



**UNIVERSIDAD DE LA FRONTERA
FACULTAD DE INGENIERÍA Y CIENCIAS
DEPARTAMENTO DE INGENIERÍA DE OBRAS CIVILES**

“Estimación de modelos de predicción de tránsito diario de bicicletas en base a datos pasivos, incorporando variables meteorológicas y contextuales para la ciudad de Temuco”

FRANCISCO JOSÉ SAN MARTÍN KLOCK

2019



**UNIVERSIDAD DE LA FRONTERA
FACULTAD DE INGENIERÍA Y CIENCIAS
DEPARTAMENTO DE INGENIERÍA DE OBRAS CIVILES**

“Estimación de modelos de predicción de tránsito diario de bicicletas en base a datos pasivos, incorporando variables meteorológicas y contextuales para la ciudad de Temuco”

**TRABAJO PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL**

PROFESOR GUÍA : MSc. MAXIMILIANO EXEQUIEL LIZANA MALDONADO

FRANCISCO JOSÉ SAN MARTÍN KLOCK

2019

**ESTIMACIÓN DE MODELOS DE PREDICCIÓN DE TRÁNSITO DIARIO
DE BICICLETAS EN BASE A DATOS PASIVOS, INCORPORANDO
VARIABLES METEOROLÓGICAS Y CONTEXTUALES PARA LA
CIUDAD DE TEMUCO**

FRANCISCO JOSÉ SAN MARTÍN KLOCK

COMISIÓN EXAMINADORA

MSc. MAXIMILIANO EXEQUIEL LIZANA MALDONADO

Profesor Guía

Juan Pablo Torres Álvarez

Académico Evaluador

Mauricio Zambrano Bigiarini

Académico Evaluador

Calificación trabajo escrito :
Calificación examen :
Calificación final :

A mi familia.

AGRADECIMIENTOS

En primer lugar, agradezco a mi familia por todo el cariño y amor que me han entregado siempre. Mi familia es la parte más importante de mi vida. Siempre me han acompañado en todo y no existen palabras para expresar todo mi agradecimiento por otorgarme la oportunidad de estudiar una carrera universitaria. Me dirijo en particular a mi padre, Marco, por enseñarme importantes valores para la vida tales como la responsabilidad, el valor del trabajo y el esfuerzo. A mi madre, Victoria, por su incondicional apoyo, amor y cariño. A mi hermana, Consuelo, por su compañía y amor. A Claudia por su cariño durante toda mi vida. A mis abuelos Roberto y Lilian quienes siempre han brindado todo su amor a sus nietos.

Agradezco a mi polola Paulina por acompañarme en todo durante este año, por las incontables instancias en que nos juntamos a avanzar en nuestros trabajos y todo el amor y cariño que me ha demostrado. Un saludo especial para mi amor.

Agradezco a mi profesor Maximiliano Lizana por otorgarme la oportunidad de trabajar con él, confiando en mis capacidades y por su constante preocupación. Su apoyo fue muy importante para elaborar esta investigación y le doy las gracias por el tiempo y la disposición para enseñar el gran mundo de la Ingeniería de Transportes. Desde que dictó las asignaturas de Ingeniería de Transportes e Ingeniería Vial, y ahora como profesor guía, ha sido lejos uno de los mejores docentes que he tenido en mi vida, sobre todo por su calidad humana y capacidad de enseñar.

Finalmente, quiero agradecer a todas aquellas personas y amigos que me han enseñado algo en la vida, brindándome la oportunidad de estar en el lugar que hoy me encuentro.

RESUMEN

El presente trabajo de investigación evalúa el efecto de variables meteorológicas y contextuales sobre el volumen de tránsito diario de bicicletas en ciclovías de la ciudad de Temuco. Para la elaboración de este estudio se han utilizado tres años de datos pasivos que comprenden el período entre el 31/07/2016 y el 31/07/2019. Dichos datos se han obtenido de contadores automáticos instalados por el Ministerio de Vivienda y Urbanismo (MINVU). Los datos referidos a variables meteorológicas y contextuales se han recolectado de diversas fuentes, destacándose la Dirección Meteorológica de Chile (DMC), la Dirección General de Aguas (DGA) y el Servicio de Impuestos Internos (SII). Se seleccionaron una serie de variables potencialmente explicativas sobre la variable de respuesta llamada conteos diarios y se realizó un análisis descriptivo de ellas, junto con una evaluación de la efectividad de los contadores automáticos. Posteriormente, se desarrollaron cuatro modelos de predicción para puntos geográficos específicos de la ciudad. Los modelos implementados, utilizan la regresión lineal múltiple, el modelo Poisson, el modelo Binomial Negativo y el algoritmo de *Machine Learning* de Bosques Aleatorios o *Random Forests*. Los resultados confirman la significancia de variables que tienen un efecto positivo sobre el número de pasadas diarias de bicicleta, como la temperatura máxima, radiación, longitud de la ciclovía, entropía en el uso de suelo, y variables que poseen un efecto negativo, como la temperatura máxima al cuadrado, precipitación y humedad. A su vez, se observó la tendencia general al alza que tiene este modo de transporte en la ciudad.

Debido a la naturaleza anidada de los datos tratados, se desarrolló un modelo multinivel de predicción que incluye todos los contadores automáticos en un modelo global. En dicho modelo, se han incluido variables de tipo contextual asociadas al uso de suelo y la longitud de la ciclovía. Estas características del entorno construido permiten explicar el uso de la bicicleta en diferentes puntos geográficos de la ciudad, a pesar de que las condiciones meteorológicas sean las mismas para un día particular.

Finalmente, los resultados del modelo multinivel pueden ser utilizados para focalizar decisiones de inversión en la ciudad de Temuco, ya que a través de esta herramienta se puede caracterizar contextualmente un punto geográfico de la ciudad, y al evaluarlo en el modelo, estimar la cantidad de ciclistas diarios que pasarían por un punto ficticio de la urbe.

ÍNDICE DE CONTENIDOS

Capítulo 1. Introducción.....	1
1.1. Descripción del problema	4
1.2. Objetivos.....	5
1.2.1. Objetivo general	5
1.2.2. Objetivos específicos.....	5
Capítulo 2. Antecedentes Generales.....	6
2.1. Relación entre meteorología y uso de la bicicleta	6
2.2. Relación entre variables contextuales y uso de la bicicleta	10
2.3. Contadores automáticos de bicicletas	11
Capítulo 3. Fundamentos Teóricos.....	14
3.1. Modelos lineales generalizados	14
3.2. Bosques aleatorios (<i>Random Forests</i>)	19
3.3. Modelos multinivel	21
Capítulo 4. Metodología.....	23
4.1. Datos	23
4.2. Selección de variables	25
4.3. Implementación de modelos.....	27
Capítulo 5. Análisis Descriptivo	29
5.1. Información general.....	29
5.2. Análisis de datos meteorológicos	32
5.3. Análisis de datos pasivos	35
5.3.1. Crecimiento de la demanda	35
5.3.2. Variabilidad en días de fin de semana	39
5.3.3. Variabilidad horaria.....	40
Capítulo 6. Resultados y Análisis	42
6.1. Evaluación de la efectividad de los contadores automáticos.....	42
6.2. Modelos de predicción puntuales	44
6.2.1. Modelo puntual Ex Vía Férrea 1	44
6.2.2. Modelo puntual Los Músicos – Los Creadores	51
6.2.3. Modelo puntual Pablo Neruda – Olimpia.....	57
6.2.4. Modelo puntual San Martín - Mackenna.....	62
6.3. Análisis de modelos puntuales.....	67
6.4. Modelo de predicción multinivel	71
6.5. Validación de los modelos.....	78

Capítulo 7. Conclusiones.....	79
7.1. Futuras aplicaciones	81
Bibliografía.....	82
Anexo A. Figuras.....	88
Anexo B. Derivación modelos lineales generalizados	135
Anexo C. Índices de segregación residencial	142

ÍNDICE DE TABLAS

Tabla 2.1. Resumen de trabajos relacionados.....	9
Tabla 3.1. Resumen de distribuciones exponenciales (Alcaide, 2015).	16
Tabla 3.2. Funciones enlace canónicas de las distribuciones de familia exponencial (Alcaide, 2015).	17
Tabla 3.3. Tipos de modelos lineales generalizados de la familia exponencial (Alcaide, 2015).	18
Tabla 4.1. Selección de variables para la implementación de modelos de predicción.	27
Tabla 5.1. Coordenadas geográficas de los contadores automáticos de Temuco.	30
Tabla 6.1. Efectividad de contadores automáticos de bicicleta en Temuco.	42
Tabla 6.2. Resumen base de datos contador Ex Vía Férrea 1.	44
Tabla 6.3. Modelo regresión lineal múltiple contador Ex Vía Férrea 1.	45
Tabla 6.4. Modelo Poisson contador Ex Vía Férrea 1.	47
Tabla 6.5. Modelo binomial negativo contador Ex Vía Férrea 1.	49
Tabla 6.6. Modelo <i>Random Forests</i> contador Ex Vía Férrea 1.	50
Tabla 6.7. Resumen de modelos desarrollados contador Ex Vía Férrea 1.	51
Tabla 6.8. Resumen base de datos contador Los Músicos – Los Creadores.	52
Tabla 6.9. Modelo regresión lineal múltiple contador Los Músicos – Los Creadores.	52
Tabla 6.10. Modelo Poisson contador Los Músicos – Los Creadores.	54
Tabla 6.11. Modelo binomial negativo contador Los Músicos – Los Creadores.	55
Tabla 6.12. Modelo <i>Random Forests</i> contador Los Músicos – Los Creadores.	56
Tabla 6.13. Resumen de modelos desarrollados contador Los Músicos – Los Creadores.	57
Tabla 6.14. Resumen base de datos contador Pablo Neruda – Olimpia.	57
Tabla 6.15. Modelo regresión lineal múltiple contador Pablo Neruda – Olimpia.	58
Tabla 6.16. Modelo Poisson contador Pablo Neruda – Olimpia.	59
Tabla 6.17. Modelo binomial negativo contador Pablo Neruda – Olimpia.	60
Tabla 6.18. Modelo <i>Random Forests</i> contador Pablo Neruda – Olimpia.	61
Tabla 6.19. Resumen de modelos desarrollados contador Pablo Neruda – Olimpia.	62
Tabla 6.20. Resumen base de datos contador San Martín – Mackenna.	63
Tabla 6.21. Modelo regresión lineal múltiple contador San Martín – Mackenna.	63
Tabla 6.22. Modelo Poisson contador San Martín – Mackenna.	64
Tabla 6.23. Modelo binomial negativo contador San Martín – Mackenna.	65
Tabla 6.24. Modelo <i>Random Forests</i> contador San Martín – Mackenna.	66
Tabla 6.25. Resumen modelos desarrollados contador San Martín – Mackenna.	67
Tabla 6.26. Importancia de las variables en los modelos <i>Random Forests</i>	71
Tabla 6.27. Modelo regresión lineal múltiple contadores Temuco.	72
Tabla 6.28. Modelo nulo solo basado en el intercepto.	72
Tabla 6.29. Efectos fijos del modelo de interceptos aleatorios.	73
Tabla 6.30. Efectos aleatorios en el modelo de interceptos aleatorios.	73
Tabla 6.31. Valores de los interceptos aleatorios.	73
Tabla 6.32. Efectos fijos de las variables de nivel uno.	74
Tabla 6.33. Efectos aleatorios del modelo multinivel con nivel uno.	75
Tabla 6.34. Efectos fijos de las variables de nivel uno y dos.	75
Tabla 6.35. Efectos aleatorios del modelo multinivel con nivel uno y dos.	75
Tabla 6.36. Coeficientes del modelo multinivel de familia binomial negativa.	77
Tabla 6.37. Porcentaje de error absoluto medio de modelos desarrollados.	78

ÍNDICE DE FIGURAS

Figura 1.1. Círculo vicioso del transporte público (Ortúzar y Willumsen, 1990).....	2
Figura 2.1. Partes del contador ZELT (New Zealand Transport Agency, 2009).	11
Figura 2.2. Esquema del contador ZELT (New Zealand Transport Agency, 2009).	12
Figura 2.3. Contador pasivo de bicicletas en Temuco.....	13
Figura 3.1. Esquema de un árbol de regresión.	20
Figura 4.1. Esquema conceptual de la metodología seguida.	23
Figura 5.1. Mapa de ubicación contadores automáticos de Temuco.....	29
Figura 5.2. Distribución de conteos diarios acumulados en Temuco.....	31
Figura 5.3. Mapa de ubicación de contadores automáticos seleccionados.....	32
Figura 5.4. Matriz de correlación de variables meteorológicas del contador Ex Vía Férrea 1. ...	33
Figura 5.5. Evolución de la demanda en ciclovías de Temuco.	35
Figura 5.6. Tasas de crecimiento [%] contadores automáticos 2018-2019.....	36
Figura 5.7. Valor absoluto de pasadas en contadores automáticos año 2019.....	37
Figura 5.8. Valores absolutos de pasadas años 2018-2019.	38
Figura 5.9. Media diaria de conteos en días de fin de semana.	39
Figura 5.10. Comparación entre la media diaria de conteos semanales y de fin de semana.	40
Figura 5.11. Variabilidad horaria de un día laboral de semana.....	41
Figura 6.1. Relación entre conteos manuales y conteos automáticos.....	43
Figura A.1. Conteos diarios [Nº Pasadas] v/s temperatura media diaria [°C].	88
Figura A.2. Conteos diarios [Nº Pasadas] v/s temperatura mínima diaria [°C].	88
Figura A.3. Conteos diarios [Nº Pasadas] v/s temperatura máxima diaria [°C]	89
Figura A.4. Conteos diarios [Nº Pasadas] v/s precipitación diaria [mm].	89
Figura A.5. Conteos diarios [Nº Pasadas] v/s radiación media diaria [W/m ²].	90
Figura A.6. Conteos diarios [Nº Pasadas] v/s humedad relativa media diaria [%].	90
Figura A.7. Conteos diarios [Nº Pasadas] v/s material particulado medio diario 10 [µg/m ³].	91
Figura A.8. Conteos diarios [Nº Pasadas] v/s material particulado medio diario 2.5 [µg/m ³].	91
Figura A.9. Histograma de temperatura media diaria [°C].	92
Figura A.10. Histograma de temperatura máxima diaria [°C].	92
Figura A.11. Histograma de temperatura mínima diaria [°C].	93
Figura A.12. Histograma de precipitación diaria [mm].	93
Figura A.13. Histograma de radiación media diaria [W/m ²].	94
Figura A.14. Histograma de material particulado medio diario 10 [µg/m ³].	94
Figura A.15. Histograma de material particulado medio diario MP 2.5 [µg/m ³].	95
Figura A.16. Modelo regresión lineal múltiple v/s fecha para el contador Ex Vía Férrea 1.....	95
Figura A.17. Predicción conteos diarios modelo regresión lineal múltiple v/s conteos diarios reales en función del tiempo para el contador Ex Vía Férrea 1.	96
Figura A.18. Predicción conteos diarios modelo regresión lineal múltiple v/s conteos diarios reales para el contador Ex Vía Férrea 1.	96
Figura A.19. Modelo Poisson v/s fecha para el contador Ex Vía Férrea 1.	97
Figura A.20. Predicción conteos diarios modelo Poisson v/s conteos diarios reales en función del tiempo para el contador Ex Vía Férrea 1.	97
Figura A.21. Predicción conteos diarios modelo Poisson v/s conteos diarios reales para el contador Ex Vía Férrea 1.	98
Figura A.22. Modelo binomial negativo v/s fecha para el contador Ex Vía Férrea 1.....	98
Figura A.23. Predicción conteos diarios modelo binomial negativo v/s conteos diarios reales en función del tiempo para el contador Ex Vía Férrea 1.....	99

Figura A.24. Predicción conteos diarios modelo binomial negativo v/s conteos diarios reales para el contador Ex Vía Férrea 1.....	99
Figura A.25. Modelo <i>Random Forests</i> v/s fecha para el contador Ex Vía Férrea 1.	100
Figura A.26. Predicción conteos diarios modelo <i>Random Forests</i> v/s conteos diarios reales en función del tiempo para el contador Ex Vía Férrea 1.....	100
Figura A.27. Predicción conteos diarios modelo <i>Random Forests</i> v/s conteos diarios reales para el contador Ex Vía Férrea 1.....	101
Figura A.28. Modelo regresión lineal múltiple v/s fecha para el contador Los Músicos – Los Creadores.	101
Figura A.29. Predicción conteos diarios modelo regresión lineal múltiple v/s conteos diarios reales en función del tiempo para el contador Los Músicos – Los Creadores.....	102
Figura A.30. Predicción conteos diarios modelo regresión lineal múltiple v/s conteos diarios reales para el contador Los Músicos – Los Creadores.....	102
Figura A.31. Modelo Poisson v/s fecha para el contador Los Músicos – Los Creadores.	103
Figura A.32. Predicción conteos diarios modelo Poisson v/s conteos diarios reales en función del tiempo para el contador Los Músicos – Los Creadores.	103
Figura A.33. Predicción conteos diarios modelo Poisson v/s conteos diarios reales para el contador Los Músicos – Los Creadores.	104
Figura A.34. Modelo binomial negativo v/s fecha para el contador Los Músicos – Los Creadores.	104
Figura A.35. Predicción conteos diarios modelo binomial negativo v/s conteos diarios reales en función del tiempo para el contador Los Músicos – Los Creadores.	105
Figura A.36. Predicción conteos diarios modelo binomial negativo v/s conteos diarios reales para el contador Los Músicos – Los Creadores.....	105
Figura A.37. Modelo <i>Random Forests</i> v/s fecha para el contador Los Músicos – Los Creadores.	106
Figura A.38. Predicción conteos diarios modelo <i>Random Forests</i> v/s conteos diarios reales en función del tiempo para el contador Los Músicos – Los Creadores.	106
Figura A.39. Predicción conteos diarios modelo <i>Random Forests</i> v/s conteos diarios reales para el contador Los Músicos – Los Creadores.	107
Figura A.40. Modelo regresión lineal múltiple v/s fecha para el contador Pablo Neruda – Olimpia.	107
Figura A.41. Predicción conteos diarios modelo regresión lineal múltiple v/s conteos diarios reales en función del tiempo para el contador Pablo Neruda – Olimpia.	108
Figura A.42. Predicción conteos diarios modelo regresión lineal múltiple v/s conteos diarios reales para el contador Pablo Neruda – Olimpia.	108
Figura A.43. Modelo Poisson v/s fecha para el contador Pablo Neruda – Olimpia.....	109
Figura A.44. Predicción conteos diarios modelo Poisson v/s conteos diarios reales en función del tiempo para el contador Pablo Neruda – Olimpia.	109
Figura A.45. Predicción conteos diarios modelo Poisson v/s conteos diarios reales para el contador Pablo Neruda – Olimpia.	110
Figura A.46. Modelo binomial negativo v/s fecha para el contador Pablo Neruda – Olimpia. .	110
Figura A.47. Predicción conteos diarios modelo binomial negativo v/s conteos diarios reales en función del tiempo para el contador Pablo Neruda – Olimpia.	111
Figura A.48. Predicción conteos diarios modelo binomial negativo v/s conteos diarios reales para el contador Pablo Neruda – Olimpia.	111
Figura A.49. Modelo <i>Random Forests</i> v/s fecha para el contador Pablo Neruda – Olimpia.....	112
Figura A.50. Predicción conteos diarios modelo <i>Random Forests</i> v/s conteos diarios reales en función del tiempo para el contador Pablo Neruda – Olimpia.	112

Figura A.51. Predicción conteos diarios modelo <i>Random Forests</i> v/s conteos diarios reales para el contador Pablo Neruda – Olimpia.	113
Figura A.52. Modelo regresión lineal múltiple v/s fecha para el contador San Martín - Mackenna.	113
Figura A.53. Predicción conteos diarios modelo regresión lineal múltiple v/s conteos diarios reales en función del tiempo para el contador San Martín – Mackenna.	114
Figura A.54. Predicción conteos diarios modelo regresión lineal múltiple v/s conteos diarios reales para el contador San Martín – Mackenna.	114
Figura A.55. Modelo Poisson v/s fecha para el contador San Martín – Mackenna.	115
Figura A.56. Predicción conteos diarios modelo Poisson v/s conteos diarios reales en función del tiempo para el contador San Martín – Mackenna.	115
Figura A.57. Predicción conteos diarios modelo Poisson v/s conteos diarios reales para el contador San Martín – Mackenna.	116
Figura A.58. Modelo binomial negativo v/s fecha para el contador San Martín – Mackenna... 116	116
Figura A.59. Predicción conteos diarios modelo binomial negativo v/s conteos diarios reales en función del tiempo para el contador San Martín – Mackenna.	117
Figura A.60. Predicción conteos diarios modelo binomial negativo v/s conteos diarios reales para el contador San Martín – Mackenna.	117
Figura A.61. Modelo <i>Random Forests</i> v/s fecha para el contador San Martín - Mackenna.	118
Figura A.62. Predicción conteos diarios modelo <i>Random Forests</i> v/s conteos diarios reales en función del tiempo para el contador San Martín – Mackenna.	118
Figura A.63. Predicción conteos diarios modelo <i>Random Forests</i> v/s conteos diarios reales para el contador San Martín – Mackenna.	119
Figura A.64. Mapa del índice de entropía por zona Encuesta Origen-Destino de Temuco.	119
Figura A.65. Mapa del índice de entropía por manzana censal de Temuco.	120
Figura A.66. Histograma de conteos diarios reales para el contador Ex Vía Férrea 1.	120
Figura A.67. Histograma de conteos diarios modelo regresión lineal múltiple para el contador Ex Vía Férrea 1.	121
Figura A.68. Histograma de conteos diarios modelo Poisson para el contador Ex Vía Férrea 1.	121
Figura A.69. Histograma de conteos diarios modelo binomial negativo para el contador Ex Vía Férrea 1.	122
Figura A.70. Histograma de conteos diarios modelo <i>Random Forests</i> para el contador Ex Vía Férrea 1.	122
Figura A.71. Diagramas de caja de modelos desarrollados para el contador Ex Vía Férrea 1. .	123
Figura A.72. Histograma de conteos diarios reales para el contador Los Músicos – Los Creadores.	123
Figura A.73. Histograma de conteos diarios modelo regresión lineal múltiple para el contador Los Músicos – Los Creadores.	124
Figura A.74. Histograma de conteos diarios modelo Poisson para el contador Los Músicos – Los Creadores.	124
Figura A.75. Histograma de conteos diarios modelo binomial negativo para el contador Los Músicos – Los Creadores.	125
Figura A.76. Histograma de conteos diarios modelo <i>Random Forests</i> para el contador Los Músicos – Los Creadores.	125
Figura A.77. Diagramas de caja de modelos desarrollados para el contador Los Músicos – Los Creadores.	126
Figura A.78. Histograma de conteos diarios reales para el contador Pablo Neruda – Olimpia. 126	126

Figura A.79. Histograma de conteos diarios modelo regresión lineal múltiple para el contador Pablo Neruda – Olimpia.	127
Figura A.80. Histograma de conteos diarios modelo Poisson para el contador Pablo Neruda – Olimpia.	127
Figura A.81. Histograma de conteos diarios modelo binomial negativo para el contador Pablo Neruda – Olimpia.	128
Figura A.82. Histograma de conteos diarios modelo <i>Random Forests</i> para el contador Pablo Neruda – Olimpia.	128
Figura A.83. Diagramas de caja de modelos desarrollados para el contador Pablo Neruda – Olimpia.	129
Figura A.84. Histograma de conteos diarios reales para el contador San Martín – Mackenna..	129
Figura A.85. Histograma de conteos diarios modelo regresión lineal múltiple para el contador San Martín – Mackenna.	130
Figura A.86. Histograma de conteos diarios modelo Poisson para el contador San Martín – Mackenna.	130
Figura A.87. Histograma de conteos diarios modelo binomial negativo para el contador San Martín – Mackenna.	131
Figura A.88. Histograma de conteos diarios modelo <i>Random Forests</i> para el contador San Martín – Mackenna.	131
Figura A.89. Diagramas de caja de modelos desarrollados para el contador San Martín – Mackenna.	132
Figura A.90. Disminución del error del algoritmo <i>Random Forests</i> en función del número de árboles.	132
Figura A.91. Predicción conteos diarios modelo multinivel v/s conteos diarios reales en función de la entropía en el uso de suelo de cada ciclovía.	133
Figura A.92. Predicción conteos diarios modelo multinivel v/s conteos diarios reales para cada contador automático.	133
Figura A.93. Curva cuantil-cuantil de los conteos diarios reales.	134
Figura A.94. Curva cuantil-cuantil de la predicción de conteos diarios del modelo multinivel.	134

CAPÍTULO 1

INTRODUCCIÓN

Capítulo 1. Introducción

En la actualidad el aumento de la tasa de motorización como fenómeno a nivel mundial y también local ha generado muchos problemas asociados a congestión vial, descenso de niveles de servicio y baja seguridad, debido a las externalidades negativas que provoca el excesivo uso de vehículos motorizados (Medina, 2012). Esta afirmación se refuerza al observar los datos de tasas de motorización a nivel mundial. En Estados Unidos, dicha tasa corresponde a un automóvil cada 1.3 habitantes, mientras que en España alcanza un vehículo cada dos habitantes (Asociación Nacional Automotriz de Chile, 2018). En Chile, según datos de la misma asociación, la tasa de motorización corresponde a un automóvil cada 3.8 personas, valor que aún se encuentra lejos de las tasas de motorización de los países anteriormente mencionados. Según De Grange (2017), el valor de dicha tasa aumenta principalmente por una variable: el incremento en el ingreso per cápita de la población. Este fenómeno ocurre actualmente en Chile y en gran parte del mundo, sin embargo, al país aún le falta duplicar su tasa de motorización en los próximos años, lo cual supone que los problemas se acrecentarán si no se toman medidas paliativas. En particular, el aumento de esta tasa en la ciudad de Temuco es del orden del 8% anual, la cual es superior al promedio nacional de los últimos 5 años (Secretaría de Planificación de Transporte, 2014).

Los aumentos en las tasas de motorización tienen como consecuencia aumentos en los tiempos de viajes y un acrecentamiento de episodios de congestión vial. Esto se suma a la lenta capacidad del sistema para adaptar la oferta vial a este crecimiento exponencial del flujo, donde el uso del automóvil disminuye la capacidad de las vías en transportar personas debido a su ineficiente uso del espacio urbano en comparación con otros modos tales como el transporte público o la bicicleta. Como referencia, el actual sistema de transporte urbano que funciona en la capital regional de La Araucanía cuenta con 10 líneas de buses con 32 servicios, 810 máquinas y 1447 colectivos en 14 líneas que operan 24 servicios, contrastado con un parque de vehículos particulares en el entorno Temuco - Padre Las Casas que supera las 83.000 unidades (Ministerio de Transportes y Telecomunicaciones, 2017). Lo anterior genera un problema denominado “Círculo vicioso del transporte público”, el cual es definido por Ortúzar y Willumsen (1990) como aquel fenómeno que deteriora los sistemas de transporte colectivo, pues al existir mayor congestión, el modo de transporte público se vuelve más lento y genera menor frecuencia durante el día y hace más cara la operación de los servicios de transporte. Esto vuelve más atractivo a los modos de transporte privados ya que estos pueden variar sus rutas, lo que va generando mayor congestión en otras vías

repitiendo este fenómeno en el tiempo. La descripción completa de este fenómeno está expuesta en el siguiente esquema conceptual.

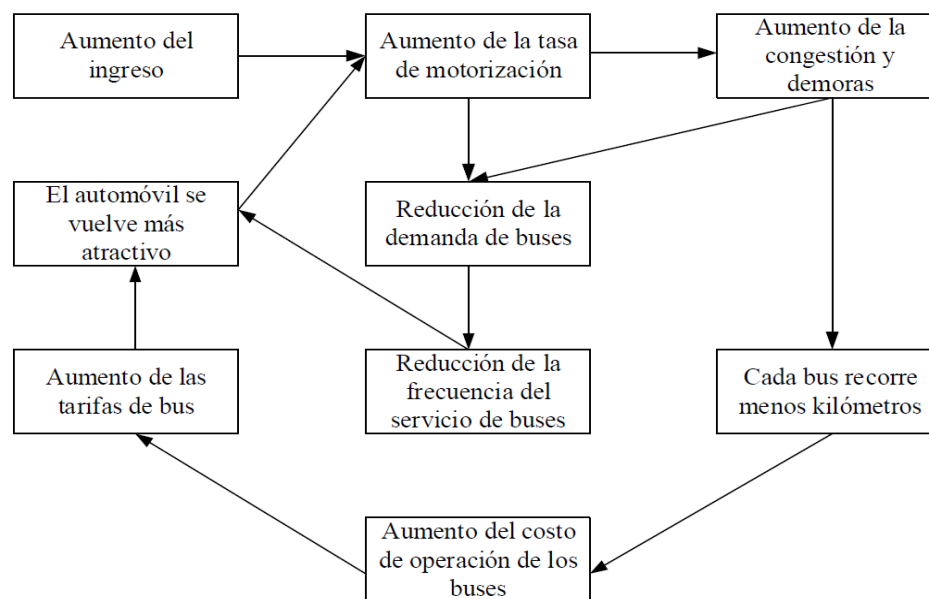


Figura 1.1. Círculo vicioso del transporte público (Ortúzar y Willumsen, 1990).

Algunas medidas o enfoques pertinentes para aminorar el efecto de este fenómeno son explicadas por algunos expertos en la materia a nivel nacional: Iván Poduje (2018), señala que hay que avanzar en tres aspectos. Uno de ellos es mejorar el transporte público, racionalizando subsidios, ampliando metro y desarrollando trenes urbanos en regiones. Asimismo, plantea invertir en infraestructura vial, pero también “elevar los costos de andar en auto”, a través de la tarificación vial y los permisos de circulación. Juan Carlos Muñoz (2018), director del Centro de Desarrollo Urbano Sustentable de la Universidad Católica (CEDEUS), sostiene que el país debe pensar en sistemas de transporte que desincentiven el uso del auto para zonas congestionadas, favoreciendo otros modos masivos como buses y metro, además de la bicicleta y caminatas. En ese sentido, es fundamental que los municipios que tienen las mayores tasas de motorización en Chile, dejen de promover soluciones basadas en el automóvil para resolver su congestión. Es clave que esas comunas destinen sus recursos a promover soluciones que usen el escaso espacio vial eficientemente. Finalmente, Alejandro Tirachini (2018), profesor de Ingeniería de Transportes del Departamento de Ingeniería Civil de la Universidad de Chile, señala que el objetivo de las políticas de transporte jamás será que la gente no se compre un automóvil, sí que el uso sea racional y se disminuya su

utilización en lugares congestionados y contaminados. Para esto se requiere un enfoque de “garrote”, que es implementar tarificación vial, tarificar y racionar estacionamientos; y de “zanahoria”, es decir, inversiones en transporte público de calidad, incentivar la caminata, la bicicleta y el uso compartido del automóvil.

Abordando dicha problemática, a lo largo del país y en específico en la ciudad de Temuco, se han desarrollado últimamente importantes proyectos de infraestructura vial asociados a mejoramiento y ampliación de vías enfocadas en automóviles, pero que han incluido en ellas la construcción de ciclovías (Ministerio de Vivienda y Urbanismo, 2018). El objetivo es diversificar la oferta de modos de transporte, incentivando el uso de vehículos no motorizados y que sean amigables con el medio ambiente tales como la bicicleta, cuya partición modal en la ciudad de Temuco es tan solo del 2% (Secretaría de Planificación de Transporte, 2014).

Diversos estudios internacionales tales como Nosal *et al.*, (2014), Schmiedeskamp y Zhao (2016) y An *et al.*, (2019) señalan que los conteos diarios de bicicleta están fuertemente relacionados con las condiciones meteorológicas de un día particular. Por lo tanto, la explicación del aumento o descenso de los conteos diarios de bicicleta puede estar explicada en primer lugar por factores como: precipitación, temperatura máxima, media y mínima, radiación, viento, contaminación ambiental, entre otros. Esta serie de variables explicaría la diferencia en el uso de la bicicleta a nivel diario, pero no sería capaz de capturar las diferencias en el uso de las diferentes ciclovías dentro de una misma ciudad. En el caso de Temuco, aunque las condiciones meteorológicas sean las mismas, existen ciclovías que se utilizan con mayor frecuencia que otras. Esta afirmación se puede concluir analizando los datos pasivos referidos a conteos automáticos obtenidos por el Ministerio de Vivienda y Urbanismo (MINVU). Hipotéticamente, la razón de esta diferencia podría ser atribuida a una serie de variables de tipo contextual asociadas a cada ciclovía, que la caracterizan a diferencia de otra en función de las características socioespaciales circundantes, las características de conectividad y de atracción de viajes, entre otras. Estudios como Lu *et al.*, (2018) desarrollan modelos de demanda directa para ciertos puntos de una ciudad en que definen variables de este tipo para poder explicar las diferencias en el uso de la infraestructura de ciclovías en una misma urbe.

Este proyecto ajusta modelos de predicción de tránsito diario de bicicletas utilizando datos pasivos de conteos automáticos en diversos puntos de la ciudad de Temuco, en el cual se incluyen variables

meteorológicas y contextuales. Para ello se cuenta con una base de datos de contadores automáticos de bicicleta, los cuales han sido instalados por el Ministerio de Vivienda y Urbanismo (MINVU) con el objetivo de cuantificar el uso que se le da a esta infraestructura. Se pondrá énfasis en identificar las variables relevantes que explican los conteos diarios de bicicleta para así generar nuevos modelos que sirvan como metodología para estimar el tránsito diario de bicicletas, y de esta forma, ser útil para futuros proyectos que se deseen construir en la ciudad de Temuco.

1.1. Descripción del problema

En Chile, cuando se busca estimar la demanda de bicicletas para evaluar la construcción de una nueva ciclovía en un cierto punto de una ciudad de tamaño medio o grande, se asume que la transferencia modal corresponde a una proporción de los viajes del modo caminata que se transferirá a la bicicleta. Para efectos de estimar esta demanda, la metodología actual considera la utilización de un valor fijo de 4.7% sobre el total de viajes diarios en caminata (Ministerio de Desarrollo Social y Secretaría de Planificación de Transporte, 2013). Esta estimación se obtuvo de un estudio en la ciudad de Copiapó y se aplica normativamente en todo el territorio nacional para ciudades de tamaño medio y grande. Un estudio realizado para una ciudad particular del norte de Chile, difícilmente será aplicable para una ciudad de la zona central o sur, debido a las notorias diferencias existentes en parámetros como la partición modal, meteorología, factores culturales, uso de suelo, entre otros. La diferencia es observable a gran nivel cuando se compara el uso que se le da a la infraestructura de ciclovías en diferentes ciudades de nuestro país, en el cual destacan las ciudades de Curicó y Chillán, las cuales poseen una partición modal de bicicleta de 12 y 8.5% respectivamente (Secretaría de Planificación de Transporte, 2019). De igual modo, existen ciudades en las cuales no se les da el mismo uso, como es el caso de Temuco, en la cual como fue mencionado con anterioridad, la partición de este modo solo alcanza el 2%.

En la ciudad de Temuco, existe además una notoria diferencia en el uso que se le da a ciertas ciclovías de la ciudad en comparación con otras. Este hecho sugiere que existe un problema en la metodología de proyección de ciclovías o que al menos no es lo suficientemente precisa como podría ser. La estimación de modelos predictivos de tránsito diario de bicicletas en base a datos pasivos, incorporando variables meteorológicas y contextuales, puede suponer un gran aporte para mejorar la metodología existente y que se aplica actualmente en la ciudad.

El presente informe se estructura en siete capítulos. En el presente capítulo se introduce al tema de investigación y se enuncian los objetivos del estudio. En el segundo capítulo se presentan los antecedentes generales basados en una revisión del estado del arte del tema. El tercer capítulo titulado fundamentos teóricos, presenta el marco conceptual de las herramientas utilizadas en este estudio. El cuarto capítulo titulado metodología, describe los procedimientos empleados en la elaboración de la memoria. En el quinto capítulo se realiza un análisis descriptivo de la muestra de datos utilizada, tanto a nivel de la variable dependiente como a nivel de las potenciales variables predictoras. El capítulo seis expone los resultados obtenidos y el análisis de estos. Finalmente, el capítulo siete enuncia las conclusiones de esta investigación.

1.2. Objetivos

1.2.1. Objetivo general

Evaluar el efecto de variables meteorológicas y contextuales en el volumen de tránsito diario de bicicletas en ciclovías de la ciudad de Temuco.

1.2.2. Objetivos específicos

- 1) Evaluar la efectividad de los contadores automáticos de bicicletas.
- 2) Elaborar una base de datos de potenciales variables y un análisis descriptivo de ésta.
- 3) Identificar las variables relevantes en la explicación de los conteos diarios de bicicleta.
- 4) Implementar modelos de predicción de tránsito diario de bicicletas, incorporando variables meteorológicas y contextuales.

CAPÍTULO 2

ANTECEDENTES GENERALES

Capítulo 2. Antecedentes Generales

En el presente capítulo se describe el estado del arte referido a la relación existente entre el uso de la bicicleta y las condiciones meteorológicas, como también la relación que tiene con variables de tipo contextual. Además, se describe el funcionamiento de los contadores automáticos de bicicletas.

2.1. Relación entre meteorología y uso de la bicicleta

La meteorología y el ciclismo están fuertemente relacionados. La evidencia empírica muestra que las malas condiciones meteorológicas tienen una relación inversa con los viajes en bicicleta (más que otros modos) debido a que los ciclistas están expuestos a los efectos de estas condiciones y son vulnerables a ellos (An *et al.*, 2019). Con el cambio climático, las condiciones meteorológicas extremas serán más frecuentes e intensas, por lo que es importante desarrollar investigación acerca de la relación existente entre las condiciones meteorológicas y el ciclismo, considerando el auge que ha tenido la bicicleta en los últimos años en el transporte urbano (Wadud, 2014).

La bicicleta ofrece una variedad de beneficios. En primer lugar, se destacan sus ventajas en la salud de aquellos que la utilizan regularmente. De igual forma, existe un beneficio económico versus otros modos de transporte, ya que su inversión inicial y su mantenimiento es mucho menor que cualquier otro modo de transporte privado. Tampoco requiere gastos en combustible, peajes, seguros o impuestos. Además, es un modo de transporte amable con el medio ambiente y ocupa un reducido espacio vial en comparación con los automóviles. Finalmente, la bicicleta supone una importante mejora en los tiempos de viaje en caso de que existan ciclovías físicamente separadas de los carriles de automóviles, evitando la congestión vial asociada a los modos de transporte motorizados (Wendel-Vos *et al.*, 2018). Por todas estas razones, este modo de transporte está siendo promovido por políticas públicas y expertos en desarrollo urbano. Para aplicar políticas públicas y mejoras en infraestructura que provoquen un mayor uso de la bicicleta, se requiere un robusto entendimiento de los factores que provocan el uso de ésta (Schmiedeskamp y Zhao, 2016).

Uno de los parámetros medibles más importantes en la investigación referida a ciclovías y uso de la bicicleta, es la estimación de tránsitos diarios de ciclistas (Roll y Proulx, 2018). El volumen diario de ciclistas en una ciudad o en un punto de una ciudad determinada, es útil para los expertos debido a que con esta medición se pueden medir cambios en la partición modal de la bicicleta,

conocer el comportamiento temporal de los viajes y obtener información para realizar una adecuada planificación en infraestructura (Nordback y Janson, 2010).

El primer estudio acerca de la relación entre variables meteorológicas y el tránsito diario de ciclistas fue desarrollado por Hanson y Hanson (1977), en el cual su principal conclusión fue la demostración del efecto de la temperatura sobre la disminución o aumento del número de ciclistas diarios. Desde ese entonces, todas las investigaciones han concluido que el aumento o descenso de los conteos diarios de ciclistas tienen una correlación positiva con temperaturas altas y precipitaciones bajas. Sin embargo, Thomas *et al.*, (2009) y Lewin (2011) concluyen que los efectos de la precipitación y la temperatura en los volúmenes diarios de bicicletas no son lineales. Es decir, el tránsito diario de bicicletas puede descender tanto con temperaturas muy bajas como con temperaturas muy altas.

La primera línea de investigación asociada a la estimación de tránsitos diarios de bicicleta, es aquella que involucra el desarrollo y aplicación de factores de ajuste. Sin embargo, estos estudios son utilizados habitualmente para estimar tránsitos diarios de ciclistas cuando se dispone de muy pocos datos, por ejemplo, cuando se realizan conteos manuales y se busca generalizar los resultados a períodos de tiempo más extensos (El Esawey, 2014).

Otra línea de investigación es aquella que busca estimar modelos que capturen la relación entre las condiciones meteorológicas y el volumen de ciclistas. Existen varios trabajos que han modelado la relación entre los volúmenes horarios y diarios de bicicleta utilizando una serie de variables identificadas, a través de modelos log-linear (Hanson y Hanson, 1997; Niemeier, 1996; Nankervis 1999; Brandenburg *et al.*, 2007; Rose *et al.*, 2011; Tin Tin *et al.*, 2012), utilizando modelos de series de tiempo (Thomas *et al.*, 2009; Lewin 2011; Gallop *et al.*, 2012) y datos de contadores automáticos empleando modelos Poisson y Binomial Negativo (Miranda-Moreno y Nosal, 2011; Nosal y Miranda-Moreno, 2014; Schmiedeskamp y Zhao, 2016).

Los resultados de estas investigaciones son generalmente consistentes unos con otros, pero el impacto varía de una ciudad o país con respecto a otro. Por ejemplo, Niemeier (1996) concluyó que los conteos de bicicleta decrecen entre un 15 a 25% en días lluviosos o días en que las temperaturas son menores que 55 °F en Washington, Estados Unidos. Nankervis (1999) examinó como las condiciones meteorológicas afectan a los viajeros de bicicleta en Melbourne, Australia. Los resultados de dicha investigación sugieren que el uso de la bicicleta está fuertemente afectado

por las temperaturas extremas, seguido por las condiciones del viento y lluvia. Miranda-Moreno y Nosal (2011) estudiaron los efectos directos inmediatos y los efectos rezagados o posteriores de las variables meteorológicas en los volúmenes horarios de tráfico de bicicletas en Montreal, Canadá. Los autores mostraron que los viajes en bicicleta aumentan a medida que la temperatura se eleva, sin embargo, cuando la temperatura excede los 28 °C y la humedad relativa es mayor que el 60%, los viajes en bicicleta decrecen entre un 11 y 20%. Lewin (2011) demostró una fuerte correlación positiva entre la temperatura y los conteos diarios. Si la temperatura aumenta más allá de 32 °C se produce un leve descenso en los conteos. Los volúmenes de conteos de bicicleta también decrecen en forma no lineal en días lluviosos o en presencia de nieve. Tin Tin *et al.*, (2012) estudió la relación entre las condiciones meteorológicas y los viajes en bicicleta en Auckland, Australia. Los volúmenes de bicicleta aumentaron en un 3.2% (horario) y 2.6% (diario) por cada incremento de 1 °C en la temperatura. De igual forma, por cada 1 mm de aumento en la precipitación, se obtiene un descenso de un 10.6% de pasadas de ciclistas (horario) y 1.5% (diario). A diferencia de otros estudios, en la investigación anterior se concluyó que la velocidad del viento afecta los volúmenes de ciclistas: por cada incremento en 1 km/hr de la velocidad del viento, los conteos horarios disminuyeron en 1.4% y en un 0.9% los volúmenes diarios. Finalmente, Gallop *et al.*, (2012) analizó la relación entre los conteos horarios de bicicleta y factores meteorológicos en Vancouver, Canadá. La precipitación tuvo un efecto negativo y significativo, causando un cambio de al menos 24% en la variación de los conteos promedios horarios de bicicleta. La temperatura también fue significativa: los conteos de bicicleta aumentaron en 1.65% con cada incremento de 1 °C sobre el promedio.

A pesar de los aportes metodológicos de todos los estudios anteriormente mencionados, su objetivo fue principalmente el análisis explicativo de las variables y la determinación de los coeficientes en sus modelos. Estimaciones de la precisión de estos modelos han sido desarrolladas solo por algunos estudios (Lindsey *et al.*, 2007; Gallop *et al.*, 2012; Wang *et al.*, 2014; Wang *et al.*, 2016).

En síntesis, actualmente existen diferentes metodologías para estimar volúmenes de tránsito de bicicletas. Estos modelos difieren en función de sus procedimientos, necesidad de datos y la precisión. Además, la mayoría de estos modelos fueron desarrollados para un número limitado de locaciones, lo cual no permite generalizar los resultados (El Esawey, 2016). A continuación se presenta la Tabla 2.1, la cual resume algunos trabajos importantes que se relacionan con el tema de

esta memoria, identificando los autores, las variables explicativas identificadas y el método utilizado para la estimación de sus modelos.

Tabla 2.1. Resumen de trabajos relacionados.

Fuente	Variables identificadas	Método
Niemeier (1996)	Presencia de lluvia, presencia de altas temperaturas, variable estacional, entre otras.	Modelo Poisson.
Hunt y Abraham (2007)	Variables descriptivas indicando uso de suelo, estacionamientos con seguridad, nivel de experiencia.	Modelo logístico de opciones relacionadas con el ciclismo.
McCahill y Garrick (2008)	Medición de elección logarítmica, densidad de población, densidad de trabajadores.	Nueva teoría utilizada para evaluar y predecir volúmenes de bicicletas en una red.
Parkin <i>et al.</i> , (2008)	Género, propiedad de automóvil, proporción de rutas fuera de la carretera, topografía.	Modelo de regresión logístico de variables socioeconómicas y físicas.
Thomas <i>et al.</i> , (2009)	Temperatura, horas de sol, precipitación, velocidad del viento, uso de ciclovías.	Estructura de dos niveles, siendo el nivel superior un modelo logarítmico lineal y el nivel inferior un modelo lineal.
Jones <i>et al.</i> , (2010)	Largo de la red de ciclovías, densidad de empleos, densidad de población.	Regresión estándar de mínimos cuadrados.
Griswold <i>et al.</i> , (2011)	Densidad de población, densidad de empleos, proximidad a autopistas centrales, edad, nivel educacional, ingreso, entre otras.	Regresión log-lineal de mínimos cuadrados.
Lewin (2011)	Temperatura máxima, presencia de lluvia, presencia de nieve, entre otras.	Regresión lineal estándar.
Rose <i>et al.</i> , (2011)	Temperatura, precipitación, presencia de vacaciones, presencia de período escolar, día de la semana.	Meteorología y otros efectos examinados con un modelo agregado a nivel diario.
Gallop <i>et al.</i> , (2012)	Temperatura, humedad relativa, velocidad del viento, visibilidad, niebla, precipitación.	Promedio móvil integrado autorregresivo utilizado para tener en cuenta los patrones de correlación en serie.
Helbich <i>et al.</i> , (2014)	Temperatura máxima diaria, velocidad del viento promedio diaria, precipitación diaria.	Modelos logit ponderados geográficamente.

Fuente	Variables identificadas	Método
Nosal y Miranda-Moreno (2014)	Temperatura, porcentaje de humedad relativa, presencia de lluvia, lluvia en la mañana, entre otras.	Modelos log-linear y Binomial Negativo.
Wadud (2014)	Precipitación, temperatura, viento, nieve, humedad relativa, horas de sol, tasa de desempleo, precio de pasaje en bus, variables dummy para cada mes.	Modelos log-linear y Binomial Negativo.
Lu <i>et al.</i> , (2018)	Densidad de población, ingreso del hogar, direcciones residenciales y no residenciales, área industrial, intersecciones, entre otras.	Regresión paso a paso.
An <i>et al.</i> , (2019)	Temperatura, precipitación, humedad relativa, viento, nieve, topografía, infraestructura, variables dummy para fin de semana, vacaciones, entre otras.	Modelos de predicción de demanda directa.

2.2. Relación entre variables contextuales y uso de la bicicleta

Existe un área de investigación que está enfocada en la estimación de modelos de demanda directa en ciertos puntos de una ciudad, en los cuales se definen variables predictoras de tipo contextual, es decir, aquellos factores que caracterizan a los puntos potencialmente generadores de demanda de bicicletas. Estas variables contextuales corresponden a aquellas que caracterizan el uso de suelo de un lugar, las características de conectividad que tiene y también indicadores socioeconómicos y demográficos de una zona. También son conocidas como variables de entorno construido. En términos generales el entorno construido corresponde a todo el ambiente físico que ha sido construido o modificado por el ser humano (Oliva, 2017). En particular, según Handy *et al.*, (2002) el entorno construido comprende diseño urbano, uso de suelo y sistema de transporte; y engloba patrones de la actividad humana dentro del ambiente físico.

En ese sentido, Lu *et al.*, (2018) desarrolló un modelo de demanda directa para modos de transporte no motorizados a nivel horario que incluye las siguientes variables dependientes: densidad de población, ingreso del hogar, direcciones residenciales y direcciones no residenciales, número de vías troncales, número de vías locales, cantidad de intersecciones, entre otras. Otros estudios en los cuales se incluyen variables de uso de suelo junto con características de transporte y conectividad (Pulughurta y Repaka 2008; Schneider *et al.*, 2009; Molino *et al.*, 2009; Griswold *et al.*, 2011;

Miranda-Moreno y Fernandes 2011; Hankey y Lindsey 2016; Wang *et al.*, 2016; Hankey *et al.*, 2017).

2.3. Contadores automáticos de bicicletas

Cuantificar el uso de la bicicleta es fundamental para entender los viajes en este modo de transporte. Los métodos de conteo varían desde los conteos temporales, manuales y conteos permanentes asociados a sensores. Los contadores automáticos son dispositivos que poseen bucles inductivos, los cuales operan detectando un cambio electromagnético que ocurre cuando una bicicleta pasa por sobre él (Nordback y Janson, 2010). En particular, en la ciudad se utiliza el sensor modelo ZELT elaborado por la empresa francesa Eco-Counter. El sistema se compone por diversas partes, las cuales son enumeradas a continuación:

- Un *logger*, el cual muestra la información de conteos y es la fuente de donde se puede descargar la información.
- Un transductor, el cual es un dispositivo de operación y contiene información referida a la sensibilidad y ajustes para la detección de conteos.
- Un set de baterías.
- Conectores de bucles, los cuales conectan el *logger* a los bucles.

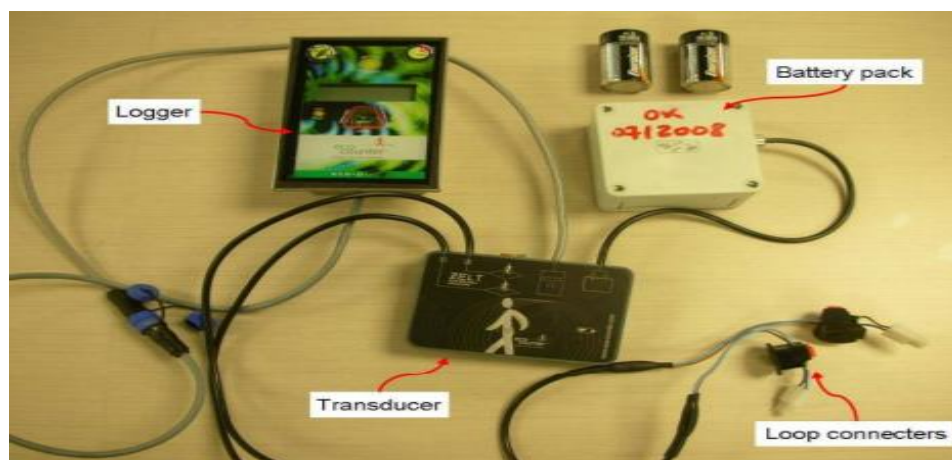


Figura 2.1. Partes del contador ZELT (New Zealand Transport Agency, 2009).

Los contadores automáticos de bicicletas realizan una detección automática que permite contar ciclistas tanto de día como de noche, en forma permanente. Estos sistemas requieren una inversión importante a diferencia de los conteos manuales, pero proveen una continuidad en los datos que otros sistemas de medición no pueden realizar (Nordback y Janson, 2010). Los datos del sistema ZELT son almacenados cada período de una hora. El contador tiene una forma de rombo o diamante de manera que el eje mayor de este polígono es perpendicular a la dirección de viaje de los ciclistas. Debido a esta razón, los contadores no deberían posicionarse en curvas. Los bucles deben estar al menos a cuatro metros de distancia de cualquier otro equipo eléctrico como semáforos u otros contadores, con el objetivo de no presentar interferencia (New Zealand Transport Agency, 2009).

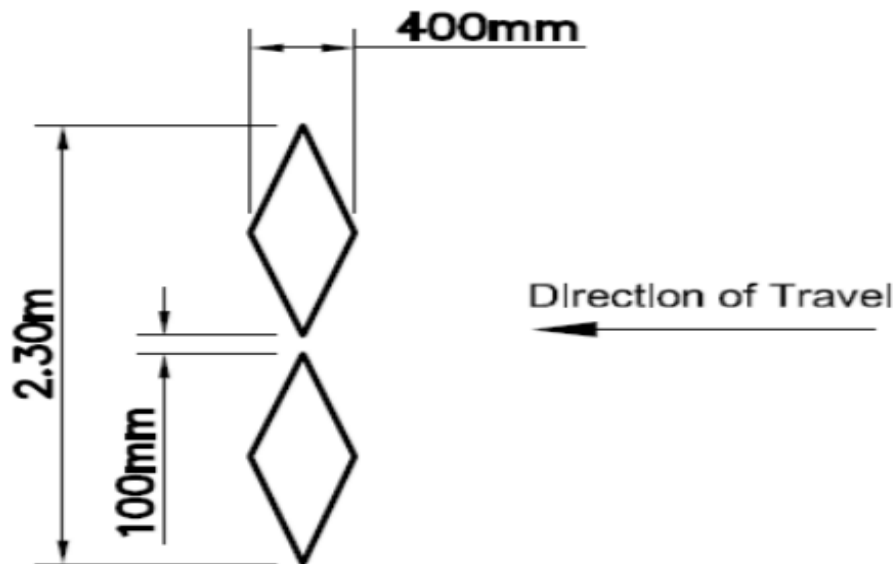


Figura 2.2. Esquema del contador ZELT (New Zealand Transport Agency, 2009).

En la Figura 2.2 es posible observar un croquis que simboliza un contador ZELT desde una vista en planta junto con sus respectivas cotas. En ella, se distinguen dos rombos, los cuales van posicionados uno en cada carril de la ciclovía, lo cual permite capturar el sentido del flujo del ciclista. El algoritmo SIRIUS desarrollado por Eco-Counter opera en base a 13 criterios de diferenciación que le permiten clasificar a los vehículos como bicicletas, incluso aquellas que son de marco de aluminio o de carbono (New Zealand Transport Agency, 2009).



Figura 2.3. Contador pasivo de bicicletas en Temuco.

