

```
write.csv(fitocu ,file = C:“/Users/darli/Desktop/fitocupantes.csv”)
plot(peatonos, type =“l” , pch=1, col=“black”, xlim= c(2000,2017),xlab =“Años”, main =“Compa-
ración estimaciones modelo peatonos versus ocupante de vehículo”)
lines(ocupante, type=“l”, col=“red”, pch=2, lwd=2)
abline(v=767, col=“blue”)
legend(“topright”, legend =c(“Peatonos”, “Ocupantes”) , pch=c(1,2), col=c(“black”, “red”), adj =
c(0,0.5))
```

```
#####
##                               Evaluación del Impacto de la ley en hospitalizaciones                               ##
#####
```

```
mod1<-glm.nb(Hospitalizaciones~Tiempo+ Autoregresiva+ harmonic(Tiempo,1,52))
mod2<-glm.nb(Hospitalizaciones~Tiempo +Intervención + Autoregresiva+ harmonic(Tiempo,1,52))
fitsin=mod1$fitted.value
fitcon=mod2$fitted.value
write.csv(fitsin,file = C:“/Users/darli/Desktop/fithospisin.csv”)
write.csv(fitcon ,file = C:“/Users/darli/Desktop/fithospicon.csv”)
plot(hospisin, type =“l” , pch=1, col=“black”, xlim= c(2013.75,2017),xlab =“Años”, main =“Com-
paración estimaciones hospitalizaciones sin intervención versus con intervención”)
lines(hospicon, type=“l”, col=“red”, pch=2, lwd=2)
abline(v=767, col=“blue”)
legend(“topright”, legend =c(“Sin intervención”, “Con intervención”) , pch=c(1,2), col=c(“black”,
“red”), adj = c(0,0.5))
```

```
#####
##                               Evaluación del Impacto de la ley en hospitalizaciones en la RM                               ##
#####
```

```
mod1<-glm.nb(Hospitalizaciones~Tiempo+ Autoregresiva+ harmonic(Tiempo,1,52))
mod2<-glm.nb(Hospitalizaciones~Tiempo+ Intervención + Autoregresiva+ harmonic(Tiempo,1,52))
fitsin=mod1$fitted.value
fitcon=mod2$fitted.value
write.csv(fitsin,file = C:“/Users/darli/Desktop/fithospisin.csv”)
write.csv(fitcon ,file = C:“/Users/darli/Desktop/fithospicon.csv”)
plot(hospisin, type =“l” , pch=1, col=“black”, xlim= c(2013.75,2017),xlab =“Años”, main =“Com-
paración estimaciones hospitalizaciones sin intervención versus con intervención”)
lines(hospicon, type=“l”, col=“red”, pch=2, lwd=2)
abline(v=767, col=“blue”)
legend(“topright”, legend =c(“Sin intervención”, “Con intervención”) , pch=c(1,2), col=c(“black”,
“red”), adj = c(0,0.5))
```