En la Figura 2.3 se observa uno de los 17 contadores automáticos tipo ZELT instalados en la ciudad de Temuco. En particular, este contador corresponde al de Javiera Carrera – Pablo Neruda 1. En esta imagen, una ciclista está avanzando en sentido poniente - oriente por el carril correspondiente a oriente - poniente, sin embargo, debido a la configuración de los rombos, la bicicleta pasa primero por el rombo más cercano al poniente, por lo que aunque el ciclista vaya por el carril incorrecto, el sistema igualmente captura la dirección correcta del flujo. Una de las desventajas que tiene este tipo de conteos, es que representa el paso de ciclistas en puntos específicos de una ciclovía o de la ciudad, no pudiendo realizar seguimiento de las rutas reales de los ciclistas. En algunos sistemas de arriendo de bicicletas como “Mobike”, el hecho de iniciar y finalizar el viaje en un estacionamiento de bicicletas, permite conocer el origen y destino de los viajes, lo cual constituye una información más valiosa debido a la naturaleza dinámica de los datos, a diferencia de los datos de contadores automáticos que solo representan una pasada por un cierto lugar. De igual forma, plataformas como “Strava Metro” utilizan GPS para almacenar cada uno de los viajes realizados por los usuarios de esta aplicación para teléfonos móviles.

CAPÍTULO 3

FUNDAMENTOS TEÓRICOS

Capítulo 3. Fundamentos Teóricos

En el siguiente capítulo, se introducen aspectos teóricos de las herramientas utilizadas en la elaboración de los modelos de predicción.

3.1. Modelos lineales generalizados

Muchos autores han tratado el problema de la modelación, es decir, el problema de representar la realidad y su variabilidad e incertidumbre a través de un modelo matemático que permita el estudio, el análisis y la comprensión de la misma, con el objetivo de predecir su futuro o simplemente de conocerla (Cox y Wermuth, 1996). El objetivo de la modelación estadística consiste en, a partir de la observación o la experimentación, explicar el comportamiento de una o más características de los individuos o elementos de una población, en base a las diferencias existentes entre las características de los individuos. La variable univariante o multivariante que se desea explicar, se conoce como variable respuesta, variable objetivo o variable dependiente, mientras que las variables en las que se desea basar la explicación se denominan variables explicativas, variables independientes o criterios (Alcaide, 2015).

Inicialmente, el planteamiento de métodos explicativos en los que se trata de explicar una variable objetivo a través de un conjunto de variables predictoras, requiere la elección de un modelo que describa la estructura de la relación entre las variables. Para el planteamiento del modelo, es importante distinguir entre el tipo de variables que intervienen (continuas, de conteo, cualitativas, entre otras) y en la clase de relaciones funcionales que se admiten para analizar la relación entre la variable dependiente y las variables explicativas. Según el tipo de variables que intervienen y de las relaciones entre ellas, se dispondrá de un conjunto de posibles modelos más o menos adecuados, capaces de explicar la realidad (Alcaide, 2015).

Generalmente, el modelo más estudiado y utilizado es el del tipo lineal, es decir, se modela la relación tratando de expresar la variable dependiente, a través de una combinación lineal de las variables predictoras. El modelo lineal clásico consiste en expresar la esperanza condicionada de la variable dependiente como combinación lineal de las variables explicativas bajo la suposición de normalidad y homocedasticidad. Esta modelación clásica presentada por Stigler (1981) se puede extender a una familia de modelos más general, propuesta por Nelder y Wedderburn (1972) y

ampliada por McCullagh y Nelder (1989), conocida como modelos lineales generalizados (MLG). Esta nueva familia permite unificar tanto los modelos con variables de respuesta categórica como numérica, y considera las distribuciones Binomial, Poisson, Hipergeométrica, Binomial Negativa, entre otras, y no únicamente la distribución Normal. Por otro lado, la esperanza μ_i , está relacionada con las variables explicativas a través de una función de enlace o *link*.

En los modelos de regresión lineal se considera el supuesto de independencia para las observaciones, sin embargo, para esta nueva familia a diferencia del modelo clásico, las distribuciones de la componente aleatoria no necesariamente son homocedásticas, es decir, no se requiere de un supuesto de homogeneidad de varianzas. Por ejemplo, en el modelo de regresión de Poisson, la varianza de la variable respuesta viene dada por el valor esperado. Por tanto, la varianza puede variar a medida que varíe el valor esperado, a diferencia del modelo clásico con distribución normal que tiene dos parámetros no relacionados, la media y la varianza que se consideran constantes para las diferentes observaciones (McCullagh y Nelder, 1989).

Para definir un modelo lineal generalizado, se requieren de tres componentes, los cuales son: la componente aleatoria, la componente sistemática y la función de enlace (Alcaide, 2015).

Sea Y la variable aleatoria objetivo o respuesta objeto de estudio y sean n las variables aleatorias independientes e idénticamente distribuidas $Y_1, \dots, Y_n$ la muestra aleatoria procedente de Y . Siendo Y denominada como componente aleatoria cuya distribución pertenece a la familia exponencial de distribuciones. Algunos miembros de la familia exponencial son las distribuciones: Normal, Binomial, Poisson, Gamma o Binomial Negativa. La distribución de una variable aleatoria Y , caracterizada por los parámetros θ y ϕ pertenece a la familia exponencial, si tiene la forma:

$$f(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\} \quad (3.1)$$

Donde $f()$ denota la función de probabilidad en caso de que Y sea una variable discreta, o la función de densidad en caso de que Y sea una variable continua. El término θ es el parámetro de localización o canónico, ϕ el parámetro escala y $a(\phi)$, $b(\theta)$ y $c(y, \phi)$ son funciones específicas de cada elemento de la familia. En la Tabla 3.1 se resumen los elementos principales que caracterizan a algunas de las distribuciones más utilizadas de la familia exponencial.

Tabla 3.1. Resumen de distribuciones exponenciales (Alcaide, 2015).

Distribución	Rango de Y	θ	$a(\theta)$	$b(\theta)$	$V(\mu)$
Binomial $B(n, p)$	$[0, n]$	$\ln(\frac{p}{1-p})$	1	$n \ln(1 + \exp(\theta))$	$np(1-p)$
Binomial Negativa $NB(p, r)$	$Ent[0, \infty)$	$\ln(1-p)$	1	$-r(\ln(1 - \exp(\theta)))$	$\frac{r(1-p)}{p^2}$
Gamma $G(\mu, v)$	$(0, \infty)$	$-1/\mu$	$1/v$	$-\ln(-\theta)$	μ^2
Normal $N(\mu, \sigma^2)$	$(-\infty, \infty)$	μ	σ^2	$\theta^2/2$	1
Poisson $P(\mu)$	$Ent[0, \infty)$	$\ln(\mu)$	1	$\exp(\theta)$	μ

Por su parte, la componente sistemática recoge la variabilidad de Y expresada a través de p variables explicativas $X_1, \dots, X_p$, denotadas por X , y de sus correspondientes parámetros $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$. La componente sistemática, también denominada predictor lineal se simboliza por η .

$$\eta = X\beta \quad (3.2)$$

El predictor lineal del modelo lineal generalizado puede incluir términos tales como las variables explicativas originales, potencias y transformaciones de estas variables dependiendo de la función enlace que se utilice. En el modelo de regresión lineal se modela el valor esperado como una combinación lineal de las variables explicativas. Sin embargo, en muchos experimentos reales, esta relación no es adecuada, por lo que es necesario la inclusión de una función que relacione el valor esperado con las variables explicativas. Esta función se denomina función de enlace y se simboliza por $g(\mu_i)$. La función enlace que transforma el valor esperado a la escala del predictor lineal es:

$$g(\mu_i) = \eta_i = X_i\beta \quad (3.3)$$

Donde X_i representa las p variables explicativas para el i -ésimo individuo con $i = 1, \dots, n$.

La elección de la función enlace no siempre resulta obvia, dado que pueden existir diferentes funciones enlace aplicables a un problema particular, de forma que se debe decidir cuál es la más apropiada en cada caso. En particular, para cada elemento de la familia exponencial de distribuciones existe una función enlace denominada canónica o natural, que consiste en relacionar el parámetro natural directamente con el predictor lineal, esto es:

$$\theta_i = \theta(\mu_i) = \eta_i = X_i\beta \quad (3.4)$$

$$g(\mu_i) = \theta(\mu_i) \quad (3.5)$$

De esta forma, para las distribuciones exponenciales se tienen las siguientes funciones enlace canónicas definidas, expuestas en la Tabla 3.2.

Tabla 3.2. Funciones enlace canónicas de las distribuciones de familia exponencial (Alcaide, 2015).

Distribución	Función Enlace Canónica
Binomial	$\theta(\mu_i) = \ln\left(\frac{\mu_i}{1 - \mu_i}\right)$
Binomial Negativa	$\theta(\mu_i) = \ln\left(\frac{\alpha\mu_i}{1 + \alpha\mu_i}\right)$
Gamma	$\theta(\mu_i) = -\frac{1}{\mu_i}$
Normal	$\theta(\mu_i) = \mu_i$
Poisson	$\theta(\mu_i) = \log(\mu_i)$

Los tipos de modelos lineales generalizados de la familia exponencial que tienen especial interés para efectos de esta memoria, se diferencian fundamentalmente en la naturaleza de la variable dependiente que desean explicar. Debido a que existen variables de tipo categórica, variables discretas de conteo o variables continuas, existen ciertas distribuciones de la familia exponencial más o menos adecuadas para abordar cada problema. Estas diferencias quedan explícitas en la Tabla 3.3 presentada a continuación.

Tabla 3.3. Tipos de modelos lineales generalizados de la familia exponencial (Alcaide, 2015).

Modelos para respuestas binarias (Bernoulli o Binomial)	Permiten considerar variables objetivo de tipo 0-1 (tener/no tener una característica, éxito/fracaso, entre otros).
Modelos para datos de conteo (Poisson o Binomial Negativa)	Permiten considerar y analizar el comportamiento de variables de conteo (número de accidentes, nacimientos, individuos de una especie, entre otros), frente a los valores del conjunto de variables explicativas.
Modelos para variables respuesta continua (Normal o Exponencial)	Permiten considerar y analizar el comportamiento de variables continuas y cuantitativas (ganancias, concentraciones de compuestos químicos, tiempo de vida, entre otros), frente a los valores del conjunto de variables explicativas.

En particular, los modelos que tienen especial importancia para esta investigación, son los denominados modelos para datos de conteo. En estos modelos, la distribución de Poisson es el modelo clásicamente utilizado, pero últimamente ha adquirido relevancia el uso de la distribución Binomial Negativa demostrado en estudios como Miranda-Moreno y Nosal (2011) y Schmiedeskamp y Zhao (2016). Se denominan variables de recuento o conteo a aquellas variables que determinan el número de sucesos o eventos que ocurren en una misma unidad de observación en un intervalo espacial o temporal definido. A partir de esta definición propuesta por Lindsey (1995), se derivan las características principales de una variable de recuento, que a diferencia de una variable cuantitativa continua, son su naturaleza discreta y no negativa. Para efectos de esta investigación, se han implementado modelos predictivos basados en las distribuciones Poisson y Binomial Negativa, con objeto de explicar la variable de recuento llamada conteos diarios de bicicletas. En el Anexo B de este documento, se detalla la derivación matemática de estos dos modelos de regresión, utilizando para ello la explicación recopilada por Alcaide (2015).

3.2. Bosques aleatorios (*Random Forests*)

Un bosque aleatorio es una combinación de árboles de decisión o de regresión formulada por Breiman (1996), que se utiliza habitualmente para tratar problemas de regresión o de clasificación. Su objetivo es predecir el valor de una variable continua Y o la pertenencia a una categoría a partir de un conjunto de variables predictoras X . Un árbol de regresión es un árbol binario que se compone de un nodo raíz, de nodos internos y de nodos terminales llamados hojas, representados por círculos en la Figura 3.1. Cada nodo interno representa un subconjunto de las observaciones y un *test* binario que resulta en la generación de dos nodos hijos R_i y R_d . El algoritmo de bosques aleatorios de regresión consiste en dividir el espacio de las variables X_i en particiones homogéneas para obtener en la salida del árbol una predicción de la variable Y en función de los valores de las variables predictoras. Por ello, en cada nodo del árbol se busca una variable X_j y un umbral S para dividir el nodo en dos nodos hijos R_i y R_d (Chardon, 2019).

$$R_i(j, S) = \{X | X_j < S\} \quad (3.6)$$

$$R_d(j, S) = \{X | X_j \geq S\} \quad (3.7)$$

El objetivo es encontrar los valores de j y S que minimizan la heterogeneidad resultante de la división de un nodo R , es decir maximizar la expresión:

$$\Delta i(s, R) = i(R) - P_i(R_i) - P_d(R_d) \quad (3.8)$$

Donde P_i y P_d son la proporción de observaciones que pertenecen a los nodos R_i y R_d respectivamente, mientras que $i(R)$ es una función que mide la “impureza” del nodo R . Las funciones más comunes para i son la entropía de Shannon, el índice de Gini o la varianza.

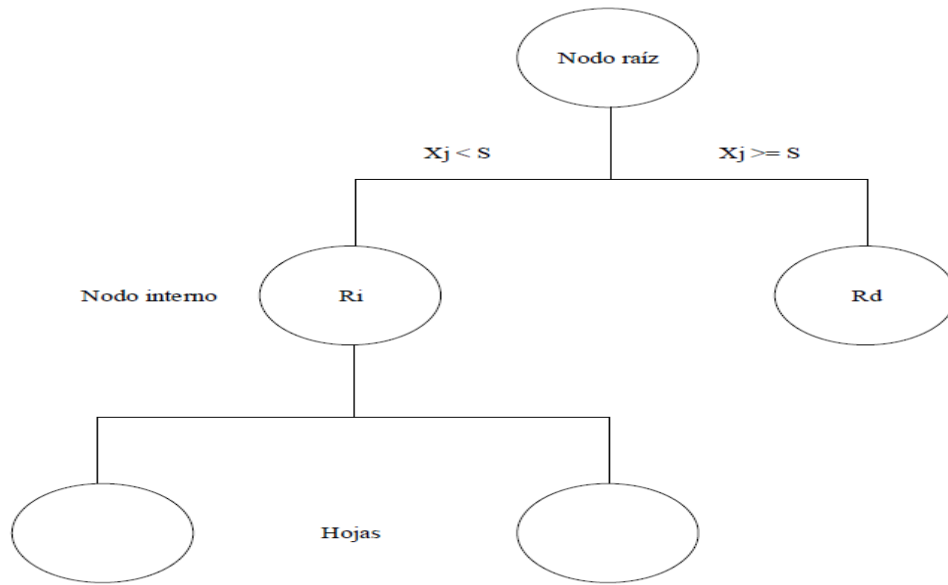


Figura 3.1. Esquema de un árbol de regresión.

El proceso de partición continúa hasta que se cumple con una condición de parada, es decir, cuando no es posible dividir las observaciones que pertenecen a un nodo en conjuntos más homogéneos, o cuando la cantidad de observaciones en el nodo alcanza un límite inferior definido por el modelador, dando lugar a un conjunto de nodos terminales (Gavilá, 2006). El árbol de regresión presenta la desventaja de ser inestable cuando existe ruido en las variables. Para proveer estabilidad al algoritmo se aplica un método de *bagging*, el cual consiste en promediar los resultados de muchos árboles de regresión para obtener un modelo muy estable y que además no presente problemas de sobreajuste. A partir de la base de datos original se generan n muestras aleatorias elaboradas según el método *bootstrap* (Efron, 1979) destinadas a construir n árboles de regresión. Al dividir los nodos de los árboles en nodos hijos se utiliza un conjunto de $m < p$ variables predictoras en vez de las p disponibles y finalmente se promedian los resultados entregados por todos los árboles para obtener la predicción final (Chardon, 2019).

A partir de un bosque aleatorio es posible obtener la importancia de las variables X_i en la predicción. Para medir la importancia de una variable X_m , Breiman (2001) propone promediar sobre los N_T árboles del bosque las impurezas $\Delta i(R)$ de los nodos donde se utiliza la variable X_m para dividir el nodo R , esto es:

$$Imp(X_m) = \frac{1}{N_T} \sum_T \sum_{R \in T: v(R)=X_m} p(R) \Delta i(R) \quad (3.9)$$

3.3. Modelos multinivel

En áreas como la investigación epidemiológica o educacional, es muy común que la estructura de los individuos en estudio esté organizada en forma jerárquica. Esto es, los individuos están agrupados en unidades de nivel más alto, que a su vez también pueden estar agrupadas en otras unidades. Ejemplos de estas estructuras son los alumnos en diferentes clases, pacientes en hospitales, entre otros. En estudios que utilizan este tipo de modelos se dispone de información acerca de los individuos como de su contexto. Por lo tanto, se puede hablar de distintas jerarquías de información disponible: el nivel 1, y por otra, el nivel 2, que corresponde al contexto o grupo al que pertenece el individuo. Los individuos pertenecientes a un mismo contexto tenderán a ser más similares en su comportamiento entre sí que respecto a los que pertenezcan a distintos contextos. Esta similitud entre los individuos dentro de los grupos establece una estructura de correlación intracontextual que impide el cumplimiento de la hipótesis de independencia sobre las que están basados los modelos de regresión tradicionales e invalida por tanto sus métodos de estimación, lo que se traduce en estimaciones incorrectas de los errores estándar (De la Cruz, 2008). Si se realiza un modelo de regresión lineal en estructuras de datos jerarquizadas y no se aplica un modelo multinivel, las pruebas estadísticas estándar se inclinan fuertemente en la suposición de la independencia de las observaciones. Si esta suposición es violada como ocurre normalmente con datos multinivel, los estimadores de los errores estándares de las pruebas estadísticas convencionales son mucho más pequeños y los resultados son falsamente significativos. Dos conceptos fundamentales que suponen un gran aporte de los modelos multinivel, son los conceptos de coeficiente fijo y coeficiente aleatorio. En los modelos de regresión clásicos los parámetros que se estiman son el intercepto y las pendientes. Desde esa perspectiva clásica, estos coeficientes se asumen como fijos, es decir, comunes a todos los sujetos y son estimados a partir de los datos. Los coeficientes aleatorios, sin embargo, son variables y se distribuyen según una función de probabilidad. En una estructura multinivel los coeficientes del primer nivel son tratados como aleatorios en el segundo nivel. En los modelos multinivel se permite a los grupos desviarse de la solución central o global, tanto en el intercepto como en la pendiente. O, lo que es lo mismo, los modelos multinivel están compuestos por dos partes, una parte común a todos los contextos, que

es la parte fija, y otra que varía y que se estima a través de la varianza en los distintos niveles. Los modelos multinivel son, en esencia, ampliaciones de los modelos de regresión lineal clásicos, de tal forma que en realidad son varios modelos lineales para cada nivel. Así, los modelos del primer nivel están relacionados con uno de segundo nivel en el que los coeficientes de regresión del nivel 1 se regresan en un segundo nivel de variables explicativas y así sucesivamente para los diferentes niveles (Murillo, 2008). Si se ejemplifica con una ecuación de regresión lineal sencilla con dos variables independientes se obtiene:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad (3.10)$$

Si se permite que el intercepto pueda tomar diferentes valores en función de un segundo nivel, la ecuación aplicada a la explicación de los conteos diarios de bicicleta de este estudio es:

$$y_{ij} = \beta_{0j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \varepsilon_{ij} \quad (3.11)$$

$$\beta_{0j} = \beta_0 + \mu_{0j} \quad (3.12)$$

Donde:

- y_i : Conteos diarios en un día i en una ciclovía j .
- ε_{ij} : Error distribuido normalmente con varianza constante e igual a $\sigma_{\varepsilon 0}^2$.
- β_{0j} : Promedio de y de la ciclovía j -ésima.
- β_0 : Promedio global del modelo.
- μ_{0j} : Efecto aleatorio asociado a la ciclovía j -ésima, de media 0 y varianza $\sigma_{\mu 0}^2$.

Los efectos de las variables predictoras pueden ser considerados como constantes o variables en función del problema a tratar. Debido a que las ciclovías de Temuco deberían tener un comportamiento similar frente a la meteorología, las variables de dicho nivel pueden asumirse constantes. Si el estudio fuera realizado utilizando más ciudades, estos efectos podrían ser considerados como aleatorios.

CAPÍTULO 4

METODOLOGÍA

Capítulo 4. Metodología

En el presente capítulo se detallan las tareas y procedimientos realizados para la elaboración de los modelos de predicción. A modo de resumen se expone en la Figura 4.1 un esquema conceptual que explica la metodología seguida.

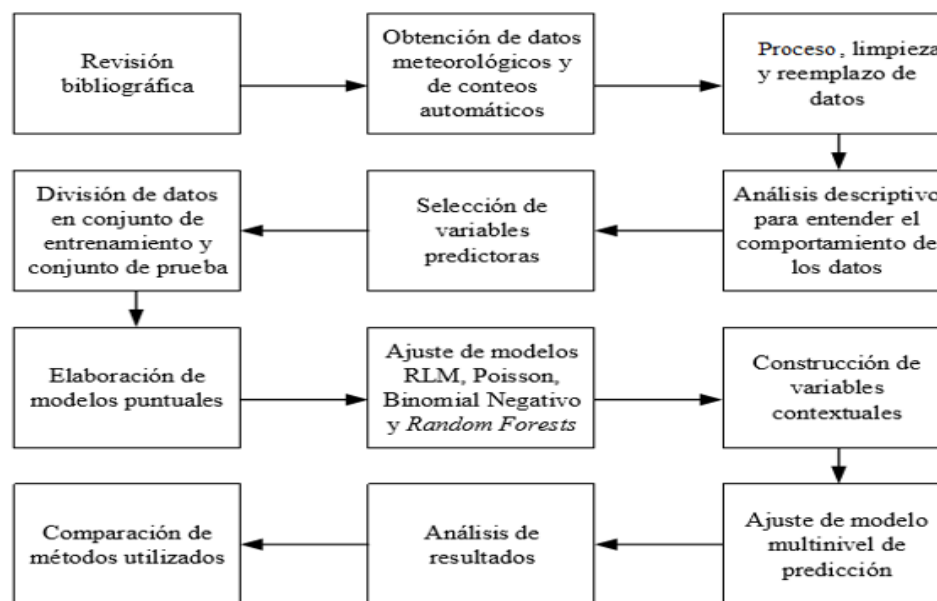


Figura 4.1. Esquema conceptual de la metodología seguida.

A continuación se realiza una descripción de los aspectos metodológicos más importantes realizados durante esta investigación.

4.1. Datos

La variable dependiente discreta y positiva que se busca explicar en este estudio corresponde a los conteos diarios de bicicletas. La información real de esta variable fue obtenida mediante los datos pasivos recolectados por contadores automáticos ubicados en la ciudad de Temuco, Chile. Las ubicaciones de los dispositivos están expuestas en el Capítulo 5. Temuco es una comuna y ciudad del sur de Chile, capital de la provincia de Cautín y de la Región de La Araucanía. Está localizada en las coordenadas $38^{\circ}45'S$ y $72^{\circ}40'O$. La ciudad tiene una población total de 282.415 personas, y un total de 104.757 viviendas (Instituto Nacional de Estadísticas, 2017). La información de

conteos fue obtenida a través de la plataforma web Eco-Visio (<https://www.eco-visio.net/ECovisio/>), servicio contratado por una oficina gubernamental (SERVIU Araucanía). La información utilizada en este estudio cubre tres años de datos en aquellos contadores que poseen dicho volumen de información (Hochstetter – O’Higgins, Hochstetter – Pucón, Pablo Neruda – Olimpia y Pablo Neruda – Javiera Carrera 2) y decrece en función del período de datos disponibles para los otros contadores. El período de datos considerado corresponde al período comprendido entre el 31/07/2016 y el 31/07/2019. Los datos referidos a variables meteorológicas fueron recolectados a través de diferentes fuentes, entre las cuales se encuentran la Dirección General de Aguas (DGA), la Dirección Meteorológica de Chile (DMC) y el Centro de Ciencia del Clima y la Resiliencia (CR2). Estos datos están disponibles en las páginas web de dichos organismos, pudiendo ser descargados en formato csv. Se descargaron las bases de datos de diversas variables meteorológicas como: humedad relativa, altura de nieve, precipitación, presión, radiación, temperatura máxima, temperatura media, temperatura mínima y viento. Las variables de calidad del aire obtenidas del Sistema de Información Nacional de Calidad del Aire (SINCA), fueron descargadas para el período mencionado anteriormente y corresponden al material particulado 2.5 y material particulado 10. La estación meteorológica empleada en este estudio es la estación Maquehue, Temuco Ad., debido al volumen de datos disponible para el período de análisis y la cercanía geográfica con la ciudad. La estación se ubica en las coordenadas -38.76° y -72.63° de latitud y longitud respectivamente. Una vez descargadas las bases de datos de las potenciales variables predictoras para el período de análisis de este estudio en forma horaria, estas fueron procesadas para ser transformadas a nivel agregado (diario). Los conteos diarios de los diversos contadores fueron agregados a una base de datos conformada por el autor, incorporándose a las variables meteorológicas en función de la fecha. Además de las variables de tipo meteorológico, fue necesario incorporar en la información recolectada aquellos días anómalos, como feriados, fines de semana o eventos que pudiesen alterar los conteos diarios de un día en condiciones normales. De igual forma, aquellos días en que existen datos faltantes en alguna variable meteorológica o de calidad del aire fueron categorizados como valores faltantes o NA. Por su parte, para la elaboración de modelos que incluyen variables de tipo contextual, se ha utilizado la base de datos referida a usos de suelo del Servicio de Impuestos Internos (SII) e información de la Encuesta Origen-Destino (Secretaría de Planificación de Transporte, 2014). La conformación de la base de datos de este proyecto fue preparada en Microsoft Excel (Microsoft Corporation, 2016) para posteriormente ser utilizada en R (R Core Team, 2019). Se procedió también a realizar un análisis

descriptivo de la muestra de datos, tanto a nivel de la variable dependiente como también sobre las potenciales variables predictoras a utilizar en los modelos. La evaluación de la confiabilidad de los datos provistos por los contadores automáticos fue realizada mediante conteos manuales en cuatro puntos de análisis. Estas mediciones manuales fueron realizadas durante dos horas en cada punto, totalizando un total de ocho horas de conteos, con el objetivo de contrastarlo posteriormente con los conteos informados por el visor Eco-Visio. Para ello, se debe asumir que el contador manual no cometió ningún error en su medición y de esta forma establecer el porcentaje de error absoluto entre ambos valores. Las mediciones de efectividad fueron realizadas el día 10/10/2019 y el día 11/10/2019, en los cuatro puntos seleccionados para la elaboración de modelos individuales que están explícitos en el capítulo siguiente. Estos registros fueron desarrollados en dos días diferentes con condiciones cercanas a la media y sin eventos extremos asociados a variables meteorológicas. Además, los días seleccionados corresponden a días laborales de semana.

4.2. Selección de variables

Debido a la correlación existente entre ciertas variables independientes y los posibles problemas de multicolinealidad que esto suele producir (Del Valle y Guerra, 2012), se decide optar por solo incluir algunas variables candidatas, de forma de no introducir ruido a los modelos. En ese sentido, en la mayoría de modelos puntuales desarrollados se presenta el siguiente *set* de variables independientes seleccionadas: La temperatura máxima diaria medida en °C fue la variable escogida para representar la temperatura. Esta decisión se fundamenta en la presencia de dicho factor en varios estudios revisados en la literatura, como el desarrollado por Schmiedeskamp y Zhao (2016). La temperatura mínima diaria no fue incluida en los modelos debido a que esta variable puede adquirir valores tanto negativos como positivos, por lo que con el objetivo de mantener la simplicidad de los modelos, se prefiere representar la temperatura por tan solo una variable. Además, al intentar incorporarla en los modelos, el signo resultante de esta variable no resulta ser el correcto. Con el fin de representar el efecto no lineal de la temperatura, se incluye una variable de temperatura máxima diaria al cuadrado, de esta forma, cuando la temperatura máxima adquiere valores muy elevados o muy bajos, los conteos diarios decrecen, hecho que se condice con la realidad. La precipitación acumulada diaria fue escogida para representar los efectos de la precipitación a nivel diario en la ciudad de Temuco. La humedad relativa diaria fue incorporada en los modelos con el objetivo de añadir efectos rezagados de la precipitación que pudiesen explicar

descensos en los conteos diarios debido a las condiciones de humedad posterior a lluvias fuertes. Dicha variable ha sido incluida en diversos estudios anteriores como Patil *et al.*, (2015) y Schmiedeskamp y Zhao (2016). Por su parte, la radiación media diaria fue incorporada con el objetivo de complementar la variable de temperatura, siendo un símil de variables como las horas de sol diarias las cuales han sido incluidas en estudios como Wadud (2014). También, se incluyen en la especificación de los modelos variables de tipo categórica, entre las cuales se destacan las variables de fin de semana y feriados. Estas variables categóricas son utilizadas como variables *dummy* o ficticias, en las cuales mediante esta herramienta es posible caracterizar la presencia o ausencia de una característica. Las variables de este tipo se incluyen en los modelos debido a su presencia en prácticamente todos los modelos revisados en el Capítulo 2, en los cuales se observan generalmente variaciones en los patrones de conteos diarios entre los días laborales de semana y días de fin de semana. La última variable que conforma el conjunto de variables predictoras es el día secuencial, de forma de explicar el crecimiento del uso de la bicicleta a lo largo del tiempo, el cual no es atribuible a las variables meteorológicas.

Con respecto a las variables de tipo contextual, se ha utilizado la longitud de cada ciclovía en la cual se ubica su respectivo contador, junto con un índice de entropía de uso de suelo. La justificación de esta elección se basa en la literatura revisada, en la cual se observa que un gran porcentaje de estudios incluidos en la Tabla 2.1 incluyen estas dos variables predictoras. El índice de entropía de uso de suelo ha sido construido a nivel de zona geográfica de la Encuesta Origen-Destino (Secretaría de Planificación de Transporte, 2014) y a nivel de manzana censal (Instituto Nacional de Estadísticas, 2017). Se utilizó el software Geo-Segregation Analyzer (Apparicio y Martori, 2014) para obtener el valor espacial del índice de entropía para Temuco, con el objetivo de implementarlo en un modelo que considere factores predictores del entorno construido. En dicho *software*, fue necesario crear un archivo vectorial (ESRI Shapefile) con los datos de uso de suelo de cada unidad geográfica, obtenidos de la base de datos del Servicio de Impuestos Internos (SII). Luego, se calculó el índice de Theil (Theil, 1972) para cada una de las zonas geográficas predefinidas en dicho archivo vectorial. El proceso de cálculo se realiza en forma automática en el *software*. Los resultados de los mapas de índice de entropía para la ciudad pueden ser consultados en la Figura A.64 y la Figura A.65, tanto a nivel de zona de Encuesta Origen-Destino (Secretaría de Planificación de Transporte, 2014), como de manzana censal (Instituto Nacional de Estadísticas, 2017). El resumen de variables seleccionadas para los diversos modelos se muestra en la Tabla 4.1.

Tabla 4.1. Selección de variables para la implementación de modelos de predicción.

Variable	Abreviación	Descripción	Unidad	Tipo
Conteos diarios	C_D	Número de bicicletas por día	[N° Pasadas/Día]	Respuesta
Temperatura máxima	T_MAX	Temperatura máxima diaria	[°C]	Meteorológica
Temperatura máxima al cuadrado	T_MAX ²	Temperatura máxima diaria al cuadrado	[°C] ²	Meteorológica
Precipitación	PRECIP	Precipitación diaria acumulada	[mm]	Meteorológica
Humedad relativa	HUM	Humedad relativa media diaria	[%]	Meteorológica
Radiación	RAD	Radiación media diaria	[W/m ²]	Meteorológica
Material particulado	MP_2.5	Material particulado medio diario MP 2.5	[µg/m ³]	Calidad del aire
Fin de semana	WEEKEND	El día es un fin de semana (Verdadero/Falso)	-	Categórica
Feriados	HOLIDAY	El día es un feriado (Verdadero/Falso)	-	Categórica
Día secuencial	DAY	Días de estudio numerados en forma secuencial	[N° día de estudio]	-
Longitud ciclovía	LONGITUD	Longitud de la ciclovía	[km]	Contextual
Índice de entropía	ENTROPY	Indicador de la entropía en el uso de suelo	[0,1]	Contextual

4.3. Implementación de modelos

Para la elaboración de modelos predictivos se empleó el ambiente estadístico R (R Core Team, 2019). Para el desarrollo de los modelos se utilizó un 80% de los datos disponibles (*training set*). El 20% de datos reales que quedan fuera de los modelos pertenecen al conjunto de prueba o *testing set*, el cual se utiliza para realizar la validación de los modelos. La división de datos en proporción 80-20% es recomendada por diversos estudios (Suthaharan, 2015). Se utilizó la función `sample.split` perteneciente al paquete `caTools` (Tuszynski, 2019) para realizar la división y la función `set.seed` para asegurar la reproducibilidad de la partición. Para la implementación de modelos de regresión lineal múltiple se empleó la función `lm`. Por su parte, para los ajustes que

emplean la regresión Poisson y la regresión Binomial Negativa, se utilizó la función `glm` y `glm.nb` respectivamente, que emplean la definición de modelos lineales generalizados, utilizando la función enlace logarítmica en ambos casos. Por su parte, para el cálculo de regresiones utilizando la técnica de bosques aleatorios o *Random Forests*, se utilizó el paquete `randomForest` (Liaw y Wiener, 2002), mientras que para la implementación de un modelo multinivel se empleó el paquete `lme4` (Bates *et al.*, 2015). Finalmente, con el objetivo realizar predicciones utilizando el conjunto de prueba, se usó la función `predict`. Los gráficos personalizados que se presentan posteriormente en el Anexo A fueron elaborados con la función `ggplot2` (Wickham, 2016).

Para ajustar modelos de predicción puntuales que expliquen los conteos diarios de ciclistas basados en predictores meteorológicos, se han seleccionado cuatro puntos representativos de la urbe, alejados geográficamente. Para ello, se ha escogido un contador correspondiente a cada macrozona de la ciudad, con excepción de aquellas macrozonas en las cuales no existen contadores automáticos. Para la zona poniente se seleccionó el contador Ex Vía Férrea 1, mientras que para la zona centro se escogió el contador San Martín – Mackenna. Por su parte, para la zona nor - poniente (Fundo el Carmen) se selecciona el contador Los Músicos – Los Creadores. Finalmente, con el objetivo de abarcar el eje de Pablo Neruda que posee un gran número de pasadas diarias, se escoge el contador Pablo Neruda – Olimpia, el cual forma parte de la macrozona centro - poniente. En la zona norte de la ciudad no existen contadores automáticos con datos disponibles en el visor Eco-Visio, lo cual imposibilita incluirlos en el análisis. El fin de realizar modelos puntuales en diferentes puntos geográficos de la ciudad es observar el impacto de las variables meteorológicas sobre los conteos diarios. Los puntos seleccionados para la elaboración de dichos modelos puntuales se presentan a continuación en el siguiente capítulo.

CAPÍTULO 5

ANÁLISIS DESCRIPTIVO

Capítulo 5. Análisis Descriptivo

En el presente capítulo se realiza un análisis descriptivo y exploratorio de los datos referidos a conteos automáticos y factores meteorológicos.

5.1. Información general

Actualmente existen 17 contadores automáticos distribuidos espacialmente en la ciudad de Temuco. En el mapa de la Figura 5.1 se muestra la ubicación espacial de estos dispositivos. En los últimos meses se han instalado otros contadores en la ciudad, sin embargo, estos aún no se encuentran disponibles en la plataforma Eco-Visio para ser utilizados.

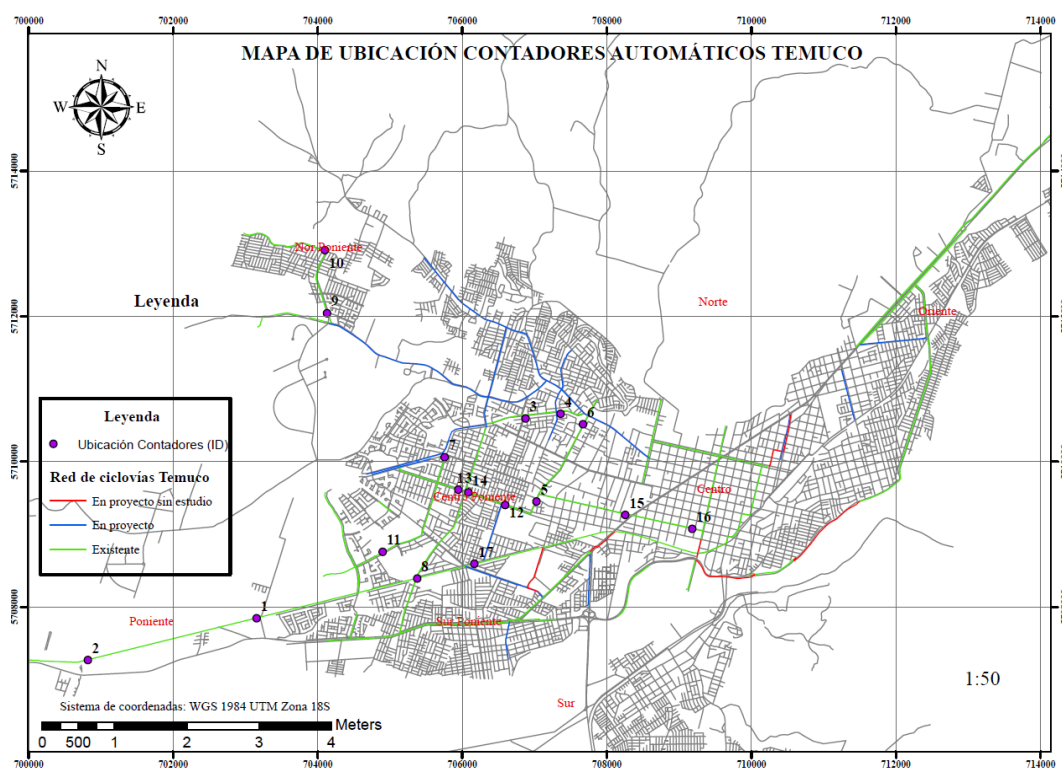


Figura 5.1. Mapa de ubicación contadores automáticos de Temuco.

La simbología representa con puntos cada uno de los contadores, junto con la red de ciclovías de Temuco, diferenciando las ciclovías existentes, en proyecto y en estudio sin proyecto. Cada uno de los contadores automáticos y sus respectivas coordenadas geográficas (Este y Norte), están resumidas en la Tabla 5.1.

Tabla 5.1. Coordenadas geográficas de los contadores automáticos de Temuco.

ID	Nombre Contador	Este	Norte
1	Ex Vía Férrea 1	703155.7	5707843.1
2	Ex Vía Férrea 2	700821.9	5707268.2
3	Gabriela Mistral - Andes	706877.1	5710592.5
4	Gabriela Mistral - Recreo	707361.6	5710657.6
5	Hochstetter - O'Higgins	707027.4	5709450.2
6	Hochstetter - Pucón	707671.9	5710514.0
7	Inés de Suarez	705756.7	5710059.6
8	Javiera Carrera - Simón Bolívar	705377.8	5708390.1
9	Los Creadores - Raúl Sáez	704130.4	5712041.7
10	Los Músicos - Los Creadores	704100.4	5712908.7
11	Los Pablos - Fundación	704899.4	5708757.9
12	Pablo Neruda - Olimpia	706593.9	5709402.6
13	Pablo Neruda - Javiera Carrera 1	705945.2	5709612.7
14	Pablo Neruda - Javiera Carrera 2	706085.2	5709575.7
15	San Martín - Caupolicán	708256.0	5709266.0
16	San Martín - Mackenna	709182.6	5709073.9
17	Simón Bolívar - Las Encinas	706167.6	5708594.0

Sistema de Coordenadas: WGS 1984 UTM Zona 18S.

Los contadores Ex Vía Férrea 1 y Ex Vía Férrea 2 comprenden el segmento de red que une la ciudad de Temuco con la localidad de Labranza. Por su parte, los contadores Los Creadores – Raúl Sáez y Los Músicos – Los Creadores se encuentran ubicados en la zona de Fundo el Carmen. En la zona centro de la ciudad, solo se cuenta con información del contador San Martín – Mackenna. Los demás contadores se encuentran en su mayoría en la zona centro - poniente de la ciudad, abarcando los ejes de Javiera Carrera, Pablo Neruda, Gabriela Mistral, Hochstetter, Los Pablos, entre otros. Estos dispositivos almacenan datos desde el momento en que se instalan, por lo cual, existen diferentes períodos de datos disponibles para la variedad de contadores de la ciudad, basados en la puesta en marcha de cada contador. De esta forma, existen puntos de la ciudad que poseen datos disponibles desde el año 2016 y otros desde el año 2017. La distribución de conteos diarios en la ciudad de Temuco en forma acumulada considerando todo el período de datos disponible utilizado en este estudio (26/07/2016 - 31/07/2019) está representada en la Figura 5.2.

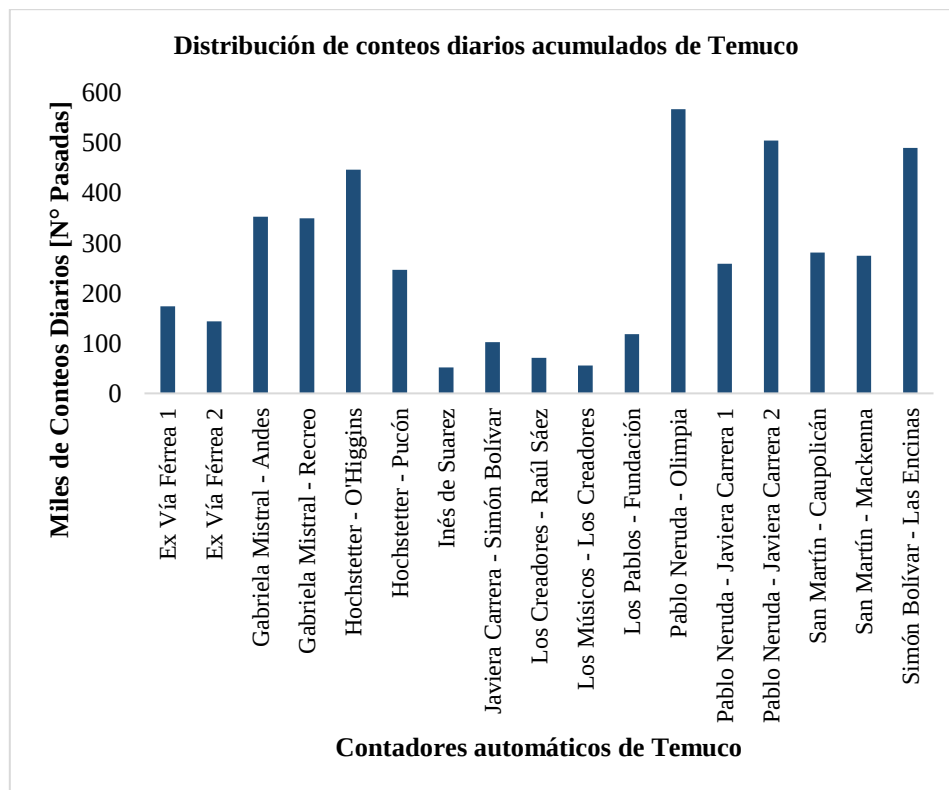


Figura 5.2. Distribución de conteos diarios acumulados en Temuco.

De la Figura 5.2 se puede observar la predominancia en valores absolutos de pasadas acumuladas en los contadores Pablo Neruda – Olimpia, Pablo Neruda – Javiera Carrera 2 y Simón Bolívar – Las Encinas. Los contadores que poseen menor número de pasadas acumuladas son Los Creadores – Raúl Sáez, Los Músicos – Los Creadores e Inés de Suárez.

Se exponen a continuación en la Figura 5.3 los puntos geográficos seleccionados para la elaboración de los modelos puntuales basados en factores meteorológicos. Estos sectores geográficos fueron seleccionados por las razones expuestas en el Capítulo 4, específicamente en la sección de implementación de los modelos.

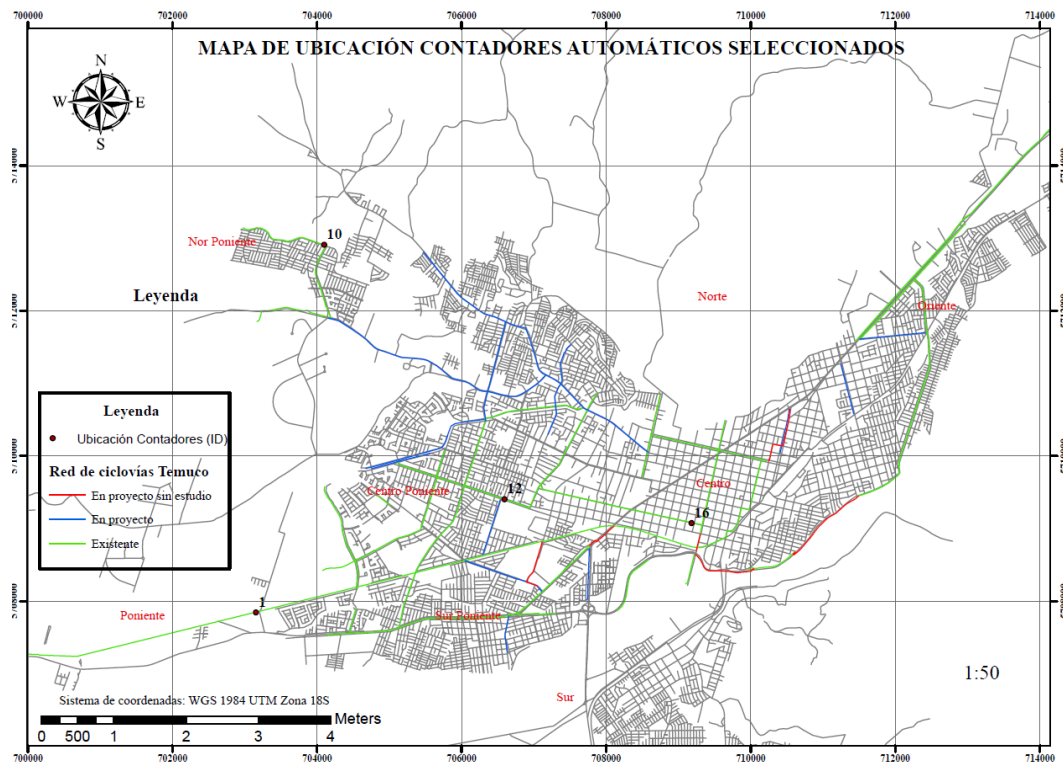


Figura 5.3. Mapa de ubicación de contadores automáticos seleccionados.

5.2. Análisis de datos meteorológicos

Como fue mencionado en el capítulo Metodología, se ha elaborado una extensa base de datos con una serie de variables explicativas que tentativamente pudiesen explicar los conteos diarios de ciclistas. Entre las múltiples variables seleccionadas se destacan la temperatura media, temperatura mínima, temperatura máxima, precipitación, radiación, humedad relativa y el material particulado, expresado por el indicador MP 2.5. Otras variables como la nubosidad asociada al cielo cubierto o el número de horas de sol diarias se han descartado durante el proceso de elaboración de la base de datos debido a la ausencia de los datos en las estaciones meteorológicas de la ciudad.

Con el fin de observar la posible correlación de las potenciales variables predictoras con los conteos diarios y la correlación entre variables explicativas, se realiza una exploración de los datos en la cual se cruzan las diferentes variables a estudiar versus los conteos automáticos diarios. Este procedimiento se lleva a cabo en cada uno de los cuatro puntos a estudiar en los modelos puntuales, exponiéndose los resultados de un contador particular en el Anexo A (Figura A.1 a Figura A.8). La Figura A.1 demuestra el efecto de la temperatura media sobre los conteos diarios. La Figura A.2

muestra que la temperatura mínima no posee una clara correlación con la variable a explicar. Por su parte, la Figura A.3 demuestra una correlación positiva de esta variable sobre los conteos diarios, de mayor magnitud que la expresada en la Figura A.1. La Figura A.4 indica que para valores de precipitaciones menores a 5 mm existe una importante variabilidad de la variable de respuesta, registrándose conteos diarios de alta y baja magnitud, lo cual sugiere que dicha diferencia debe ser explicada por otras variables. La Figura A.5 refleja un comportamiento similar al de las variables de temperatura media y temperatura máxima sobre los conteos diarios. A su vez, la Figura A.6 muestra que la humedad tiene un claro efecto negativo sobre la variable dependiente. Finalmente, tanto la Figura A.7 como la Figura A.8 poseen un comportamiento similar al mostrado en la Figura A.4, sobre todo cuando los valores de dichas variables de contaminación ambiental son bajos. El cruce de variables para los demás contadores seleccionados posee un comportamiento similar a este contador, pero con coeficientes de correlación levemente diferentes debido al dispar número de datos disponibles en cada caso. A modo de resumen, se muestra la Figura 5.4 que representa la matriz de correlación de las potenciales variables explicativas a utilizar en este estudio. Esta relación permite conocer a través del coeficiente de correlación de Pearson, la relación con la variable objetivo y otros predictores.

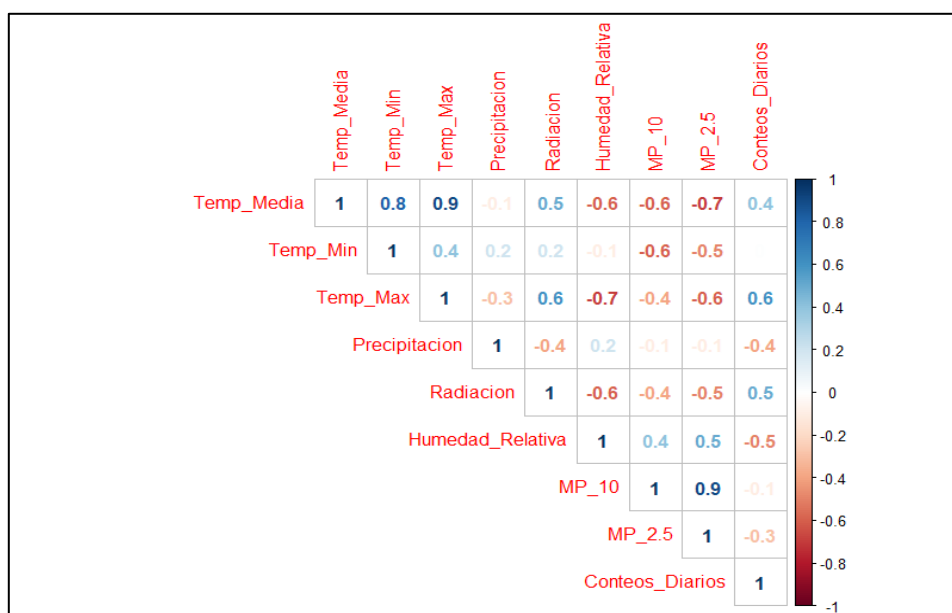


Figura 5.4. Matriz de correlación de variables meteorológicas del contador Ex Vía Férrea 1.

La Figura 5.4 resume los valores del coeficiente de correlación lineal de Pearson entre las variables explicativas. La paleta de colores de la simbología representa con color azul aquellos valores en que el coeficiente de correlación tiene signo positivo, es decir, aquellas combinaciones en que existe una correlación positiva entre las variables. De igual forma, en color rojo se destacan aquellas relaciones negativas existentes entre cada par de factores. Se destaca la relación positiva entre los diferentes tipos de temperaturas. Esto sugiere que la alta correlación de las variables pudiese traer problemas si se intenta modelar con los tres factores de temperatura simultáneamente, por lo tanto, se debe optar por una variable que represente dicho efecto. También es clara la alta correlación entre las dos variables de material particulado (0.9), por lo cual, debido a que estas variables explican el mismo fenómeno, se debe escoger una de dichas variables. Las diferencias entre que tan significativa es esta relación quedan determinadas por la magnitud del coeficiente de correlación de Pearson, el cual fluctúa en el intervalo $[-1,1]$. En ese sentido, las relaciones cuyo coeficiente de correlación son mayores, son conformadas por la relación entre la temperatura máxima y la temperatura media (0.9), junto con la relación entre el material particulado MP 2.5 y el material particulado MP 10 (0.9). Por su parte, las correlaciones negativas de mayor magnitud son aquellas relaciones entre la temperatura máxima y la humedad relativa (-0.7) y la relación entre la temperatura media y el material particulado MP 2.5 (-0.7). En la Figura A.9 se representa el histograma de temperatura media. La distribución de frecuencias de esta variable demuestra que este factor sigue una distribución normal, concentrándose la mayoría de datos entre los 10 y 15 °C. La Figura A.10 expone el histograma de temperatura máxima. Este gráfico de frecuencia sigue al igual que el caso anterior una distribución normal pero con un leve sesgo o asimetría hacia la derecha. La Figura A.11 describe el histograma de temperatura mínima. En ella, se observa una distribución normal en la cual existe presencia de valores negativos, debido a la naturaleza de la variable en la cual pueden existir valores de temperatura bajo cero grados Celsius. En la Figura A.12 se exhibe el histograma de precipitación acumulada. La mayoría de valores se encuentran bajo los 5 mm. La Figura A.13 expresa el histograma de la variable radiación, la cual no posee una distribución característica como las variables predecesoras. La mayoría de los valores se sitúan entre los 500 y 1000 W/m². Finalmente, tanto la Figura A.14 como la Figura A.15, los cuales representan los histogramas de material particulado MP 10 y MP 2.5 respectivamente, presentan una distribución claramente asimétrica hacia a la derecha, similar a una distribución Binomial Negativa. La mayoría de valores en estos factores se concentran bajo los 50 µg/m³.

5.3. Análisis de datos pasivos

Todos los países y ciudades que estudian el comportamiento de los ciclistas lo hacen con la finalidad de promover el uso, justificar y evaluar proyectos, optimizar diseños, mejorar la seguridad, alcanzar mejores niveles de servicio, gestionar políticas públicas, entre otras razones. En Chile, monitorear el comportamiento de los ciclistas ya es un estándar, así definido por el Manual de Vialidad Cicloinclusiva del MINVU (Urbanismo y Territorio, 2016). A continuación, se realiza una descripción de los datos pasivos referentes a los contadores automáticos tanto a nivel general como también específicamente para los puntos seleccionados para la estimación de modelos puntuales de predicción.

5.3.1. Crecimiento de la demanda

Una de las principales inquietudes que surge al analizar la información de datos pasivos, es comprobar si existe crecimiento en el uso del modo bicicleta en la ciudad. Para responder a dicha pregunta, se han graficado los cuatro puntos individuales correspondientes a los cuatro contadores automáticos seleccionados y se ha aplicado un promedio móvil semanal para suavizar la curva. En la gráfica de la Figura 5.5, se puede concluir que este ascenso no es uniforme en todos los puntos.

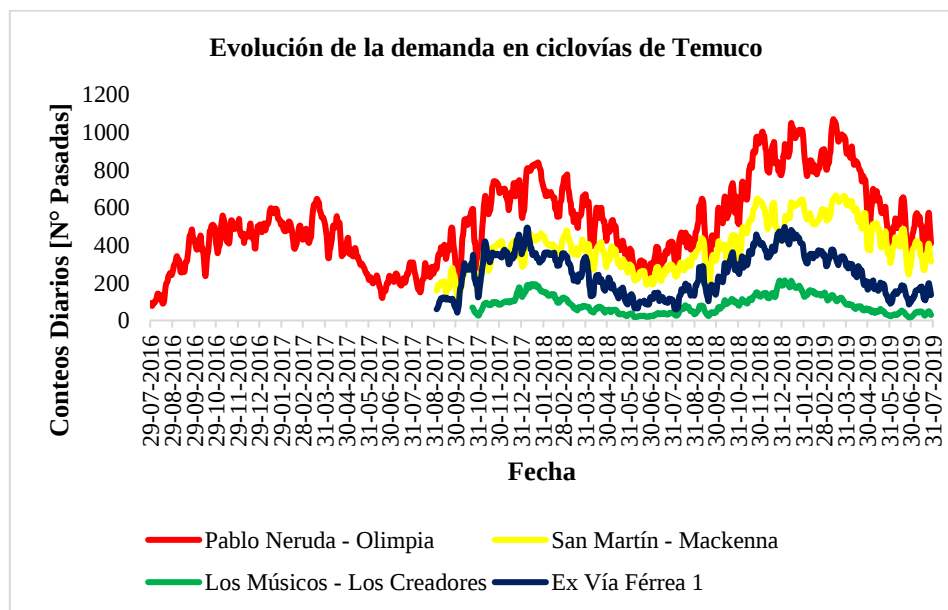


Figura 5.5. Evolución de la demanda en ciclovías de Temuco.

Al analizar el gráfico, se observa un crecimiento del número de pasadas diarias en los contadores Pablo Neruda – Olimpia y San Martín – Mackenna. A su vez, en los contadores Ex Vía Férrea 1 y Los Músicos – Los Creadores, el valor se mantiene relativamente constante en el período analizado. Se puede verificar como los conteos diarios aumentan en la estación de verano y descienden de forma abrupta durante invierno, siguiendo un comportamiento sinusoidal. De los puntos analizados en este estudio, el contador Pablo Neruda – Olimpia es el que mayor número de pasadas y crecimiento anual registra. Sin embargo, las tasas de crecimiento de las ciclovías de la ciudad no son uniformes. Con el objetivo de mostrar dicha diferencia, se presenta un mapa con las tasas de crecimiento de los 17 contadores automáticos de la ciudad de Temuco en el período comprendido entre los años 2018 y 2019. De esta forma se puede identificar espacialmente los polos de crecimiento más grandes de la ciudad.

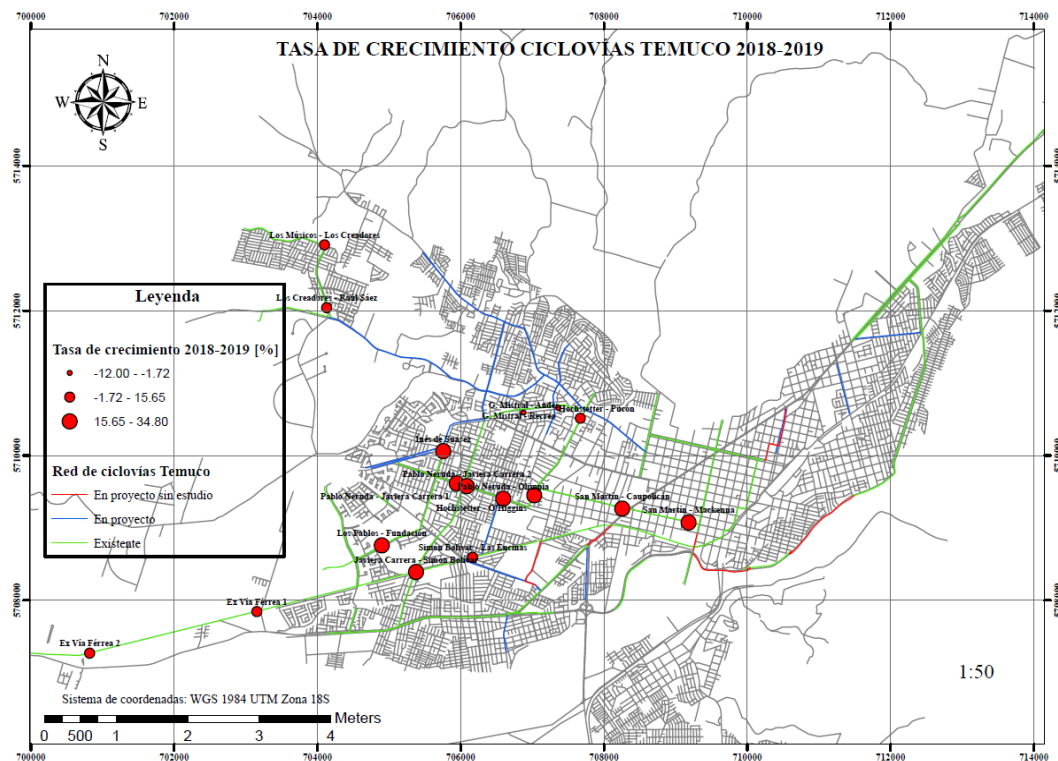


Figura 5.6. Tasas de crecimiento [%] contadores automáticos 2018-2019.

En el mapa anterior se identifican con diferentes tamaños de puntos los diversos intervalos de la tasa de crecimiento anual. Se destaca el gran aumento en comparación con el año anterior de los contadores Javiera Carrera – Simón Bolívar, Inés de Suarez, Pablo Neruda – Javiera Carrera 1 y

San Martín – Caupolicán, con tasas de crecimiento superiores al 30%. La mayoría de contadores se sitúan en el intervalo de crecimiento comprendido entre 15 y 30%, mientras que los contadores Gabriela Mistral – Andes y Gabriela Mistral – Recreo registran tasas de decrecimiento en el período considerado. La razón de este descenso en el número de pasadas puede ser atribuida a factores externos, debido a los proyectos de infraestructura vial que se han desarrollado en las cercanías de estos contadores en los últimos meses.

La Figura 5.7 representa a través de un mapa los valores absolutos de pasadas en el año 2019 en cada uno de los contadores. La simbología de puntos de color azul es creciente en función del aumento del número de pasadas.

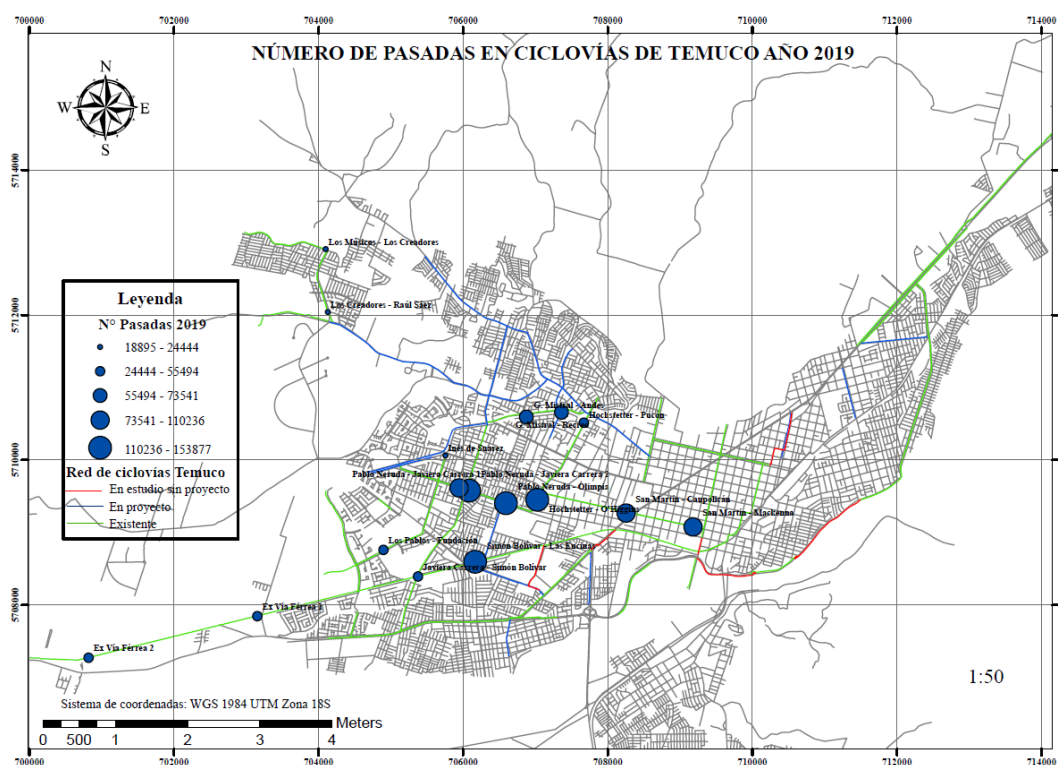


Figura 5.7. Valor absoluto de pasadas en contadores automáticos año 2019.

Si bien las tasas de crecimiento en ciertos puntos pueden ser elevadas, dicho valor no refleja si efectivamente dichas ciclovías se utilizan con alta demanda en el número absoluto de pasadas. En el mapa anterior se muestran las grandes diferencias en valor absoluto de los conteos diarios. Actualmente, el contador que registra mayor número de pasadas en el presente estudio es Simón Bolívar – Las Encinas, con 153877 pasadas acumuladas entre el 1/01/2019 y el 31/07/2019. El

contador con menor uso es Los Músicos – Los Creadores con 18895 pasadas acumuladas en el mismo período. Este mapa demuestra espacialmente que los conteos absolutos se concentran en el sector centro - poniente de la ciudad, específicamente en los ejes de Pablo Neruda, Javiera Carrera, Simón Bolívar y San Martín, descendiendo a medida que se alejan de esta zona.

Si se compara el número de conteos diarios del año 2019 con el año 2018 en el mismo período de tiempo, es posible identificar la evolución de la demanda, como se muestra en la Figura 5.8, del cual es posible extraer la tasa de crecimiento anual mostrada anteriormente en la Figura 5.7.

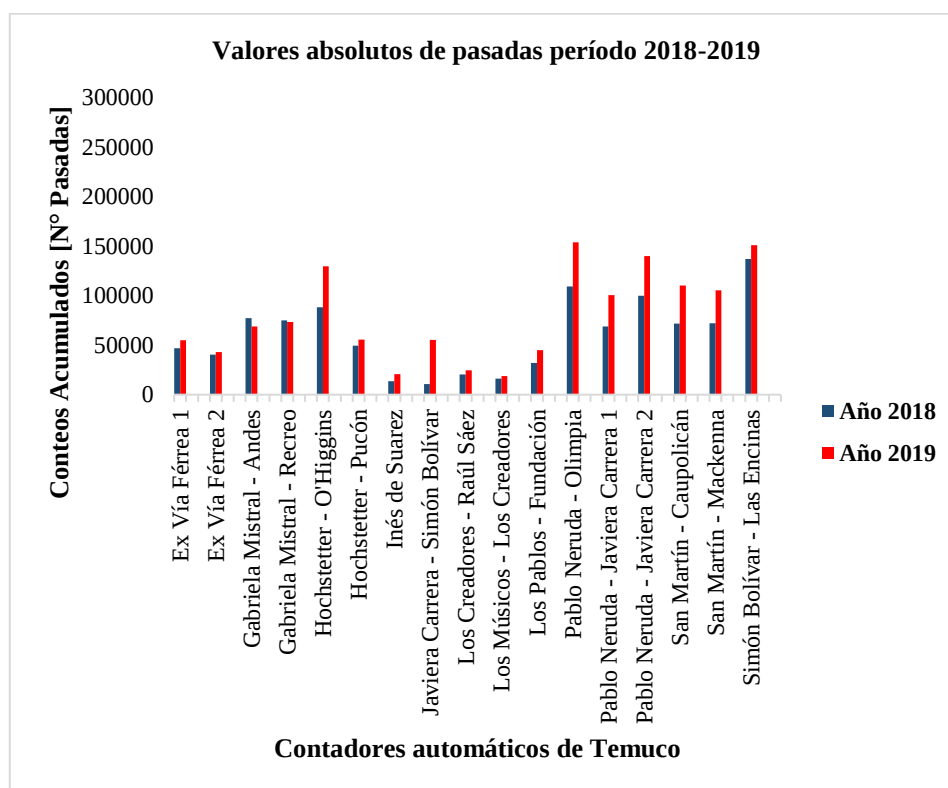


Figura 5.8. Valores absolutos de pasadas años 2018-2019.

De la Figura 5.8, se desprende el aumento sostenido del número de pasadas en la mayoría de contadores de la ciudad, con la excepción de los contadores Gabriela Mistral – Andes y Gabriela Mistral – Recreo, como fue mencionado con anterioridad en la Figura 5.6.

Otro aspecto importante que genera variabilidad, es el uso de las ciclovías en días de fin de semana. Para el caso de estudio presente, la Figura 5.9 presenta el uso de la infraestructura de ciclovías en días de fin de semana. Al analizar el número de pasadas, se pueden identificar aquellos puntos de

la ciudad que poseen mayor uso los días de fin de semana que los días de semana, pudiendo explicar un comportamiento o uso de la ciclovía recreacional.

5.3.2. Variabilidad en días de fin de semana

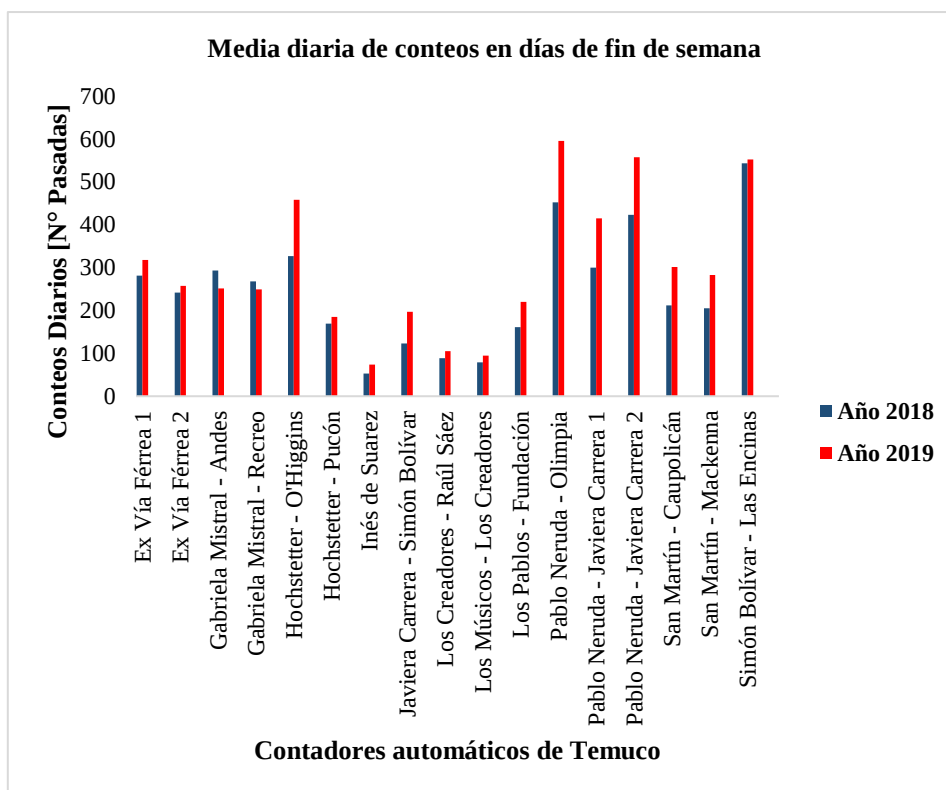


Figura 5.9. Media diaria de conteos en días de fin de semana.

La media diaria de pasadas solo en días de fin de semana sigue un comportamiento similar al gráfico anteriormente presentado (Figura 5.8) con un aumento en el número de conteos diarios, solo disminuyendo en aquellos contadores ya explicados con anterioridad. Por su parte, en la Figura 5.10 se presenta el valor de la media diaria para cada uno de los contadores de la ciudad, a nivel de día laboral de semana y de fin de semana, de manera de comparar su comportamiento en ambos casos.

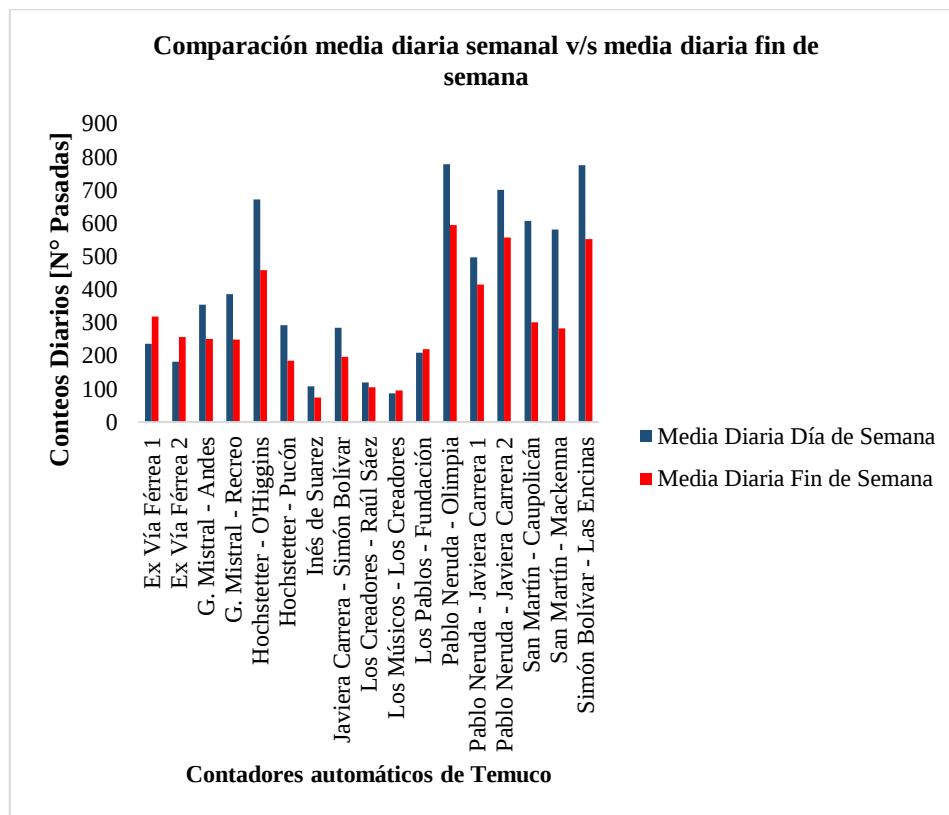


Figura 5.10. Comparación entre la media diaria de conteos semanales y de fin de semana.

Si se analiza la Figura 5.10, se puede concluir que cuatro contadores registran medias diarias mayores los días de fin de semana que los días de semana. Estos contadores son Ex Vía Férrea 1, Ex Vía Férrea 2, Los Músicos – Los Creadores y Los Pablos – Fundación. Esto sugiere que la dinámica de viajes y los motivos de estos, son diferentes en estos puntos de la ciudad.

5.3.3. Variabilidad horaria

Por último, en la Figura 5.11 se muestra la variabilidad a nivel horario de los conteos automáticos, con el objetivo de conocer los patrones de viaje de la ciudad en este modo de transporte. Con dicho fin, se han graficado el número de pasadas en forma horaria observadas en un día laboral de semana en cada uno de los cuatro puntos de análisis. Se ha seleccionado un día representativo, con ausencia de precipitaciones, y con variables meteorológicas cercanas a la media de cada factor (temperatura máxima, humedad relativa y radiación), con el objetivo de representar fielmente un día común de viaje en Temuco. Además, para efectos de la gráfica presentada, se han sumado los registros de ambos sentidos del flujo.

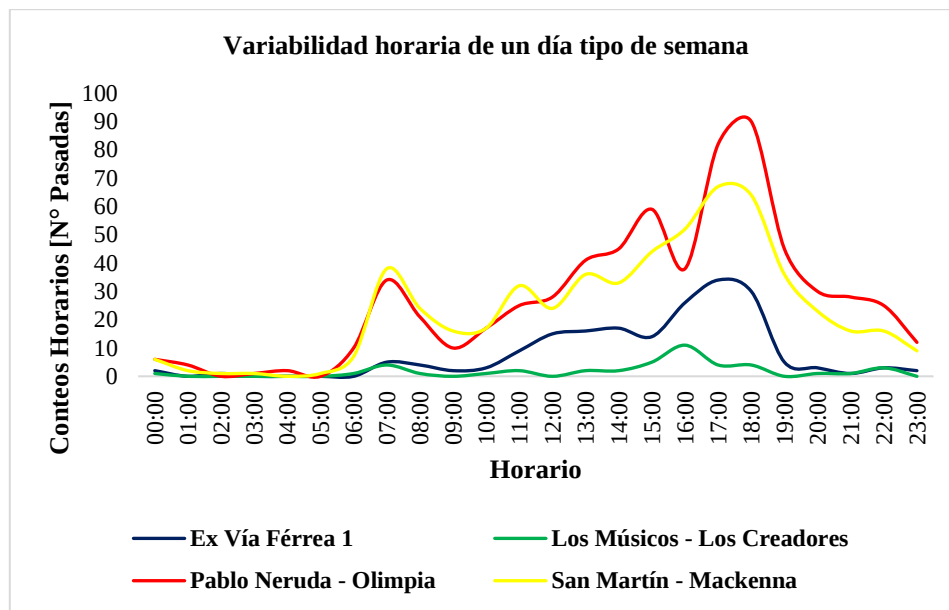


Figura 5.11. Variabilidad horaria de un día laboral de semana.

La Figura 5.11 representa la variabilidad horaria de un día tipo de semana. Es posible identificar las diferencias en el número de pasadas al igual que en la información anterior junto con la posibilidad de conocer el patrón horario de viajes en bicicleta en estos cuatro puntos de análisis. Se presenta un *peak* importante entre 6:00 y 8:00 de la mañana y posteriormente otro *peak* en horario de la tarde entre 16:00 y 19:00, los cuales coinciden con el horario que corresponde a los viajes obligatorios por motivos de trabajo. Resulta interesante destacar el hecho de que las ciclovías suelen tener mayor uso durante las horas punta de la tarde por sobre las horas de punta mañana. Esto significa que los ciclistas utilizan las ciclovías con mayor frecuencia para regresar de sus trabajos, lo que indica que para llegar a ellos utilizan con mayor frecuencia las vías destinadas a automóviles. Entre las 10:00 y las 15:00 se observan ascensos y descensos, pero con una tendencia general al alza.

CAPÍTULO 6

RESULTADOS Y ANÁLISIS

Capítulo 6. Resultados y Análisis

El siguiente capítulo presenta cada uno de los resultados obtenidos en esta memoria. Se evalúa la efectividad de los contadores automáticos y además se presentan los resultados tanto a nivel de los modelos predictivos puntuales basados en factores meteorológicos, como del modelo multinivel que incluye factores meteorológicos y variables contextuales. Finalmente, se presenta un análisis de los modelos ajustados junto con la validación de éstos.

6.1. Evaluación de la efectividad de los contadores automáticos

El primer objetivo específico de esta investigación es evaluar la efectividad de los contadores automáticos de bicicleta de la ciudad de Temuco. Para realizar dicha acción y determinar el porcentaje de precisión con el cual se trabaja, fue necesario comparar los resultados informados por el visor Eco-Visio y los conteos manuales realizados en cada punto de análisis. Los resultados están resumidos en la Tabla 6.1.

Tabla 6.1. Efectividad de contadores automáticos de bicicleta en Temuco.

Contador	Conteo Manual	Conteo Automático	Error (%)
Ex Vía Férrea 1	40	34	15.0%
Los Músicos – Los	28	25	10.7%
Creadores	14	12	14.3%
Pablo Neruda –	14	13	7.2%
Olimpia	66	60	9.1%
San Martín -	80	71	11.3%
Mackenna	59	52	11.9%
	52	48	9.4%

Error promedio = 11.1%

De la Tabla 6.1 se puede desprender que los contadores automáticos subestiman el número de ciclistas que pasan en forma horaria y por consiguiente en forma diaria en la ciudad de Temuco. Si bien la precisión general es del orden del 90%, en todas las mediciones realizadas los contadores automáticos informaron menos conteos de los reales a través de su plataforma Eco-Visio. La razón de esta diferencia puede ser atribuida a la configuración de sensibilidad con la cual fueron instalados estos dispositivos, la cual puede traer como consecuencia que algunas bicicletas no sean

contabilizadas. También, dicha diferencia puede deberse a problemas que ocurren cuando dos ciclistas pasan por los dos carriles al mismo tiempo sobre el contador, hecho que ocurría habitualmente en los contadores con más frecuentación. Estos problemas son mencionados por Nordback y Janson (2010) en su estudio de efectividad de contadores automáticos en la ciudad de Boulder, Estados Unidos. La Figura 6.1 representa la curva de dispersión entre los conteos manuales y los conteos automáticos realizados en los puntos de interés para la realización de los modelos puntuales. A pesar de la existencia de una discrepancia entre el valor real y el valor del dato pasivo, la relación entre ambos datos resulta ser prácticamente lineal, con un alto valor del coeficiente de determinación. La efectividad de un 90% que subestima los conteos reales de ciclistas coincide con estudios internacionales realizados para el mismo modelo de contador, como muestra el estudio realizado por New Zealand Transport Agency (2009).

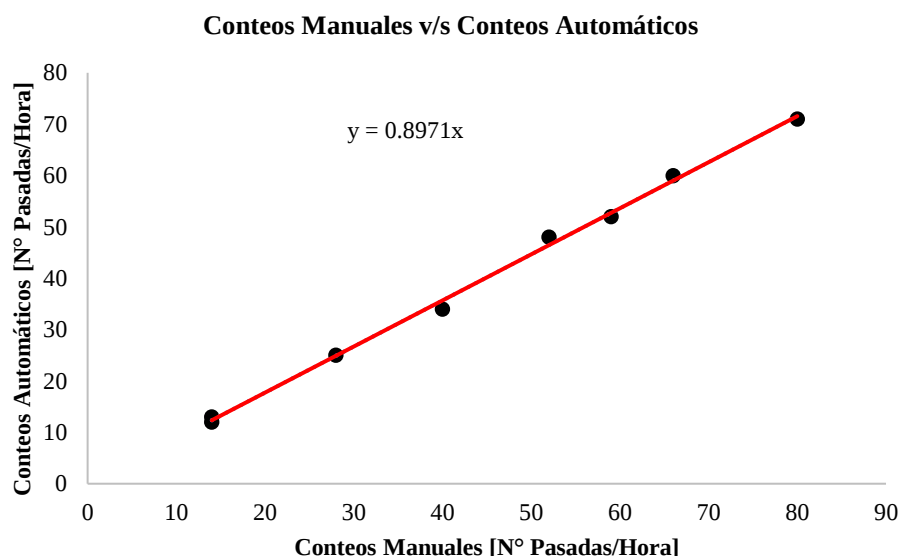


Figura 6.1. Relación entre conteos manuales y conteos automáticos.

6.2. Modelos de predicción puntuales

Con el fin de evaluar el efecto de factores meteorológicos en los conteos diarios de ciclistas en la ciudad de Temuco, se desarrollan una serie de modelos puntuales que simulan el número de conteos diarios en ciertos lugares espaciales de la ciudad. Para ello, se han seleccionado cuatro puntos de la ciudad distantes geográficamente unos con otros. Los contadores utilizados para la elaboración de modelos de predicción puntuales son los siguientes: Ex Vía Férrea 1, Los Músicos – Los Creadores, Pablo Neruda – Olimpia y San Martín – Mackenna. Se utilizan cuatro modelos diferentes para cada punto de análisis. Dichos ajustes basan en la regresión lineal múltiple, la regresión Poisson, la regresión Binomial Negativa y el algoritmo de regresión de *Machine Learning* de Bosques Aleatorios o *Random Forests*.

6.2.1. Modelo puntual Ex Vía Férrea 1

El contador Ex Vía Férrea 1 se sitúa en el comienzo de la ciclovía que conecta la ciudad de Temuco con la localidad de Labranza. La ciclovía Temuco - Labranza fue construida en su primera etapa durante el año 2008 y se ubica en la ex vía ferroviaria que conectaba antiguamente estos dos puntos. La Tabla 6.2 muestra el resumen de la variable dependiente y de cada una de las potenciales variables meteorológicas predictoras consideradas para un total de 651 días.

Tabla 6.2. Resumen base de datos contador Ex Vía Férrea 1.

Resumen	TEMP_ MEDIA	TEMP_ _MIN	TEMP_ _MAX	PRECIP	RAD	HUM	MP_10	MP_2.5	C_D
Mínimo	0.20	-3.80	2.70	0.00	37.20	25.30	7.00	1.00	7.00
Cuartil 1	8.60	2.75	14.35	0.00	551.40	75.20	21.00	6.00	123.00
Mediana	11.60	6.20	18.00	0.10	848.20	82.70	30.00	13.00	236.00
Media	11.35	5.96	18.74	2.69	801.00	81.21	42.82	27.91	250.60
Cuartil 3	14.50	9.20	22.55	1.90	1062.00	88.60	54.00	38.00	355.50
Máximo	24.70	16.30	41.10	74.80	1356.50	99.70	225.00	194.00	732.00

En este resumen, se presentan los valores mínimos, máximos, la media, mediana y los valores referentes al primer y tercer cuartil de la muestra, para cada una de las variables de la base de datos. La media de conteos diarios es de aproximadamente 250 pasadas, mientras que la mediana corresponde a 236. La Figura A.66 que representa el histograma de los conteos diarios para este

contador, muestra que la distribución de frecuencias es similar a una distribución Binomial Negativa, con un sesgo o asimetría hacia la derecha.

6.2.1.1. Modelo Regresión Lineal Múltiple Contador Ex Vía Férrea 1

El primer modelo desarrollado para el contador Ex Vía Férrea 1, es un modelo de regresión lineal múltiple. Para la estimación de este modelo, se utilizan todas las variables explicativas recolectadas en la base de datos, para posteriormente eliminar aquellas variables que tengan alta correlación entre ellas o no sean significativas. El criterio arbitrario para eliminar una variable es que su valor p sea mayor o igual al 5%. El método anterior fue realizado implementando una función en R que realiza el procedimiento conocido como Eliminación Hacia Atrás o *Backward Elimination*. Esta herramienta se basa en estimar un primer modelo con todas las variables recolectadas, para luego eliminar paso a paso una variable en función de su grado de significancia (Groißböck *et al.*, 2004). Sin embargo, al realizar este proceso es necesario tener en cuenta que el hecho de introducir variables que tienen alta correlación entre ellas, puede introducir problemas de ruido en la significancia estadística de dichas variables (Lizana, 2017). La Tabla 6.3 muestra la especificación del modelo de regresión múltiple para este contador.

Tabla 6.3. Modelo regresión lineal múltiple contador Ex Vía Férrea 1.

Variables	Estimador	Error Estándar	Valor t	Valor p	Significancia
INTERCEPTO	-93.26	64.59	-1.44	0.14	
TEMP_MAX	28.12	3.27	8.59	< 2e-16	***
TEMP_MAX ²	-0.45	0.07	-5.98	4.07e-09	***
PRECIP	-4.34	0.56	-7.69	7.15e-14	***
HUM	-1.73	0.58	-2.97	0.003	**
RAD	0.11	0.01	6.35	4.75e-10	***
WEEKEND	90.39	8.00	11.28	< 2e-16	***
HOLIDAY	121.26	20.45	5.93	5.59e-09	***
DAY	0.07	0.02	3.89	0.0001	***

Código de significancia: 0 '***' 0.001 '**' 0.01 '*' '.' 0.1

R² ajustado = 0.69

Se puede observar que todas las variables explicativas son altamente significativas ya que poseen un valor p menor o igual al 5%. El valor del coeficiente de determinación ajustado del presente Estimación de modelos de predicción de tránsito diario de bicicletas en base a datos pasivos, incorporando variables meteorológicas y contextuales para la ciudad de Temuco

modelo es de 0.69. Los signos de las variables explicativas son los siguientes: temperatura máxima (+), temperatura máxima al cuadrado (-), precipitación (-), humedad (-), radiación (+), fin de semana (+), feriados (+) y día secuencial (+). Los signos de las variables explicativas son correctos, ya que la temperatura máxima y la radiación tienen una relación positiva con los conteos diarios y a su vez, la precipitación y la humedad tienen una relación negativa o que inhibe los conteos diarios. Por su parte, las variables dicotómicas de fin de semana y feriados refuerzan los conteos diarios en esta zona espacial de la ciudad, lo cual significa que existiendo las mismas condiciones meteorológicas, el número de pasadas será mayor durante los días de fin de semana o feriados. Finalmente, la variable de día secuencial explica positivamente el aumento sostenido en los conteos diarios a lo largo del tiempo, cambio que obedece a factores externos a las variables meteorológicas.

Al evaluar los datos reales en el modelo desarrollado se obtiene una predicción de conteos diarios. En la Figura A.16, se observa en puntos de color rojo la predicción de conteos diarios del modelo de regresión lineal múltiple para el contador Ex Vía Férrea 1. La curva que conforma la predicción de puntos mostrada sigue una curva sinusoidal, en el cual los conteos diarios aumentan en período estival y decrecen abruptamente durante invierno. Este comportamiento es clásico en la literatura, destacando el estudio realizado por Fournier *et al.*, (2017). Un detalle importante que se observa en la gráfica, es la presencia de predicciones de signo negativo. Este hecho viola la restricción de que la variable dependiente de conteo sea cero o de signo positivo, siendo uno de los grandes problemas que posee este tipo de modelo. El problema surge debido a la naturaleza de la variable dependiente, la cual, al ser una variable de recuento o conteo, solo está conformada por valores positivos. La regresión lineal múltiple no posee dicha restricción, por lo cual, no es la mejor opción para explicar dicha variable objetivo. La Figura A.17 ilustra con puntos de color rojo las predicciones del modelo de regresión lineal múltiple para este contador (al igual que la Figura A.16) junto con los valores de los conteos diarios reales simbolizados con puntos de color verde. Se observa una buena primera aproximación, sin embargo, el modelo no es capaz de capturar los valores extremos y además predice valores negativos. En la Figura A.18 se representa un gráfico que explica las diferencias entre la predicción de conteos diarios y los conteos diarios reales para este modelo. Cada dato simbolizado por puntos negros tiene asociado en el eje de las abscisas los conteos diarios reales, mientras que la predicción está asociada al eje de las ordenadas. La línea azul representa la línea de tendencia, mientras que la sombra de la línea simboliza el intervalo de

confianza al 95%. Este tipo de gráficos realiza una comparación inmediata en los ejes x-y para las predicciones y los datos reales, siendo una herramienta de visualización más potente. Se desprende que el modelo ajustado solo agrupa algunas predicciones en la cercanía de la sombra de la línea, hecho que confirma que el modelo no es el más adecuado. Los valores negativos en la predicción constituyen los principales valores atípicos u *outliers* para este modelo. La Figura A.67 demuestra que el histograma de frecuencias posee un sesgo hacia la izquierda y que no se asemeja al histograma de conteos reales mostrado anteriormente en la Figura A.66.

6.2.1.2. Modelo Poisson Contador Ex Vía Férrea 1

El segundo modelo de regresión implementado para este contador es aquel basado en el modelo lineal generalizado que utiliza la distribución Poisson. Se realiza un proceso análogo al anterior (*Backward Elimination*) para obtener el mejor modelo posible, el cual queda explícito en la Tabla 6.4.

Tabla 6.4. Modelo Poisson contador Ex Vía Férrea 1.

Variables	Estimador	Error Estándar	Valor z	Valor p	Significancia
INTERCEPTO	3.073	0.05	57.76	< 2e-16	***
TEMP_MAX	0.2017	0.003	58.67	< 2e-16	***
TEMP_MAX²	-0.003767	0.00007	-49.92	< 2e-16	***
PRECIP	-0.04632	0.0009	-49.75	< 2e-16	***
HUM	-0.005769	0.0004	-13.74	< 2e-16	***
RAD	-0.00042	0.00001	27.00	< 2e-16	***
MP_10	-0.0004032	0.0005	-3.31	0.0009	***
WEEKEND	0.3412	0.005	58.59	< 2e-16	***
HOLIDAY	0.4822	0.01	34.51	< 2e-16	***
DAY	0.0003166	0.00001	19.90	< 2e-16	***

Código de significancia: 0 '***' 0.001 '**' 0.01 '*' '.' 0.1

Desviación Residual = 10031 en 510 grados de libertad

AIC = 13757

Los signos de las variables independientes son los siguientes: temperatura máxima (+), temperatura máxima al cuadrado (-), precipitación (-), humedad (-), radiación (+), material particulado (-), fin de semana (+), feriados (+) y día secuencial (+). Los signos del presente modelo son los correctos

y coinciden con los signos del modelo anterior expresado en la Tabla 6.3. El presente modelo resulta ser el más adecuado dentro de la familia de modelos Poisson desarrollados, ya que todas las variables poseen alta significancia estadística y además el criterio de información de Akaike (AIC) es el menor entre la variedad de modelos probados. El estadístico desviación, bajo las hipótesis del modelo correcto, sigue una distribución chi-cuadrado con 510 grados de libertad. Este estadístico tiene un valor de 10031, el cual, si se evalúa en la siguiente expresión:

$$\frac{\text{Desviación Residual}}{\text{Grados de Libertad}} = \frac{10031}{510} = 19.67 > 1 \quad (6.1)$$

Se obtiene que la relación es igual a 19.67. Si dicho valor es mayor que 1, indica una fuerte sobredispersión de los datos. Los modelos de regresión Poisson sobredispersos algunas veces conducen a la confusión de que algunas variables explicativas contribuyen significativamente al modelo, cuando esto no ocurre realmente (Alcaide, 2015). La Figura A.19 representa la predicción de conteos diarios bajo el modelo Poisson en función del tiempo. En la Figura A.20 se simboliza la predicción del modelo y los conteos diarios reales, con simbología de puntos rojos y verdes respectivamente. Se observa que a diferencia del modelo de regresión lineal múltiple, el modelo Poisson no predice valores de signo negativo en la variable dependiente, lo cual supone una importante mejora en la estimación del modelo con respecto al desarrollado con anterioridad. Esto ocurre debido a la naturaleza y restricciones matemáticas que posee la distribución de Poisson, las cuales lo hacen más adecuado para modelos de este tipo en que se busca explicar variables de conteo. La Figura A.21 representa un gráfico que explica las diferencias entre la predicción de conteos diarios y los conteos diarios reales para este modelo. Al igual que en la Figura A.19 y Figura A.20, se verifica que el modelo solo predice valores positivos. Si se compara con la Figura A.18, se observa que la cantidad de puntos se agrupa en mayor grado a la línea de tendencia, por lo que sin dudas el modelo resulta ser mucho más adecuado que el anterior. El histograma de valores predichos bajo este modelo expresado en la Figura A.68 resulta ser similar al histograma de valores reales expresado en la Figura A.66.

6.2.1.3. Modelo Binomial Negativo Contador Ex Vía Férrea 1

Siguiendo las sugerencias de la literatura, se estima un modelo de regresión utilizando la distribución Binomial Negativa, con el objetivo de evitar la sobredispersión de los datos que desarrollan habitualmente los modelos tipo Poisson.

Tabla 6.5. Modelo binomial negativo contador Ex Vía Férrea 1.

Variables	Estimador	Error Estándar	Valor z	Valor p	Significancia
INTERCEPTO	3.41	0.26	13.06	< 2e-16	***
TEMP_MAX	0.15	0.013	11.49	< 2e-16	***
TEMP_MAX ²	-0.002	0.0003	-8.48	< 2e-16	***
PRECIP	-0.038	0.002	-15.10	< 2e-16	***
HUM	-0.006	0.002	-2.57	0.01	*
RAD	0.0004	0.00007	6.68	2.33e-11	***
WEEKEND	0.346	0.03	10.73	< 2e-16	***
HOLIDAY	0.61	0.08	7.47	7.66e-14	***
DAY	0.0004	0.00008	5.36	7.99e-08	***

Código de significancia: 0 '***' 0.001 '**' 0.01 '*' '.' 0.1

Desviación Residual = 550.77 en 511 grados de libertad

AIC = 5893.10

2 x Log-Verosimilitud = -5873.14

En la Tabla 6.5 se observa que todas las variables tienen un grado importante de significancia, con un valor p menor o igual al 5%. Los signos de las variables son los siguientes: temperatura máxima (+), temperatura máxima al cuadrado (-), precipitación (-), humedad (-), radiación (+), fin de semana (+), feriados (+) y día secuencial (+). El criterio de información de Akaike (AIC) es de 5893.10, magnitud que corresponde al menor valor hallado para todos los modelos desarrollados bajo este tipo de regresión. En la *ecuación (6.2)* se evalúa la sobredispersión de los datos, la cual resulta ser aproximadamente 1, lo cual verifica que los datos no presentan sobredispersión a diferencia de lo expresado en la *ecuación (6.1)* para el modelo Poisson.

$$\frac{\text{Desviación Residual}}{\text{Grados de Libertad}} = \frac{550.77}{511} = 1.07 \cong 1 \quad (6.2)$$

La Figura A.22 representa la predicción de conteos diarios bajo el modelo Binomial Negativo. En ella, es posible verificar que el modelo solo predice valores estrictamente positivos al igual que el modelo Poisson. En la Figura A.23 se representa la predicción del modelo y los conteos diarios reales, con simbología de puntos rojos y verdes respectivamente. Se destaca el hecho de que este modelo, si bien tiene una forma y rendimiento similar al modelo Poisson, es capaz de capturar algunos valores atípicos extremos presentes en los meses de verano, y como demuestra la ecuación (6.2), no se presenta sobredispersión en los datos. La Figura A.24 representa el gráfico de comparación x-y para la variable predicha y los datos reales de la variable. El histograma de la predicción de conteos diarios queda explícito en la Figura A.69, presentando un comportamiento asimétrico hacia la derecha, siendo similar al histograma de frecuencias del modelo Poisson.

6.2.1.4. Modelo *Random Forests* Contador Ex Vía Férrea 1

A continuación se implementa un método diferente a los abordados habitualmente en la literatura. El método utilizado en esa ocasión incorpora un algoritmo de *Machine Learning*, denominado Bosques Aleatorios o *Random Forests*.

Tabla 6.6. Modelo *Random Forests* contador Ex Vía Férrea 1.

Tipo de <i>Random Forest</i>	Regresión
Número de árboles	500
Número de variables probadas en cada división	3
Media de los residuos al cuadrado	600.21
% de Variables explicadas	97.27

En la Tabla 6.6 se presenta el resumen del modelo. Se ha probado variando el número de árboles y los hiperparámetros del algoritmo, obteniéndose el modelo anterior. Después de cierto número de árboles, el resultado no cambia significativamente. Una de las desventajas de algunos modelos de aprendizaje automático es que funcionan para el usuario como una “caja negra”, es decir, no se tiene una interpretación certera de las variables que participan en el modelo y el usuario no conoce exactamente como el algoritmo realiza sus divisiones internas (Osinga, 2018). La Figura A.25 representa la predicción de conteos diarios bajo el modelo *Random Forests*. En la Figura A.26, se simboliza la predicción de conteos diarios y los conteos diarios reales, con simbología de puntos rojos y verdes respectivamente. Se observa que, a diferencia de los tres modelos anteriores, el

Estimación de modelos de predicción de tránsito diario de bicicletas en base a datos pasivos, incorporando variables meteorológicas y contextuales para la ciudad de Temuco

algoritmo *Random Forests* representa de forma más precisa la variable a explicar, pudiendo capturar la forma de la curva y en su mayoría los valores extremos. Es importante mencionar que, para efectos gráficos de la simbología presentada, se muestra el comportamiento de los modelos sobre el conjunto de datos bajo el cual fue entrenado (*training set*). Al realizar predicciones con el conjunto de prueba (*testing set*), el cual está conformado por datos que no fueron incluidos en el entrenamiento del modelo, el resultado es similar. Debido a que el conjunto de entrenamiento posee el 80% de los datos, se prefiere mostrar el conjunto que posee mayor volumen de información. En la gráfica representada por la Figura A.27 es posible evidenciar que la relación entre la predicción y los datos reales es mucho mayor que sus modelos predecesores, existiendo una relación prácticamente lineal entre las variables. Al comparar la Figura A.70 con la Figura A.66 se demuestra que el histograma de la predicción de este modelo es el más similar a la distribución de frecuencias reales. A continuación, en la Tabla 6.7 se expone un resumen de los modelos desarrollados para este contador, mientras que en la Figura A.71 se exponen los diagramas de caja referidos a las predicciones de este contador.

Tabla 6.7. Resumen de modelos desarrollados contador Ex Vía Férrea 1.

Modelos	Mínimo	Cuartil 1	Mediana	Media	Cuartil 3	Máximo
Real	7.00	121.80	233.00	249.40	355.20	732.00
RLM	-277.20	161.50	267.40	249.40	342.20	497.40
P	2.72	140.54	238.45	249.42	341.47	636.33
BN	5.01	142.34	232.28	247.96	339.58	711.16
RF	30.36	130.54	228.64	248.89	353.01	588.22

6.2.2. Modelo puntual Los Músicos – Los Creadores

El contador Los Músicos – Los Creadores se localiza en el sector nor - poniente de la ciudad de Temuco, específicamente en el sector de Fundo el Carmen. Este sector a diferencia de los otros contadores seleccionados para el desarrollo de modelos puntuales, registra menores valores de conteos diarios. La Tabla 6.8 muestra el resumen de la base de datos con la cual se realizan los modelos de predicción, para un total de 606 días, cantidad de datos levemente menor que los utilizados para la elaboración del modelo puntual anterior. El histograma de frecuencias de datos reales para este punto geográfico queda simbolizado en la Figura A.72, el cual presenta al igual

que el contador anterior, un sesgo hacia la derecha y una forma similar a una distribución Binomial Negativa.

Tabla 6.8. Resumen base de datos contador Los Músicos – Los Creadores.

Resumen	TEMP_ MEDIA	TEMP_ MIN	TEMP_ MAX	PRECIP	RAD	HUM	MP_10	MP_2.5	C_D
Mínimo	0.20	-3.80	2.70	0.00	37.20	25.30	7.00	1.00	2.00
Cuartil 1	8.62	2.80	14.60	0.00	546.80	74.62	21.00	6.00	38.00
Mediana	11.90	6.30	18.45	0.10	848.20	82.60	30.00	13.00	77.00
Media	11.52	6.09	18.97	2.63	799.70	81.12	43.21	28.15	85.83
Cuartil 3	14.67	9.30	22.88	1.90	1067.90	88.90	54.00	38.00	124.00
Máximo	24.70	16.30	41.10	74.80	1356.50	99.70	225.00	194.00	283.00

Al observar el resumen, se evidencia la diferencia en los conteos diarios con respecto al contador anterior, ya que en este caso la media de conteos diarios es de 86 pasadas, en contraste con las 250 del contador Ex Vía Férrea 1.

6.2.2.1. Modelo Regresión Lineal Múltiple Contador Los Músicos – Los Creadores

El primer modelo desarrollado para el contador Los Músicos – Los Creadores, es un modelo de regresión lineal múltiple. Se presenta a continuación el modelo final para este contador.

Tabla 6.9. Modelo regresión lineal múltiple contador Los Músicos – Los Creadores.

Variables	Estimador	Error Estándar	Valor <i>t</i>	Valor <i>p</i>	Significancia
INTERCEPTO	45.66	28.42	1.60	0.10	•
TEMP_MAX	5.69	1.52	3.72	0.0002	***
TEMP_MAX²	-0.06	0.03	-1.79	0.07	•
PRECIP	-1.04	0.26	-3.99	7.39e-05	***
HUM	-1.17	0.24	-4.73	2.89e-06	***
RAD	0.05	0.007	7.61	1.36e-13	***
WEEKEND	6.14	3.55	1.72	0.08	•
DAY	0.02	0.009	2.58	0.01	*

Código de significancia: 0 '***' 0.001 '**' 0.01 '*' '.' 0.1

R² ajustado = 0.63

En este modelo a diferencia del anterior, la variable feriados no resulta ser estadísticamente significativa. Además, la variable de fin de semana es significativa al 10%. Por su parte, el Estimación de modelos de predicción de tránsito diario de bicicletas en base a datos pasivos, incorporando variables meteorológicas y contextuales para la ciudad de Temuco

coeficiente de determinación ajustado para el modelo es de 0.63. Los signos de las variables explicativas son los siguientes: temperatura máxima (+), temperatura máxima al cuadrado (-), precipitación (-), humedad (-), radiación (+), fin de semana (+) y día secuencial (+). Las variables poseen los mismos signos que en el modelo de regresión lineal múltiple asociado al contador Ex Vía Férrea 1, sin embargo, las magnitudes de los estimadores difieren en valor entre un contador y otro. En la Figura A.28 y Figura A.29 se observa que debido a la diferencia en los datos disponibles en comparación con el contador anterior, la curva está truncada en su inicio. La Figura A.29 ilustra con puntos de color rojo las predicciones del modelo de regresión lineal múltiple para este contador, junto con los valores de los conteos diarios reales simbolizados con puntos de color verde. Se observa una buena primera aproximación, sin embargo, el modelo no es capaz de capturar los valores extremos y además predice valores negativos, produciéndose un fenómeno similar al contador anterior. En el gráfico representado por la Figura A.30 se representa la relación lineal existente entre la variable dependiente y la variable resultante de la modelación que se desea explicar. Al compararlo con la Figura A.18 se observa la similitud que tienen los modelos de regresión lineal múltiple en diferentes contadores, presentándose en ambos valores negativos y una forma similar de la línea de tendencia, lo cual refleja que en la globalidad de los modelos, las variables meteorológicas explican de forma similar los conteos diarios en diferentes ubicaciones espaciales asociados a contadores. Sin embargo, en igualdad de condiciones meteorológicas, estos modelos no son capaces de explicar las diferencias en volúmenes diarios de ciclistas en ciertos lugares geográficos con respecto a otros, hecho que sugiere que dicha diferencia está determinada por otro tipo de variables. La Figura A.73 presenta el histograma de predicciones bajo este modelo, el cual presenta un sesgo hacia la izquierda y una distribución diferente al histograma de datos reales presentados anteriormente.

6.2.2.2. Modelo Poisson Contador Los Músicos – Los Creadores

Siguiendo los pasos realizados para el contador Ex Vía Férrea 1, el segundo modelo de regresión implementado para este contador es el modelo Poisson, resultando el modelo final expresado en la Tabla 6.10.

Tabla 6.10. Modelo Poisson contador Los Músicos – Los Creadores.

Variables	Estimador	Error Estándar	Valor z	Valor p	Significancia
INTERCEPTO	2.35	0.08	27.37	< 2e-16	***
TEMP_MAX	0.17	0.005	29.33	< 2e-16	***
TEMP_MAX²	-0.003	0.0001	-24.20	< 2e-16	***
PRECIP	-0.03	0.001	-22.96	< 2e-16	***
HUM	-0.01	0.0006	-15.38	< 2e-16	***
RAD	0.0007	0.00002	30.20	< 2e-16	***
WEEKEND	0.08	0.01	7.91	2.39e-15	***
HOLIDAY	0.10	0.02	4.17	3.01e-05	***
DAY	0.0003	0.00002	11.85	< 2e-16	***

Código de significancia: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ ‘.’ 0.1

Desviación Residual = 5136.90 en 487 grados de libertad

AIC = 8152.40

Los signos de las variables son los siguientes: temperatura máxima (+), temperatura máxima al cuadrado (-), precipitación (-), humedad (-), radiación (+), fin de semana (+), feriados (+) y día secuencial (+). El criterio de información de Akaike (AIC) es de 8152.40 siendo el menor valor obtenido para todos los modelos ajustados que utilizan esta regresión en este contador. Si se evalúa la sobredispersión de los datos se obtiene:

$$\frac{\text{Desviación Residual}}{\text{Grados de Libertad}} = \frac{5136.90}{487} = 10.55 > 1 \quad (6.3)$$

La ecuación (6.3) confirma que la relación establecida tiene un valor mayor a 1, lo cual significa que la muestra de datos presenta sobredispersión, siendo más adecuado emplear otro modelo. Se observa en los gráficos representados por la Figura A.31 y la Figura A.32 que, a diferencia del modelo de regresión lineal múltiple, el modelo Poisson no predice valores de conteos diarios de carácter negativo, lo cual supone una importante mejora en la estimación del modelo con respecto al desarrollado con anterioridad. La Figura A.33 representa un gráfico que explica las diferencias entre la predicción de conteos diarios y los conteos diarios reales para este modelo. Al igual que en la Figura A.31 y Figura A.32, se verifica que el modelo solo predice valores positivos. La Figura A.74 simboliza el histograma de predicciones, el cual no posee una distribución clara y no se asemeja al histograma de conteos diarios reales (ver Figura A.72).

Estimación de modelos de predicción de tránsito diario de bicicletas en base a datos pasivos, incorporando variables meteorológicas y contextuales para la ciudad de Temuco

6.2.2.3. Modelo Binomial Negativo Contador Los Músicos – Los Creadores

El modelo Binomial Negativo para el contador Los Músicos – Los Creadores tiene por resultado un modelo con ocho variables significativas. Los signos de las variables independientes son: temperatura máxima (+), temperatura máxima al cuadrado (-), precipitación (-), humedad (+), radiación (+), fin de semana (+), feriados (+) y día secuencial (+). La Tabla 6.11 muestra la especificación del modelo.

Tabla 6.11. Modelo binomial negativo contador Los Músicos – Los Creadores.

Variables	Estimador	Error Estándar	Valor z	Valor p	Significancia
INTERCEPTO	3.04	0.30	9.91	< 2e-16	***
TEMP_MAX	0.14	0.0001	8.22	< 2e-16	***
TEMP_MAX ²	-0.002	0.01	-6.06	1.36e-09	***
PRECIP	-0.03	0.0003	-10.62	< 2e-16	***
HUM	-0.01	0.002	-5.01	5.21e-07	***
RAD	0.0006	0.00008	7.76	8.48e-15	***
WEEKEND	0.11	0.03	3.02	0.002	**
HOLIDAY	0.20	0.08	2.26	0.02	*
DAY	0.0002	0.0001	2.46	0.01	*

Código de significancia: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ ‘.’ 0.1

Desviación Residual = 522.88 en 487 grados de libertad

AIC = 4700.80

2 x Log-Verosimilitud = -4680.81

El valor del criterio de información de Akaike (AIC) es de 4700.80, y la evaluación de la siguiente relación confirma que la muestra no posee sobredispersión, por lo que resulta ser una mejor opción que el modelo de regresión tipo Poisson.

$$\frac{\text{Desviación Residual}}{\text{Grados de Libertad}} = \frac{522.88}{487} = 1.07 \cong 1 \quad (6.4)$$

La Figura A.34 muestra en simbología de puntos de color rojo la predicción de conteos diarios del modelo Binomial Negativo para este contador. Se evidencia la ausencia de valores negativos en la variable dependiente junto con la forma sinusoidal característica de los conteos diarios de bicicletas asociado a la meteorología. La Figura A.35 presenta en el mismo gráfico la información de la Estimación de modelos de predicción de tránsito diario de bicicletas en base a datos pasivos, incorporando variables meteorológicas y contextuales para la ciudad de Temuco

Figura A.34 en conjunto con los conteos diarios reales simbolizados por puntos de color verde. El gráfico representado por la Figura A.36 representa la predicción de conteos diarios del modelo Binomial Negativo en contraste con los conteos diarios reales. La Figura A.75 refleja que el histograma de frecuencias de la predicción no posee una clara distribución, sin embargo, tiene una mayor similitud que el histograma de predicciones del modelo Poisson con respecto a los datos reales.

6.2.2.4. Modelo *Random Forests* Contador Los Músicos – Los Creadores

El modelo de Bosques Aleatorios o *Random Forests* aplicado al contador Los Músicos – Los Creadores queda resumido en la Tabla 6.12.

Tabla 6.12. Modelo *Random Forests* contador Los Músicos – Los Creadores.

Tipo de <i>Random Forest</i>	Regresión
Número de árboles	500
Número de variables probadas en cada división	3
Media de los residuos al cuadrado	90.69
% de Variables explicadas	97.38

En este modelo se ha logrado un 97.38% de explicación de las variables de la base de datos. La Figura A.37 simboliza la estimación de conteos diarios que realiza el modelo *Random Forests* para el presente contador en función del tiempo. Se puede notar que el modelo no predice valores negativos y representa de forma adecuada el comportamiento real de la variable respuesta. Por su parte, en la Figura A.38 se representa la predicción del modelo *Random Forests* y los conteos diarios reales, con simbología de puntos rojos y verdes respectivamente. Se observa que a diferencia de los modelos anteriores, el modelo *Random Forests* representa de forma más verídica los conteos diarios reales, pudiendo capturar la forma de la curva y en su mayoría los valores extremos. La Figura A.39 relaciona cada dato de la predicción de conteos diarios con los conteos diarios reales. La Figura A.76 muestra que el histograma de frecuencias de la predicción es prácticamente idéntico al histograma de datos reales. Finalmente, la Tabla 6.13 resume los resultados de todos los modelos desarrollados para este contador, mientras que en la Figura A.77 se exponen los diagramas de caja referidos a las predicciones de este contador.

Tabla 6.13. Resumen de modelos desarrollados contador Los Músicos – Los Creadores.

Modelos	Mínimo	Cuartil 1	Mediana	Media	Cuartil 3	Máximo
Real	2.00	38.00	79.00	88.42	126.25	283.00
RLM	-72.03	51.20	93.64	88.42	125.50	209.28
P	1.41	43.21	83.76	88.42	129.35	198.52
BN	1.58	44.80	83.45	88.00	126.12	247.18
RF	9.01	40.39	79.18	88.01	126.83	222.22

6.2.3. Modelo puntual Pablo Neruda – Olimpia

El contador Pablo Neruda – Olimpia se localiza en el sector centro - poniente de la ciudad de Temuco. Este contador se caracteriza por poseer el mayor número de conteos diarios del grupo de contadores seleccionados. La ciclovía de Pablo Neruda se encuentra sobre el bandejón central de la avenida y fue inaugurada el año 2016. La Tabla 6.14 muestra el resumen de los datos disponibles para este contador, los cuales totalizan 1017 observaciones diarias.

Tabla 6.14. Resumen base de datos contador Pablo Neruda – Olimpia.

Resumen	TEMP_MEDIA	TEMP_MIN	TEMP_MAX	PRECIP	RAD	HUM	MP_10	MP_2.5	C_D
Mínimo	-0.10	-4.30	2.70	0.00	37.20	25.30	7.00	1.00	18.00
Cuartil 1	8.70	3.00	14.10	0.00	560.00	75.20	21.00	6.00	329.00
Mediana	11.70	6.20	18.20	0.10	835.60	82.70	30.00	14.00	504.00
Media	11.43	6.06	18.72	2.92	799.10	81.27	42.10	27.34	525.80
Cuartil 3	14.50	9.10	22.70	2.20	1057.00	88.20	52.00	37.00	695.00
Máximo	24.70	16.80	41.10	74.80	1368.50	99.70	225.00	204.00	1212.00

En este resumen, se sintetizan aspectos descriptivos de la base de datos, como valores máximos y mínimos, la media, valor central y valores de los primer y tercer cuartil para cada variable a analizar. La media de conteos diarios en este dispositivo de conteo es de aproximadamente 526 pasadas diarias. El histograma referido a la variable objetivo se presenta en la Figura A.78.

6.2.3.1. Modelo Regresión Lineal Múltiple Contador Pablo Neruda – Olimpia

El primer modelo desarrollado para el contador de Pablo Neruda - Olimpia, es un modelo de regresión lineal múltiple al igual que en los contadores anteriores. En la Tabla 6.15 se presenta la especificación del modelo para este contador.

Estimación de modelos de predicción de tránsito diario de bicicletas en base a datos pasivos, incorporando variables meteorológicas y contextuales para la ciudad de Temuco

Tabla 6.15. Modelo regresión lineal múltiple contador Pablo Neruda – Olimpia.

Variables	Estimador	Error Estándar	Valor <i>t</i>	Valor <i>p</i>	Significancia
INTERCEPTO	-21.26	81.53	-0.26	0.79	
TEMP_MAX	47.90	4.433	10.80	< 2e-16	***
TEMP_MAX²	-0.75	0.104	-7.24	1.05e-12	***
PRECIP	-9.31	0.672	-13.85	< 2e-16	***
HUM	-3.52	0.707	-4.98	7.48e-07	***
RAD	0.08	0.02	3.89	1.06e-04	***
WEEKEND	-111.86	9.92	-11.27	< 2e-16	***
HOLIDAY	-113.07	21.97	-5.14	3.35e-07	***
DAY	0.44	0.01	28.72	< 2e-16	***

Código de significancia: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ ‘.’ 0.1

R² ajustado = 0.75

En este modelo las variables estadísticamente significativas tienen los siguientes signos: temperatura máxima (+), temperatura máxima al cuadrado (-), precipitación (-), humedad (-), radiación (+), fin de semana (-), feriados (-) y día secuencial (+). El coeficiente de determinación ajustado equivale a 0.75. En este modelo, las variables de fin de semana y feriados poseen signo negativo, a diferencia de los dos contadores predecesores. Este hecho se condice con la realidad, en la cual, los conteos diarios de esta zona de la ciudad disminuyen en días de fin de semana o feriados, como fue mostrado en Capítulo 5. Si se grafica la predicción de conteos diarios del modelo de regresión lineal múltiple en función del tiempo, se obtiene la curva expuesta en la Figura A.40. En dicha imagen, se observa la forma sinusoidal característica de estos gráficos. También es posible notar que, según la predicción de este modelo, las curvas sinusoidales tienen mayor amplitud de un año a otro, presentando un crecimiento sostenido de los conteos diarios en función del tiempo. La inclusión de la variable día secuencial al igual que en estudios como Schmiedeskamp y Zhao (2016), permite reflejar la tendencia global al alza de las pasadas diarias en la ciudad, lo cual quiere decir que en base a factores no meteorológicos, el uso de la bicicleta se está expandiendo en esta parte de la urbe, aumentando el número de viajes en este modo de transporte. Dicho aumento puede atribuirse a la elección modal de las personas, es decir, personas que se han sumado al uso de la bicicleta por sobre otros modos a pesar de que las condiciones meteorológicas no hayan cambiado.

Al igual que en los gráficos anteriores de este tipo, la Figura A.42 representa en cada coordenada *x* e *y* la relación existente entre los conteos diarios reales y la predicción de conteos diarios. La

relación se vuelve más exacta a medida que la curva se aproxima a una relación lineal de 45° entre ambas variables. En este caso, podemos notar que el modelo no es la mejor alternativa para la predicción de los conteos diarios reales, debido fundamentalmente a la predicción de valores de carácter negativo. El histograma para este modelo expresado en la Figura A.79 refleja dichos problemas.

6.2.3.2. Modelo Poisson Contador Pablo Neruda - Olimpia

El segundo modelo de regresión implementado para este contador es un modelo tipo Poisson, el cual queda expresado en la Tabla 6.16.

Tabla 6.16. Modelo Poisson contador Pablo Neruda – Olimpia.

Variables	Estimador	Error Estándar	Valor z	Valor p	Significancia
INTERCEPTO	4.46	0.02	156.82	< 2e-16	***
TEMP_MAX	0.14	0.001	80.56	< 2e-16	***
TEMP_MAX²	-0.002	0.00003	-66.17	< 2e-16	***
PRECIP	-0.03	0.0003	-90.93	< 2e-16	***
HUM	-0.004	0.0002	-18.42	< 2e-16	***
RAD	0.0001	0.000007	19.78	< 2e-16	***
WEEKEND	-0.21	0.003	-61.92	< 2e-16	***
HOLIDAY	-0.20	0.008	-24.73	< 2e-16	***
DAY	0.0008	0.000005	159.42	< 2e-16	***

Código de significancia: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ ‘.’ 0.1

Desviación Residual = 22243 en 804 grados de libertad

AIC = 28729

Los signos de las variables del modelo resultante son los siguientes: temperatura máxima (+), temperatura máxima al cuadrado (-), precipitación (-), humedad (-), radiación (+), fin de semana (-), feriados (-) y día secuencial (+). La ecuación (6.5) demuestra la sobredispersión de los datos en este modelo.

$$\frac{\text{Desviación Residual}}{\text{Grados de Libertad}} = \frac{22243}{804} = 27.67 > 1 \quad (6.5)$$

La Figura A.43 representa la predicción de conteos diarios bajo el modelo Poisson. En la Figura A.44 se representa la predicción del modelo y los conteos diarios reales, con simbología de puntos rojos y verdes respectivamente. La Figura A.45 muestra una relación lineal bastante acertada entre la variable dependiente y la variable resultante del modelo. Es importante recordar que al desarrollar un modelo estadístico, normalmente se busca explicar en forma simplificada la realidad a través de herramientas matemáticas y no replicarla exactamente, por lo que modelos predictivos de este tipo son suficientes para asegurar cierta precisión. La Figura A.80 representa el histograma de la variable respuesta obtenida por las predicciones bajo este modelo.

6.2.3.3. Modelo Binomial Negativo Contador Pablo Neruda - Olimpia

Los signos de las variables independientes bajo este modelo son: temperatura máxima (+), temperatura máxima al cuadrado, precipitación (-), humedad (-), radiación (+), fin de semana (-), feriados (-) y día secuencial (+), tal cual como se esperaba previamente. La Tabla 6.17 muestra la especificación del modelo.

Tabla 6.17. Modelo binomial negativo contador Pablo Neruda – Olimpia.

Variables	Estimador	Error Estándar	Valor z	Valor p	Significancia
INTERCEPTO	4.67	0.18	25.88	< 2e-16	***
TEMP_MAX	0.11	0.009	11.48	< 2e-16	***
TEMP_MAX ²	-0.001	0.0002	-8.31	< 2e-16	***
PRECIP	-0.02	0.001	-19.09	< 2e-16	***
HUM	-0.004	0.001	-3.01	0.00256	***
RAD	0.0002	0.00004	4.61	3.95e-06	**
WEEKEND	-0.22	0.02	-10.05	< 2e-16	***
HOLIDAY	-0.21	0.04	-4.47	7.67e-06	***
DAY	0.0009	0.00003	26.58	< 2e-16	***

Código de significancia: 0 '***' 0.001 '**' 0.01 '*' '.' 0.1

Desviación Residual = 843.45 en 804 grados de libertad

AIC = 10272

2 x Log-Verosimilitud = -10252.13

$$\frac{\text{Desviación Residual}}{\text{Grados de Libertad}} = \frac{843.45}{804} = 1.05 \cong 1 \quad (6.6)$$

La ecuación (6.6) demuestra que los datos no presentan sobredispersión. Es importante mencionar que la función enlace logarítmica de la distribución Binomial Negativa tiene por resultado que los valores de los estimadores tengan distinto orden de magnitud que los obtenidos en los modelos de regresión lineal múltiple, por lo cual se debe realizar una transformación previa para poder evaluar de igual forma los datos. En R, los valores ajustados se almacenan en la tabla *fitted.values* luego de ejecutar la modelación, por lo que no es necesario aplicar la función enlace en forma manual a cada uno de los valores. Sin embargo, si se desea evaluar un día ficticio bajo este modelo, es necesario aplicar la transformación derivada de la especificación de la distribución Binomial Negativa para obtener los valores correctos de predicción para la variable de conteos diarios. En definitiva, los coeficientes presentados anteriormente son los valores de los coeficientes correspondientes a explicar la variable logaritmo natural de los conteos diarios. Si se grafica la predicción de conteos diarios realizada por el modelo Binomial Negativo luego de entrenar el modelo de regresión con el conjunto de entrenamiento (*training set*) en función del tiempo disponible de datos, se obtiene la gráfica expuesta en la Figura A.46. La Figura A.47 muestra que el modelo Binomial Negativo, dada su naturaleza, permite pronosticar solo valores positivos. En los gráficos anteriores se evidencia un rendimiento del modelo similar al modelo Poisson, con la ventaja de no presentar el problema de sobredispersión. La Figura A.48 representa en el eje de las ordenadas los valores de conteos diarios para el modelo Binomial Negativo, mientras que en el eje de las abscisas los valores de los conteos diarios reales. La Figura A.81 refleja que el histograma de frecuencias de este modelo es similar a la distribución de los datos reales.

6.2.3.4. Modelo *Random Forests* Contador Pablo Neruda - Olimpia

Finalmente, se aplica el modelo de Bosques Aleatorios o *Random Forests* al conjunto de datos disponibles para el contador de Pablo Neruda – Olimpia. La especificación del modelo queda explícita en la Tabla 6.18, indicando el número de árboles de regresión utilizados y el % de variables que se logra explicar en la modelación.

Tabla 6.18. Modelo *Random Forests* contador Pablo Neruda – Olimpia.

Tipo de <i>Random Forest</i>	Regresión
Número de árboles	500
Número de variables probadas en cada división	3

Estimación de modelos de predicción de tránsito diario de bicicletas en base a datos pasivos, incorporando variables meteorológicas y contextuales para la ciudad de Temuco

Tipo de <i>Random Forest</i>	Regresión
Media de los residuos al cuadrado	933.66
% de Variables explicadas	98.59

Esta estimación es capaz de capturar fielmente el crecimiento del modo de transporte en el sector centro – poniente de la ciudad y en su mayoría los valores extremos, hechos que se desprenden de la Figura A.49. La densidad de puntos que relacionan los datos reales y la predicción mostrada en la Figura A.51 se ajusta muy bien a la curva de tendencia. La Figura A.82 demuestra que la predicción arroja un histograma prácticamente idéntico al desarrollado por los conteos reales. La Tabla 6.19 muestra el resumen estadístico de los modelos de predicción desarrollados para este contador, mientras que en la Figura A.83 se exponen los diagramas de caja referidos a las predicciones de este contador.

Tabla 6.19. Resumen de modelos desarrollados contador Pablo Neruda – Olimpia.

Modelos	Mínimo	Cuartil 1	Mediana	Media	Cuartil 3	Máximo
Real	18.00	336.00	514.00	529.50	699.00	1212.00
RLM	-508.50	377.60	545.30	529.50	671.80	1012.00
P	17.90	359.80	506.80	529.50	667.70	1165.60
BN	27.80	351.40	502.30	529.30	661.90	1218.50
RF	71.65	343.81	509.12	529.02	687.42	1128.44

6.2.4. Modelo puntual San Martín - Mackenna

El contador San Martín – Mackenna se localiza en el sector céntrico de la ciudad de Temuco. Es el único contador en este sector de la ciudad, lo cual es importante para generar variabilidad espacial en el estudio y para comparar los resultados de los modelos puntuales entre sí. La Tabla 6.20 muestra el resumen de los datos disponibles para este contador. La Figura A.84 representa el histograma de la variable conteos diarios para este contador, la cual no posee una distribución característica como sus predecesoras. Se observa mayor aleatoriedad en la distribución de frecuencias en el número de pasadas diarias, a diferencia de los histogramas de los contadores predecesores.

Tabla 6.20. Resumen base de datos contador San Martín – Mackenna.

Resumen	TEMP_ MEDIA	TEMP_ _MIN	TEMP_ _MAX	PRECIP	RAD	HUM	MP_10	MP_2.5	C_D
Mínimo	0.20	-3.80	2.70	0.00	37.20	25.30	7.00	1.00	27.00
Cuartil 1	8.60	2.75	14.35	0.00	551.40	75.20	21.00	6.00	243.50
Mediana	11.60	6.20	18.00	0.10	848.20	82.70	30.00	13.00	374.00
Media	11.35	5.96	18.74	2.69	801.00	81.21	42.82	27.91	398.70
Cuartil 3	14.50	9.20	22.55	1.90	1062.00	88.60	54.00	38.00	536.00
Máximo	24.70	16.30	41.10	74.80	1356.50	99.70	225.00	194.00	859.00

En este resumen, se sintetizan aspectos descriptivos de la base de datos, como valores máximos y mínimos, la media, valor central y valores de los primer y tercer cuartil para cada variable a analizar. La media de conteos alcanza aproximadamente las 400 pasadas diarias.

6.2.4.1. Modelo Regresión Lineal Múltiple Contador San Martín - Mackenna

La Tabla 6.21 representa la especificación del modelo de regresión múltiple para este contador.

Tabla 6.21. Modelo regresión lineal múltiple contador San Martín – Mackenna.

Variables	Estimador	Error Estándar	Valor <i>t</i>	Valor <i>p</i>	Significancia
INTERCEPTO	56.049	72.059	0.778	0.437	
TEMP_MAX	34.643	3.407	10.167	< 2e-16	***
TEMP_MAX²	-0.596	0.082	-7.255	1.5e-12	***
PRECIP	-7.177	0.637	-11.264	< 2e-16	***
HUM	-1.374	0.632	-2.172	0.0303	*
WEEKEND	-250.834	9.243	-27.137	< 2e-16	***
HOLIDAY	-239.911	23.478	-10.218	< 2e-16	***
DAY	0.422	0.022	18.958	< 2e-16	***

Código de significancia: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ ‘.’ 0.1

R² ajustado = 0.76

En este modelo, las variables significativas son: temperatura máxima (+), temperatura máxima al cuadrado (-), precipitación (-), humedad (-), fin de semana (-), feriados (-) y día secuencial (+). El coeficiente R² ajustado es de 0.76, valor similar al coeficiente del modelo puntual Pablo Neruda - Olimpia. En este modelo, al igual que en el contador Pablo Neruda – Olimpia, el efecto de las variables temporales es negativo en los conteos diarios, hecho contrario al comportamiento

presentado en el contador Ex Vía Férrea 1 y en el contador Los Músicos – Los Creadores. La Figura A.53 refleja que las predicciones se ajustan relativamente bien a los conteos diarios reales dado su alto nivel del coeficiente de determinación, sin embargo, si se observan los valores extremos presentes en la curva sinusoidal, se desprende que el modelo no predice adecuadamente dichos valores. La Figura A.54 confirma las deficiencias de este tipo de modelos cuando se trabaja con variables dependientes de conteo, debido a que ciertas combinaciones de los coeficientes y los valores de las variables sumado al valor del intercepto, pueden concluir en valores negativos para dicha variable objetivo, como es el caso de algunos datos predichos por este modelo. La Figura A.85 presenta el histograma de la predicción de conteos diarios de este ajuste.

6.2.4.2. Modelo Poisson Contador San Martín - Mackenna

El segundo modelo de regresión implementado para este contador es el modelo Poisson, expuesto en la Tabla 6.22.

Tabla 6.22. Modelo Poisson contador San Martín – Mackenna.

Variables	Estimador	Error Estándar	Valor z	Valor p	Significancia
INTERCEPTO	4.85	0.03	128.09	< 2e-16	***
TEMP_MAX	0.10	0.001	51.57	< 2e-16	***
TEMP_MAX²	-0.001	0.00004	-40.15	< 2e-16	***
PRECIP	-0.02	0.0004	-53.43	< 2e-16	***
HUM	-0.002	0.0003	-8.29	< 2e-16	***
WEEKEND	-0.72	0.005	-122.61	< 2e-16	***
HOLIDAY	-0.79	0.01	-44.38	< 2e-16	***
DAY	0.001	0.00001	88.58	< 2e-16	***

Código de significancia: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ ‘.’ 0.1

Desviación Residual = 10708 en 512 grados de libertad

AIC = 14725

Los signos de las variables son los siguientes: temperatura máxima (+), temperatura máxima al cuadrado (-), precipitación (-), humedad (-), fin de semana (-), feriados (-) y día secuencial (+). La calidad relativa del modelo presentado queda explicada por el criterio de información de Akaike (AIC), el cual corresponde a 14725, siendo el menor valor obtenido para la combinación de

variables probadas para este modelo. La evaluación de la sobredispersión de los datos es presentada a continuación.

$$\frac{\text{Desviación Residual}}{\text{Grados de Libertad}} = \frac{10708}{512} = 20.91 > 1 \quad (6.7)$$

Por lo tanto, según la ecuación (6.7) el modelo posee sobredispersión. En la Figura A.55 y la Figura A.56 se puede notar un mejor comportamiento de este modelo en comparación con el modelo de regresión lineal múltiple, sobre todo al estimar valores extremos de días de verano. La Figura A.57 representa el gráfico que demuestra las diferencias entre la predicción de conteos diarios y los conteos diarios reales para este modelo. Se observa una mejor relación lineal entre las variables si se compara con el caso anterior (ver Figura A.54). La Figura A.86 representa el histograma de las predicciones de esta estimación.

6.2.4.3. Modelo Binomial Negativo Contador San Martín - Mackenna

Los signos de las variables predictoras del modelo Binomial Negativo son: temperatura máxima (+), temperatura máxima al cuadrado (-), precipitación (-), humedad (-), fin de semana (-), feriados (-) y día secuencial (+). La especificación del modelo está contenida en la Tabla 6.23.

Tabla 6.23. Modelo binomial negativo contador San Martín – Mackenna.

Variables	Estimador	Error Estándar	Valor z	Valor p	Significancia
INTERCEPTO	4.91	0.20	23.69	< 2e-16	***
TEMP_MAX	0.09	0.009	9.93	< 2e-16	***
TEMP_MAX ²	-0.001	0.0002	-7.34	2.12e-13	***
PRECIP	-0.02	0.001	-12.61	< 2e-16	***
HUM	-0.003	0.001	-1.68	< 2e-16	***
WEEKEND	-0.72	0.02	-27.15	0.0921	.
HOLIDAY	-0.83	0.06	-12.12	< 2e-16	***
DAY	0.001	0.00006	17.41	< 2e-16	***

Código de significancia: 0 '***' 0.001 '**' 0.01 '*' '.' 0.1

Desviación Residual = 538.71 en 512 grados de libertad

AIC = 6256.60

2 x Log-Verosimilitud = -6238.59

Estimación de modelos de predicción de tránsito diario de bicicletas en base a datos pasivos, incorporando variables meteorológicas y contextuales para la ciudad de Temuco

$$\frac{\text{Desviación Residual}}{\text{Grados de Libertad}} = \frac{538.71}{512} = 1.05 \cong 1 \quad (6.8)$$

La evaluación de la sobredispersión expresada en la ecuación (6.8) confirma que la muestra de datos no presenta este fenómeno. A su vez, la variable de fin de semana es estadísticamente significativa para un valor p menor o igual al 10%. El valor del criterio de información de Akaike (AIC) es de 6256.60. Al igual que en los casos anteriores, el modelo Binomial Negativo resulta ser la opción más adecuada dentro de los modelos lineales generalizados debido a sus restricciones en la variable dependiente y también el hecho de que no exista la restricción de igualdad entre la media y la varianza como sí ocurre en los modelos de Poisson. La Figura A.87 representa el histograma de frecuencias de la predicción de conteos diarios.

6.2.4.4. Modelo *Random Forests* Contador San Martín - Mackenna

Finalmente, se aplica el modelo de Bosques Aleatorios o *Random Forests* al conjunto de datos disponibles para el contador San Martín - Mackenna. La descripción del modelo se presenta en la Tabla 6.24.

Tabla 6.24. Modelo *Random Forests* contador San Martín – Mackenna.

Tipo de <i>Random Forest</i>	Regresión
Número de árboles	500
Número de variables probadas en cada división	3
Media de los residuos al cuadrado	616.78
% de Variables explicadas	98.38

La Figura A.61 demuestra el resultado de la predicción bajo el modelo de Bosques Aleatorios, en el cual la curva se distribuye en forma más suave que en los modelos lineales generalizados, manteniéndose la forma sinusoidal de ésta. La Figura A.62 refuerza la afirmación que a simple vista, los algoritmos de *Machine Learning* poseen aplicaciones muy diversas, y en particular, algoritmos de regresión como los aplicados en este modelo, presentan a priori un comportamiento superior a los tradicionales, hecho que también se refuerza al evidenciar la semejanza entre el histograma de datos reales versus el histograma de predicciones presentado en la Figura A.88. Como último resultado asociado a los modelos puntuales se presenta la Figura A.63, la cual

Estimación de modelos de predicción de tránsito diario de bicicletas en base a datos pasivos, incorporando variables meteorológicas y contextuales para la ciudad de Temuco

demuestra que el ajuste de este modelo es muy similar a un ajuste lineal, por lo que resulta sumamente eficiente en la representación de los conteos diarios. El resumen estadístico de los modelos desarrollados para este contador se presenta en la Tabla 6.25, mientras que en la Figura A.89 se exponen los diagramas de caja referidos a las predicciones de este contador.

Tabla 6.25. Resumen modelos desarrollados contador San Martín – Mackenna.

Modelos	Mínimo	Cuartil 1	Mediana	Media	Cuartil 3	Máximo
Real	27.00	249.80	382.00	404.80	545.20	833.00
RLM	-117.40	275.80	425.40	404.80	525.70	703.00
P	63.58	246.82	392.06	404.77	510.12	769.56
BN	79.74	249.92	389.91	405.10	512.46	787.12
RF	79.71	253.83	371.69	404.68	540.89	774.95

6.3. Análisis de modelos puntuales

En el siguiente apartado se presenta el análisis de resultados referidos a los modelos puntuales y el impacto de cada uno de los factores meteorológicos en los conteos diarios.

En el contador Ex Vía Férrea 1 se encontró que para el modelo de regresión lineal múltiple, una variación de 1 °C en la temperatura máxima produce un aumento de un 4.78% de los conteos diarios. Sin embargo, si la temperatura máxima en un día cualquiera resulta ser menor que 5.4 [°C], manteniendo todas las otras variables constantes, se producen conteos diarios de signo negativo, lo cual no se condice con la realidad. Por otra parte, un aumento de 1 mm en la precipitación diaria arroja un descenso de un 2.18% en los conteos diarios. El problema análogo de valores negativos surge cuando la precipitación acumulada en el día sobrepasa los 55 mm, punto en cual comienzan a producirse predicciones de signo negativo manteniendo constantes las demás variables. Si se analiza un día de fin de semana en este contador, el modelo arroja que el número de pasadas aumentan en un 39.13%, lo cual sugiere que esta ciclovía se utiliza en mayor medida los días de fin de semana en forma recreativa. Si el día es un feriado y además es fin de semana, los conteos diarios ascienden en un 91.74%. El impacto que tiene la humedad se observa cuando se eleva su valor desde un 80% a un 100%, afectando negativamente las pasadas diarias en un 14.35%. En el modelo Poisson para este contador, una variación de 1 °C en la temperatura máxima, provoca un ascenso en los conteos diarios de un 5.76%, y cuando esta variable sobrepasa los 33 °C se provoca un descenso en las pasadas de un 4.78%. Esta disminución es mayor a medida que la temperatura

Estimación de modelos de predicción de tránsito diario de bicicletas en base a datos pasivos, incorporando variables meteorológicas y contextuales para la ciudad de Temuco

se incrementa sobre dicho umbral, debido al efecto no lineal que tiene esta variable en el modelo. Con respecto a la lluvia, un aumento de 1 mm en la precipitación provoca un descenso de un 4.33% en la cantidad de ciclistas diarios. En este tipo de modelo para este contador, la variable material particulado resulta ser estadísticamente significativa, y un aumento en 10 $[\mu\text{g}/\text{m}^3]$ provoca un descenso del 0.1% de los Conteos Diarios. Esto sugiere que esta variable no tiene tanta importancia en el modelo como las anteriores. En días de fin de semana las pasadas diarias aumentan en un 40.86% y si el día resulta ser feriado y a la vez fin de semana, estas aumentan en un 127.88%. Por su parte, bajo el modelo Binomial Negativo, un aumento de 1 °C en la temperatura máxima provoca un ascenso de un 5.52% de la variable dependiente y por cada 1 mm de precipitación extra, el número de pasadas decrece en un 3.64%. La radiación aporta al alza en un 5.02% por cada 100 $[\text{W}/\text{m}^2]$ sobre la media. En días de fin de semana, los conteos diarios aumentan un 41.70% manteniendo las mismas condiciones meteorológicas constantes que hubiesen ocurrido en un día de semana. Si el día es feriado y fin de semana a la vez, las pasadas diarias crecen abruptamente en un 161.30%, lo cual confirma el uso de fin de semana que tiene esta ciclovía.

En el contador Los Músicos – Los Creadores, caracterizado por ser un punto con baja conectividad a la red y cuyo entorno construido está fundamentalmente conformado por propiedades habitacionales, se encontró a través del modelo de regresión lineal múltiple que un aumento en 1 °C provoca un ascenso de un 3.79% en los conteos diarios de dicha ciclovía. Por su parte, 1 mm de precipitación extra provocan un descenso de 1.28% en el número de pasadas. A diferencia del contador Ex Vía Férrea 1, en este sitio se encontró significancia estadística en la variable de fin de semana pero no en la variable de feriados. En días de fin de semana los conteos diarios aumentan en un 7.69%, valor considerablemente menor que en el caso anterior. Para el caso del modelo Poisson, el aumento en una unidad de temperatura máxima trae consigo un ascenso de 5.33% en los conteos diarios, valor que es creciente hasta el límite de 28.20 °C, punto en el cual comienzan a descender debido a la excesiva temperatura. El aumento de 1 mm en la precipitación tiene por consecuencia una disminución de 4.16% en la variable objetivo. Un ascenso de la humedad desde un 80 a un 100% trae consigo un descenso de 21.3% del número de pasadas manteniendo constante las demás variables. En días de fin de semana los conteos diarios se incrementan en promedio un 7.40% y en días de fin de semana que a su vez son feriados, el aumento totaliza un 20%. Por su parte, la variable día secuencial demuestra que en el primer día de estudio bajo condiciones meteorológicas promedio se totalizaban 69 pasadas diarias, mientras que bajo los mismos factores

meteorológicos en el último día de estudio los conteos suben a 82, lo cual se traduce en un 18.84% de aumento no atribuible a las variables meteorológicas, sino a un aumento en el uso del modo de transporte sujeto a otras variables que pudiesen estar relacionadas con cambios en la elección modal de los habitantes. El modelo Binomial Negativo tiene por resultado que un aumento en 1 °C se traduce en un ascenso de un 4.10% en los conteos. Por su parte, un aumento en 1 mm de precipitación es perjudicial en un 4.28% sobre la variable que se desea explicar. El aumento en fin de semana es de 12.32% mientras que en días de fin de semana y feriados el aumento combinado corresponde a 36.98%.

En contador Pablo Neruda – Olimpia, punto caracterizado por su cercanía a polos atractivos de comercio, es el sitio con mayores conteos diarios de los puntos seleccionados. Se encontró a través de la regresión lineal múltiple que un aumento de 1 °C de temperatura máxima trae por consecuencia un aumento de 3.30% en los conteos. De igual forma, un aumento de 1 mm en la precipitación disminuye los conteos diarios en 1.86%. Si el día corresponde a un fin de semana los conteos diarios descienden en un 25.86% según este modelo, lo cual significa que esta ciclovía se utiliza más en días de semana bajo las mismas condiciones meteorológicas. Este hecho diferencia a este punto de los dos contadores anteriores. También, si el día corresponde a un feriado y un fin de semana a la vez, los conteos descienden en un 70.31%. Bajo el modelo Poisson, un aumento en 1 °C en la temperatura máxima trae consigo un aumento de 4.14%, y un incremento de 1 mm en la precipitación provoca un descenso de un 1.80%. En días de fin de semana los conteos diarios descienden en 24.26% y en días feriados que coinciden con días de fin de semana el descenso aumenta a un 53.17%. Un aumento en la radiación de 100 W/m² aumenta las pasadas en 1.57%, y un aumento de la humedad desde 80 a 100% decrece los conteos en 8.33%. En el modelo Binomial Negativo, el aumento en una unidad de temperatura máxima provoca un aumento de 3.80% hasta el límite de 29.30 °C, punto en el cual los conteos diarios comienzan a descender en este contador. Lluvias acumuladas sobre 20 mm perjudican en más de un 70% los conteos diarios. En días de fin de semana el número de pasadas baja en 24.75%. La diferencia entre el primer día de datos y el último, arroja que bajo las mismas condiciones climáticas, los conteos diarios en este contador han aumentado en 89.55% por razones no atribuibles a cambios en los factores meteorológicos. Los prácticamente tres años de datos en este contador reflejan que la tasa de crecimiento ha sido de alrededor de 30% anual, siendo uno de los polos de crecimiento más importantes en la ciudad,

hecho que fue confirmado anteriormente en el análisis descriptivo presentado en el Capítulo 5 que confirma al sector centro – poniente como la zona de mayor uso de la infraestructura de ciclovías.

Finalmente, para el contador céntrico de San Martín – Mackenna, se concluye que para el modelo de regresión lineal múltiple un aumento en 1 °C provoca un ascenso de 2.30% en los conteos. Por su parte, un aumento en 1 mm de precipitación induce un descenso de conteos de 1.47%. Si el día corresponde a un fin de semana, el número de pasadas diarias descienden en un 52.84%. Con respecto al modelo Poisson para este contador, se halló que un aumento de 1 °C en la temperatura máxima trae por consecuencia un aumento de un 3.09%, mientras que un aumento de 1 mm en la precipitación produce un descenso del número de pasadas diarias de un 2.65%. El efecto de la temperatura es creciente hasta los 28.20 °C, momento en el cual esta variable es superada por la variable temperatura máxima al cuadrado, lo cual provoca que el efecto global de la temperatura sea negativo sobre los conteos diarios. Este hecho se condice con la realidad de la literatura revisada, en la cual, luego de cierto umbral de temperatura, se provoca el descenso del uso diario de la bicicleta, como es fundamentado en Schmiedeskamp y Zhao (2016). Bajo este modelo, si el día corresponde a un fin de semana, los conteos decrecen abruptamente en un 51.43%, mientras que si el día es fin de semana y además feriado, el descenso totaliza un 78.15%. El hecho de que este contador esté situado en la ciclovía de San Martín, uniendo el sector centro – poniente de la ciudad con el sector centro, induce la idea de que esta ciclovía es utilizada en forma obligatoria debido a trabajo, ya que su conectividad es fundamentalmente entre una zona comercial con una zona residencial. Finalmente, bajo el modelo Binomial Negativo, se encuentra lo siguiente: un aumento de la temperatura en 1 °C provoca un aumento de 3.02% en el número de pasadas, mientras que un aumento en la precipitación en 1 mm induce un descenso de 2.22%. El efecto negativo de la variable fin de semana es de un 51.59%, mientras que el efecto combinado de la variable anterior sumada con la variable feriados es de 79.09%.

Con respecto a los modelos realizados mediante el algoritmo de Bosques Aleatorios, se presenta la siguiente tabla resumen exponiendo la importancia que el algoritmo le asigna a las variables predictoras utilizando para ello el método de la impureza de los nodos.

Tabla 6.26. Importancia de las variables en los modelos *Random Forests*.

Variables	Ex Vía Férrea 1	Los Músicos – Los Creadores	Pablo Neruda – Olimpia	San Martín - Mackenna
TEMP_MAX	1575786	232702	5767434	907341
TEMP_MAX²	1707071	197929	5341798	871388
PRECIP	587135	65880	4024161	970784
RAD	703469	156207	1742679	344416
HUM	709276	149048	1282704	317321
WEEKEND	257192	4663	852787	2743342
MP_10	227705	34724	446690	308215
HOLIDAY	36359	1467	74257	205968
DAY	201827	41589	27463543	2243091

La tabla anterior evidencia que el algoritmo otorga mayor importancia relativa a las variables de temperatura y precipitación. Sin embargo, en el modelo de San Martín – Mackenna, las variables de fin de semana y de día secuencial adquieren mayor protagonismo. A modo de ilustración, se presenta la Figura A.90, la cual representa la disminución del error de este algoritmo de regresión a medida que se agregan más árboles de decisión.

6.4. Modelo de predicción multinivel

Debido a la naturaleza jerárquica y anidada de los datos tratados, se desarrolla un modelo multinivel de predicción. Las variables meteorológicas pueden explicar la variación de conteos diarios a través de los coeficientes que se obtienen en los modelos, pero son incapaces de explicar por qué en ciertos puntos o ciclovías de la ciudad existe un mayor volumen de ciclistas diarios que en otros. Hipotéticamente dicha diferencia pudiese ser atribuida a variables socioespaciales asociadas a cada ciclovía, que permiten diferenciarlas entre ellas. Para la elaboración de este modelo se han utilizado los datos en común disponibles para los 17 contadores de la ciudad a nivel agregado (diario). A modo de prueba, se realiza una regresión lineal múltiple con todos los contadores de la ciudad, con el objetivo de posteriormente comparar los resultados obtenidos con los del modelo multinivel. El modelo de regresión lineal múltiple que incluye los datos de todos los contadores automáticos de la ciudad está presentado en la Tabla 6.27.

Tabla 6.27. Modelo regresión lineal múltiple contadores Temuco.

Variables	Estimador	Error Estándar	Valor <i>t</i>	Valor <i>p</i>	Significancia
INTERCEPTO	-313.4	36.88	-8.49	< 2e-16	***
TEMP_MAX	27.10	1.74	15.50	< 2e-16	***
TEMP_MAX²	-0.46	0.03	-11.67	< 2e-16	***
RAD	0.08	0.009	8.84	< 2e-16	***
PRECIP	-6.10	0.32	-18.64	< 2e-16	***
HUM	-1.76	0.33	-5.26	1.50e-07	***
WEEKEND	-96.03	4.64	-20.69	< 2e-16	***
DAY	0.16	0.01	8.97	< 2e-16	***
HOLIDAY	-82.66	10.47	-7.90	3.26e-15	***
ENTROPY	792.80	12.01	66.01	< 2e-16	***
LONGITUD	101.10	3.83	26.40	< 2e-16	***

Código de significancia: 0 '****' 0.001 '***' 0.01 '**' '.' 0.1
R² ajustado = 0.55

Siguiendo lo indicado por las lecturas de los autores Knowles (2015), Buchanan (2017) y Crowson (2019) se realiza el siguiente procedimiento. En primer lugar, se implementa un modelo nulo, solo basado en el intercepto, el cual es presentado en la Tabla 6.28.

Tabla 6.28. Modelo nulo solo basado en el intercepto.

Variables	Estimador	Error Estándar	Valor <i>t</i>	Valor <i>p</i>	Significancia
Intercepto	340.28	3.11	109.37	< 2e-16	***

Código de significancia: 0 '****' 0.001 '***' 0.01 '**' '.' 0.1
AIC = 94997.70

Luego de obtener el modelo nulo, procede estimar un modelo con interceptos aleatorios que definan el efecto de las ciclovías, medidos a través de un parámetro que exprese la variabilidad de dichos grupos. Según De la Cruz (2008), en los modelos multinivel, las unidades que definen agrupamientos son vistos como efectos aleatorios, de esta forma, algunas de dichas unidades pueden ser escuelas, centros de salud, domicilios, entre otros. Estos efectos aleatorios se traducen en un modelo de interceptos aleatorios que van a tomar en cuenta la variabilidad entre agrupamientos. En la información otorgada por Knowles (2015), se indica que el siguiente paso luego de estimar una regresión lineal múltiple a modo de prueba, es ajustar un modelo de

Estimación de modelos de predicción de tránsito diario de bicicletas en base a datos pasivos, incorporando variables meteorológicas y contextuales para la ciudad de Temuco

interceptos aleatorios mediante una variable de agrupamiento, que en caso del estudio de dicho autor corresponde a las escuelas. A su vez, Buchanan (2017) y Crowson (2019) estiman un modelo nulo solo basado en el intercepto, para posteriormente implementar un modelo de interceptos aleatorios basándose en alguna variable que agrupe las observaciones en la unidad que se requiera. En el caso del presente estudio, la variable que permite distinguir las observaciones y agruparlas, es el nombre de cada contador automático. En ese sentido, el efecto fijo del modelo de interceptos aleatorios queda expresado en la Tabla 6.29.

Tabla 6.29. Efectos fijos del modelo de interceptos aleatorios.

Variables	Estimador	Error Estándar	Valor t	Valor p	Significancia
Intercepto	340.28	45.10	7.54	< 2e-16	***

Código de significancia: 0 '***' 0.001 '**' 0.01 '*' '.' 0.1

AIC = 90057.30

Por su parte, los efectos aleatorios definen una desviación estándar del intercepto de efecto fijo anterior, la cual queda expresada en la Tabla 6.30.

Tabla 6.30. Efectos aleatorios en el modelo de interceptos aleatorios.

Efectos Aleatorios	Valor	Residuos
Desviación Estándar	185.74	177.39

Esto significa que para el valor fijo del intercepto correspondiente a todos los contadores automáticos, existe un efecto aleatorio que explica la variabilidad del intercepto asociado a cada contador. De esta forma, los valores para los interceptos de cada ciclovía del modelo de interceptos aleatorios quedan resumidos en la Tabla 6.31.

Tabla 6.31. Valores de los interceptos aleatorios.

Nombre Contador	Valor del Intercepto
Ex Vía Férrea 1	243.54
Ex Vía Férrea 2	195.22
Gabriela Mistral - Andes	336.48
Gabriela Mistral - Recreo	339.09
Hochstetter - O'Higgins	532.05

Estimación de modelos de predicción de tránsito diario de bicicletas en base a datos pasivos, incorporando variables meteorológicas y contextuales para la ciudad de Temuco

Nombre Contador	Valor del Intercepto
Hochstetter - Pucón	241.50
Inés de Suarez	84.06
Javiera Carrera - Simón Bolívar	242.09
Los Creadores - Raúl Sáez	105.90
Los Músicos - Los Creadores	82.06
Los Pablos - Fundación	187.84
Pablo Neruda - Olimpia	636.10
Pablo Neruda - Javiera Carrera 1	413.37
Pablo Neruda - Javiera Carrera 2	583.53
San Martín - Caupolicán	445.84
San Martín - Mackenna	435.36
Simón Bolívar - Las Encinas	680.70

El modelo nulo presenta un valor para el criterio AIC de 94997.70, mientras que el modelo de interceptos aleatorios tiene un valor de 90057.30, reflejando que la segunda estimación presenta un mejor comportamiento, por lo cual, según lo indicado por Buchanan (2017) es necesario realizar un análisis multinivel.

Según el trabajo realizado por Crowson (2019), el siguiente paso es incorporar el nivel uno de variables predictoras. El nivel uno corresponde a la serie de variables meteorológicas diarias. Los resultados de incorporar este nivel de variables están expuestos en la siguiente tabla, donde se presentan los efectos fijos de las variables correspondientes al nivel uno.

Tabla 6.32. Efectos fijos de las variables de nivel uno.

Variables	Estimador	Error Estándar	Valor <i>t</i>	Valor <i>p</i>	Significancia
INTERCEPTO	114.59	52.10	2.19	0.02	*
TEMP_MAX	27.10	1.26	21.50	< 2e-16	***
TEMP_MAX²	-0.46	0.02	-16.19	< 2e-16	***
RAD	0.08	0.007	12.26	< 2e-16	***
PRECIP	-6.10	0.23	-25.86	< 2e-16	***
HUM	-1.76	0.24	-7.29	< 2e-16	***
WEEKEND	-96.02	3.34	-28.71	< 2e-16	***
DAY	0.16	0.01	12.45	< 2e-16	***
HOLIDAY	-82.66	7.54	-10.95	< 2e-16	***

Código de significancia: 0 '***' 0.001 '**' 0.01 '*' '.' 0.1

AIC = 85199.84

Por su parte, el efecto aleatorio asociado a la variable de grupo, queda explícito en la Tabla 6.33.

Tabla 6.33. Efectos aleatorios del modelo multinivel con nivel uno.

Efectos Aleatorios	Valor	Residuos
Desviación Estándar	185.85	123.96

El criterio de información de Akaike (AIC) para el modelo que incorpora el nivel uno de variables explicativas es de 85199.84, mejorando la estimación anterior. Luego, se añade el nivel dos de variables predictoras, asociados al contexto de cada ciclovía. Estas variables son muy importantes ya que definen las diferencias socioespaciales entre los contadores de la ciudad. El modelo que añade las variables contextuales de nivel dos se presenta en la Tabla 6.34.

Tabla 6.34. Efectos fijos de las variables de nivel uno y dos.

Variables	Estimador	Error Estándar	Valor <i>t</i>	Valor <i>p</i>	Significancia
INTERCEPTO	-313.39	105.60	-2.96	0.003	***
TEMP_MAX	27.10	1.26	21.50	< 2e-16	***
TEMP_MAX²	-0.46	0.02	-16.19	< 2e-16	***
RAD	0.08	0.007	12.26	< 2e-16	***
PRECIP	-6.10	0.23	-25.85	< 2e-16	***
HUM	-1.76	0.24	-7.29	< 2e-16	***
WEEKEND	-96.02	3.34	-28.70	< 2e-16	***
DAY	0.16	0.01	12.45	< 2e-16	***
HOLIDAY	-82.66	7.54	-10.95	< 2e-16	***
ENTROPY	792.77	166.91	4.74	0.0003	***
LONGITUD	101.13	53.23	1.89	0.07	.

Código de significancia: 0 '***' 0.001 '**' 0.01 '*' '.' 0.1

AIC = 85188.77

A su vez, el efecto aleatorio otorgado por la variable de agrupamiento es el siguiente:

Tabla 6.35. Efectos aleatorios del modelo multinivel con nivel uno y dos.

Efectos Aleatorios	Valor	Residuos
Desviación Estándar	119.20	123.96

Al llegar a este punto, se ha ajustado un modelo multinivel que posee un nivel uno de variables predictoras asociadas a factores meteorológicos y un nivel dos que caracteriza el contexto. Para ello, se consideró que los coeficientes de las variables son fijos en todos los *cluster*. Si se desea complejizar más aun el modelo, se puede recurrir a modelos que varíen sus coeficientes en cada grupo, sin embargo, debido a que las condiciones meteorológicas tienen un impacto similar sobre los conteos diarios en forma espacial en la ciudad de Temuco, se asumen como efectos fijos. Las únicas variables que podrían variar su comportamiento entre grupo podrían ser las variables temporales, tal como fue demostrado en los modelos puntuales, en los cuales, algunos contadores responden positivamente a las variables de fin de semana y otros presentan un efecto negativo en los conteos diarios gracias a dicha variable. Para efectos de mantener la simplicidad en el ajuste, se decide mantener el modelo multinivel de dos jerarquías con sus efectos fijos (*fixed effects*), pero aplicando una transformación para utilizar la distribución Binomial Negativa y de esta forma, respetar la restricción de la variable respuesta de conteo. En la Tabla 6.36 se expresan los coeficientes fijos del modelo multinivel desarrollado, el cual posee un valor del criterio de información de Akaike (AIC) de 85188.77, siendo el menor valor entre todos los ajustes realizados.

Al comparar la regresión lineal múltiple que utiliza los datos de todos los contadores con el modelo multinivel, se verifica que en la regresión lineal múltiple no se respeta la independencia de las observaciones, lo que trae como consecuencia que los errores estándar sean más pequeños y por lo tanto los valores t sean muy altos, otorgando una significancia estadística que no es real. Al ejecutar el modelo multinivel, los valores de los t estadísticos descienden notablemente, como puede observarse en la Tabla 6.34. La inclusión de variables contextuales permite caracterizar espacialmente cada ciclovía, entregando una variabilidad a los modelos en una jerarquía de datos diferente. Las variables de entorno construido de entropía en el uso de suelo y longitud de la ciclovía resultan ser estadísticamente significativas. A continuación, en la Tabla 6.36 se presentan los valores de los coeficientes del modelo multinivel de familia Binomial Negativa.

Tabla 6.36. Coeficientes del modelo multinivel de familia binomial negativa.

NOMBRE	INTER CEPTO	ENTROPY	LONGI TUD	TEMP_ MAX	PRECIP	TEMP_ MAX ²	RAD	HUM	WEEKEND	HOLIDAY	DAY
Ex Vía Férrea 1	3.510	2.842	0.398	0.086	-0.028	-0.0016	0.0003	-0.006	-0.240	-0.173	0.0006
Ex Vía Férrea 2	3.183	2.842	0.398	0.086	-0.028	-0.0016	0.0003	-0.006	-0.240	-0.173	0.0006
Gabriela Mistral – Andes	3.651	2.842	0.398	0.086	-0.028	-0.0016	0.0003	-0.006	-0.240	-0.173	0.0006
Gabriela Mistral – Recreo	3.308	2.842	0.398	0.086	-0.028	-0.0016	0.0003	-0.006	-0.240	-0.173	0.0006
Hochstetter – O’Higgins	4.157	2.842	0.398	0.086	-0.028	-0.0016	0.0003	-0.006	-0.240	-0.173	0.0006
Hochstetter – Pucón	2.847	2.842	0.398	0.086	-0.028	-0.0016	0.0003	-0.006	-0.240	-0.173	0.0006
Inés de Suarez	2.457	2.842	0.398	0.086	-0.028	-0.0016	0.0003	-0.006	-0.240	-0.173	0.0006
Javiera Carrera – Simón Bolívar	3.190	2.842	0.398	0.086	-0.028	-0.0016	0.0003	-0.006	-0.240	-0.173	0.0006
Los Pablos – Fundación	3.546	2.842	0.398	0.086	-0.028	-0.0016	0.0003	-0.006	-0.240	-0.173	0.0006
Los Músicos – Los Creadores	2.724	2.842	0.398	0.086	-0.028	-0.0016	0.0003	-0.006	-0.240	-0.173	0.0006
Los Creadores – Raúl Sáez	2.656	2.842	0.398	0.086	-0.028	-0.0016	0.0003	-0.006	-0.240	-0.173	0.0006
Pablo Neruda – Javiera Carrera 1	2.861	2.842	0.398	0.086	-0.028	-0.0016	0.0003	-0.006	-0.240	-0.173	0.0006
Pablo Neruda – Javiera Carrera 2	3.202	2.842	0.398	0.086	-0.028	-0.0016	0.0003	-0.006	-0.240	-0.173	0.0006
Pablo Neruda – Olimpia	3.502	2.842	0.398	0.086	-0.028	-0.0016	0.0003	-0.006	-0.240	-0.173	0.0006
San Martín – Caupolicán	3.215	2.842	0.398	0.086	-0.028	-0.0016	0.0003	-0.006	-0.240	-0.173	0.0006
San Martín – Mackenna	3.025	2.842	0.398	0.086	-0.028	-0.0016	0.0003	-0.006	-0.240	-0.173	0.0006
Simón Bolívar – Las Encinas	3.508	2.842	0.398	0.086	-0.028	-0.0016	0.0003	-0.006	-0.240	-0.173	0.0006

Estimación de modelos de predicción de tránsito diario de bicicletas en base a datos pasivos, incorporando variables meteorológicas y contextuales para la ciudad de Temuco

La Figura A.91 representa la predicción de la variable dependiente y los datos reales en función de la entropía en el uso de suelo de cada ciclovía. Se observa que la predicción es bastante acertada y se puede distinguir como se ordenan los grupos asociados a cada contador. La Figura A.92 representa el gráfico de dispersión entre los conteos diarios reales y la predicción del presente modelo. Finalmente, la Figura A.93 simboliza la curva cuantil-cuantil (gráfico Q-Q) de los datos reales, la cual refleja que los datos no se ajustan adecuadamente a una distribución normal. La Figura A.94 representa la misma curva pero para las predicciones del modelo.

6.5. Validación de los modelos

En este apartado se muestra la validación de los modelos desarrollados. El procedimiento empleado corresponde a una validación cruzada, evaluando en los modelos desarrollados los datos que han quedado fuera del entrenamiento de ellos. Luego de dicha evaluación, se calculó el porcentaje de error absoluto medio (MAPE) entre los datos reales y las predicciones. El resumen del porcentaje de error de cada modelo desarrollado queda definido en la Tabla 6.37.

Tabla 6.37. Porcentaje de error absoluto medio de modelos desarrollados.

Modelos	Ex Vía Férrea 1	Los Músicos – Los Creadores	Pablo Neruda - Olimpia	San Martín - Mackenna
Regresión Lineal Múltiple	35.82%	48.04%	24.96%	27.08%
Modelo Poisson	28.47%	36.49%	23.44%	22.94%
Modelo Binomial Negativo	28.55%	36.64%	23.65%	22.98%
<i>Random Forests</i>	0.07%	0.08%	0.05%	0.07%

La tabla anterior demuestra que el error desciende progresivamente desde los primeros modelos estimados hasta el último. En particular, el error obtenido con el modelo *Random Forests* resulta ser muy bajo. A pesar de este notable resultado, la desventaja es que si se desea evaluar un dato ficticio para obtener una predicción de conteos diarios, se debe realizar computacionalmente, pues este tipo de ajuste no arroja coeficientes como los modelos anteriores, de manera de realizar una predicción en forma inmediata. Con respecto a los modelos Poisson y Binomial Negativo, se observa un rendimiento muy similar entre ellos, debido a la naturaleza común de las distribuciones.

CAPÍTULO 7

CONCLUSIONES

Capítulo 7. Conclusiones

- Esta investigación es útil para clarificar las relaciones existentes entre factores meteorológicos y contextuales sobre los conteos diarios de ciclistas en la ciudad de Temuco. En general, todos los modelos puntuales mostraron que tanto las variables de temperatura máxima diaria, temperatura máxima diaria al cuadrado, precipitación, humedad, radiación, fin de semana, feriados y día secuencial son estadísticamente significativas en la explicación de la variable respuesta.
- Esta memoria contribuye a la literatura existente demostrando el uso de diversos modelos para la estimación de conteos diarios de bicicletas. El presente estudio complementa la literatura referida a la bicicleta para la ciudad de Temuco, en la cual destaca el estudio desarrollado por Eltit (2011). Para ello, se ha utilizado una metodología similar a estudios internacionales y cuenta con el sustento de haberse realizado en algunos casos con tres años de datos. En el modelo multinivel, se concluye que las variables del entorno construido (longitud de la ciclovía e índice de entropía de uso de suelo) son estadísticamente significativas y además tienen un efecto positivo sobre la variable a explicar.
- Si bien este estudio se realiza a nivel agregado, resulta ser una buena primera aproximación a estudios de este tipo, obteniendo valores de R^2 ajustado superiores a 0.75 en el caso de los modelos puntuales. La Comisión de Estudios Habitacionales y Urbanos realizó en el año 2017 y 2018 informes acerca de los contadores de bicicletas y su relación con variables meteorológicas (Ministerio de Vivienda y Urbanismo, 2017), sin embargo, el documento solo se limita a realizar cruces de variables versus el número de pasadas diarias, identificando el efecto por separado de dichos factores y no agrupándolas en un modelo matemático. Dicho estudio fue realizado con un grupo de datos menor y también agrupando todas las ciclovías del país, por lo que el efecto de dichas variables se generaliza a todo el territorio nacional, perdiendo aplicabilidad local.
- De la evaluación de la efectividad de los contadores automáticos, se concluye que estos dispositivos logran capturar de forma adecuada la mayoría de pasadas diarias de ciclistas en la ciudad, totalizando un error aproximado de 10% en los conteos, por lo cual, se establece que su precisión es de aproximadamente 90%.

- La elaboración de la base de datos de este estudio y las variables recolectadas en ella, fue realizada utilizando la literatura internacional aplicable a la ciudad y los criterios de selección ya mencionados. Se concluye que en caso de poseer mayor cantidad de variables meteorológicas en la estación utilizada, se podría realizar un análisis más completo con el objetivo de incorporar más variables a los modelos y determinar si son significativas en la explicación de la variable respuesta.
- Entre los resultados de esta investigación se encuentra el efecto relativamente constante en forma espacial que poseen las variables meteorológicas en los diferentes contadores de la ciudad. Esto quiere decir que los valores de los coeficientes de dichas variables son similares en diferentes puntos geográficos de la ciudad, lo que significa que la meteorología impacta de forma equivalente a la ciudad de Temuco en sus diferentes sectores. Sin embargo, las variables categóricas presentan distinto comportamiento entre un lugar de la ciudad y otro. Este comportamiento dispar puede estar asociado al uso obligatorio o recreacional que se le da a ciertas vías.
- Sobre los modelos implementados, se concluye que la regresión lineal múltiple no constituye una buena aproximación debido a que realiza algunas predicciones de signo negativo, no respetando las restricciones de la variable de respuesta. El modelo Poisson mejora esta problemática pero posee el problema de sobredispersión en los datos, por lo cual, el modelo Binomial Negativo constituye la mejor opción entre los modelos lineales generalizados aplicados a modelos puntuales.
- La técnica de bosques aleatorios o *Random Forests* resulta ser la herramienta más precisa en la explicación de la variable respuesta, pero no es capaz de establecer coeficientes de manera de evaluarlos inmediatamente en una expresión matemática, lo cual constituye una importante desventaja en comparación con los modelos lineales generalizados.
- El modelo multinivel agrupa todos los contadores automáticos en un modelo global, incorporando además variables contextuales. Este ajuste resulta ser más preciso que una regresión lineal múltiple que agrupe todos los dispositivos, debido a que al ejecutar dicho modelo se estaría violando el principio de independencia de las observaciones, obteniendo magnitudes para los valores t que resultan muy grandes, disminuyendo los errores estándar de cada variable.

- De los resultados del análisis descriptivo se desprende que el uso de la infraestructura de ciclovías en la ciudad está en un constante aumento general, sobre todo en los meses de verano. En aquellos meses en que la meteorología afecta los conteos diarios, no se observa un crecimiento sostenido de un año a otro, hecho que sugiere que la mayoría de nuevos ciclistas que realizan pasadas se suman al uso estival de este modo de transporte y en menor grado al uso invernal.
- De los mapas presentados en el análisis descriptivo, se concluye que algunas ubicaciones de los contadores automáticos no son las más adecuadas, sobre todo en el contador de Inés de Suarez, debido a que se encuentra cerca de San Martín en un punto en que no existe conectividad a la red de ciclovías, lo cual puede afectar el número de pasadas en dicho punto. El autor sugiere instalar los dispositivos en la mitad de cada ciclovía de forma de establecer un punto representativo.

7.1. Futuras aplicaciones

- Modelos de este tipo pueden ser utilizados para estimar cambios en el uso de la bicicleta debido al cambio climático. A través de los modelos predictivos, los municipios podrían estimar como cambiarían los volúmenes diarios de ciclistas modificando valores de temperatura, precipitación, humedad, entre otros. También, es posible evaluar un punto ficticio en el modelo una vez caracterizado contextualmente.
- Si bien el control de factores meteorológicos escapa de la capacidad de aquellas personas que desarrollan políticas públicas, este estudio sugiere que planificadores de transporte y las políticas públicas en general desarrollen estrategias que ayuden a mitigar los efectos naturales que ocurren durante los meses de invierno en la ciudad. Futuras investigaciones podrían enfocarse en determinar si algún tipo de programa o infraestructura podría aminorar los efectos del clima en el uso de la bicicleta.
- Una futura línea de investigación podría abarcar el desarrollo de un modelo a nivel horario, con el objetivo de conocer los patrones horarios de los viajes en la ciudad. De igual forma, se sugiere incluir mayor cantidad de variables del entorno construido en la implementación de un modelo multinivel, siendo ésta una de las limitantes más grandes de este estudio, junto con el hecho de no incluir el sentido del flujo de las pasadas.

Bibliografía

- Alcaide, M. (2015). "Modelo de Regresión Binomial Negativa". Trabajo para optar al grado en Matemáticas. Universidad de Sevilla, Sevilla.
- An, R., Zahnow, R., Pojani, D., Corcoran, J. (2019). "Weather and cycling in New York: The case of Citibike". *Journal of Transport Geography*. **77**: 97-112.
- Apparicio, P., Martori, J.C. (2014). "Geo-Segregation Analyzer: una herramienta para el análisis de la segregación residencial". *Asociación Española de Ciencia Regional*. 1-19.
- Apparicio, P., Petkevitch, V., Charron, M. (2008). "Segregation Analyzer: a C#.Net application for calculating residential segregation indices". *Cybergeog: European Journal of Geography*. **414**.
- Asociación Nacional Automotriz de Chile. (2018). "Tasa de motorización: en Chile hay un auto cada 3.8 personas". <https://www.diarioconcepcion.cl/economia-y-negocios/2018/06/12/tasa-de-motorizacion-en-chile-hay-un-auto-cada-38-personas.html>. Concepción. Visitada el 10 de Agosto de 2019.
- Bates, D., Maechler M., Bolker B., Walker S. (2015). "Fitting Linear Mixed-Effects Models Using lme4". *Journal of Statistical Software*. **67**(1): 1-48.
- Brandenburg, C., Matzarakis, A., Arnberger, A. (2007). "Weather and cycling - A first approach to the effects of weather conditions on cycling". *Meteorological Applications*. **14**(1): 61-67.
- Breiman, L. (1996). "Bagging Predictors". *Machine Learning*. **24**: 123-140.
- Breiman, L. (2001). "Random Forests". *Machine Learning*. **45**: 5-32.
- Buchanan, E. (2017). "Multilevel Models Lecture". <http://statstools.com/learn/advanced-statistics/>. Visitada el 12 de Noviembre de 2019.
- Chardon, I. (2019). "Predicción de rentas en Santiago de Chile utilizando algoritmos de aprendizaje automático". Memoria para optar al título de Ingeniero Civil. Universidad de Chile, Santiago.
- Cox, D.R., Wermuth, N. (1996). "Multivariate Dependencies: Models, Analysis and Interpretation". Chapman and Hall, New York. ISBN: 978-0412754104.
- Crowson, M. (2019). "Multilevel modeling in R with 'lme4' package". <https://www.youtube.com/watch?v=LzAwEKrn2Mc>. Visitada el 13 de Noviembre de 2019.
- De Grange, L. (2017). "Chile superará 5 millones de vehículos en 2017: motorización sigue lejos de países desarrollados". <http://www.economiaynegocios.cl/noticias/noticias.asp?id=341770>. Santiago. Visitada el 11 de Septiembre de 2019.
- De la Cruz, F. (2008). "Modelos multinivel". *Revista Peruana de Epidemiología*. **12**(3): 1-8.
- Del Valle, J., Guerra, C. (2012). "La Multicolinealidad en modelos de Regresión Lineal Múltiple". *Revista Ciencias Técnicas Agropecuarias*. **21**(4): 80-83.

- Efron, B. (1979). "Bootstrap methods: Another look at the jackknife". *The Annals of Statistics*. 7(1): 1-26.
- El Esawey, M. (2014). "Estimation of Annual Average Daily Bicycle Traffic with Adjustment Factors". *Transportation Research Record: Journal of the Transportation Research Board*. 2443(1): 106-114.
- El Esawey, M. (2016). "Toward a Better Estimation of Annual Average Daily Bicycle Traffic: Comparison of Methods for Calculating Daily Adjustment Factors". *Transportation Research Record: Journal of the Transportation Research Board*. 2593(1): 28-36.
- Eltit, V. (2011). "Transporte Urbano No Motorizado: El Potencial de la Bicicleta en la Ciudad de Temuco". *Invi*. 26(72): 153-184.
- Fournier, N., Christofa, E., Knodler, M. (2017). "A sinusoidal model for seasonal bicycle demand estimation". *Transportation Research Part D: Transport and Environment*. 50: 154-169.
- Gallop, C., Tse, C., Zhao, J. (2012). "A Seasonal Autoregressive Model of Vancouver Bicycle Traffic Using Weather Variables". *91st Transportation Research Board Annual Meeting, Washington, D.C.*
- Griswold, J.B., Medury, A., Schneider, J.R. (2011). "Pilot models for estimating bicycle intersection volumes". *Transportation Research Record: Journal of the Transportation Research Board*. 2247(1): 1-7.
- Groißböck, W., Lughofer, E., Klement, E.P. (2004). "A Comparison of Variable Selection Methods with the Main Focus on Orthogonalization". *Soft Methodology and Random Information Systems: Advances in Soft Computing*. 26: 479-486.
- Handy, S., Boarnet, M., Erwing, R., Killingsworth, R. (2002). "How the built environment affects physical activity: Views from urban planning". *American Journal of Preventive Medicine*. 23(2): 64-73.
- Hankey, S., Lindsey, G. (2016). "Facility-demand Models of Peak Period Pedestrian and Bicycle Traffic: Comparison of Fully Specified and Reduced-Form Models". *Transportation Research Record: Journal of the Transportation Research Board*. 2586(1): 48-58.
- Hankey, S., Lu, T., Mondschein, A., Buehler, R. (2017). "Spatial models of active travel in small communities: Merging the goals of traffic monitoring and direct-demand modeling". *Journal of Transport & Health*. 7: 149-159.
- Hanson, S., Hanson, P. (1977). "Evaluating the Impact of Weather on Bicycle Use". *Transportation Research Record*. 629: 43-48.
- Helbich, M., Bocker, L., Dijst, M. (2014). "Geographic Heterogeneity in Cycling Under Various Weather Conditions: Evidence from Greater Rotterdam". *Journal of Transport Geography*. 38: 38-47.
- Hilbe, J. (1993). "Log-negative binomial regression as a generalized linear model". Arizona State University, Tempe.

Hunt, J.D., Abraham, J.E. (2007). "Influences on Bicycle Use". *Transportation: Planning Policy*. **34**(4): 453-470.

Instituto Nacional de Estadísticas. (2017). "Censo de Población y Vivienda 2017".

Jones, M., Ryan, S., Donlon, J., Ledbetter, L., Ragland, D., Arnold, L. (2010). "Seamless Travel: Measuring Bicycle and Pedestrian Activity in San Diego County and its Relationship to Land Use, Transportation, Safety, and Facility Type". *California Partners for Advanced Transportation Technology Program: Institute of Transportation Studies*. University of California, Berkeley.

Knowles, J. (2015). "Getting Started with Mixed Effects Models in R". <https://www.jaredknowles.com/journal/2013/11/25/getting-started-with-mixed-effect-models-in-r>. Visitada el 12 de Noviembre de 2019.

Lewin, A. (2011). "Temporal and Weather Impacts on Bicycle Volumes". *90th Transportation Research Board Annual Meeting, Washington, D.C.*

Liaw, A., Wiener, M. (2002). "Classification and Regression by randomForest". *R News*. **2**(3): 18-22.

Lindsey, G., Wilson, J., Rubchinskaya, E., Yang, J., Han, Y. (2007). "Estimating Urban Trail Traffic: Methods for Existing and Proposed Trails". *Landscape and Urban Planning*. **81**(4): 299-315.

Lindsey, J. (1995). "Modelling Frequency and Count Data". Oxford University Press, Oxford. ISBN: 978-0198523314.

Lizana, M. (2017). "Relaciones entre uso de tiempo, indicadores sociales, gasto y patrones de movilidad asociados a actividades no obligatorias: Caso de estudio en Concepción, Chile". Tesis para optar al grado de Magister en Ciencias de la Ingeniería con mención en Ingeniería Civil. Universidad de Concepción, Concepción.

Lu, T., Mondschein, A., Buehler, R., Hankey, S. (2018). "Adding temporal information to direct-demand models: Hourly estimation of bicycle and pedestrian traffic in Blacksburg, VA". *Transportation Research Part D: Transport and Environment*. **63**(1): 244-260.

Massey, D., Denton, N. (1988). "The Dimensions of Residential Segregation". *Social Forces*. **67**(2): 281-315.

McCahill, C., Garrick, N. (2008). "The Applicability of Space Syntax to Bicycle Facility Planning". *Transportation Research Record: Journal of the Transportation Research Board*. **2074**(1): 46-51.

McCullagh, P., Nelder, J. (1989). "Generalized Linear Models". *Monographs on statistics and applied probability*. **37**.

Medina, S. (2012). "La importancia de reducción del uso del automóvil en México: Tendencias de motorización, del uso del automóvil y de sus impactos". Instituto de Políticas para el Transporte y Desarrollo México.

Microsoft Corporation. (2016). "Microsoft Excel". <https://office.microsoft.com/excel>.

- Ministerio de Desarrollo Social, Secretaría de Planificación de Transporte. (2009). "Construcción red de ciclovías, Copiapó".
- Ministerio de Desarrollo Social, Secretaría de Planificación de Transporte. (2013). "Metodología para la formulación y evaluación de planes maestros de ciclo-rutas".
- Ministerio de Transportes y Telecomunicaciones. (2017). "Situación del Transporte Regional". http://www.temucotemueve.cl/ttm/?page_id=659. Visitado el 1 de Agosto de 2019.
- Ministerio de Vivienda y Urbanismo. (2018). "Proyectos de infraestructura vial y urbana conurbación Temuco-Padre las Casas".
- Ministerio de Vivienda y Urbanismo. (2017). "Análisis de Contadores de Ciclovías: Resultados preeliminares". Comisión de Estudios Habitacionales y Urbanos.
- Miranda-Moreno, L.F., Fernandes, D. (2011). "Modeling of Pedestrian Activity at Signalized Intersections". *Transportation Research Record: Journal of the Transportation Research Board*. **2264**(1): 74-82.
- Miranda-Moreno, L.F., Nosal, T. (2011). "Weather or Not to Cycle: Temporal Trends and Impact of Weather on Cycling in an Urban Environment". *Transportation Research Record: Journal of the Transportation Research Board*. **2247**(1): 42-52.
- Molino, J.A., Kennedy J., Johnson, P.L., Beuse, P.A., Emo, A.K., Do, A. (2009). "Pedestrian and Bicyclist Exposure to Risk: Methodology for Estimation in an Urban Environment". *Transportation Research Record: Journal of the Transportation Research Board*. **2140**(1): 145-156.
- Murillo, J. (2008). "Los modelos multinivel como herramienta para la investigación educativa". *Revista Internacional de Investigación en Educación*. **1**(1): 45-62.
- Nankervis, M. (1999). "The effect of weather and climate on bicycle commuting". *Transportation Research Part A: Policy and Practice*. **33**(6): 417-431.
- Nelder, J.A., Wedderburn R.W. (1972). "Generalized Linear Models". *Journal of the Royal Statistical Society*. **135**(3): 370-384.
- New Zealand Transport Agency. (2009). "Continuous cycle counting trial".
- Niemeier, D.A. (1996). "Longitudinal analysis of bicycle count variability: Results and modeling implications". *Journal of Transportation Engineering*. **122**(3): 200-206.
- Nordback, F., Janson, B. (2010). "Automated Bicycle Counts: Lessons from Boulder, Colorado". *Transportation Research Record: Journal of the Transportation Research Board*. **2190**(1): 11-18.
- Nosal, T., Miranda-Moreno, L.F. (2014). "The effect of weather on the use of North American bicycle facilities: A multi-city analysis using automatic counts". *Transportation Research Part A: Policy and Practice*. **66**(1): 213-225.
- Nosal, T., Miranda-Moreno, L.F., Krstulic, Z. (2014). "Incorporating weather: Comparative analysis of annual average daily bicyclist traffic estimation methods". *Transportation Research Record: Journal of the Transportation Research Board*. **2468**(1): 100-110.

- Oliva, I. (2017). "Rol del entorno construido en la frecuencia del uso de la bicicleta como modo de transporte". Tesis para optar al grado de Magister en Ciencias de la Ingeniería. Pontificia Universidad Católica de Chile, Santiago.
- Ortúzar, J. D., Willumsen, L. G. (1990). "Modelling Transport". Wiley, Chichester. ISBN: 978-0-470-76039-0.
- Osinga, J. (2018). "Uncovering the "black box": A study on how to make machine learning techniques more interpretable in an application to loan default prediction". Erasmus University Rotterdam, Rotterdam.
- Parkin, J., Wardman, M., Page, M. (2008). "Estimations of the Determinants of Bicycle Mode Share for the Journey to Work Using Census Data". *Transportation*. **35**(1): 93-109.
- Patil, A., Musale, K., Rao, P. (2015). "Bike Share Demand Prediction using RandomForests". *International Journal of Innovative Science, Engineering & Technology*. **2**(4): 1218-1223.
- Pulughurta, S., Repaka, S. (2008). "Assesment of Models to Measure Pedestrian Activity at Signalized Intersections". *Transportation Research Record: Journal of the Transportation Research Board*. **2073**(1): 39-48.
- R Core Team. (2019). "R: A language and environment for statistical computing". <https://www.R-project.org>.
- Roll, J., Proulx, F. (2018). "Estimating Annual Average Daily Bicycle Traffic without Permanent Counter Stations". *Transportation Research Record: Journal of the Transportation Research Board*. **2672**(43): 145-153.
- Rose, G., Ahmed, F., Figliozzi, M., Jakob, C. (2011). "Quantifying and comparing effects of weather on bicycle demand in Melbourne, Australia, and Portland, Oregon". *90th Transportation Research Board Annual Meeting, Washington, D.C.*
- Schmiedeskamp, P., Zhao, P. (2016). "Estimating Daily Bicycle Counts in Seattle, Washington, from Seasonal and Weather Factors". *Transportation Research Record: Journal of the Transportation Research Board*. **2593**(1): 94-102.
- Schneider, R., Arnold, L., Ragland, D. (2009). "Methodology for Counting Pedestrians at Intersections: Use of Automated Counters to Extrapolate Weekly Volumes from Short Manual Counts". *Transportation Research Record: Journal of the Transportation Research Board*. **2140**(1): 1-12.
- Secretaría de Planificación de Transporte. (2014). "Actualización Plan de Transporte Temuco y desarrollo de anteproyecto: I Etapa". TRASA Ingeniería Ltda.
- Stigler, S. (1981). "Gauss and the invention of least squares". *The Annals of Statistics*. **9**(3): 465-474.
- Suthaharan, S. (2015). "Machine Learning and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning". Springer, Boston. ISBN: 978-1-4899-7641-3.

- Theil, H. (1972). "Statistical decomposition analysis: with applications in the social and administrative sciences". *Studies in mathematical and managerial economics*. **14**(1): 325-328.
- Thomas, T., Jaarsma, R., Tutert, B. (2009). "Temporal Variations of Bicycle Demand in the Netherlands: Influence of Weather on Cycling". *88th Transportation Research Board Annual Meeting, Washington, D.C.*
- Tin Tin, S., Woodward, A., Robinson, E., Ameratunga, S. (2012). "Temporal, seasonal and weather effects on cycle volume: an ecological study". *Environmental Health*. **11**(1): 1-9.
- Tuszynski, J. (2019). "caTools: Tools: moving window statistics, GIF, Base64, ROC, AUC, etc.".
- Urbanismo y Territorio. (2016). "Plan de monitoreo Antofagasta en bicicleta: La importancia de contar". Informe de Consultoría.
- Wadud, Z. (2014). "Cycling in a changed climate". *Journal of Transport Geography*. **35**(1): 12-20.
- Wang, J., Hankey, S., Wu, X., Lindsey, G. (2016). "Monitoring and Modeling of Urban Trail Traffic: Validation of Direct Demand Models in Minneapolis, Minnesota, and Columbus, Ohio". *Transportation Research Record: Journal of the Transportation Research Board*. **2593**(1): 47-59.
- Wang, X., Lindsey, G., Hankey, S., Hoff, K. (2014). "Estimating mixed-mode urban trail traffic using negative binomial regression models". *Journal of Urban Planning and Development*. **140**(1).
- Wendel-Vos, W., Giesbers, H., Kruize, H. (2018). "Cycling in the Netherlands". National Institute for Public Health and the Environment.
- Wickham, H. (2016). "ggplot2: Elegant Graphics for Data Analysis". Springer-Verlog, New York. ISBN: 978-0-387-98141-3.

ANEXOS

Anexo A. Figuras

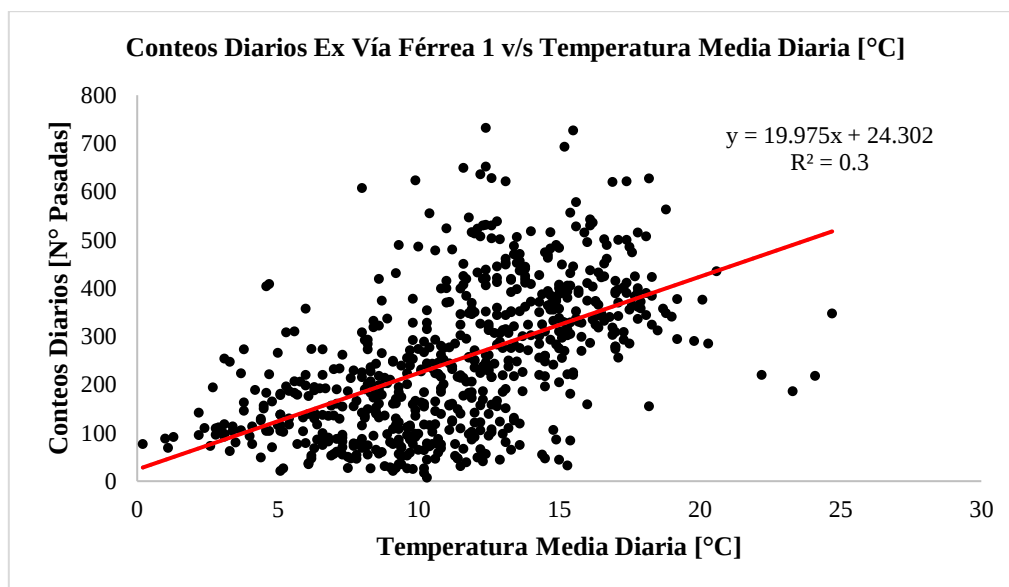


Figura A.1. Conteos diarios [N° Pasadas] v/s temperatura media diaria [°C].

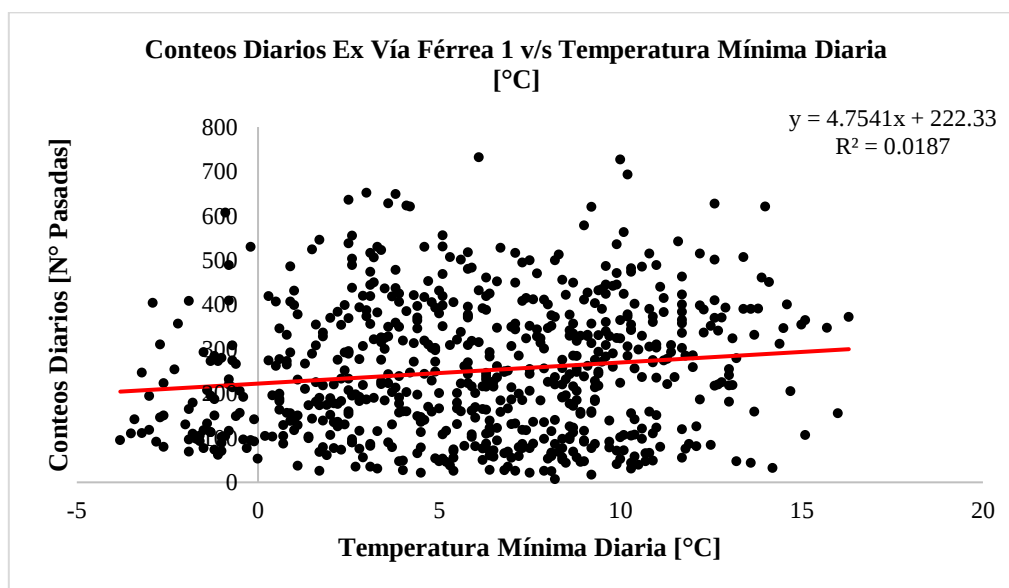


Figura A.2. Conteos diarios [N° Pasadas] v/s temperatura mínima diaria [°C].

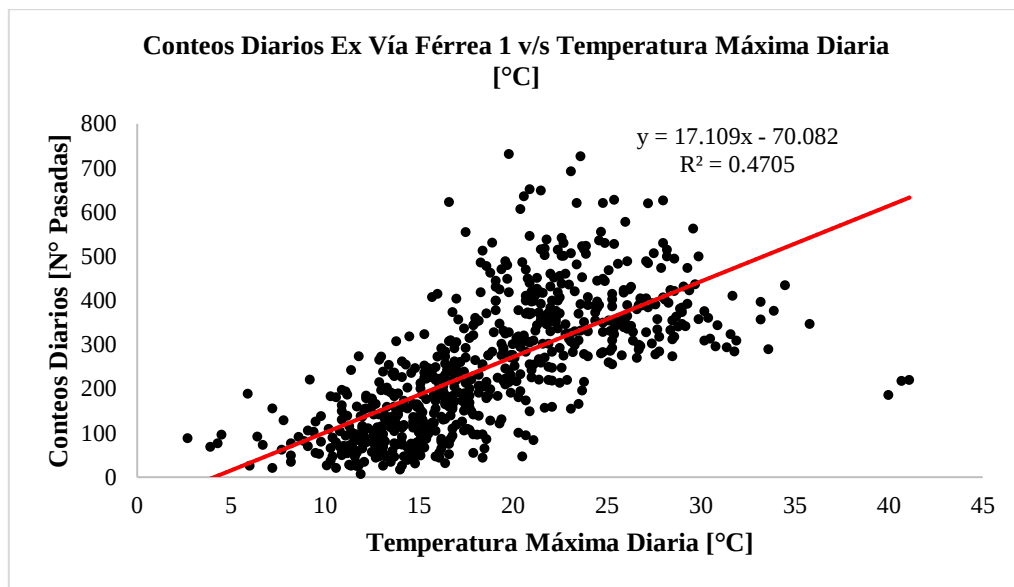


Figura A.3. Conteos diarios [Nº Pasadas] v/s temperatura máxima diaria [°C]

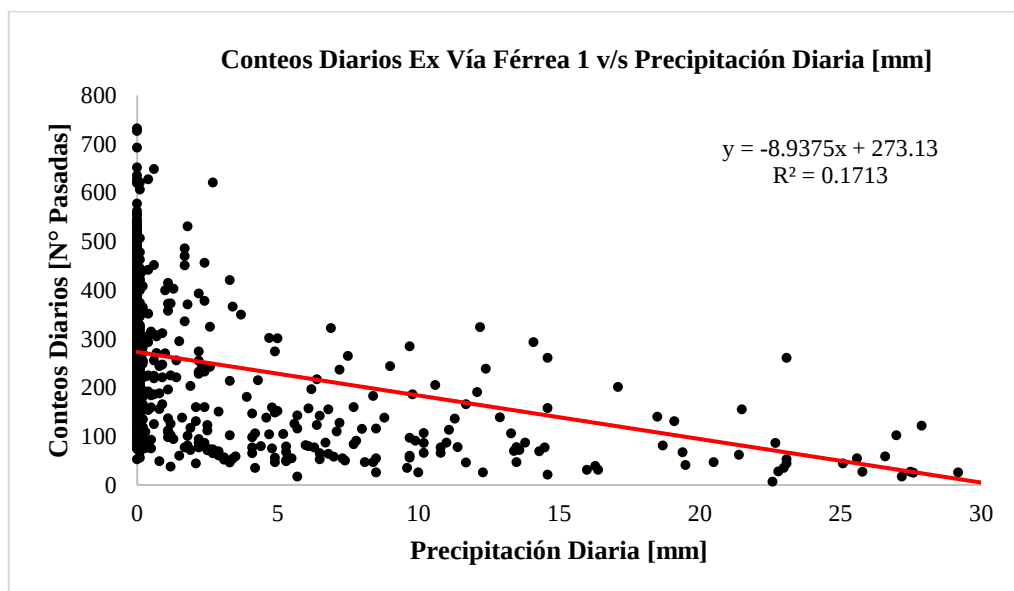


Figura A.4. Conteos diarios [Nº Pasadas] v/s precipitación diaria [mm].

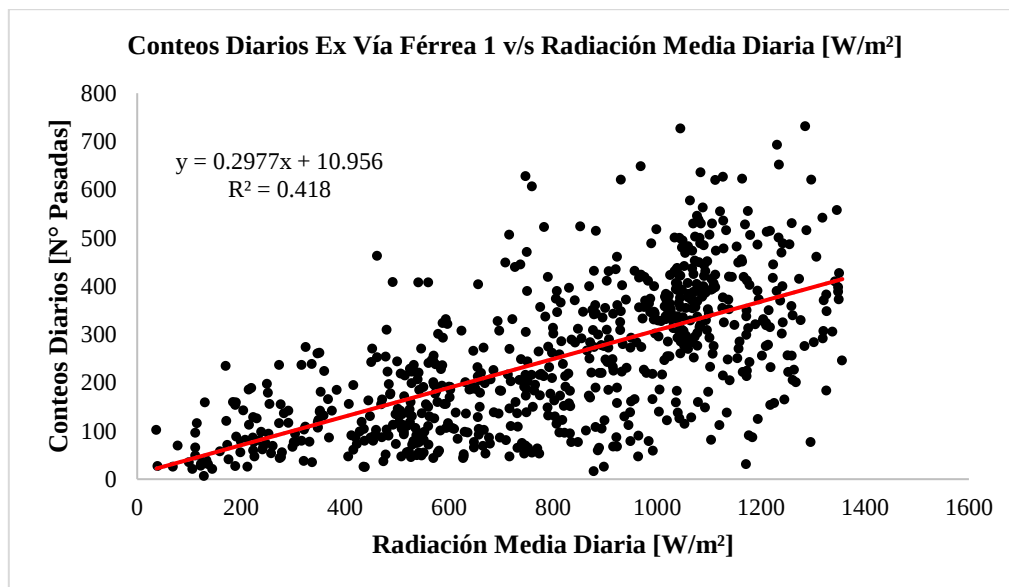


Figura A.5. Conteos diarios [N° Pasadas] v/s radiación media diaria [W/m²].

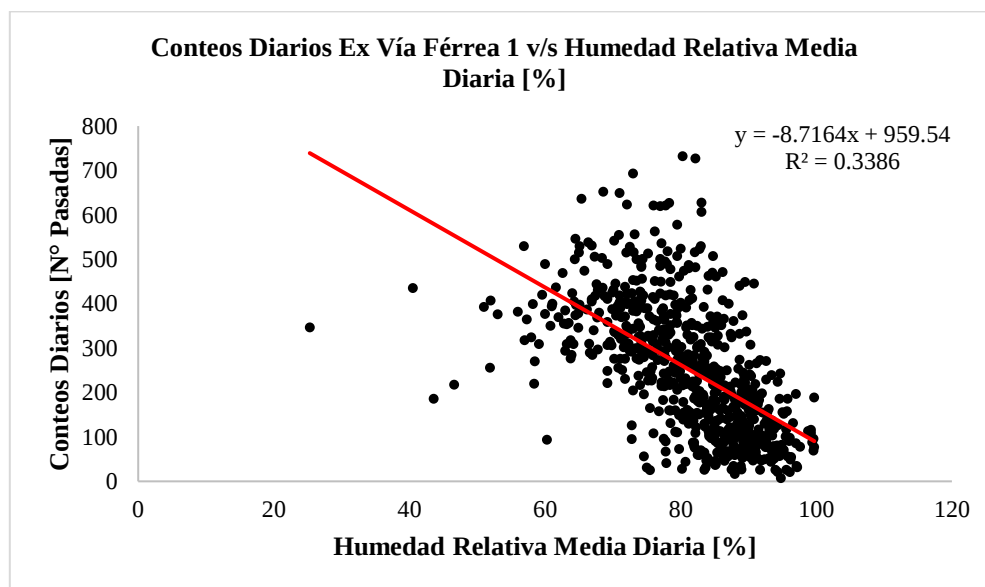


Figura A.6. Conteos diarios [N° Pasadas] v/s humedad relativa media diaria [%].

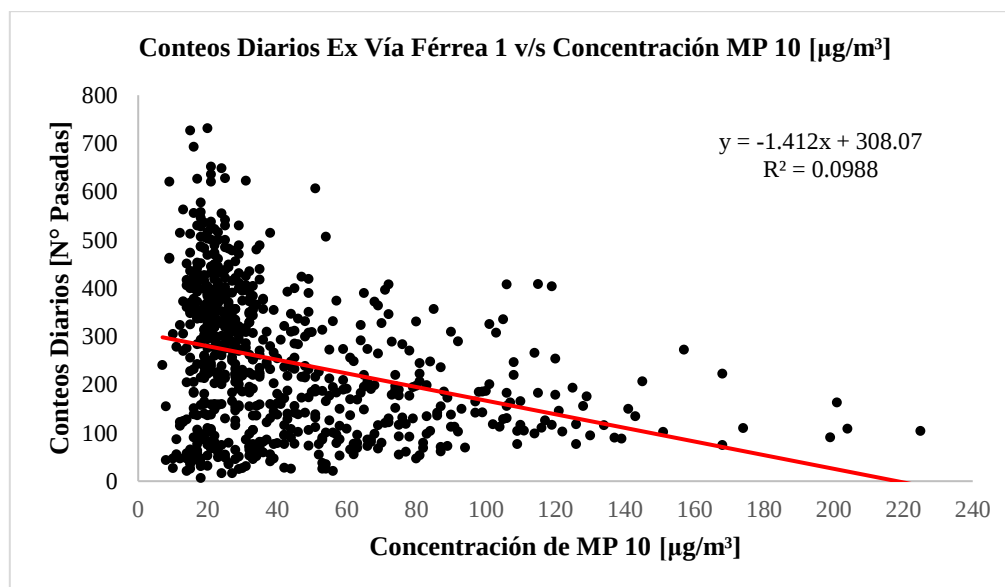


Figura A.7. Conteos diarios [Nº Pasadas] v/s material particulado medio diario 10 [$\mu\text{g}/\text{m}^3$].

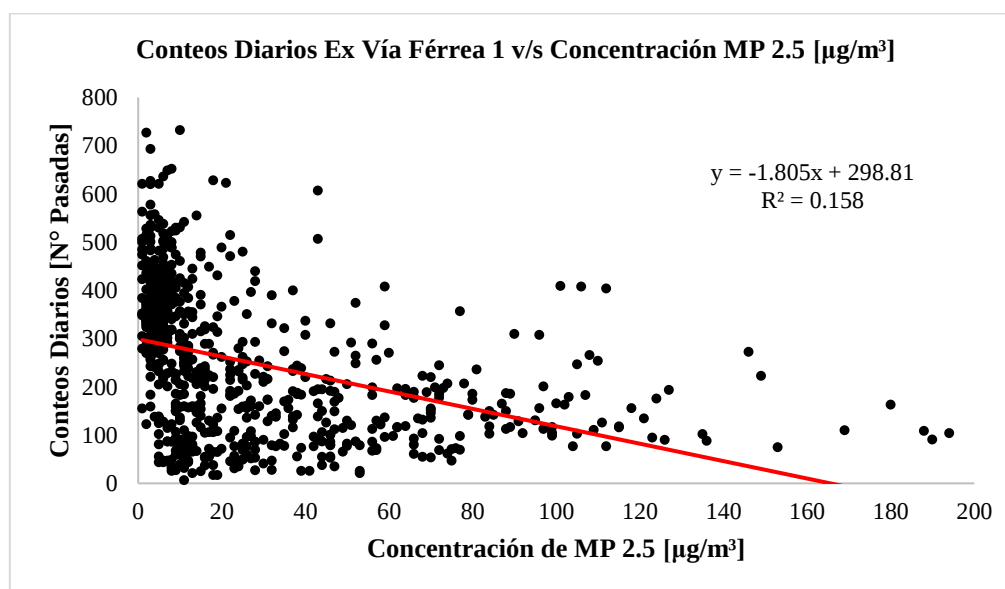


Figura A.8. Conteos diarios [Nº Pasadas] v/s material particulado medio diario 2.5 [$\mu\text{g}/\text{m}^3$].

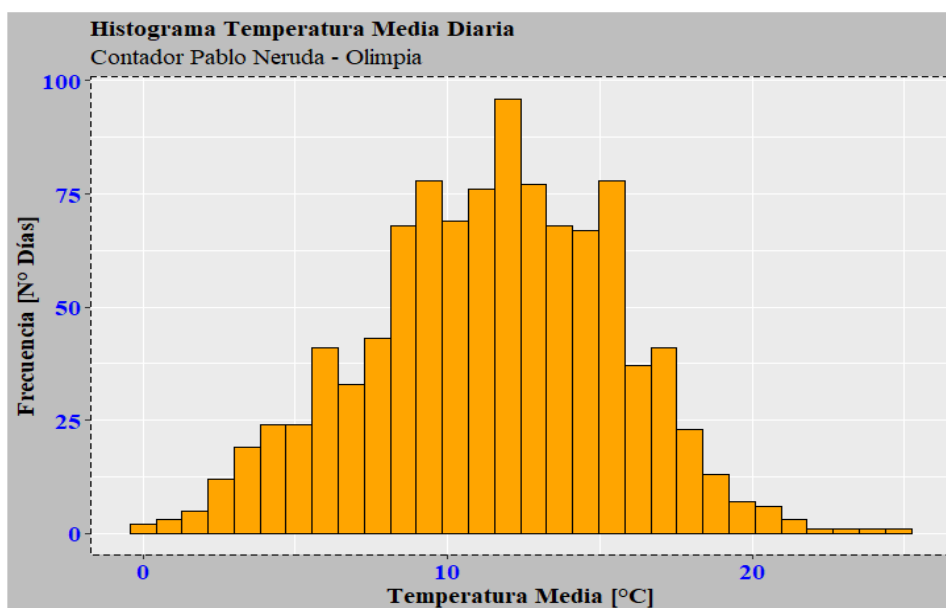


Figura A.9. Histograma de temperatura media diaria [°C].

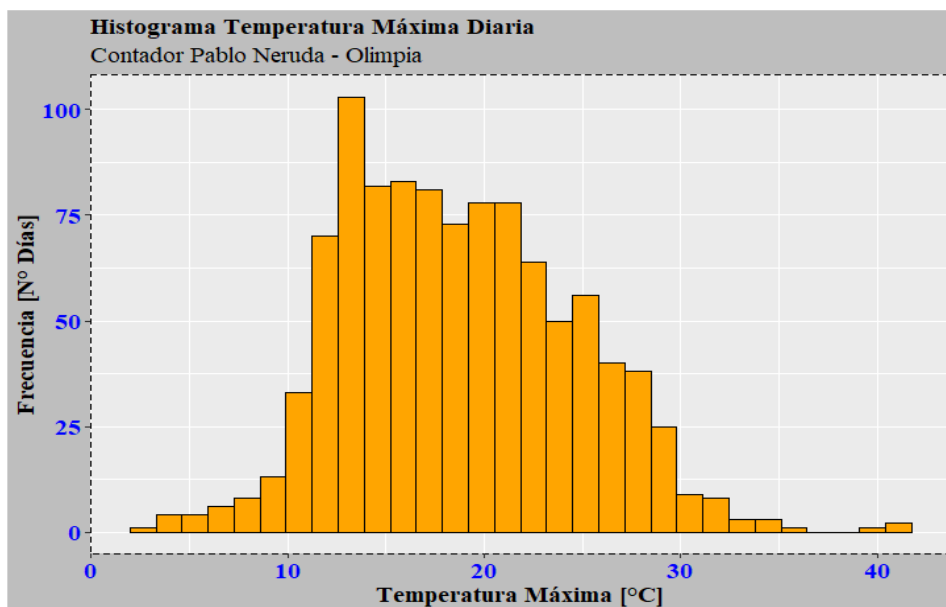


Figura A.10. Histograma de temperatura máxima diaria [°C].

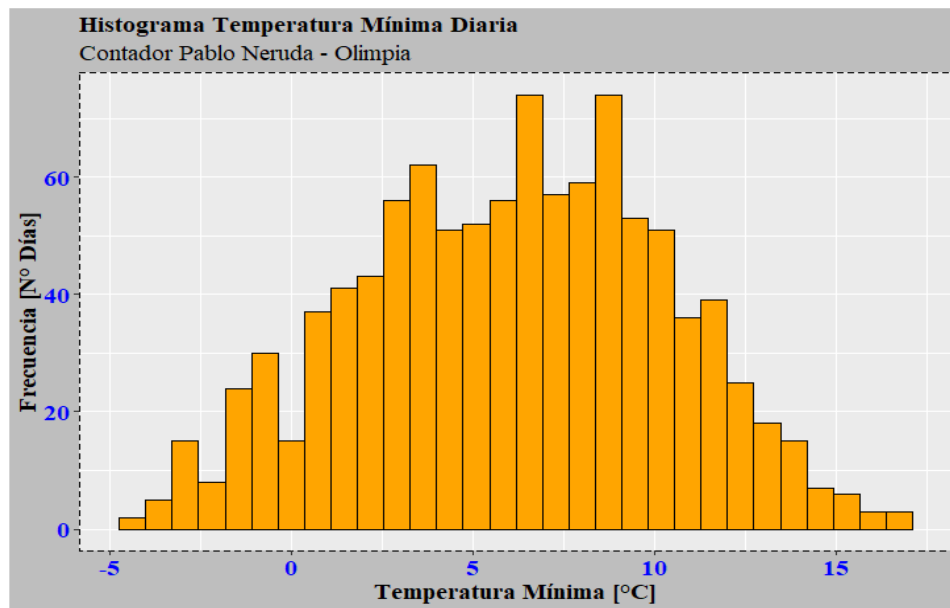


Figura A.11. Histograma de temperatura mínima diaria [°C].

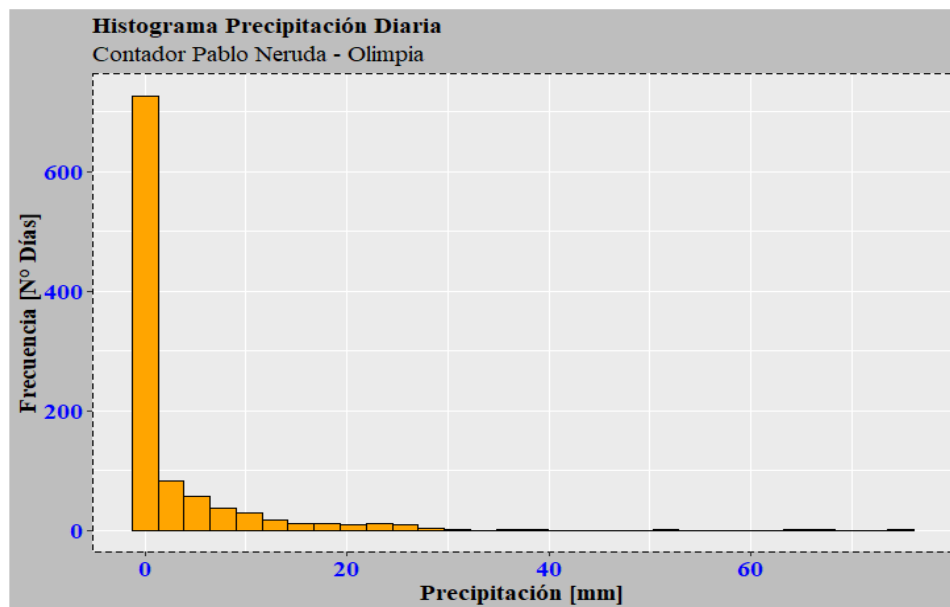


Figura A.12. Histograma de precipitación diaria [mm].

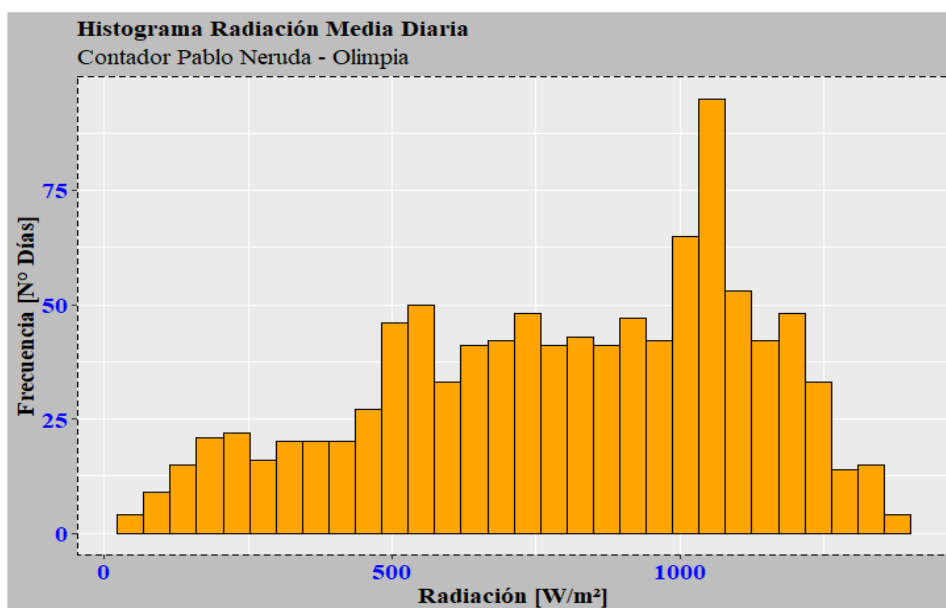


Figura A.13. Histograma de radiación media diaria [W/m^2].

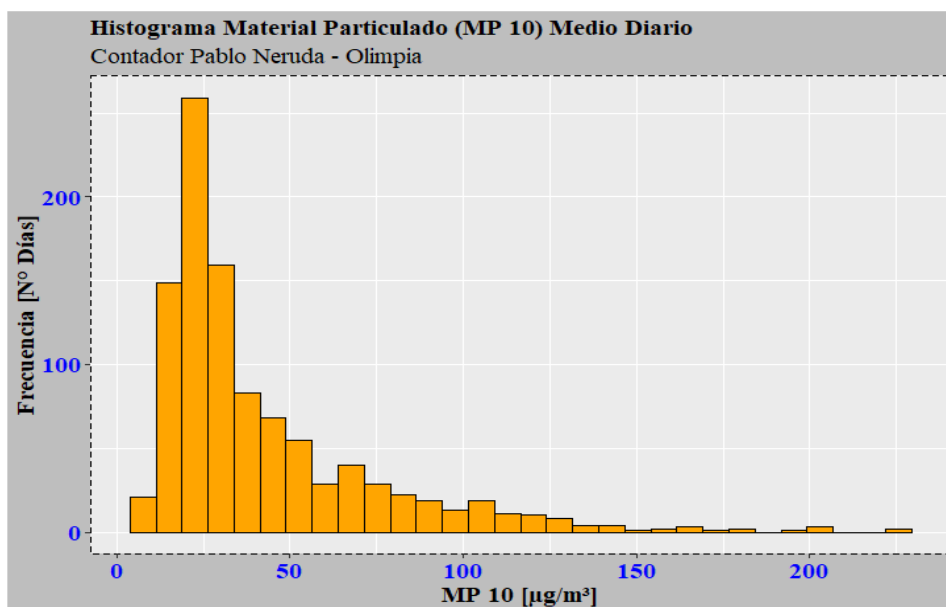


Figura A.14. Histograma de material particulado medio diario 10 [$\mu\text{g/m}^3$].

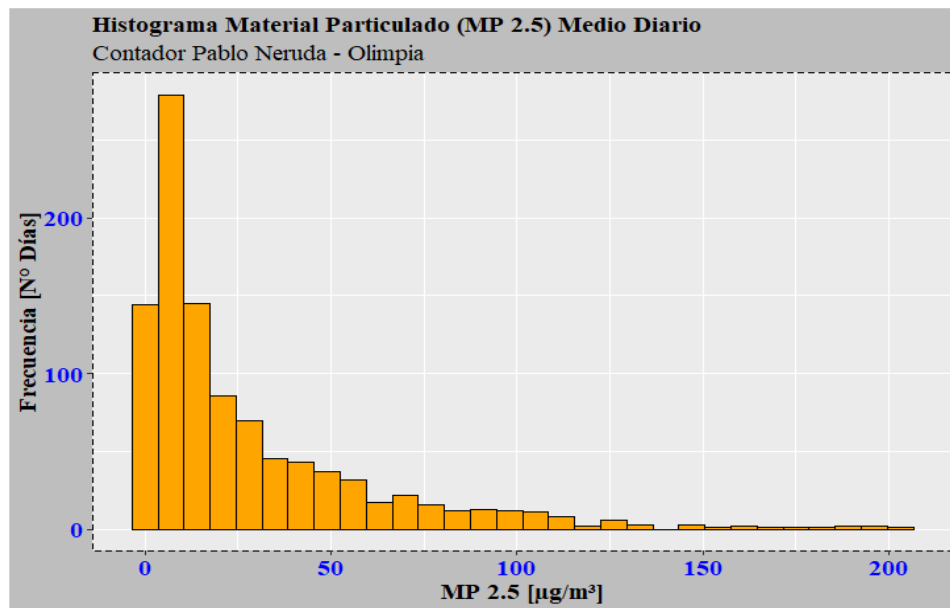


Figura A.15. Histograma de material particulado medio diario MP 2.5 [$\mu\text{g}/\text{m}^3$].

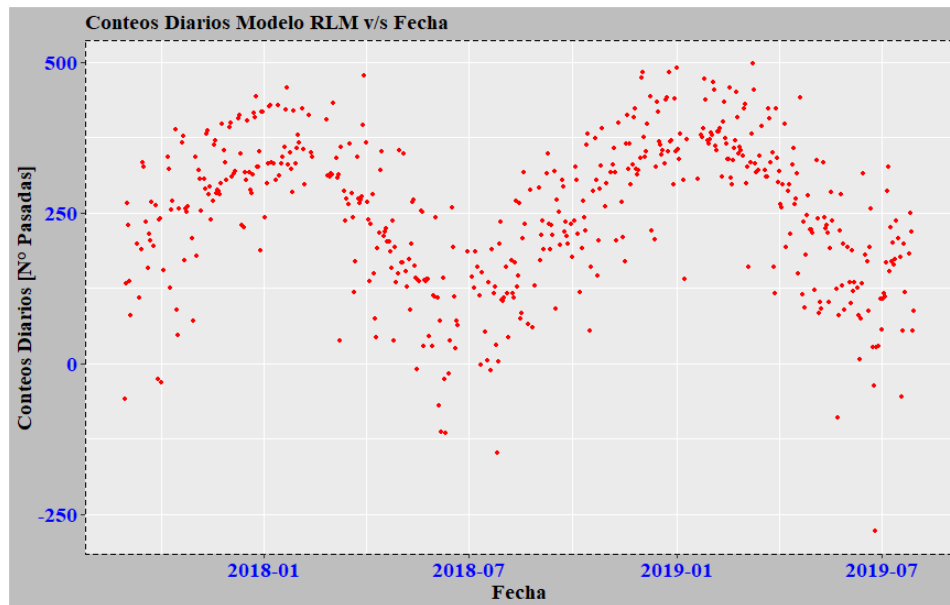


Figura A.16. Modelo regresión lineal múltiple v/s fecha para el contador Ex Vía Férrea 1.

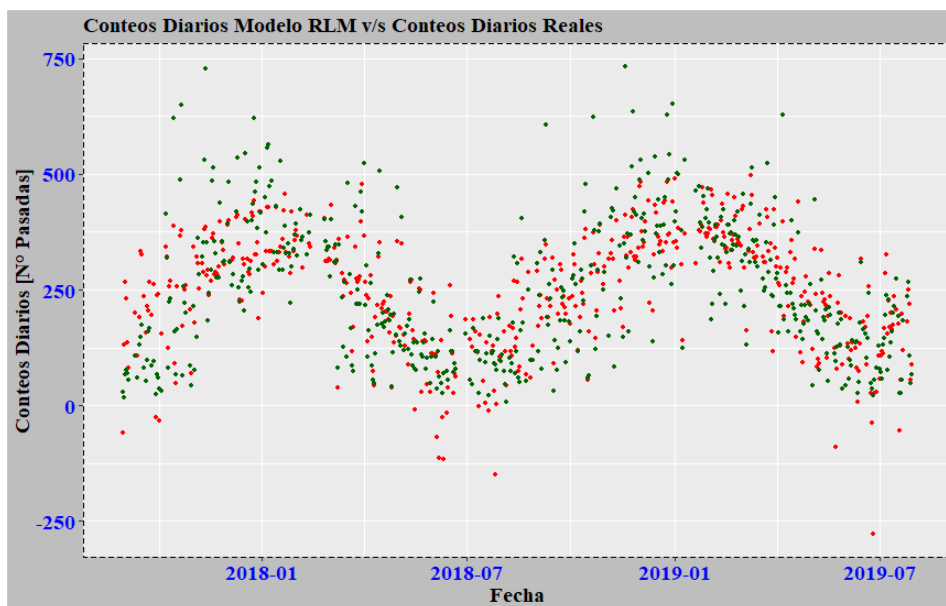


Figura A.17. Predicción conteos diarios modelo regresión lineal múltiple v/s conteos diarios reales en función del tiempo para el contador Ex Vía Férrea 1.

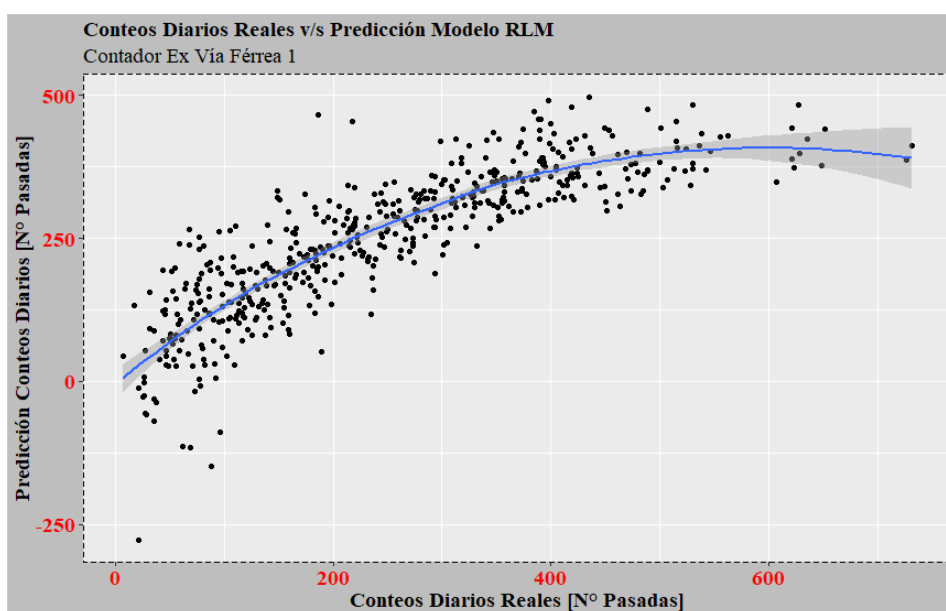


Figura A.18. Predicción conteos diarios modelo regresión lineal múltiple v/s conteos diarios reales para el contador Ex Vía Férrea 1.

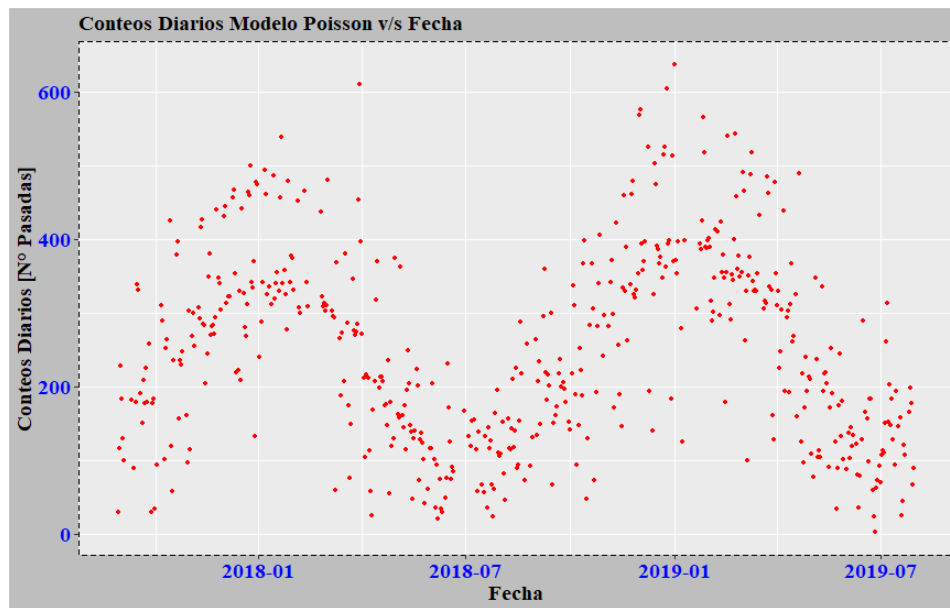


Figura A.19. Modelo Poisson v/s fecha para el contador Ex Vía Férrea 1.

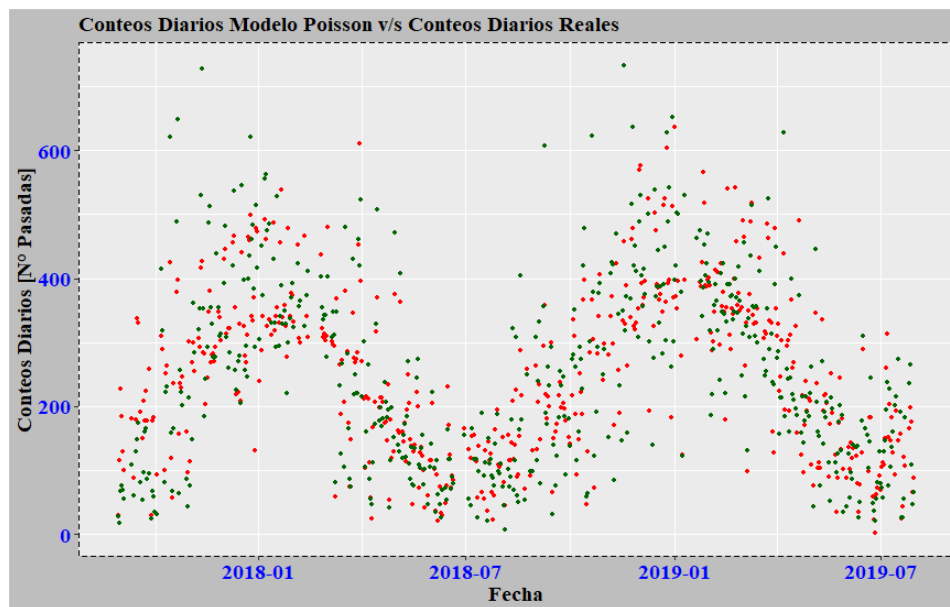


Figura A.20. Predicción conteos diarios modelo Poisson v/s conteos diarios reales en función del tiempo para el contador Ex Vía Férrea 1.

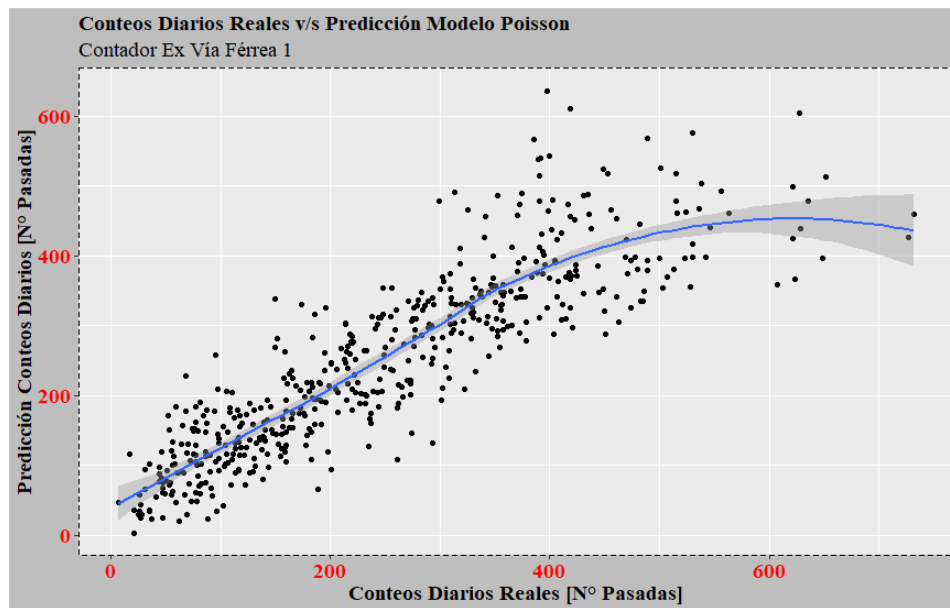


Figura A.21. Predicción conteos diarios modelo Poisson v/s conteos diarios reales para el contador Ex Vía Férrea 1.

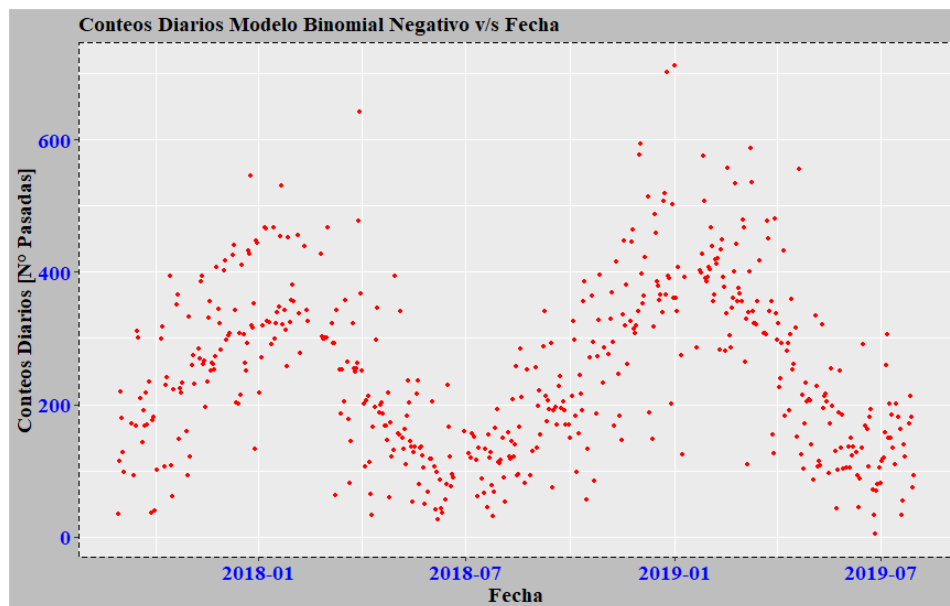


Figura A.22. Modelo binomial negativo v/s fecha para el contador Ex Vía Férrea 1.

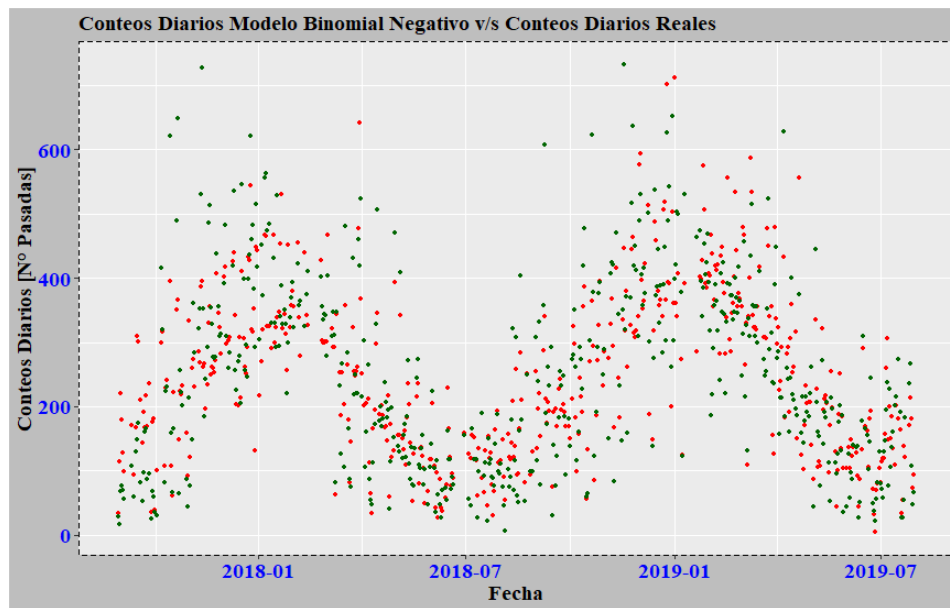


Figura A.23. Predicción conteos diarios modelo binomial negativo v/s conteos diarios reales en función del tiempo para el contador Ex Vía Férrea 1.

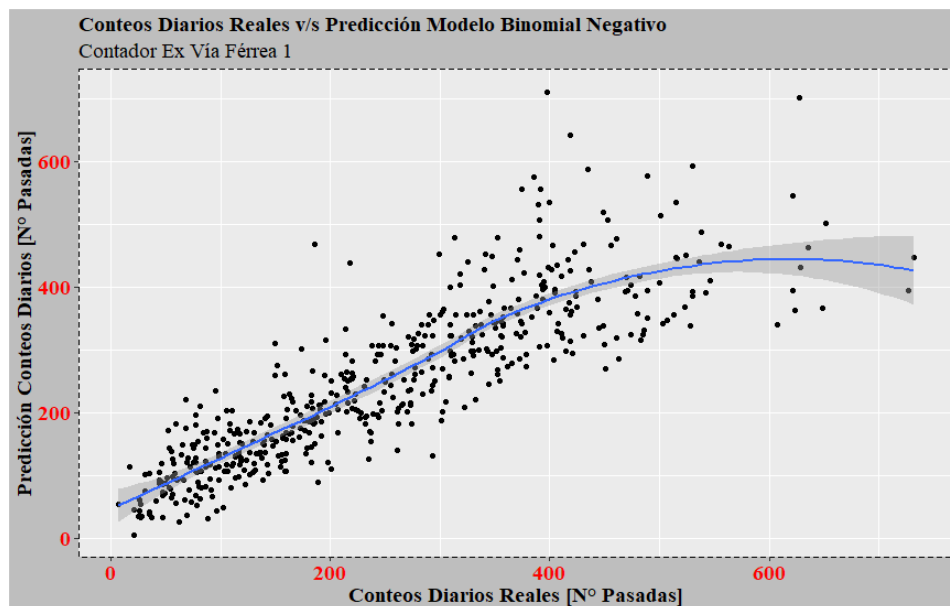


Figura A.24. Predicción conteos diarios modelo binomial negativo v/s conteos diarios reales para el contador Ex Vía Férrea 1.

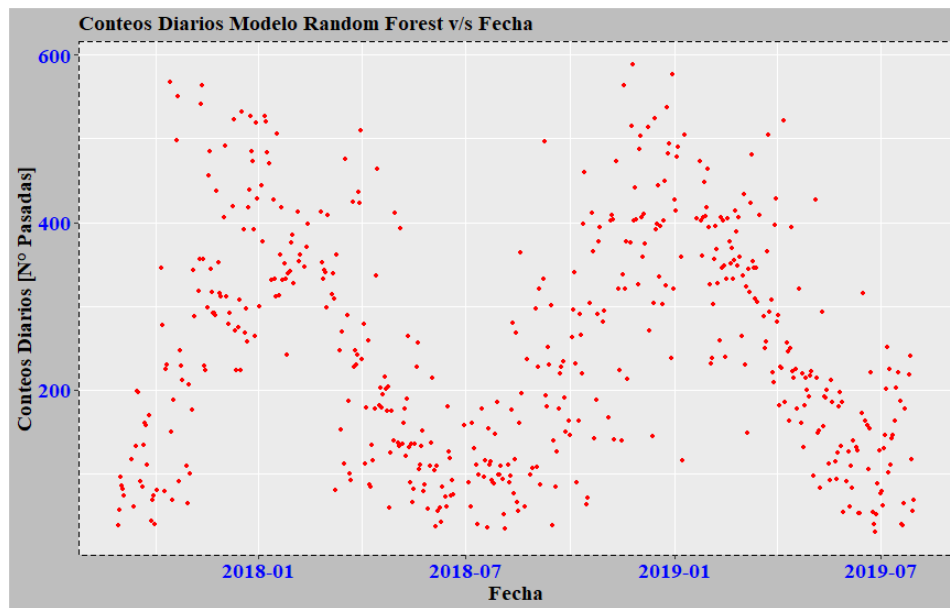


Figura A.25. Modelo *Random Forests* v/s fecha para el contador Ex Vía Férrea 1.

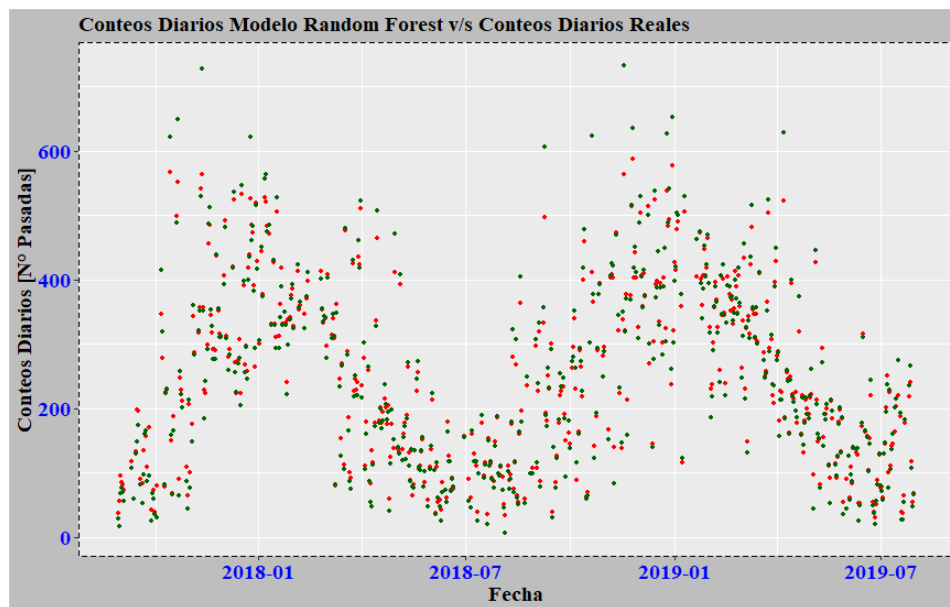


Figura A.26. Predicción conteos diarios modelo *Random Forests* v/s conteos diarios reales en función del tiempo para el contador Ex Vía Férrea 1.

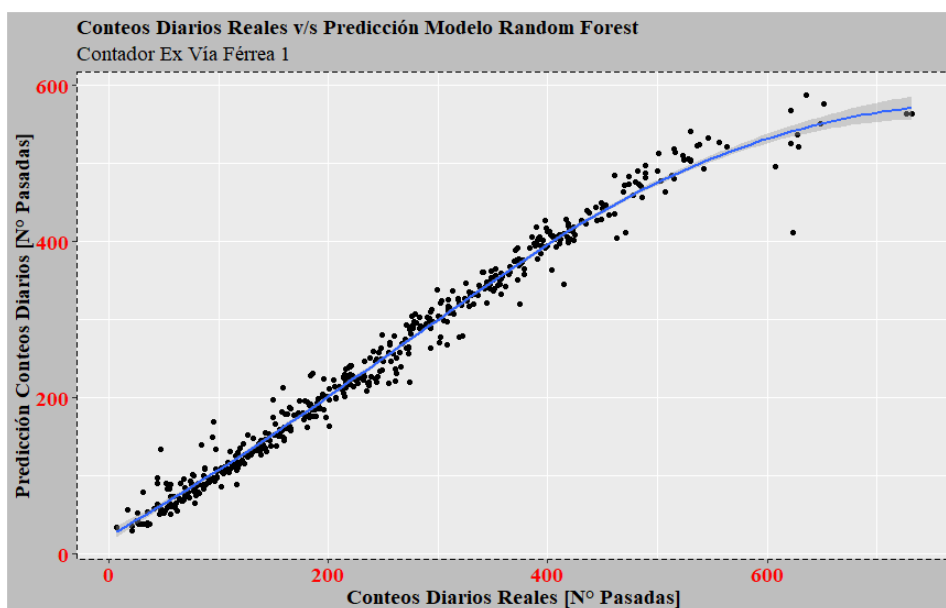


Figura A.27. Predicción conteos diarios modelo *Random Forests* v/s conteos diarios reales para el contador Ex Vía Férrea 1.

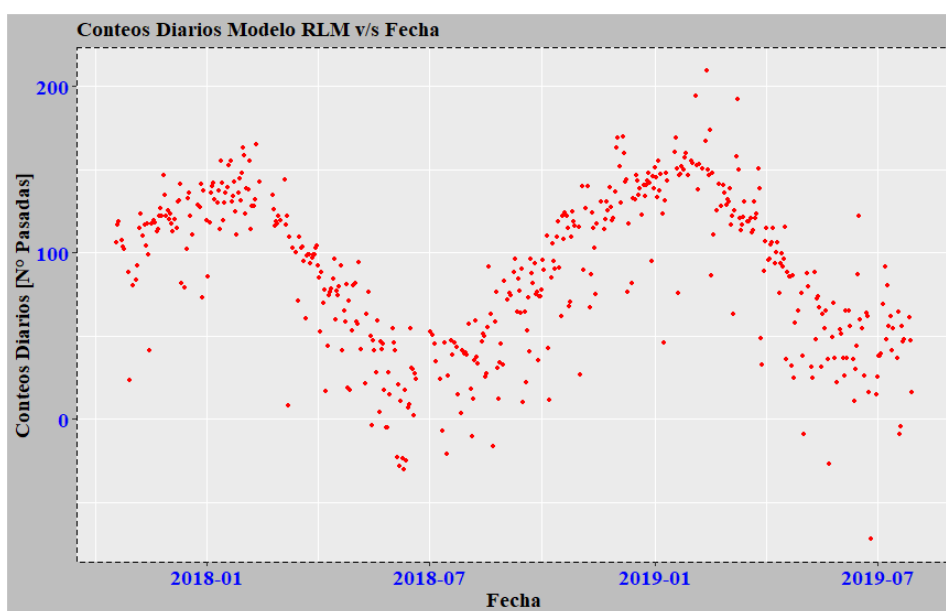


Figura A.28. Modelo regresión lineal múltiple v/s fecha para el contador Los Músicos – Los Creadores.

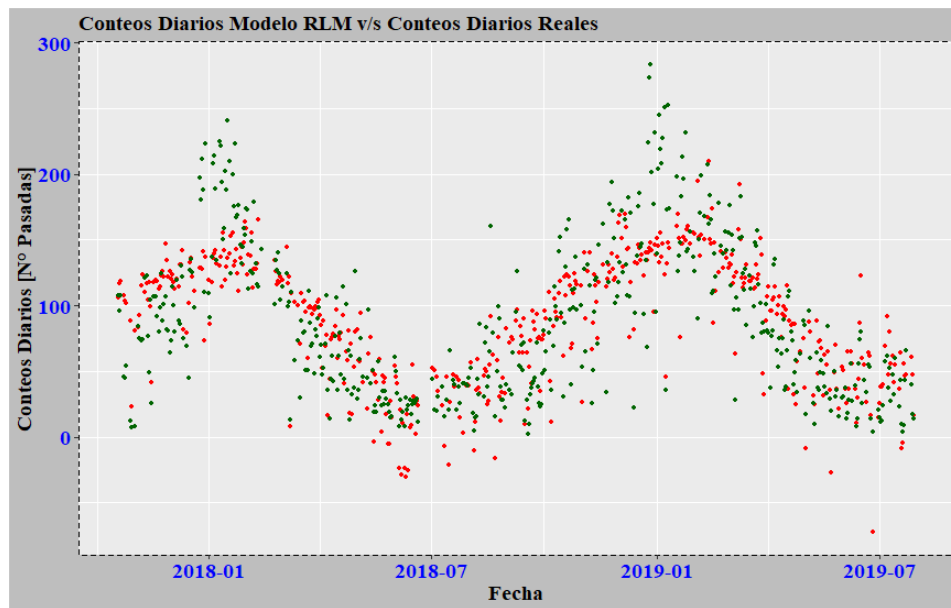


Figura A.29. Predicción conteos diarios modelo regresión lineal múltiple v/s conteos diarios reales en función del tiempo para el contador Los Músicos – Los Creadores.

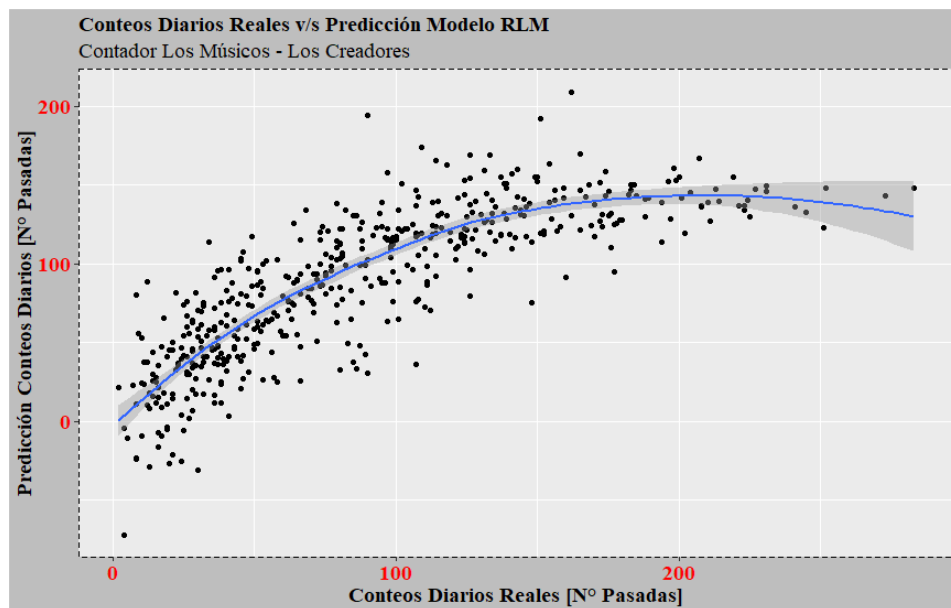


Figura A.30. Predicción conteos diarios modelo regresión lineal múltiple v/s conteos diarios reales para el contador Los Músicos – Los Creadores.

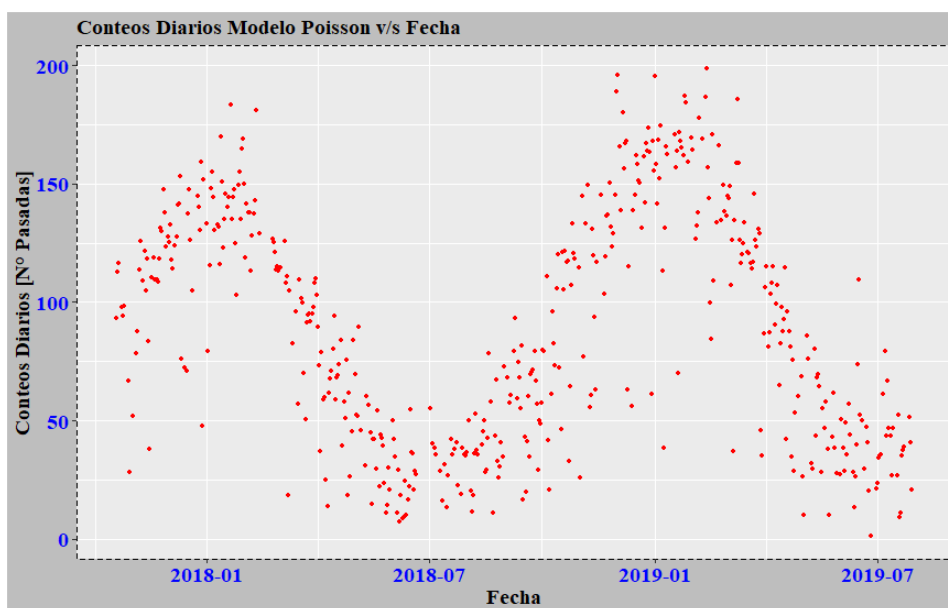


Figura A.31. Modelo Poisson v/s fecha para el contador Los Músicos – Los Creadores.

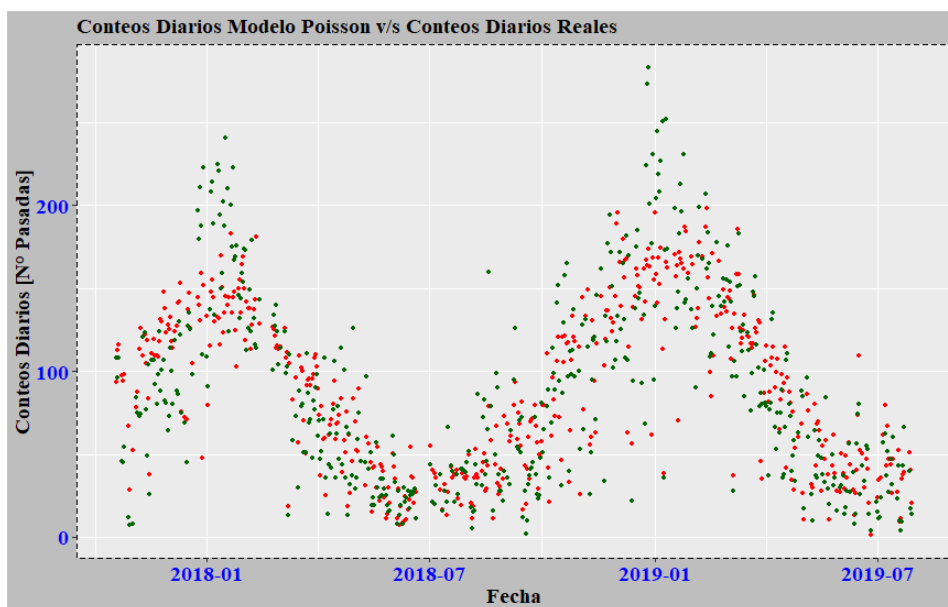


Figura A.32. Predicción conteos diarios modelo Poisson v/s conteos diarios reales en función del tiempo para el contador Los Músicos – Los Creadores.

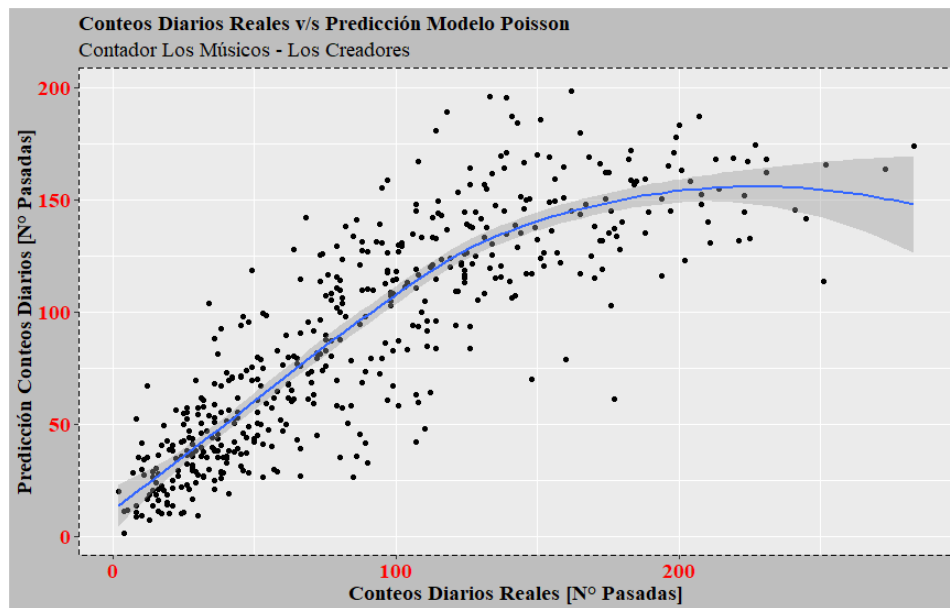


Figura A.33. Predicción conteos diarios modelo Poisson v/s conteos diarios reales para el contador Los Músicos – Los Creadores.

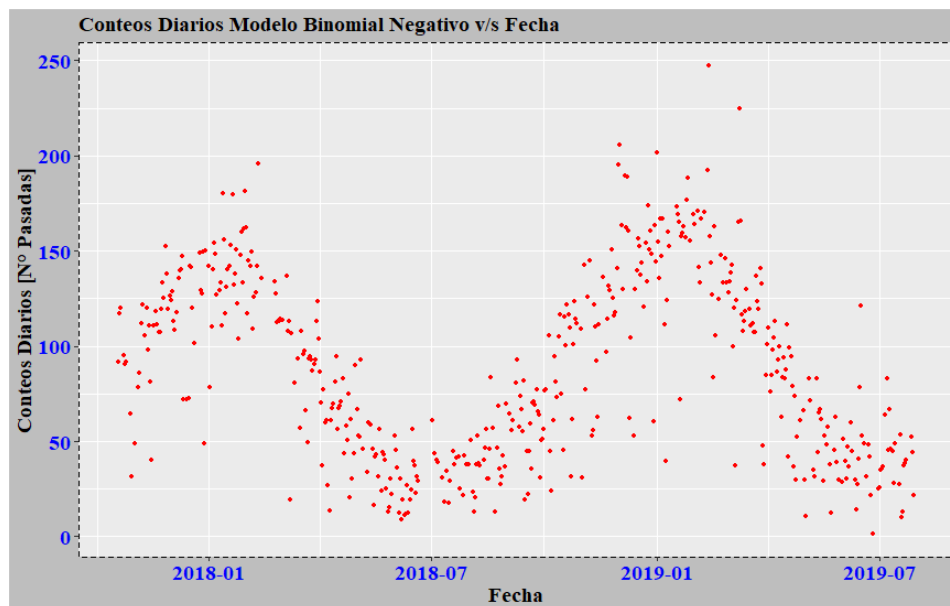


Figura A.34. Modelo binomial negativo v/s fecha para el contador Los Músicos – Los Creadores.

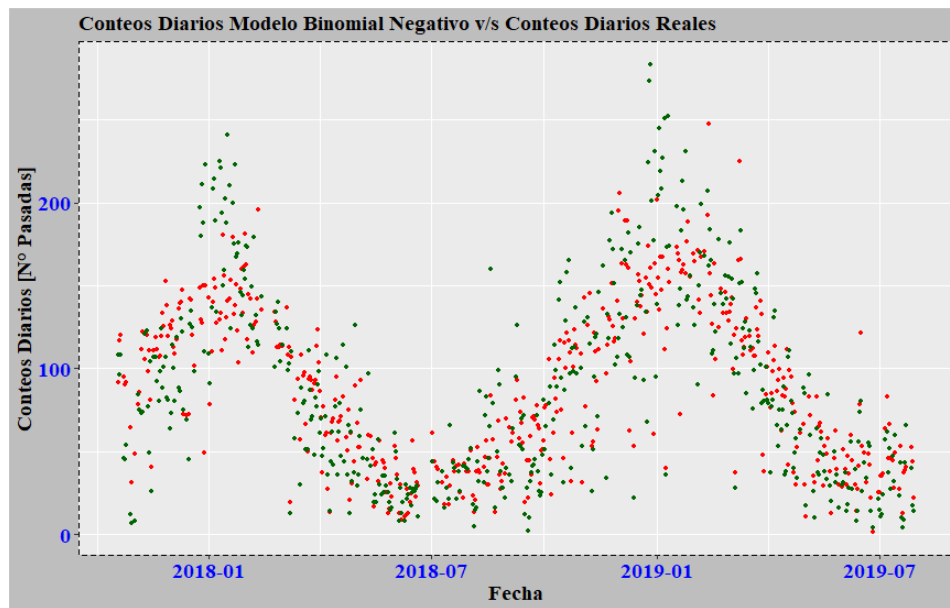


Figura A.35. Predicción conteos diarios modelo binomial negativo v/s conteos diarios reales en función del tiempo para el contador Los Músicos – Los Creadores.

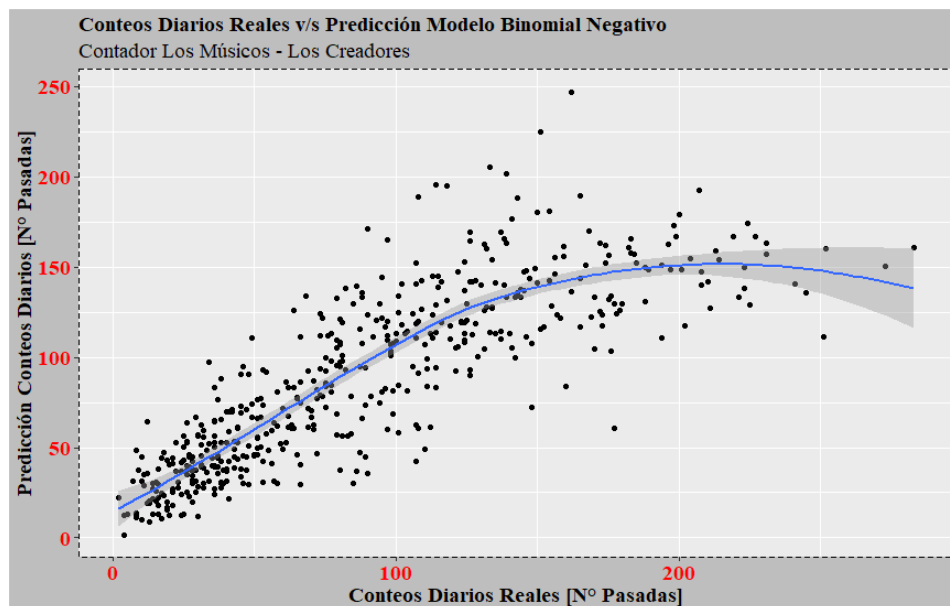


Figura A.36. Predicción conteos diarios modelo binomial negativo v/s conteos diarios reales para el contador Los Músicos – Los Creadores.

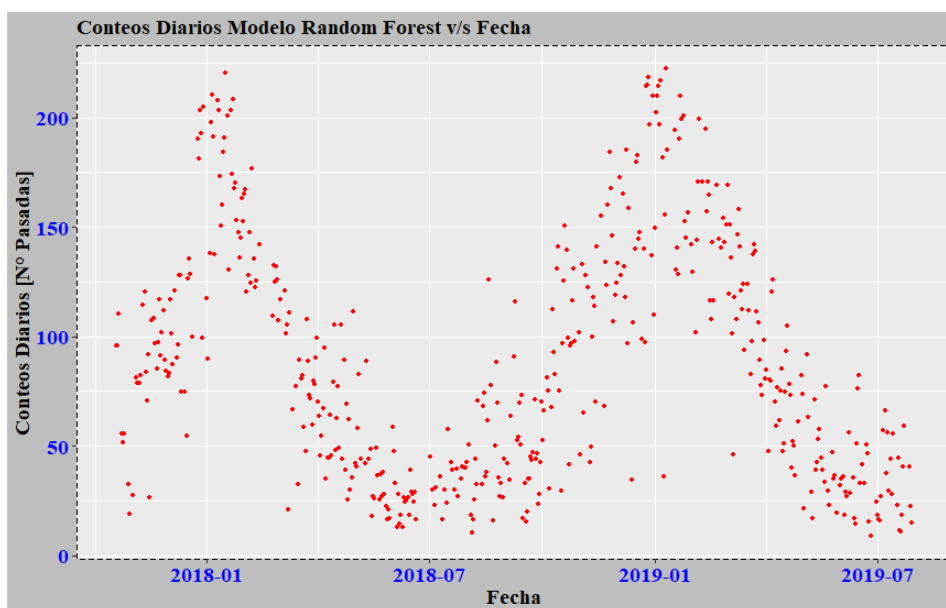


Figura A.37. Modelo *Random Forests* v/s fecha para el contador Los Músicos – Los Creadores.

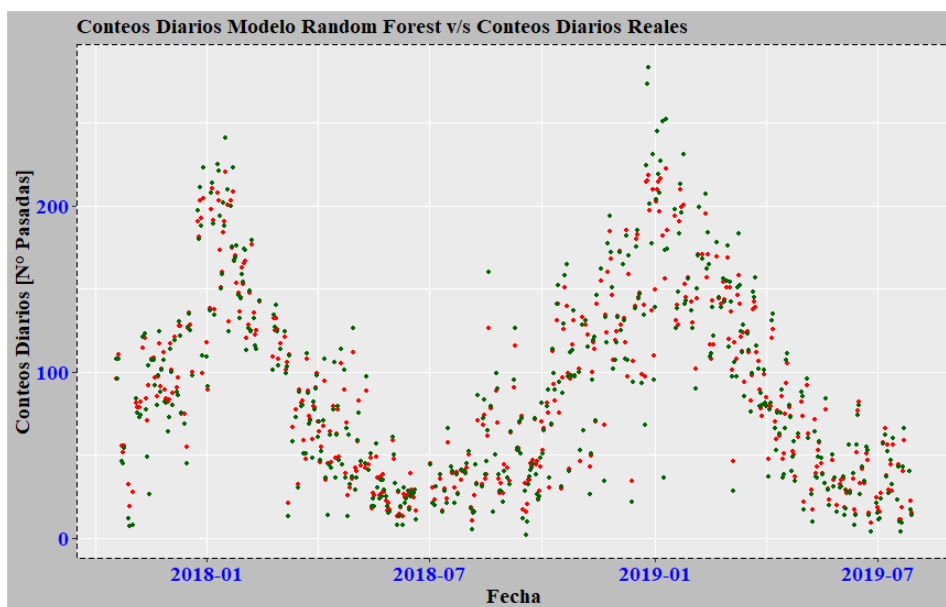


Figura A.38. Predicción conteos diarios modelo *Random Forests* v/s conteos diarios reales en función del tiempo para el contador Los Músicos – Los Creadores.

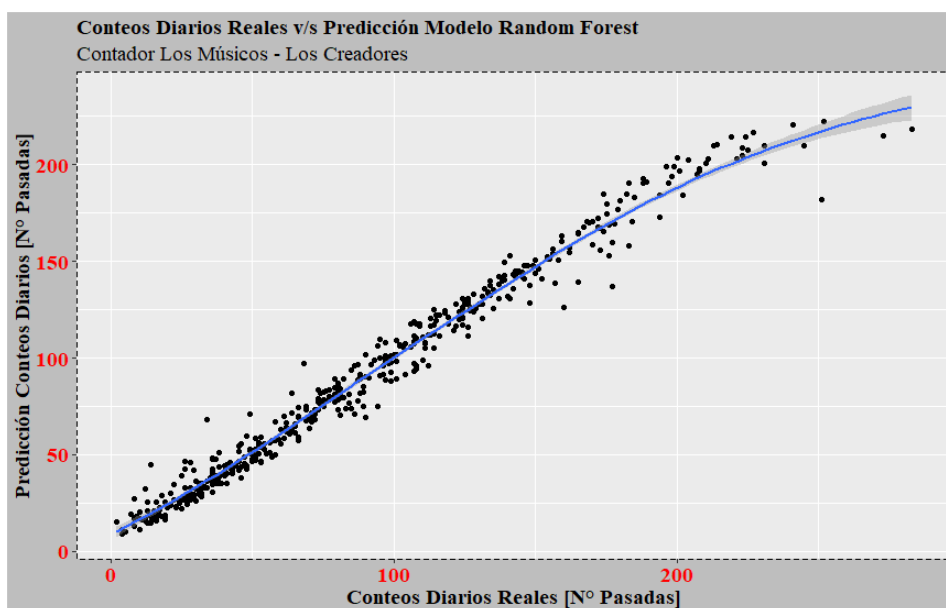


Figura A.39. Predicción conteos diarios modelo *Random Forests* v/s conteos diarios reales para el contador Los Músicos – Los Creadores.

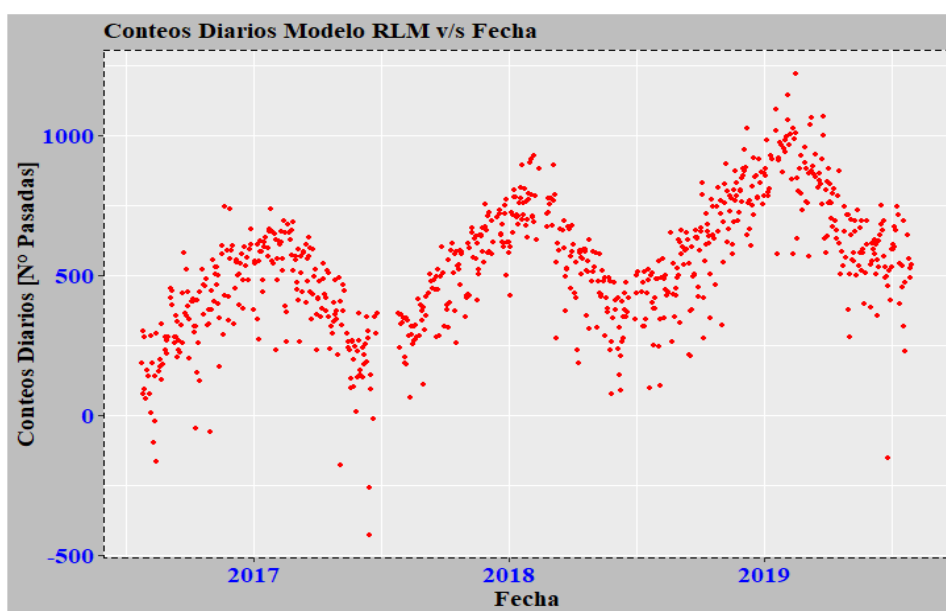


Figura A.40. Modelo regresión lineal múltiple v/s fecha para el contador Pablo Neruda – Olimpia.

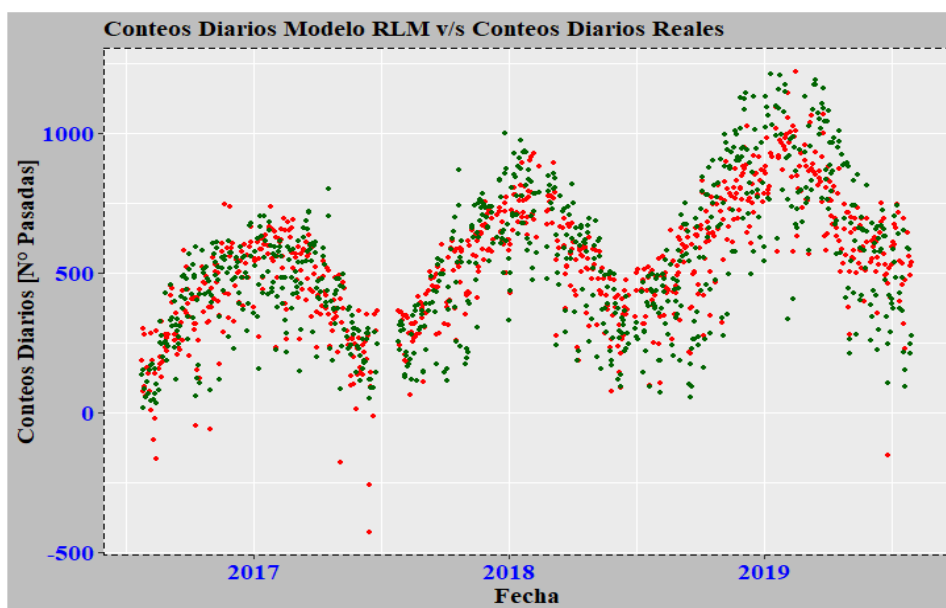


Figura A.41. Predicción conteos diarios modelo regresión lineal múltiple v/s conteos diarios reales en función del tiempo para el contador Pablo Neruda – Olimpia.

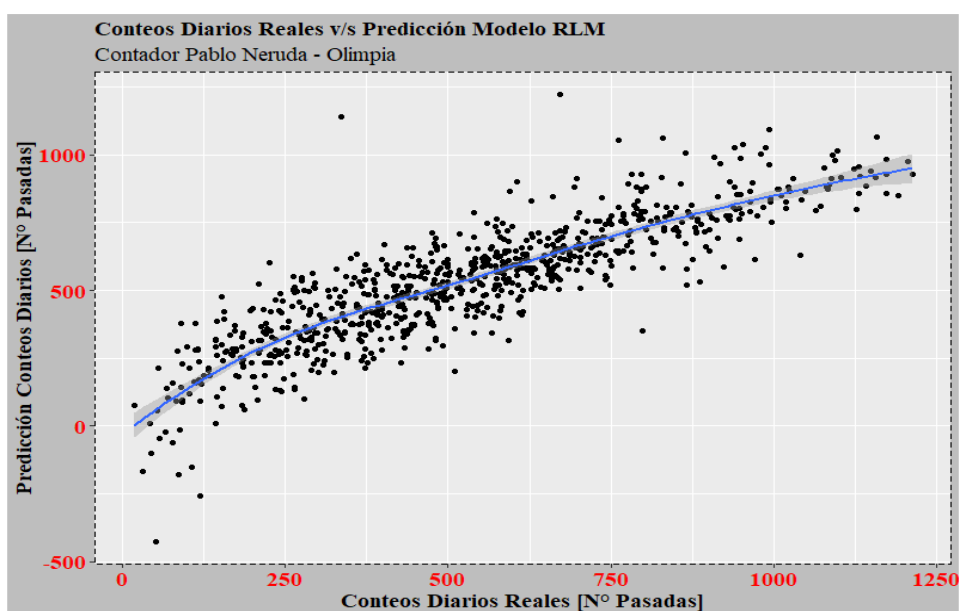


Figura A.42. Predicción conteos diarios modelo regresión lineal múltiple v/s conteos diarios reales para el contador Pablo Neruda – Olimpia.

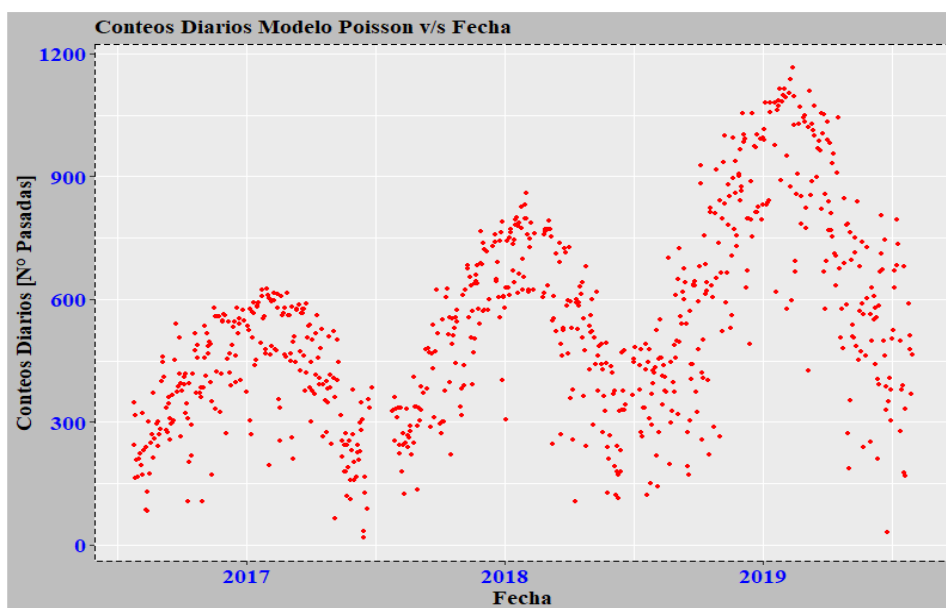


Figura A.43. Modelo Poisson v/s fecha para el contador Pablo Neruda – Olimpia.

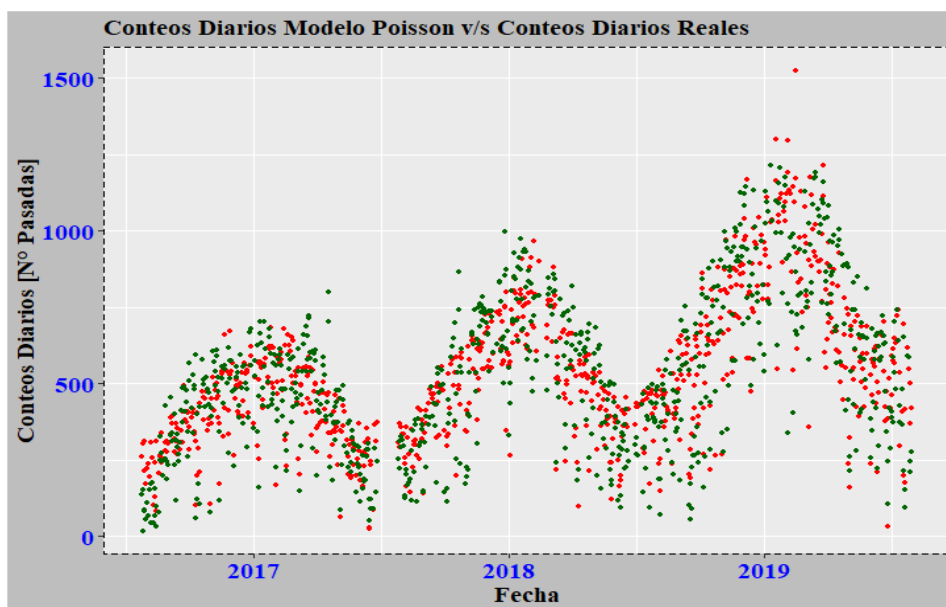


Figura A.44. Predicción conteos diarios modelo Poisson v/s conteos diarios reales en función del tiempo para el contador Pablo Neruda – Olimpia.

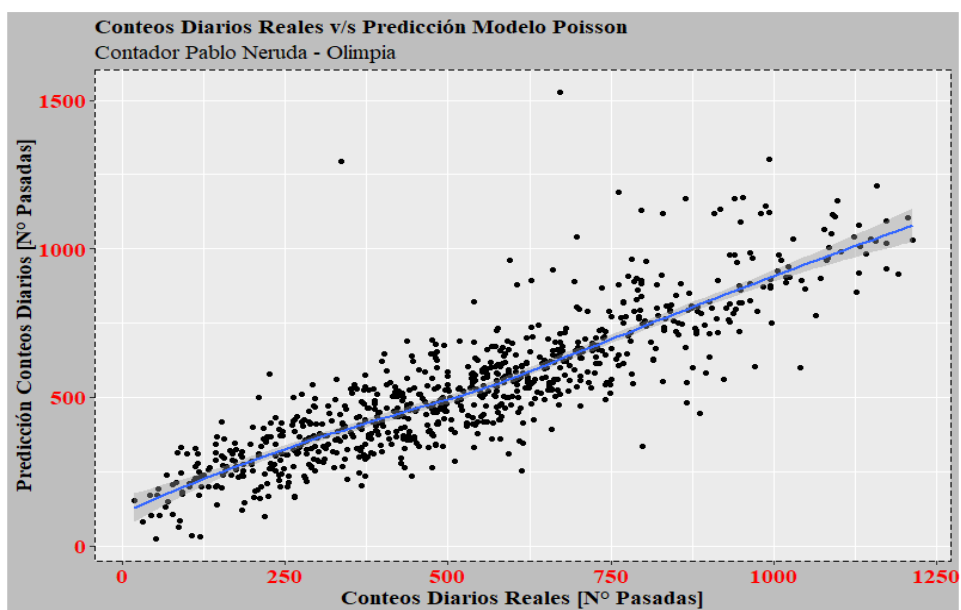


Figura A.45. Predicción conteos diarios modelo Poisson v/s conteos diarios reales para el contador Pablo Neruda – Olimpia.

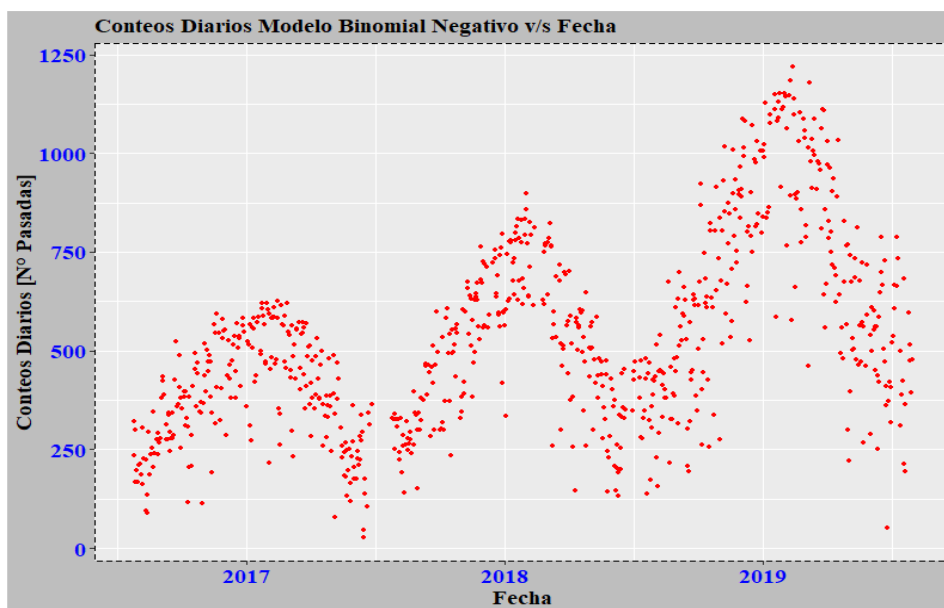


Figura A.46. Modelo binomial negativo v/s fecha para el contador Pablo Neruda – Olimpia.

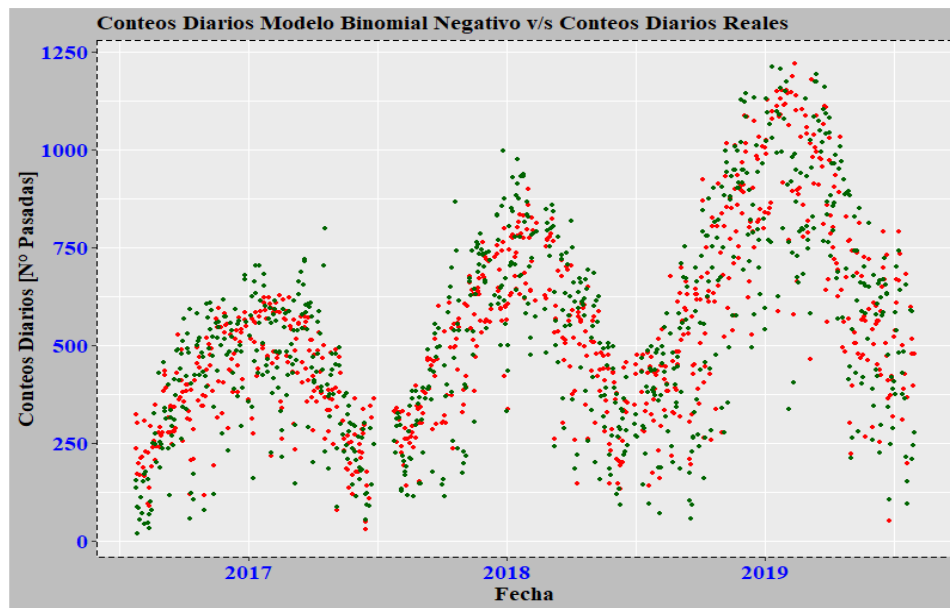


Figura A.47. Predicción conteos diarios modelo binomial negativo v/s conteos diarios reales en función del tiempo para el contador Pablo Neruda – Olimpia.

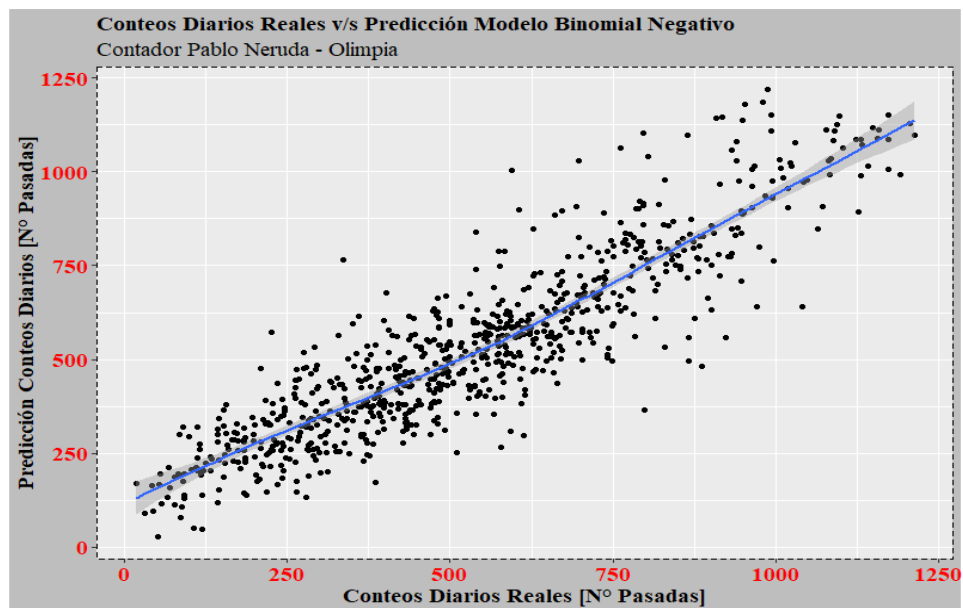


Figura A.48. Predicción conteos diarios modelo binomial negativo v/s conteos diarios reales para el contador Pablo Neruda – Olimpia.

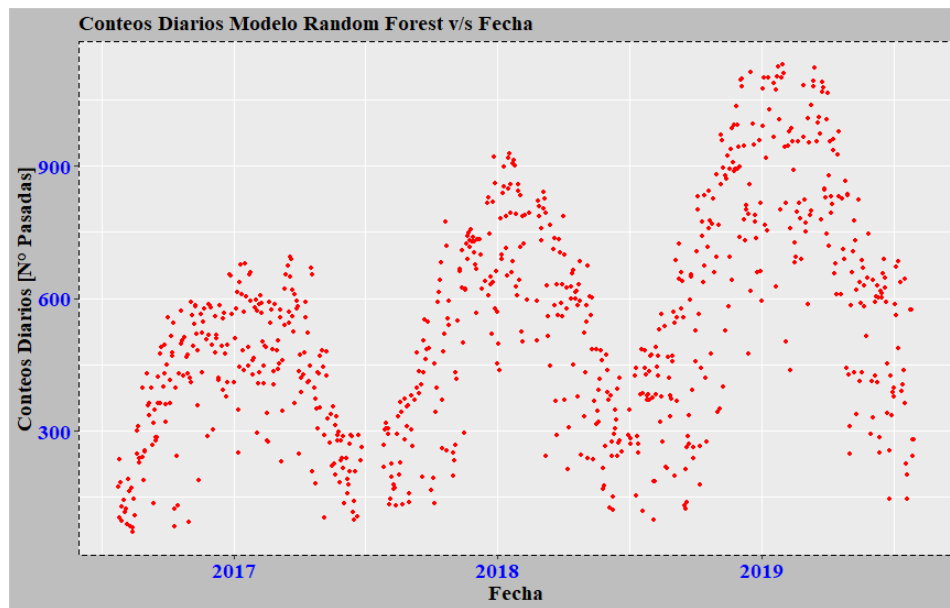


Figura A.49. Modelo *Random Forests* v/s fecha para el contador Pablo Neruda – Olimpia.

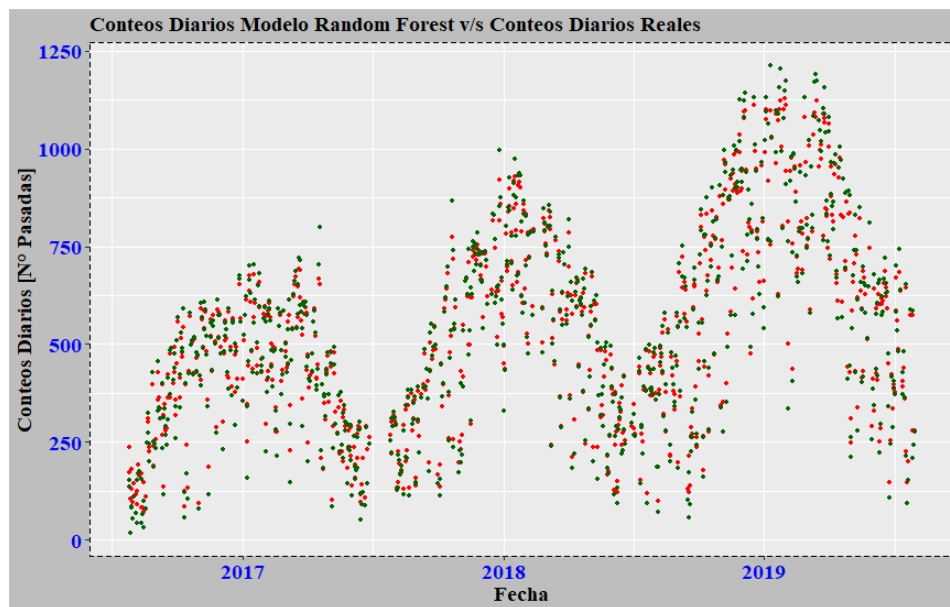


Figura A.50. Predicción conteos diarios modelo *Random Forests* v/s conteos diarios reales en función del tiempo para el contador Pablo Neruda – Olimpia.

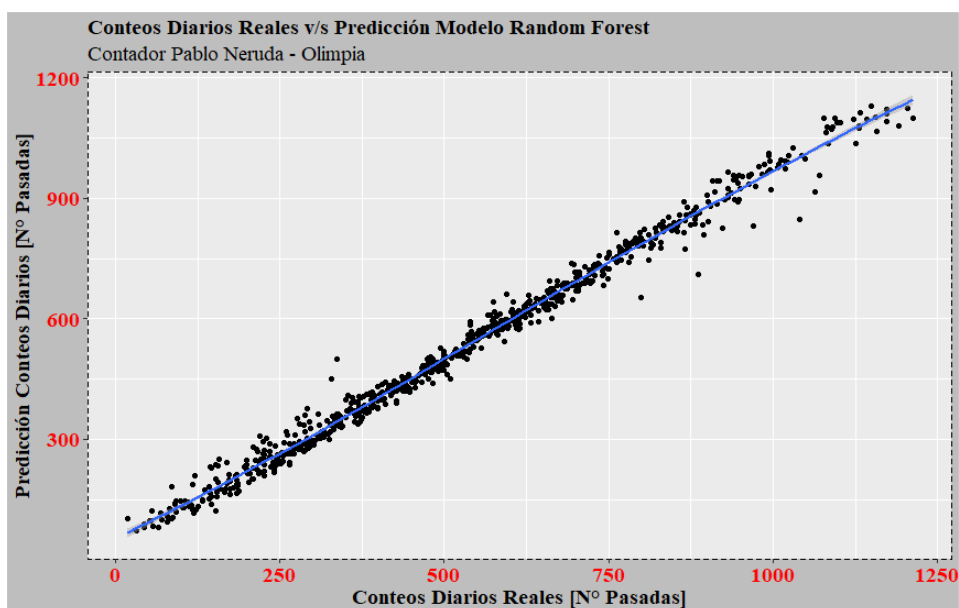


Figura A.51. Predicción conteos diarios modelo *Random Forests* v/s conteos diarios reales para el contador Pablo Neruda – Olimpia.

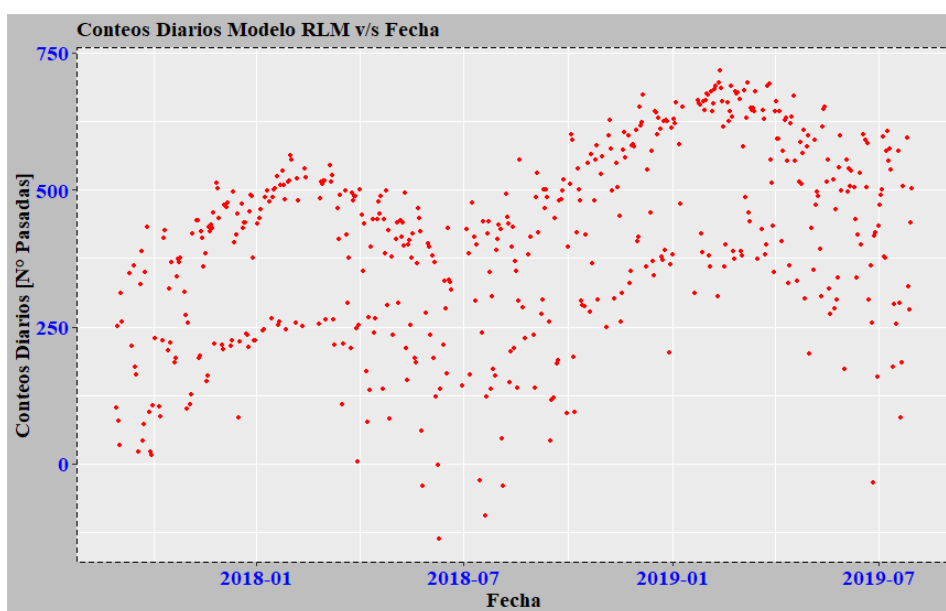


Figura A.52. Modelo regresión lineal múltiple v/s fecha para el contador San Martín - Mackenna.

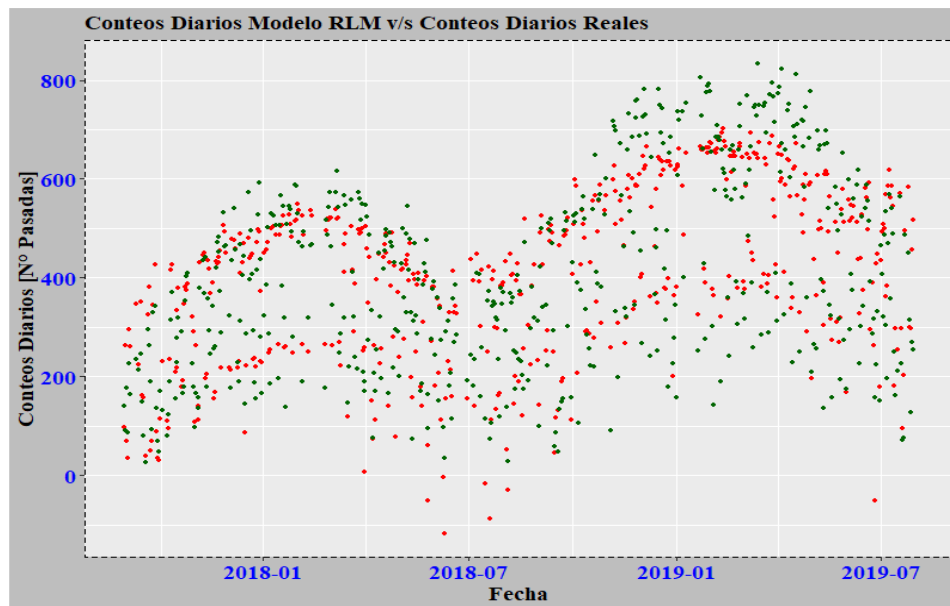


Figura A.53. Predicción conteos diarios modelo regresión lineal múltiple v/s conteos diarios reales en función del tiempo para el contador San Martín – Mackenna.

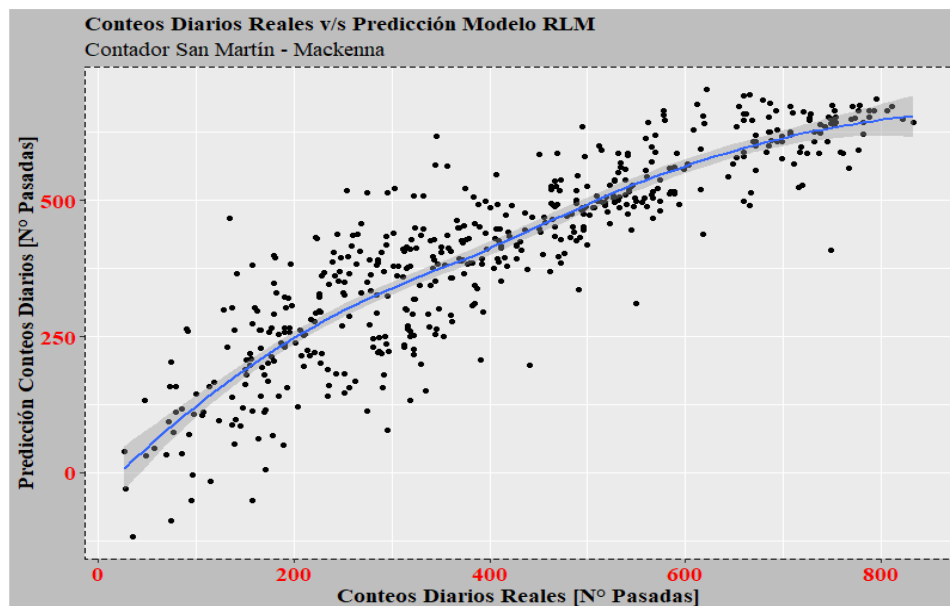


Figura A.54. Predicción conteos diarios modelo regresión lineal múltiple v/s conteos diarios reales para el contador San Martín – Mackenna.

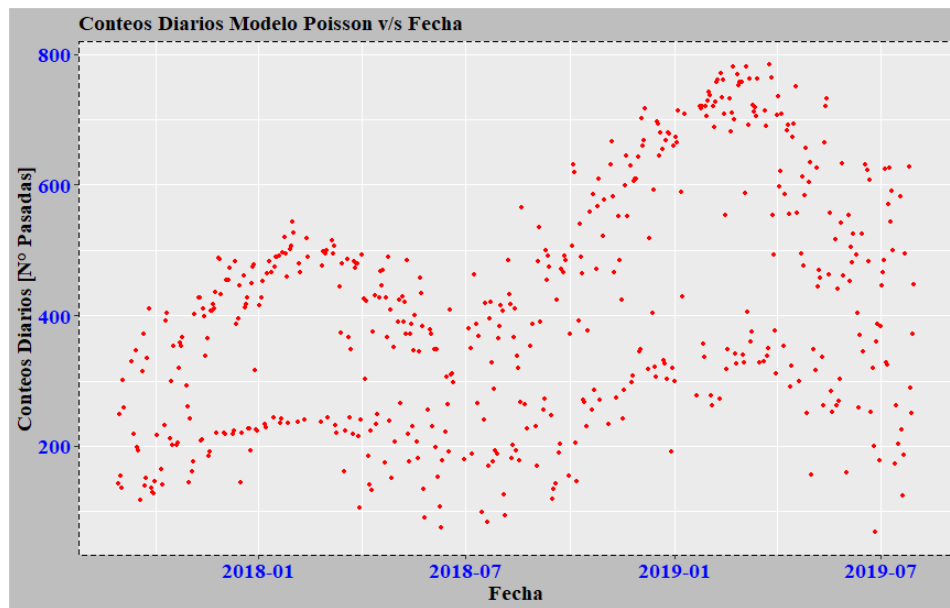


Figura A.55. Modelo Poisson v/s fecha para el contador San Martín – Mackenna.

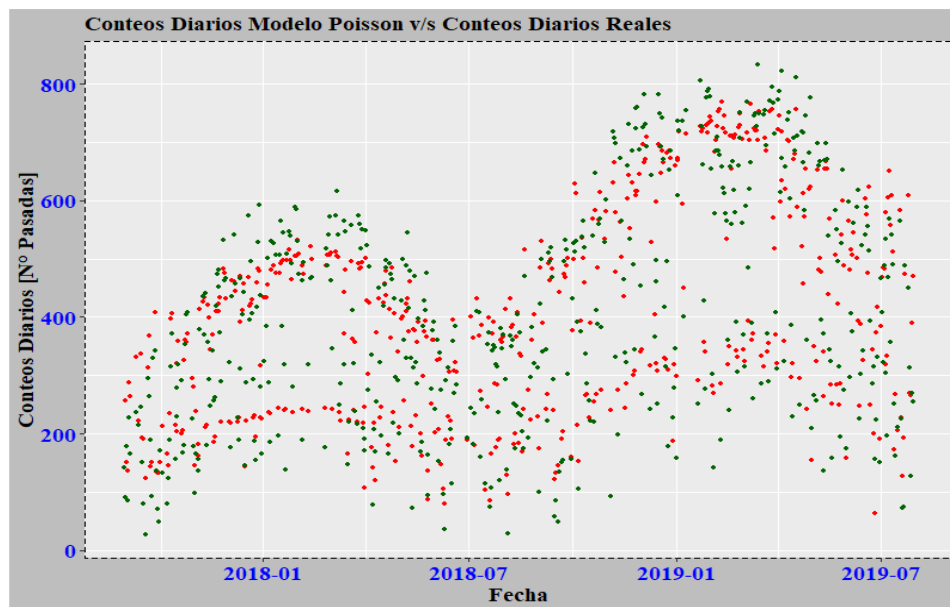


Figura A.56. Predicción conteos diarios modelo Poisson v/s conteos diarios reales en función del tiempo para el contador San Martín – Mackenna.

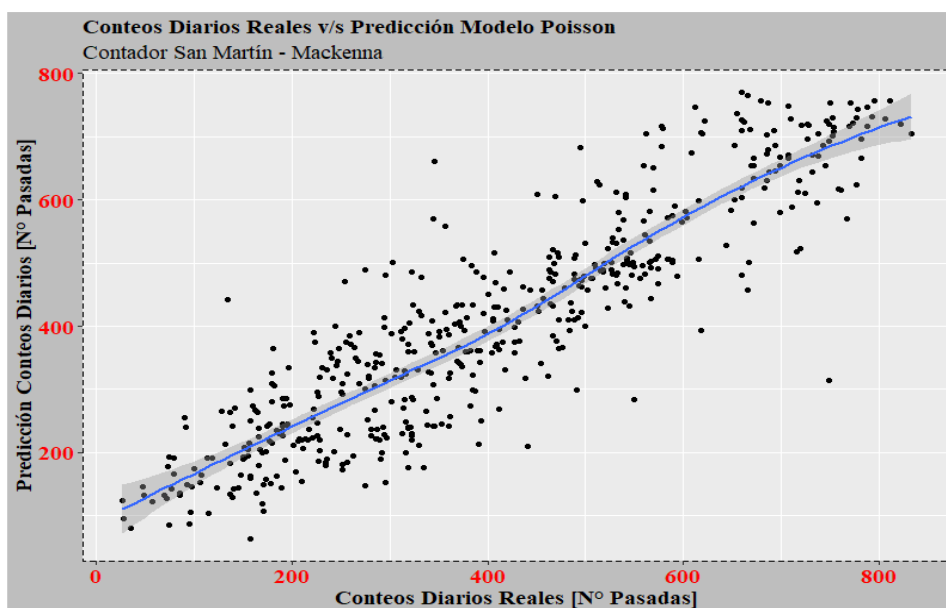


Figura A.57. Predicción conteos diarios modelo Poisson v/s conteos diarios reales para el contador San Martín – Mackenna.

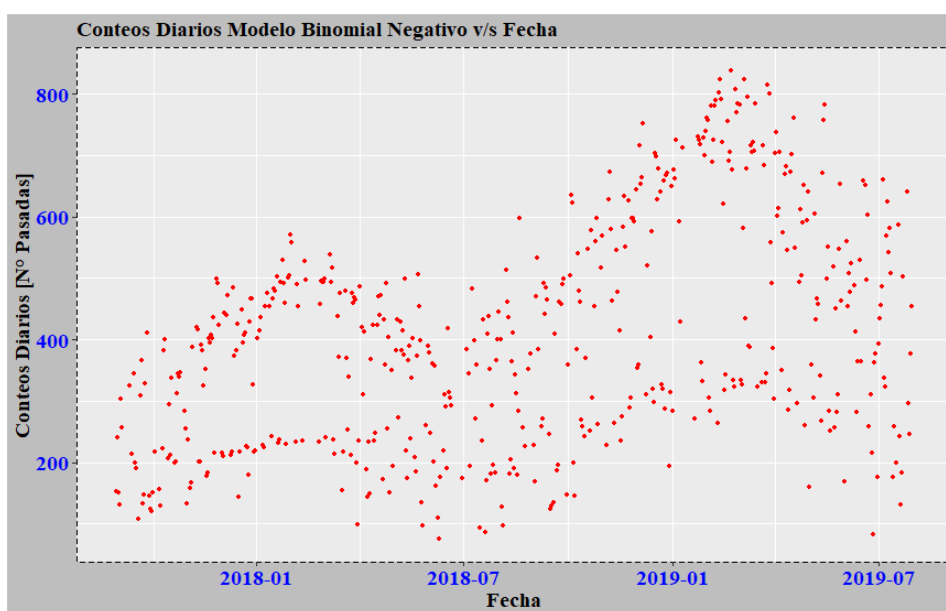


Figura A.58. Modelo binomial negativo v/s fecha para el contador San Martín – Mackenna.

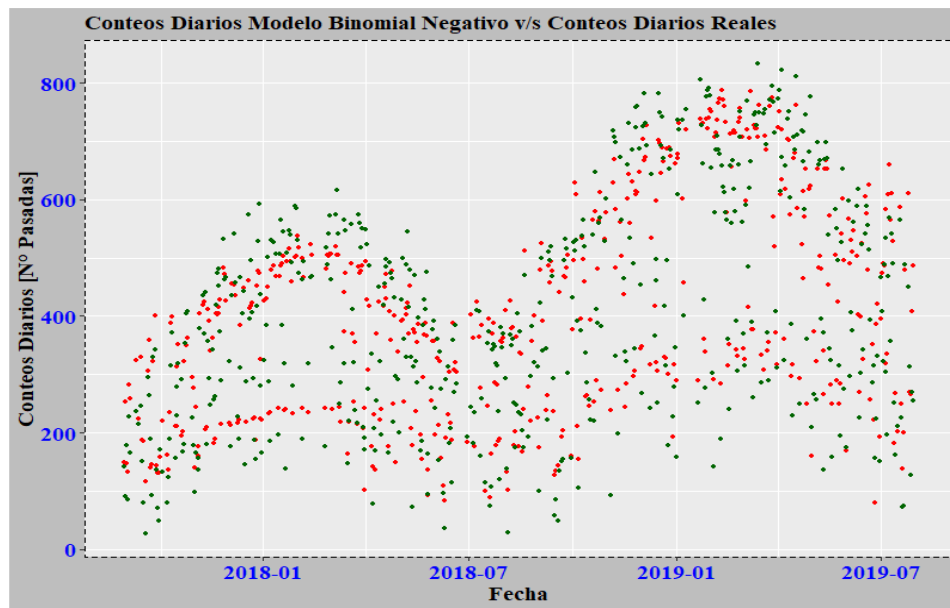


Figura A.59. Predicción conteos diarios modelo binomial negativo v/s conteos diarios reales en función del tiempo para el contador San Martín – Mackenna.

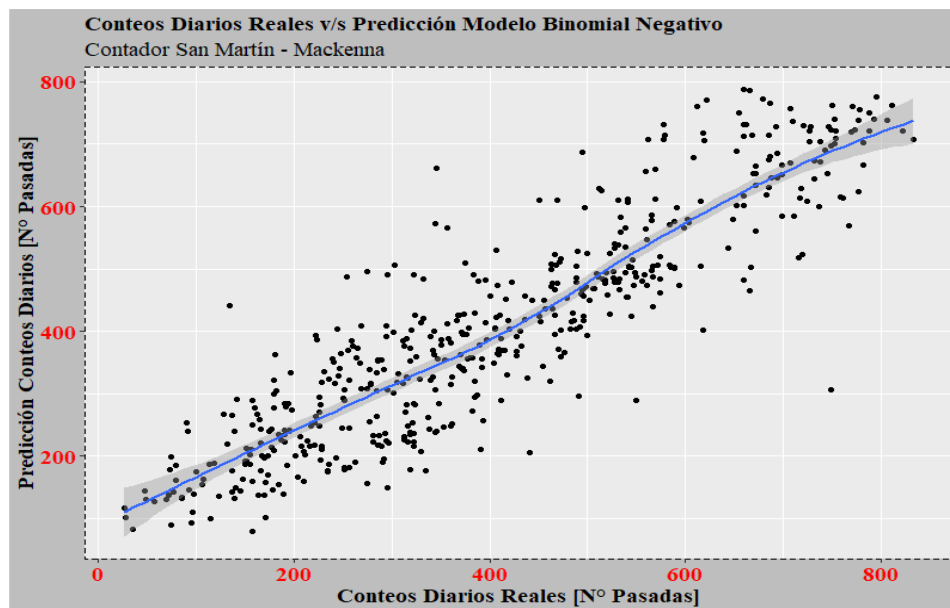


Figura A.60. Predicción conteos diarios modelo binomial negativo v/s conteos diarios reales para el contador San Martín – Mackenna.

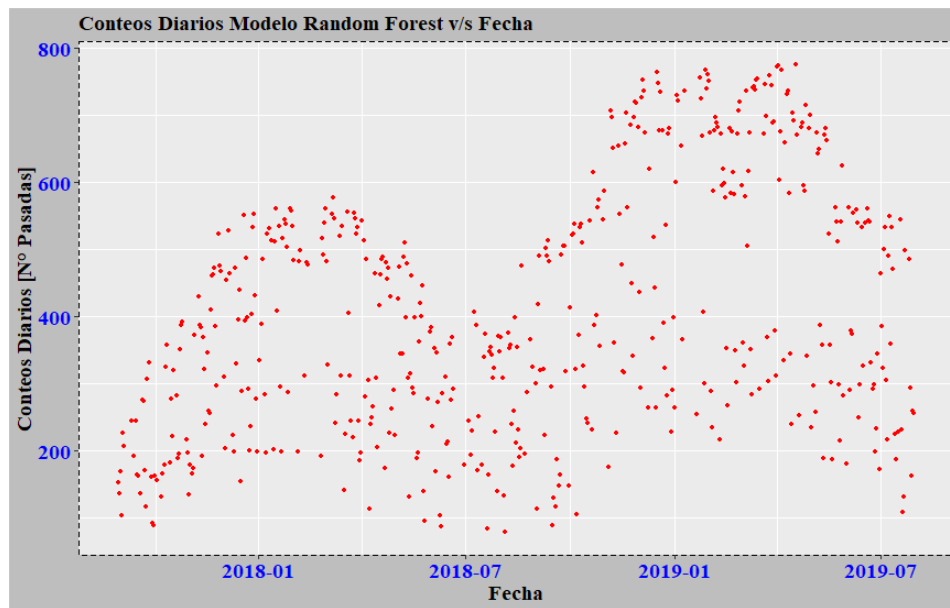


Figura A.61. Modelo *Random Forests* v/s fecha para el contador San Martín - Mackenna.

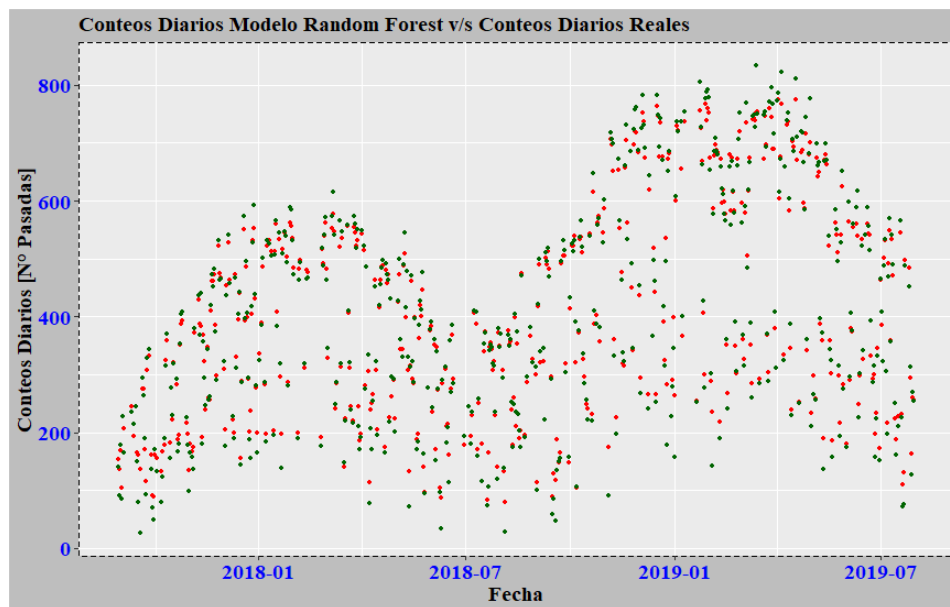


Figura A.62. Predicción conteos diarios modelo *Random Forests* v/s conteos diarios reales en función del tiempo para el contador San Martín – Mackenna.

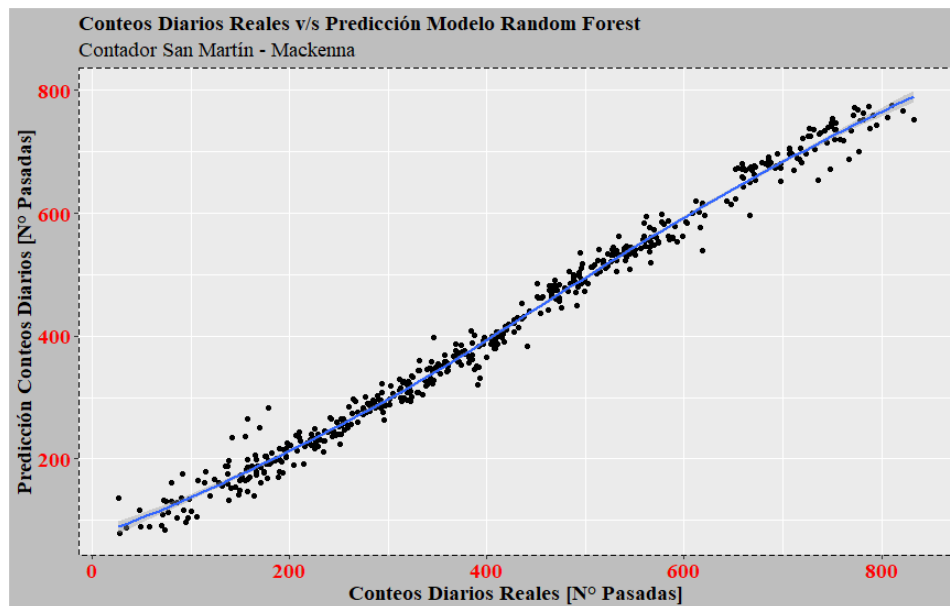


Figura A.63. Predicción conteos diarios modelo *Random Forests* v/s conteos diarios reales para el contador San Martín – Mackenna.

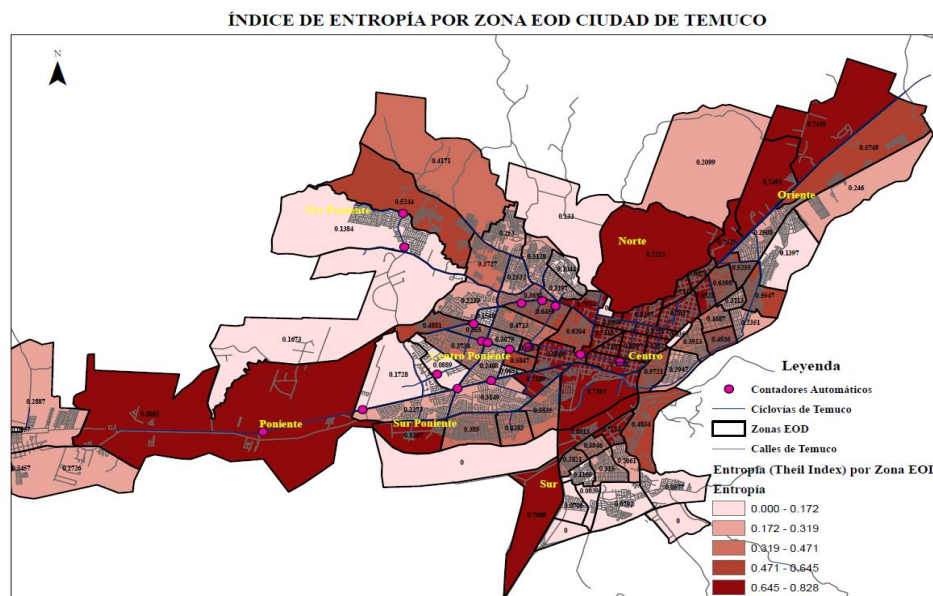


Figura A.64. Mapa del índice de entropía por zona Encuesta Origen-Destino de Temuco.

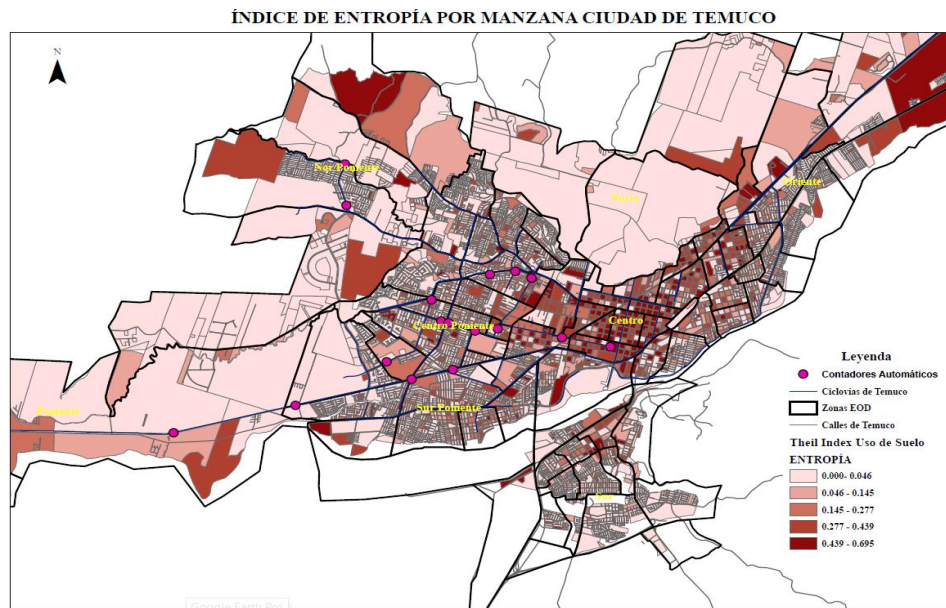


Figura A.65. Mapa del índice de entropía por manzana censal de Temuco.

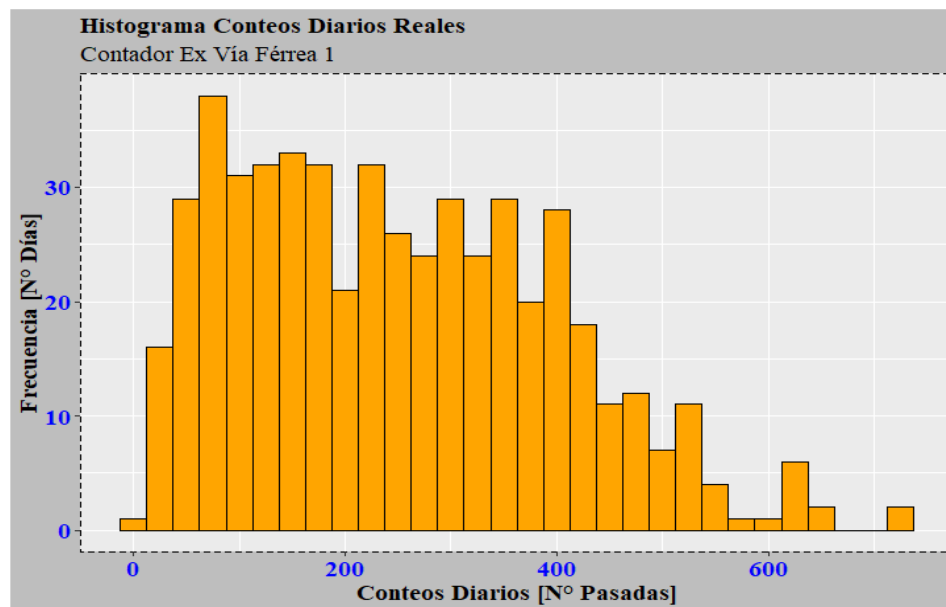


Figura A.66. Histograma de conteos diarios reales para el contador Ex Vía Férrea 1.

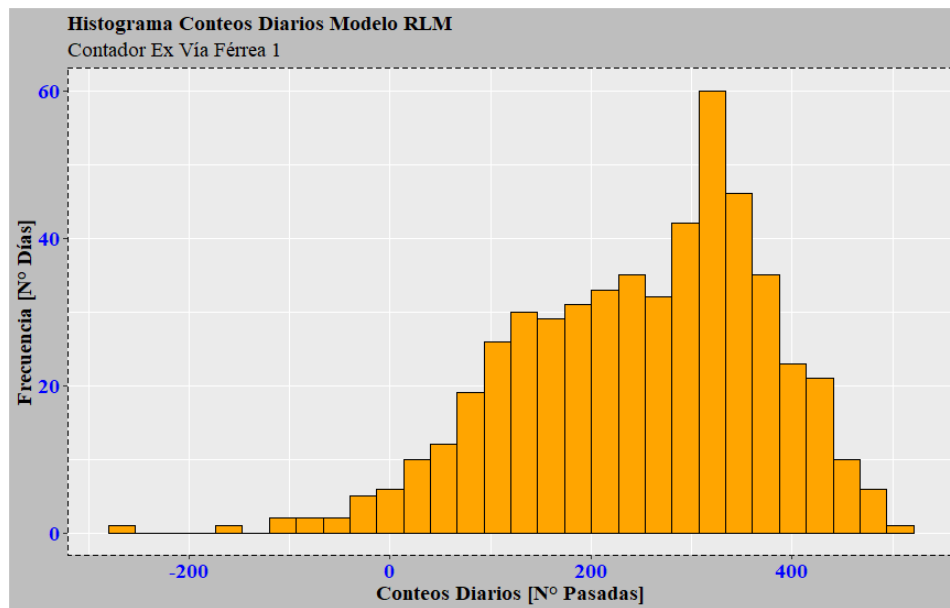


Figura A.67. Histograma de conteos diarios modelo regresión lineal múltiple para el contador Ex Vía Férrea 1.

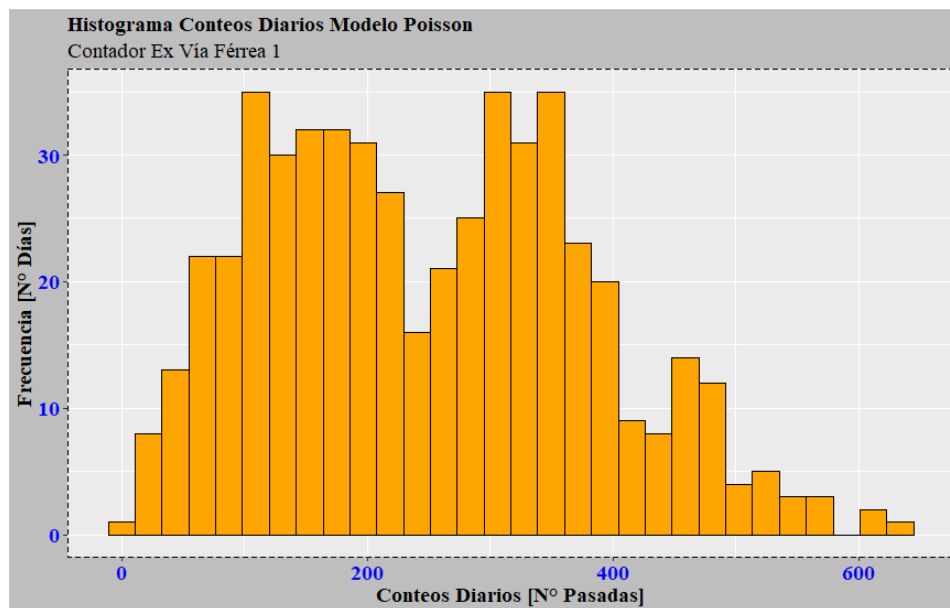


Figura A.68. Histograma de conteos diarios modelo Poisson para el contador Ex Vía Férrea 1.

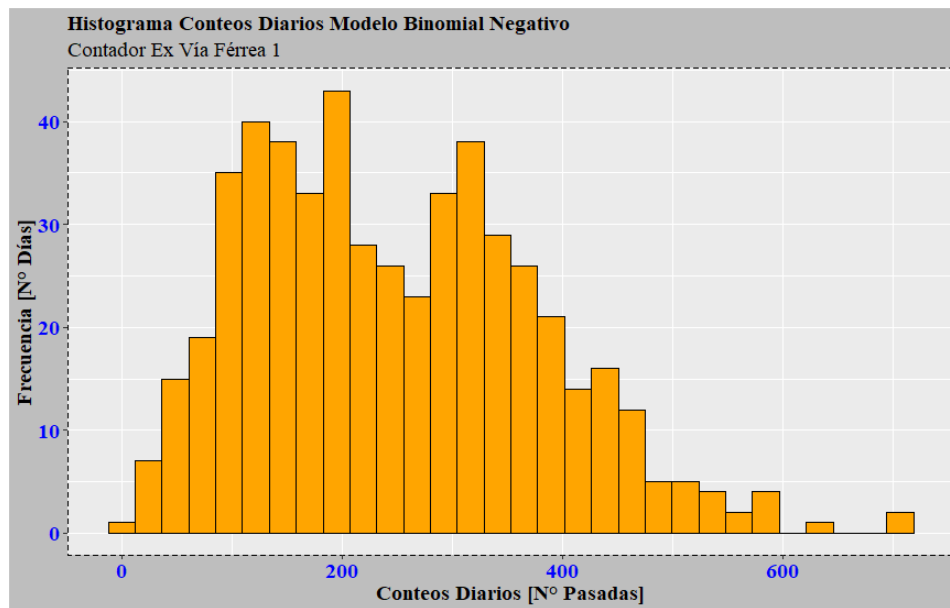


Figura A.69. Histograma de conteos diarios modelo binomial negativo para el contador Ex Vía Férrea 1.

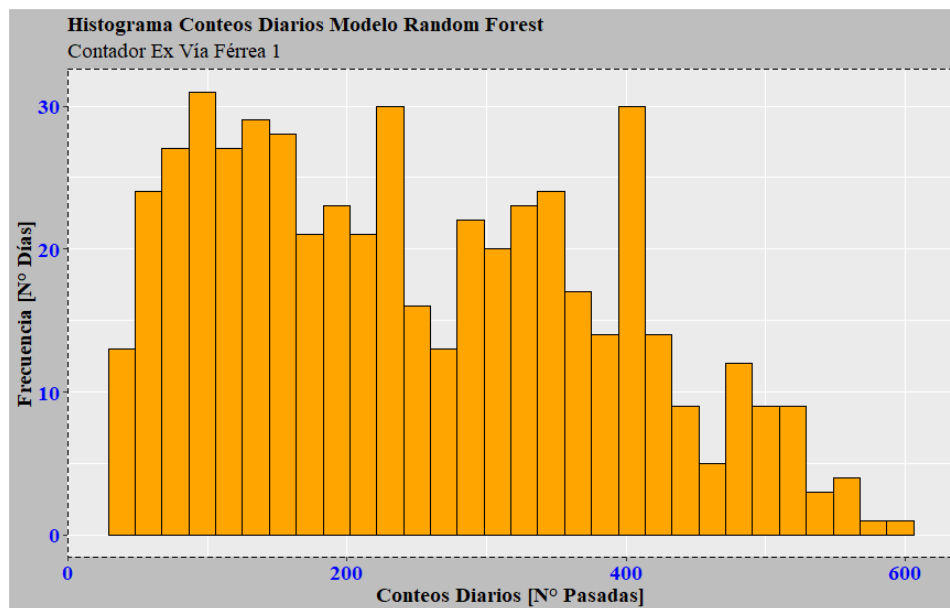


Figura A.70. Histograma de conteos diarios modelo *Random Forests* para el contador Ex Vía Férrea 1.

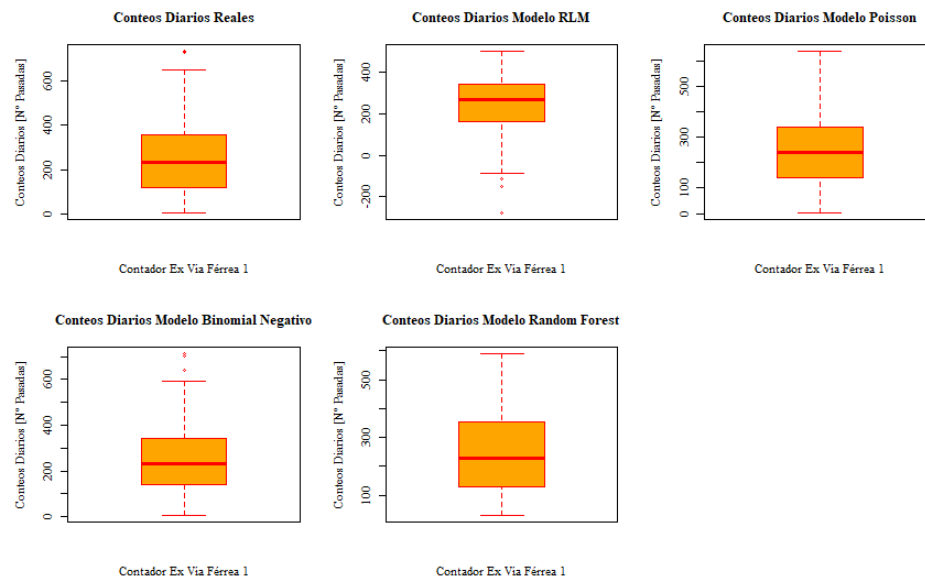


Figura A.71. Diagramas de caja de modelos desarrollados para el contador Ex Vía Férrea 1.

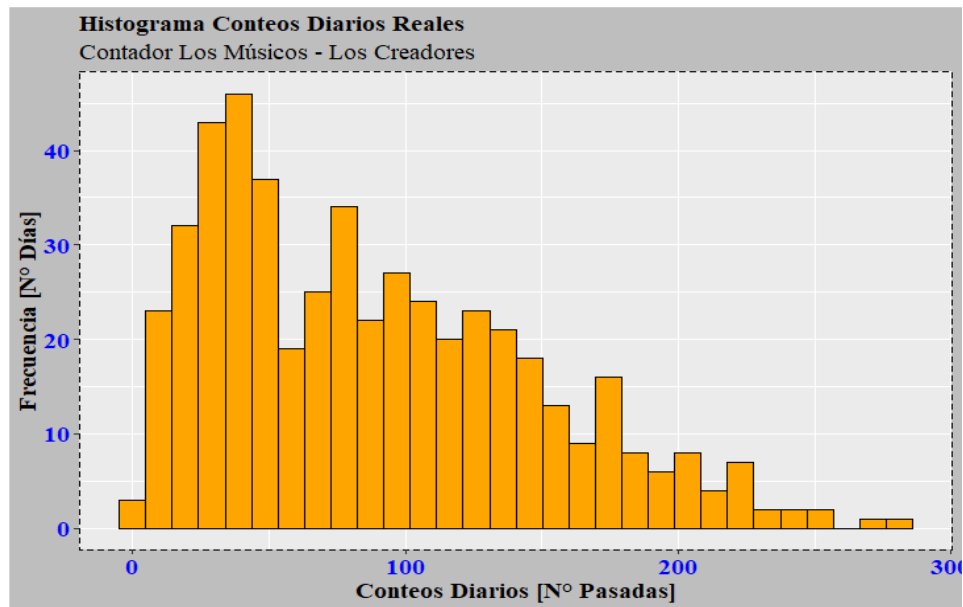


Figura A.72. Histograma de conteos diarios reales para el contador Los Músicos – Los Creadores.

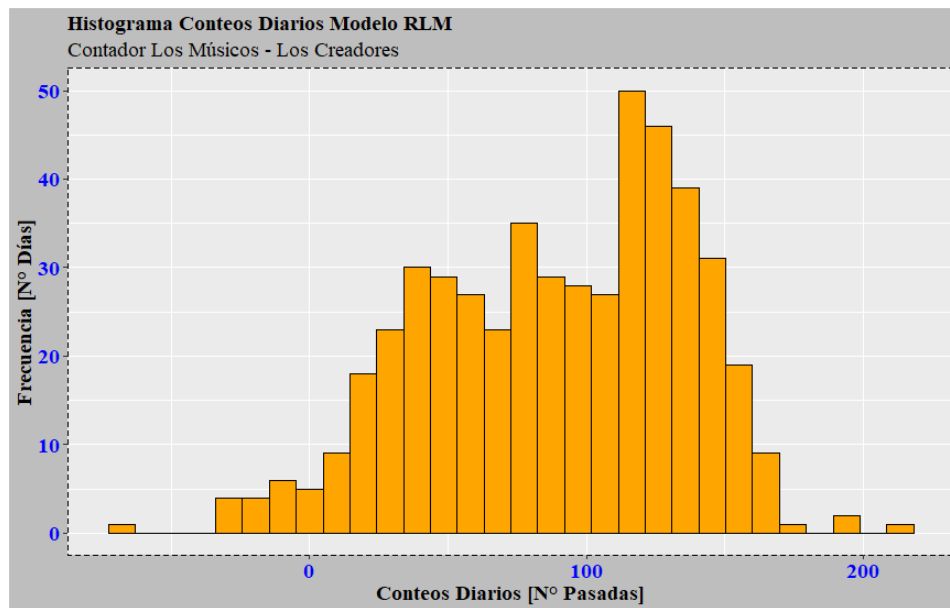


Figura A.73. Histograma de conteos diarios modelo regresión lineal múltiple para el contador Los Músicos – Los Creadores.

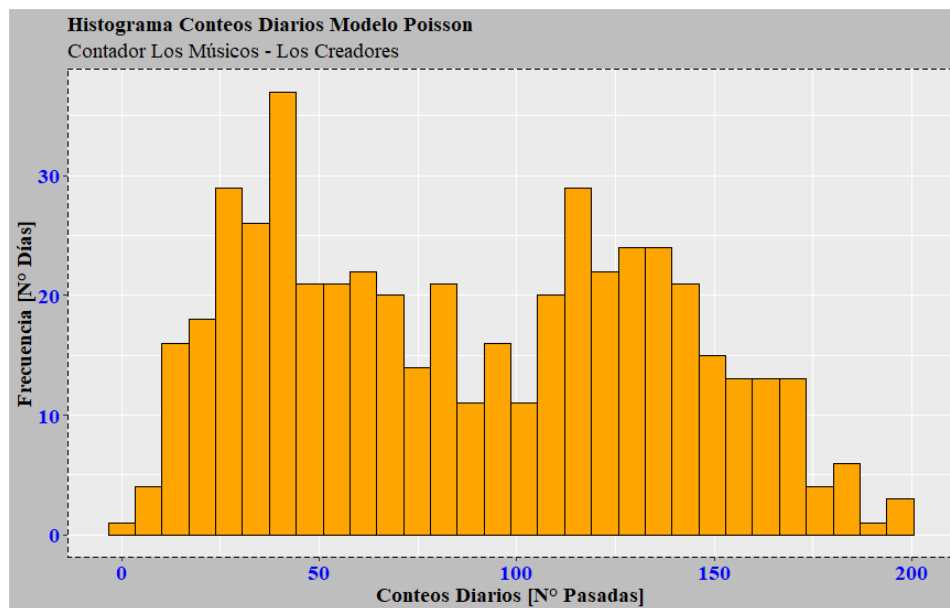


Figura A.74. Histograma de conteos diarios modelo Poisson para el contador Los Músicos – Los Creadores.

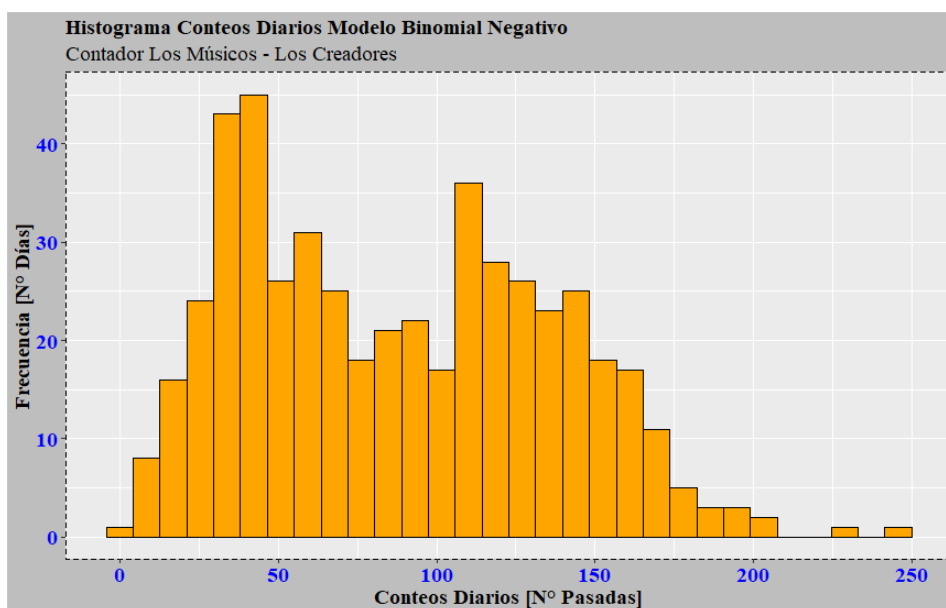


Figura A.75. Histograma de conteos diarios modelo binomial negativo para el contador Los Músicos – Los Creadores.

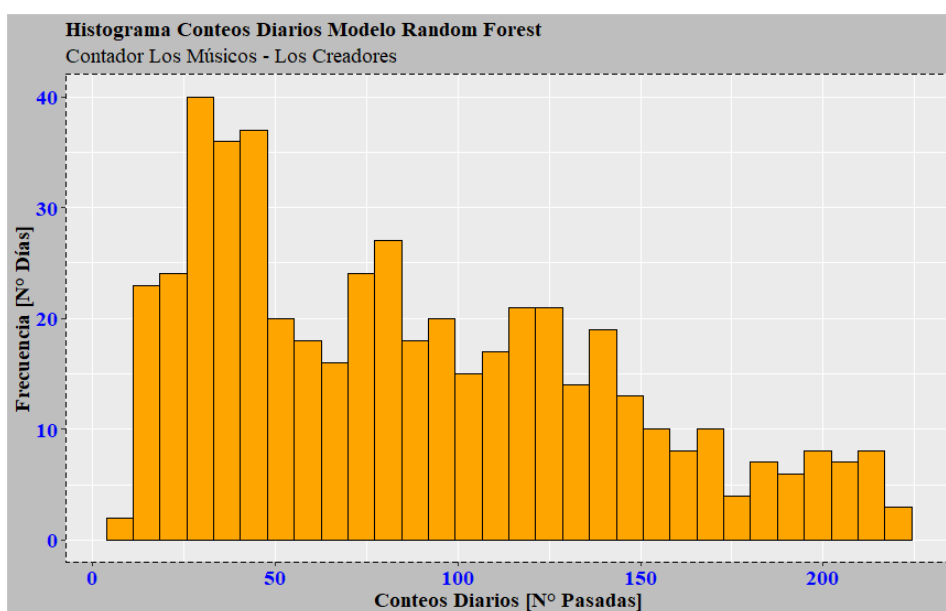


Figura A.76. Histograma de conteos diarios modelo *Random Forests* para el contador Los Músicos – Los Creadores.

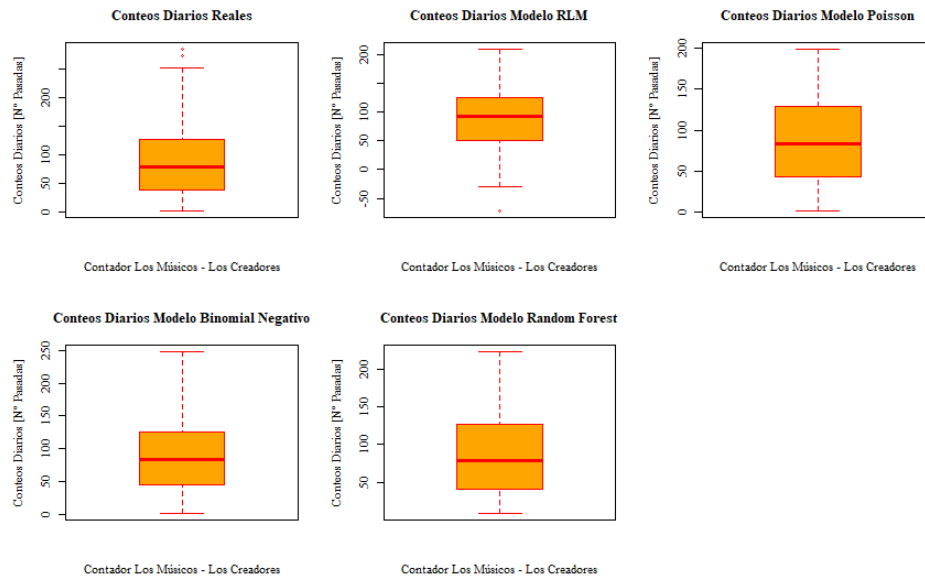


Figura A.77. Diagramas de caja de modelos desarrollados para el contador Los Músicos – Los Creadores.

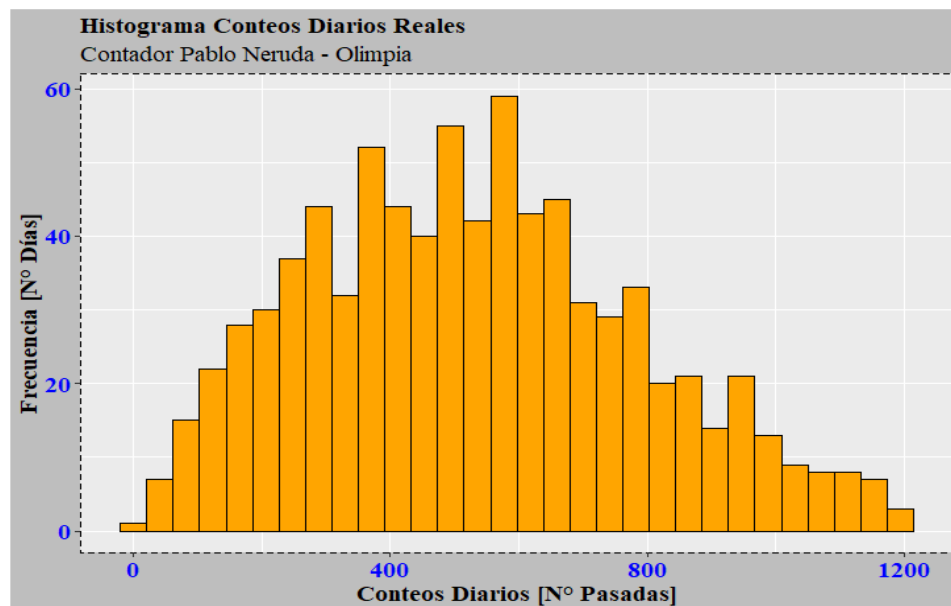


Figura A.78. Histograma de conteos diarios reales para el contador Pablo Neruda – Olimpia.

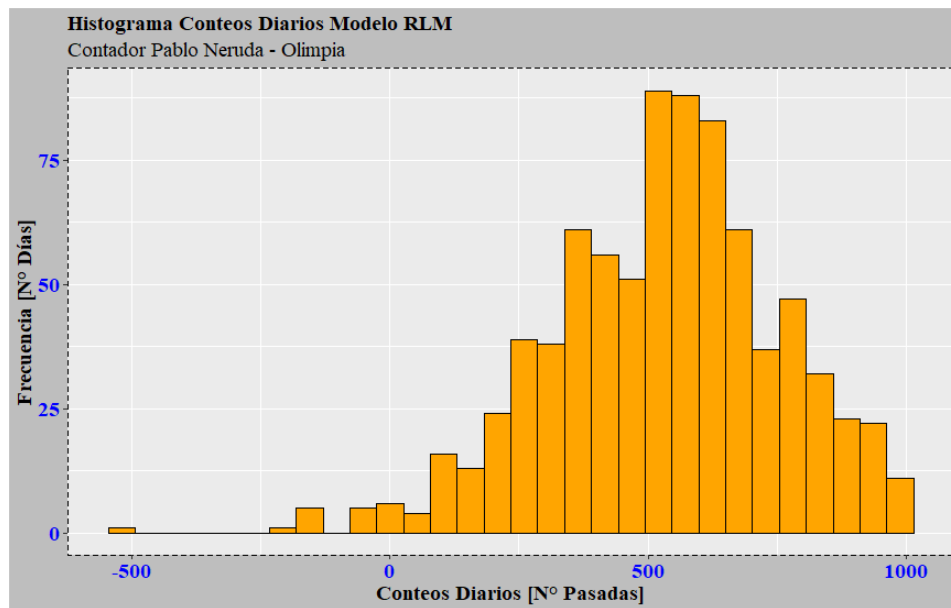


Figura A.79. Histograma de conteos diarios modelo regresión lineal múltiple para el contador Pablo Neruda – Olimpia.

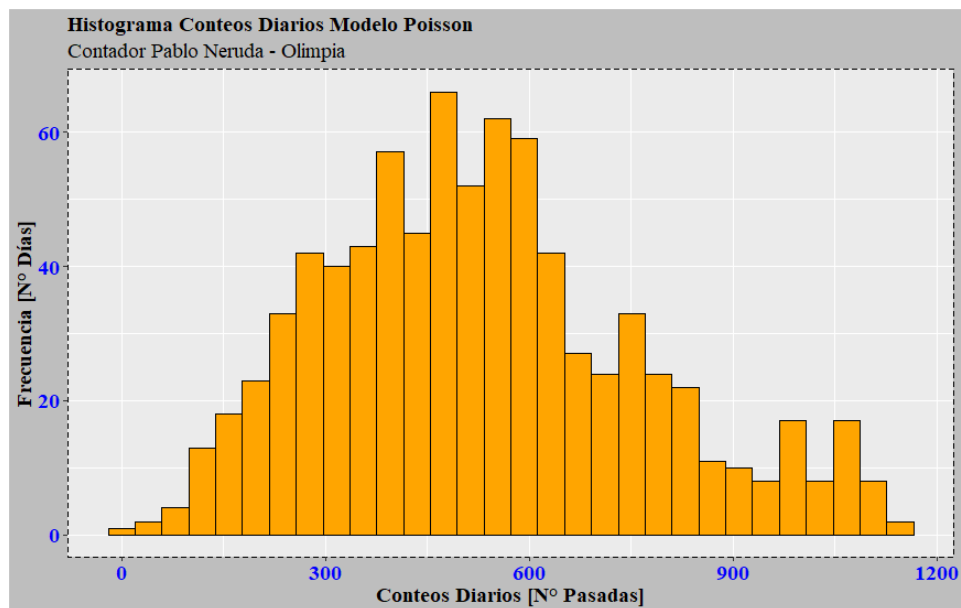


Figura A.80. Histograma de conteos diarios modelo Poisson para el contador Pablo Neruda – Olimpia.

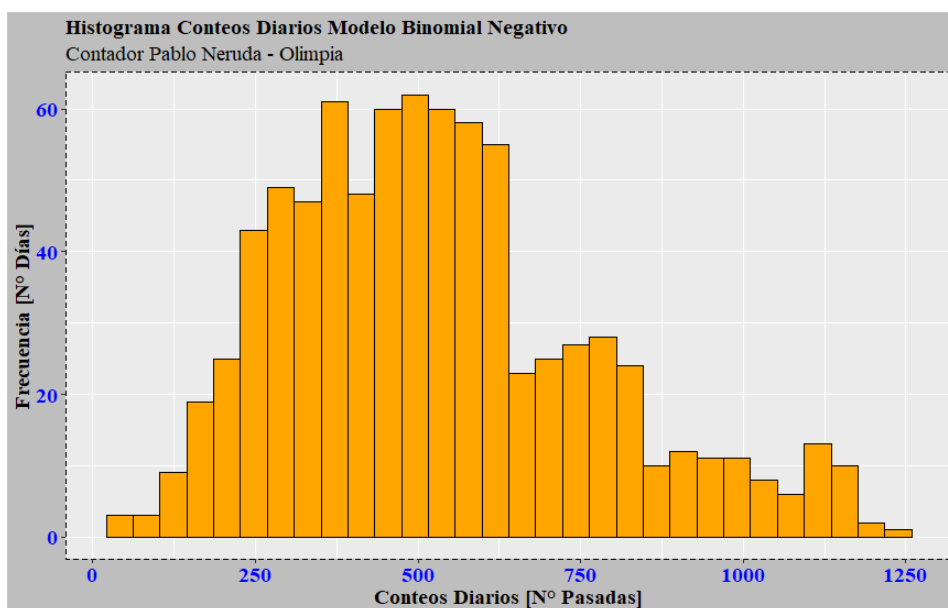


Figura A.81. Histograma de conteos diarios modelo binomial negativo para el contador Pablo Neruda – Olimpia.

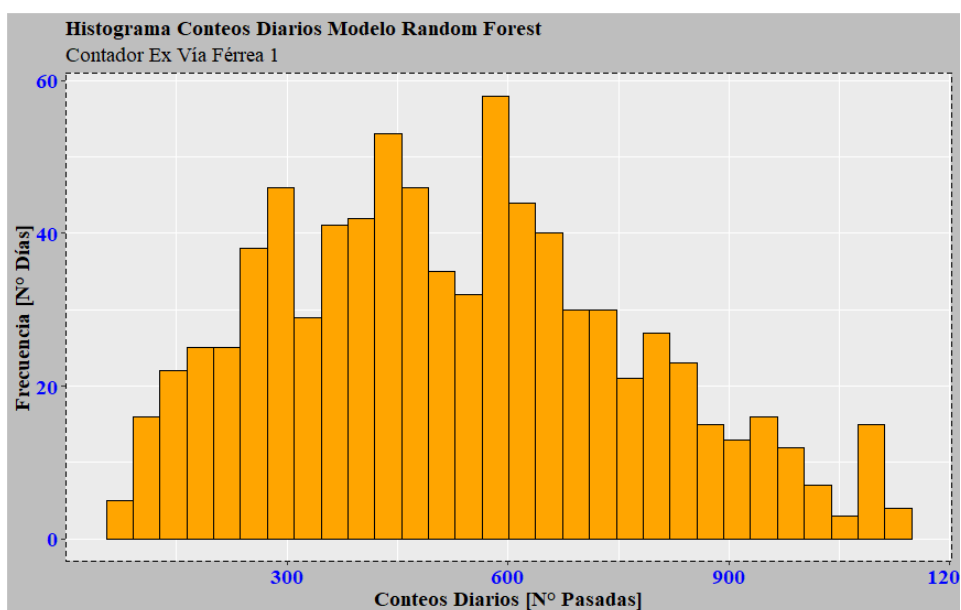


Figura A.82. Histograma de conteos diarios modelo *Random Forests* para el contador Pablo Neruda – Olimpia.

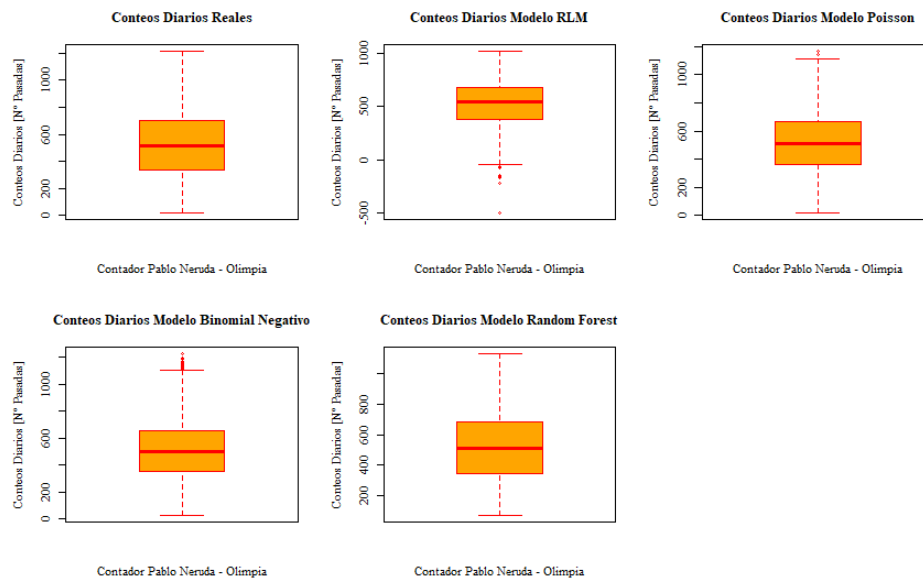


Figura A.83. Diagramas de caja de modelos desarrollados para el contador Pablo Neruda – Olimpia.

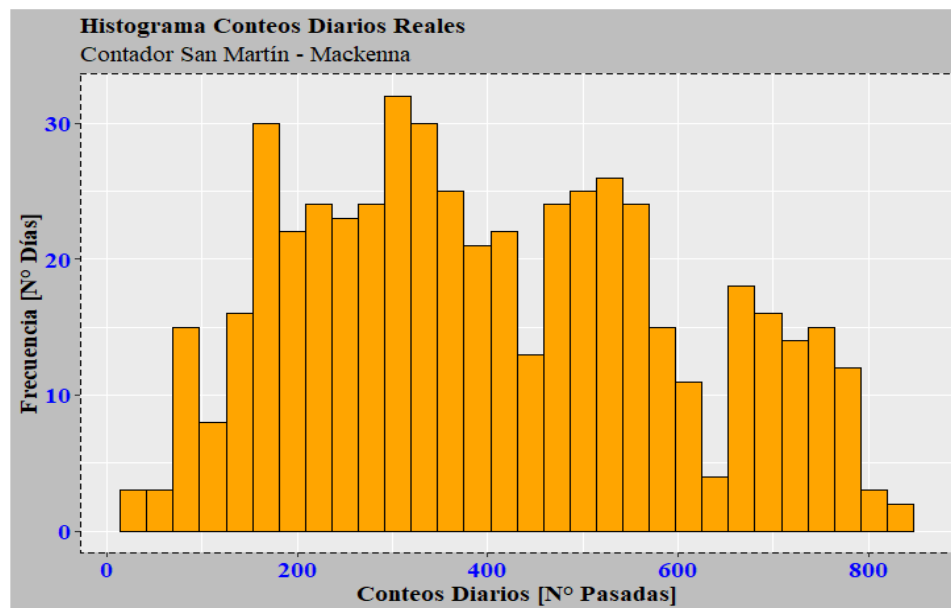


Figura A.84. Histograma de conteos diarios reales para el contador San Martín – Mackenna.

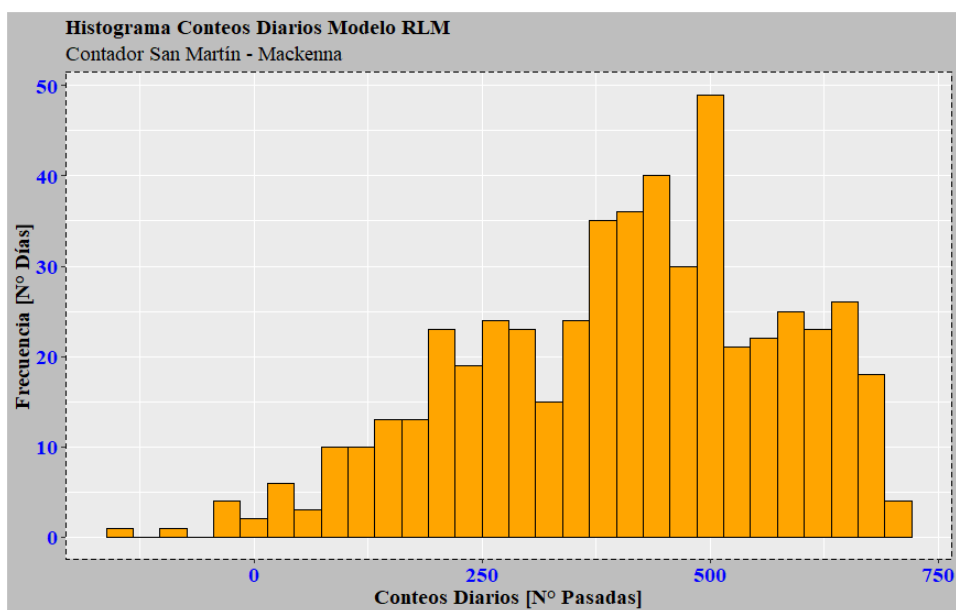


Figura A.85. Histograma de conteos diarios modelo regresión lineal múltiple para el contador San Martín – Mackenna.

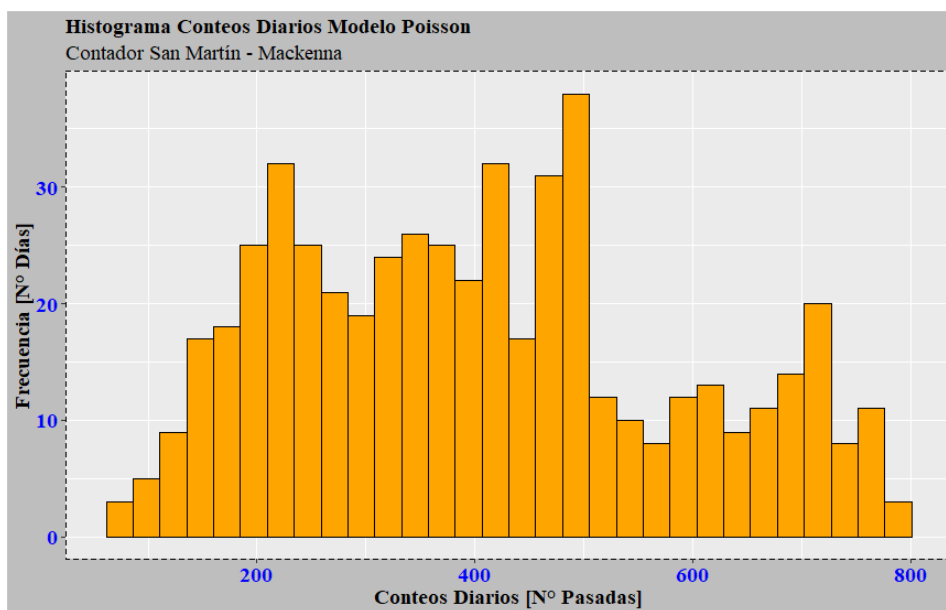


Figura A.86. Histograma de conteos diarios modelo Poisson para el contador San Martín – Mackenna.

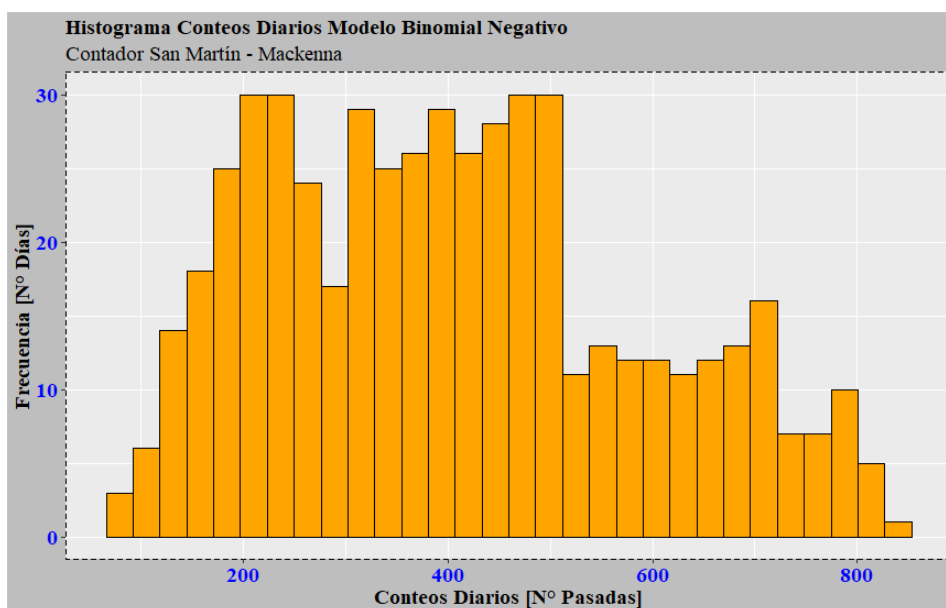


Figura A.87. Histograma de conteos diarios modelo binomial negativo para el contador San Martín – Mackenna.

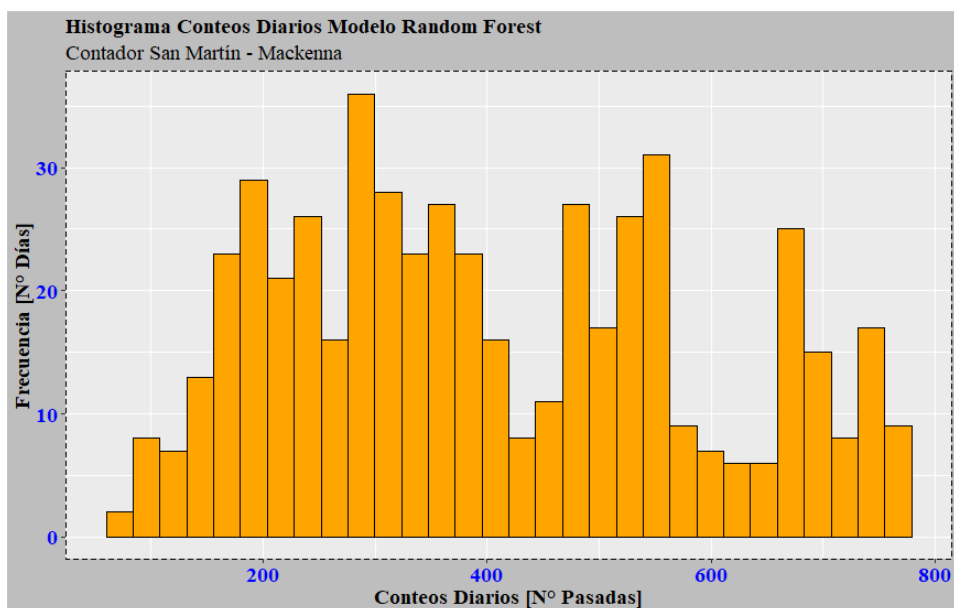


Figura A.88. Histograma de conteos diarios modelo *Random Forests* para el contador San Martín – Mackenna.

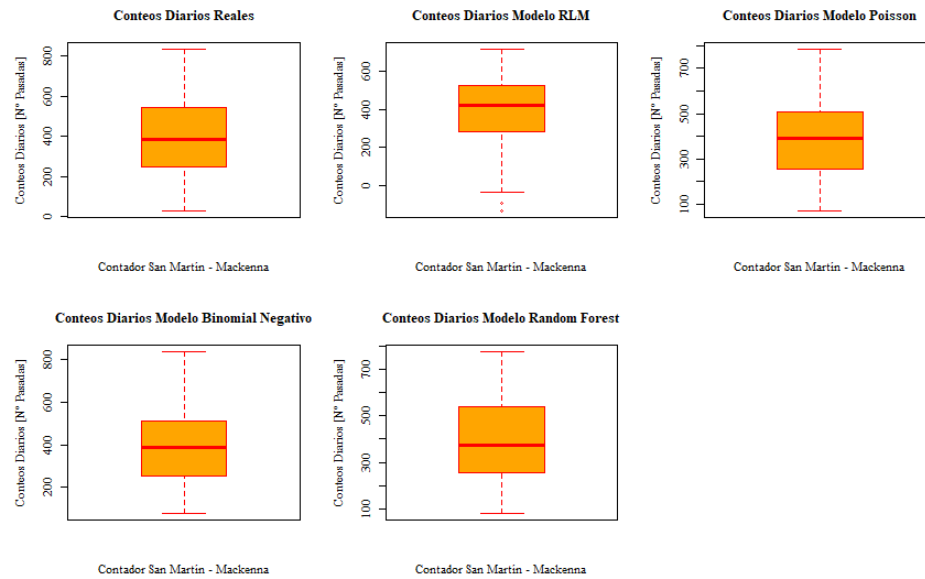


Figura A.89. Diagramas de caja de modelos desarrollados para el contador San Martín – Mackenna.

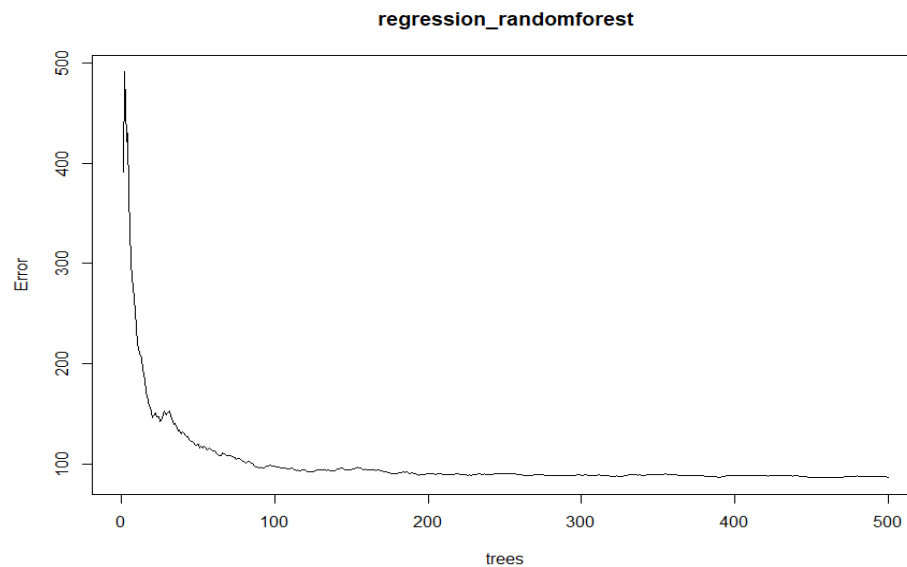


Figura A.90. Disminución del error del algoritmo *Random Forests* en función del número de árboles.

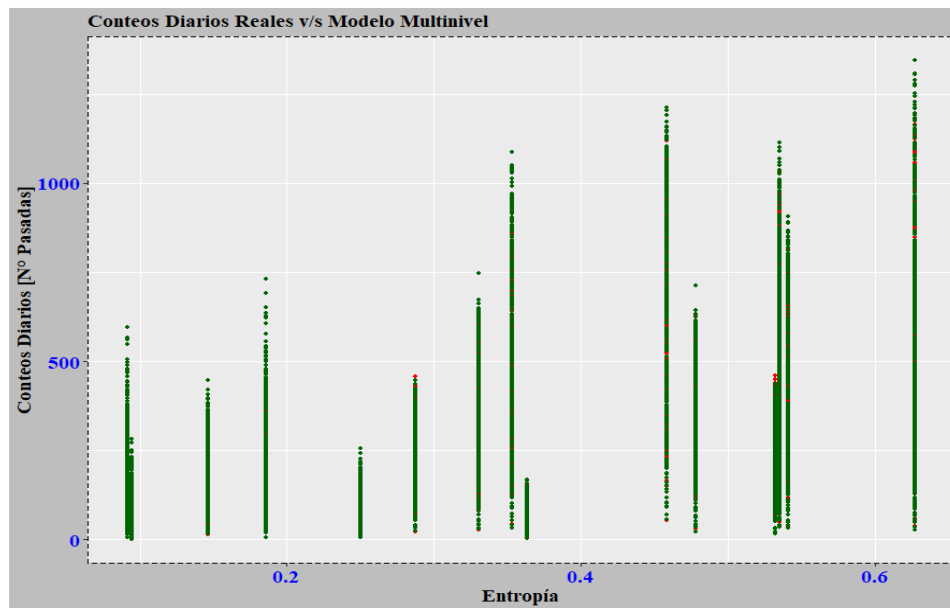


Figura A.91. Predicción conteos diarios modelo multinivel v/s conteos diarios reales en función de la entropía en el uso de suelo de cada ciclovía.

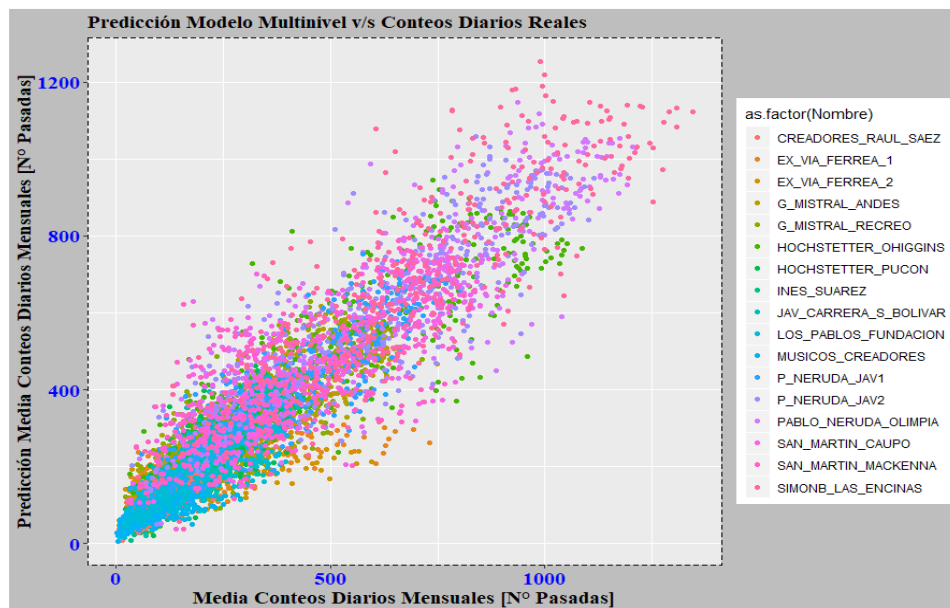


Figura A.92. Predicción conteos diarios modelo multinivel v/s conteos diarios reales para cada contador automático.

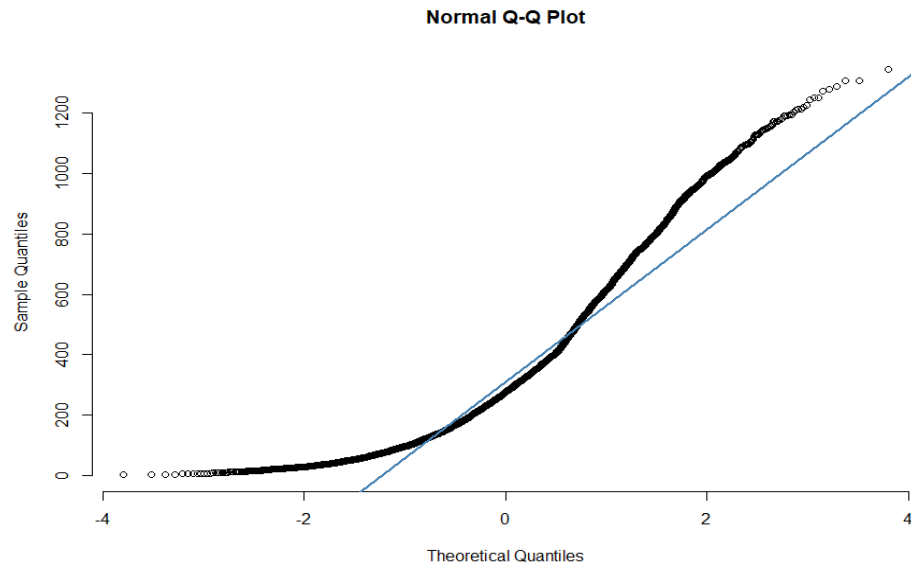


Figura A.93. Curva cuantil-cuantil de los conteos diarios reales.

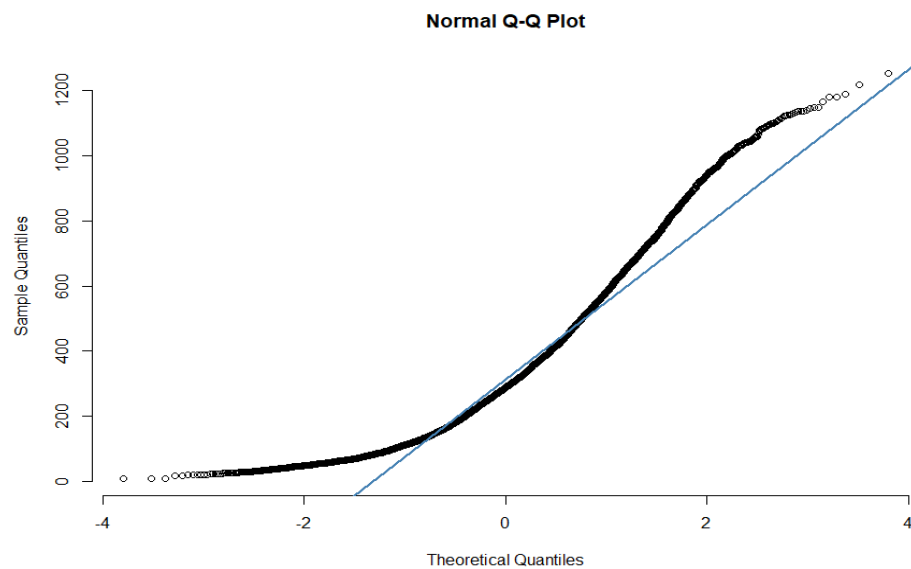


Figura A.94. Curva cuantil-cuantil de la predicción de conteos diarios del modelo multinivel.

Anexo B. Derivación modelos lineales generalizados

$$\text{Sea } Y \sim P(\mu)$$

Con función de probabilidad:

$$P(Y = y) = \frac{e^{-\mu} \mu^y}{y!} \quad y = 0, 1, 2, \dots \quad (\text{B.1})$$

De parámetro $\mu > 0$.

La media y la varianza vienen dadas por:

$$E(Y) = \mu \quad (\text{B.2})$$

$$V(Y) = \mu \quad (\text{B.3})$$

La igualdad entre la media y la varianza es conocida como la equidispersión de la distribución de Poisson. Es a través de la construcción de un modelo, en que la variable independiente sigue una distribución Poisson, donde se puede especificar el parámetro μ_i como una forma funcional de las variables explicativas $X_1, \dots, X_p$. La especificación más habitual es una exponencial lineal con el fin de garantizar que $\mu_i > 0$, es decir, se hace uso de la función canónica para la formación del modelo.

$$\mu_i = \exp(x_i \beta) \quad (\text{B.4})$$

$$E(Y_i/x_i) = \exp(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) \quad (\text{B.5})$$

Por tanto, la distribución de Poisson condicionada a las variables explicativas X viene dada por:

$$P(Y_i = y_i/x_i) = \frac{e^{-\mu_i(x_i)} \mu_i(x_i)^{y_i}}{y_i!} \quad y_i = 0, 1, 2, \dots \quad (\text{B.6})$$

Donde:

$$E(Y_i/x_i) = \mu_i(x_i) = \mu(x_{i1}, \dots, x_{ip}) = \exp(\beta_0 + \dots + \beta_p x_{ip}) \quad (\text{B.7})$$

El modelo de regresión de Poisson se caracteriza por ser un modelo heterocedástico, es decir, aquel en que las varianzas de las perturbaciones no son constantes, por lo tanto, la variabilidad es diferente para cada observación. Además, tiene la propiedad de equidispersión, o igualdad entre la varianza y la media. A pesar de que el modelo de referencia en el estudio de variables de conteo es el modelo de regresión de Poisson, éste presenta varios problemas a la hora de tratar con datos en que la media y la varianza condicionadas no son iguales, es decir, cuando los datos presentan sobredispersión. Una forma de eliminar la restricción de igualdad entre media y varianza del modelo de regresión de Poisson es especificar una distribución que permita una modelación más flexible. En ese sentido, el modelo paramétrico estándar para datos de conteo con presencia de sobredispersión, es decir, en casos en que $V(Y_i/X_i) \geq E(Y_i/X_i)$, es el modelo de regresión Binomial Negativo.

Sea la variable de conteo Y_i con $i = 1, \dots, n$ la cual sigue una distribución Binomial Negativa de parámetros r y p , de modo que $Y_i = BN(r, p)$, con función de probabilidad dada por:

$$P(Y_i = y_i) = \binom{y_i + r - 1}{r - 1} p^r (1 - p)^{y_i} \quad (\text{B.8})$$

Donde $0 < p < 1$, $r > 0$, $y_i = 0, 1, 2 \dots$

El valor esperado y la varianza vienen dados por:

$$E(Y_i) = \frac{r(1 - p)}{p} \quad V(Y_i) = \frac{r(1 - p)}{p^2} \quad (\text{B.9})$$

Estableciéndose entre ellos la siguiente relación:

$$V(Y_i) = \frac{1}{p} E(Y_i) \quad (\text{B.10})$$

Como $0 < p < 1$, se verifica que $V(Y_i) > E(Y_i)$, lo cual justifica la aptitud de esta distribución para modelar datos con presencia de sobredispersión.

La derivación del modelo Binomial Negativo puede ser llevada a cabo a partir de una regresión Poisson o como modelo lineal generalizado. El método que se presenta a continuación, es la forma más usual de derivar cualquier miembro de la familia de modelos lineales generalizados. La distribución Binomial Negativa es un miembro de la familia exponencial siempre que el parámetro de dispersión o heterogeneidad sea introducido en la distribución como una constante, por tanto, se puede usar como distribución de la componente aleatoria del modelo lineal generalizado. Al igual que el caso anterior, se presenta la derivación sugerida por Alcaide (2015). La forma de la función de probabilidad Binomial Negativa con la que se comienza se expresa como:

$$\text{Sea } Y \sim BN(r, p)$$

$$P(Y = y) = \binom{y+r-1}{r-1} p^r (1-p)^y \quad y = 0, 1, 2, \dots \quad (\text{B.11})$$

El cual se puede expresar como miembro de la familia exponencial negativa con la siguiente estructura:

$$P(Y = y) = \exp \left\{ y \ln(1-p) + r(\ln(p)) + \ln \binom{y+r-1}{r-1} \right\} \quad (\text{B.12})$$

Si se aplican los resultados obtenidos anteriormente se obtiene:

$$\theta = \ln(1-p) \Rightarrow p = 1 - \exp(\theta) \quad (\text{B.13})$$

$$b(\theta) = -r \ln(p) \Rightarrow r(1 - \exp(\theta)) \quad (\text{B.14})$$

$a(\phi)$, es el valor escalar que vale 1.

La primera y segunda derivada de $b(\theta)$ con respecto a θ dan origen a la función media y varianza respectivamente, presentadas en la ecuación (B.15) y ecuación (B.16).

$$b'(\theta) = \frac{\partial b}{\partial p} \frac{\partial p}{\partial \theta} = -\frac{r}{p} \{-(1-p)\} = \frac{r(1-p)}{p} = \mu \quad (\text{B.15})$$

$$b''(\theta) = \frac{\partial^2 b}{\partial p^2} \left(\frac{\partial p}{\partial \theta} \right)^2 + \frac{\partial b}{\partial p} \frac{\partial^2 p}{\partial \theta^2} = \frac{r}{p^2} (1-p)^2 = \frac{r(1-p)}{p^2} = \sigma^2 \quad (\text{B.16})$$

Por lo tanto:

$$V(\mu) = r(1-p)/p^2 \quad (\text{B.17})$$

Expresando p y r en términos de μ y α se obtiene que:

$$\begin{aligned} \frac{1-p}{\alpha p} &= \mu \\ \frac{1-p}{p} &= \alpha \mu \\ p &= \frac{1}{1 + \alpha \mu} \end{aligned} \quad (\text{B.18})$$

Donde $\alpha = \frac{1}{r}$.

Dados los valores definidos de μ y α , se puede expresar nuevamente la función de probabilidad de la distribución Binomial Negativa tal que:

$$P(Y = y) = \binom{y + 1/(\alpha - 1)}{1/(\alpha - 1)} \left(\frac{1}{1 + \alpha \mu} \right)^{1/\alpha} \left(\frac{\alpha \mu}{1 + \alpha \mu} \right)^y \quad (\text{B.19})$$

A partir de la ecuación anterior se obtienen los siguientes términos en función de μ y α :

$$b'(\theta) = \frac{\partial b}{\partial u} \frac{\partial u}{\partial \theta} = \frac{1}{1 + \alpha \mu} \mu (1 + \alpha \mu) = \mu \quad (\text{B.20})$$

$$b''(\theta) = V(\mu) = \frac{\partial^2 b}{\partial \mu^2} \left(\frac{\partial \mu}{\partial \theta} \right)^2 + \frac{\partial b}{\partial \mu} \frac{\partial^2 \mu}{\partial \theta^2} = \mu + \alpha \mu^2 \quad (\text{B.21})$$

El modelo de Regresión Binomial Negativa se deriva a partir de la función enlace, donde se parametriza la relación entre la media y las variables predictoras. Esta función viene dada por:

$$g(\mu) = \theta = \ln\left(\frac{\alpha\mu}{1 + \alpha\mu}\right) = -\ln\left(\frac{1}{\alpha\mu + 1}\right) \quad (\text{B.22})$$

Usando la función enlace canónica el modelo tiene la siguiente forma:

$$\begin{aligned} g(\mu) &= \eta = x\beta \\ \ln\left(\frac{1}{\alpha\mu + 1}\right) &= \eta = x\beta \end{aligned} \quad (\text{B.23})$$

Desde un punto de vista teórico, el enlace canónico representa una simplificación del estudio del modelo, pero desde el punto de vista aplicado la mayoría de los investigadores proponen el enlace logarítmico, obteniéndose el siguiente modelo:

$$\begin{aligned} g(\mu) &= \eta = x\beta \\ \ln(\mu) &= \eta = x\beta \end{aligned} \quad (\text{B.24})$$

Bajo esta función enlace se obtiene el modelo tradicional de Regresión Binomial Negativa. Este caso particular de modelo Binomial Negativo fue llamado Log-Binomial Negativo debido al enlace elegido (Hilbe, 1993).

El método más usual para la estimación de los parámetros en los modelos de recuento o conteo es el método de Máxima Verosimilitud. Debido a la complejidad de las ecuaciones de verosimilitud obtenidas, se debe recurrir a métodos numéricos para la resolución de las mismas. La mayoría de modelos utilizan el método de Newton-Rhapson o el método Fisher Scoring. Cuando el modelo de regresión Binomial Negativa es deducido como miembro de la familia de modelos lineales generalizados la estimación toma la forma del método Fisher Scoring, que permite la estimación de solo el parámetro media, μ o $\exp(x\beta)$, por lo que el parámetro de dispersión α debe de ser incluido en el algoritmo como una constante conocida. La función de log-verosimilitud de la

variante de regresión Binomial Negativa proveniente de un modelo lineal generalizado es la siguiente:

$$\mathcal{L}(\beta, \alpha) = \sum_{i=1}^n \left[-\ln(y_i!) + \sum_{j=1}^{y_i} \ln(\alpha y_i + 1 - \alpha_j) + y_i \ln(\mu_i) - \left(\frac{1}{\alpha} + y_i \right) \ln(1 + \alpha \mu_i) \right] \quad (\text{B.25})$$

Tanto el vector de parámetros del modelo (β) como el parámetro de dispersión (α) pueden ser estimados mediante el método de máxima verosimilitud conservando las propiedades teóricas de los estimadores obtenidos por este método. Las funciones score vienen dadas por:

$$s(\beta, \alpha) = \left\{ \frac{\partial \mathcal{L}(\beta, \alpha)}{\partial \beta}, \frac{\partial \mathcal{L}(\beta, \alpha)}{\partial \alpha} \right\} \quad (\text{B.26})$$

Los estimadores de máxima verosimilitud $\hat{\beta}$ y $\hat{\alpha}$ son los valores de β y α que maximizan $\mathcal{L}(\beta, \alpha)$ sobre el rango válido donde se definen los parámetros. Las ecuaciones de verosimilitud para encontrar los estimadores de máxima verosimilitud son las siguientes:

$$\frac{\partial \mathcal{L}(\beta, \alpha)}{\partial \beta} = \sum_{i=1}^n \left(x_i' \frac{y_i - \mu_i}{1 + \alpha \mu_i} \right) = 0 \quad (\text{B.27})$$

$$\frac{\partial \mathcal{L}(\beta, \alpha)}{\partial \alpha} = \sum_{i=1}^n x_i' \left\{ \sum_{j=1}^n \frac{y_i - j}{\alpha y_i + 1 - \alpha_j} + \frac{1}{\alpha} \left[\frac{\ln(1 + \alpha \mu_i)}{\alpha} - \frac{\mu_i(1 + \alpha y_i)}{1 + \alpha \mu_i} \right] \right\} = 0 \quad (\text{B.28})$$

La matriz Hessiana viene dada por:

$$H(\beta, \alpha) = \begin{bmatrix} \frac{\partial^2 \mathcal{L}(\beta, \alpha)}{\partial \beta \beta'} & \frac{\partial^2 \mathcal{L}(\beta, \alpha)}{\partial \beta \partial \alpha} \\ \frac{\partial^2 \mathcal{L}(\beta, \alpha)}{\partial \alpha \partial \beta} & \frac{\partial^2 \mathcal{L}(\beta, \alpha)}{\partial \alpha^2} \end{bmatrix} \quad (\text{B.29})$$

Por su parte, la matriz de información de Fisher esperada es:

$$I_e(\beta, \alpha) = E[-H(\beta, \alpha)] \quad (\text{B.30})$$

De igual forma, la matriz de información observada en la muestra corresponde a:

$$I_{obs}(\beta, \alpha) = -H(\beta, \alpha) \quad (\text{B.31})$$

Además, se definen los términos de la matriz Hessiana:

$$\frac{\partial^2 \mathcal{L}(\beta, \alpha)}{\partial \beta \beta'} = \sum_{i=1}^n x_i' x_i \mu_i \frac{1 + \alpha y_i}{(1 + \alpha \mu_i)^2} \quad (\text{B.32})$$

$$\frac{\partial^2 \mathcal{L}(\beta, \alpha)}{\partial \beta \partial \alpha} = \left[\frac{\partial^2 \mathcal{L}(\beta, \alpha)}{\partial \beta \partial \alpha} \right]' = \sum_{i=1}^n x_i' \mu_i \frac{y_i - \mu_i}{(1 + \alpha \mu_i)^2} \quad (\text{B.33})$$

$$\frac{\partial^2 \mathcal{L}(\beta, \alpha)}{\partial \alpha^2} = \sum_{i=1}^n \left\{ - \sum_{j=1}^{y_i} \left(\frac{y_i - j}{\alpha y_i + 1 - \alpha_j} \right)^2 + \frac{1}{\alpha^2} \left[\frac{2\mu_i + 3\alpha\mu_i^2 + \alpha^2\mu_i^2 y_i}{(1 + \alpha)^2} - \frac{2\ln(1 + \alpha\mu_i)}{\alpha} \right] \right\} \quad (\text{B.34})$$

Para obtener los estimadores de máxima verosimilitud siguiendo el método de máxima verosimilitud se hace uso del algoritmo iterativo de Fisher Scoring, el cual se define como:

$$\beta_{k+1} = \beta_k + \frac{s(\beta_k)}{I_e(\beta_k)} \quad (\text{B.35})$$

El algoritmo de Fisher Scoring es una simplificación del método de máxima verosimilitud de Newton-Rhapson. Este algoritmo es usado tradicionalmente para la estimación de modelos lineales generalizados y permite una estimación consistente únicamente del parámetro β . El parámetro de dispersión se incluye como una constante conocida.

Anexo C. Índices de segregación residencial

Los índices de segregación residencial se han utilizado en el ámbito de estudios que abarcan el área de investigación llamada modelos de demanda directa y variables de entorno construido. Actualmente los llamados índices de segregación residencial son una técnica muy utilizada en estudios urbanos. Estos clasificadores permiten comparar la distribución de alguna variable en un área metropolitana, de grupos de población que presentan diferentes orígenes étnicos, lugares de nacimiento, lengua materna, niveles de renta, entre otros. También son utilizados para comparar los resultados de diferentes ciudades y realizar análisis en diferentes momentos del tiempo (Apparicio *et al.*, 2008). Un estudio particular que utiliza índices de segregación residencial y de entorno construido en forma complementaria a las variables meteorológicas y temporales es el estudio realizado por An *et al.*, (2019). En dicho estudio se utilizan variables de entorno construido como por ejemplo la longitud de cada ciclovía y un índice de uso de suelo basado en el índice de entropía de Shannon. Las cinco dimensiones de la segregación residencial definidos por Massey y Denton (1988) los clasifican en: igualdad, exposición, concentración, agrupamiento y centralización. En particular, dentro de la dimensión de igualdad, se ubica el Índice de entropía propuesto por Theil (1972), el cual es una medida de la desigualdad basada en la entropía de Shannon. Si bien en primera instancia fue desarrollado para medir y comparar la distribución espacial de la renta, actualmente tiene un uso generalizado para comparar la distribución inequitativa del uso de suelo que se presenta en las ciudades. El índice de entropía de Theil se define de la siguiente forma:

$$T_T = \sum_{i=1}^m s_i T_i + \sum_{i=1}^m s_i \ln\left(\frac{\bar{x}_i}{\mu}\right) \quad (\text{C.1})$$

Teniendo en cuenta que:

$$s_i = \frac{N_i}{N} \frac{\bar{x}_i}{\mu} \quad (\text{C.2})$$

Donde:

s_i : Es la proporción del ingreso en el grupo i.

- N_i : Es la población del grupo i.
- N : Es la población total.
- T_i : Es el índice de Theil para ese subgrupo.
- x_i : Es el ingreso promedio del grupo i.
- μ : Es el ingreso promedio de la población.