

Estimación del consumo promedio de combustible en los automóviles de combustión interna, en función de otras variables

Victoria Cao, Mauricio Serrano, Jose Valverde¹

victoria.cao@ucr.ac.cr, mauricio.serrano@ucr.ac.cr, jose.valverde@ucr.ac.cr

RESUMEN

Existen varios factores mecánicos y humanos asociados al consumo de combustible en los automóviles de combustión interna. El análisis de la combinación de esos factores puede contribuir a la construcción de un modelo estadístico que permita analizar la incidencia, en el consumo del combustible, de factores como la distancia recorrida, la ubicación de la carretera o camino en un entorno social, las características del automóvil, el estado del tiempo al conducir, entre otros. El presente trabajo busca determinar cuáles factores son los que contribuyen en el consumo promedio de combustible de los automóviles particulares. Se analizan todos esos factores mediante la confección de un modelo de regresión lineal. Se encuentra una alta incidencia de las variables predictoras en el consumo del combustible. Algunas relaciones entre los factores y el consumo de combustible coinciden con resultados realizados en otros experimentos, mientras que otras relaciones describen fenómenos distintos a los hallados por otros especialistas.

PALABRAS CLAVE: consumo de combustible, distancia recorrida, modelo de regresión lineal, ANOVA.

INTRODUCCIÓN

El estudio del consumo de combustibles fósiles en los automóviles tiene gran importancia para el planeta, debido a factores socioeconómicos y ambientales, entre otros. Por ello, la reducción en el uso de estos combustibles se ha convertido en un tema de gran interés internacional. En el presente trabajo se estudiaron variables que influyen en el consumo de combustible de automóviles particulares en Costa Rica. Se tomarán en cuenta variables relacionadas con la mecánica de los automóviles y variables ambientales que inciden directamente en el consumo del combustible. Para llevar a cabo el estudio, se aplicarán los fundamentos y características del modelo de regresión lineal sobre un conjunto de datos de al menos 50 automóviles particulares que circulan en Costa Rica.

La Organización de las Naciones Unidas, en el contexto del cambio climático, indica que la utilización de este tipo de combustibles genera considerables emisiones de gases tipo

¹ Estudiantes de la Escuela de Estadística de la Universidad de Costa Rica

invernadero, los cuales provocan la retención del calor del sol y la elevación consecuente de las temperaturas a nivel mundial (Organización de las Naciones Unidas [ONU] 2023).

Por otra parte, las frecuentes variaciones en el precio del petróleo afectan directamente el bolsillo de los conductores de vehículos. De acuerdo con datos publicados en el diario El Financiero de Costa Rica, según un estudio del año 2022, en el país hubo un consumo total de combustible de 3 449 millones de litros, tomando en cuenta los tres principales combustibles, a saber, la gasolina regular, la gasolina superior (super) y el diesel (ARESEP, 2022, como se citó en El Financiero, 2023).

Se plantea como objetivo general determinar cuáles factores contribuyen en el consumo promedio de combustible de los automóviles. Asimismo, se establece como objetivo específico analizar y determinar si existe relación entre la distancia recorrida, año del modelo del automóvil, el peso adicional, uso de aire acondicionado, marca, estado del tiempo al conducir, tipo de carretera, sexo y edad del conductor y el consumo de combustible. Durante la etapa de definición de los objetivos se contó con el criterio experto de la Dra. Leonora De Lemos, docente e investigadora de la Escuela de Ingeniería Mecánica de la Universidad de Costa Rica.

Se espera observar y comprobar que los vehículos que han recorrido distancias más largas, han utilizado el aire acondicionado con mayor frecuencia y tienen un peso adicional significativo presenten un mayor consumo promedio de combustible.

METODOLOGÍA

Se definió como unidad de observación el automóvil. Se confeccionó una pequeña encuesta en la que se solicitaron datos mecánicos de los automóviles, datos de consumo en un período específico y datos del contexto de los viajes realizados durante ese período. Como variable respuesta se definió el consumo de litros de combustible promedio de los automóviles. Se realizó el estudio para 49 automóviles de la zona urbana de la Gran Área Metropolitana. Debido a la renuencia de algunas personas conductoras, de brindar información sobre sus vehículos, se seleccionaron 49 automóviles por conveniencia. Para garantizar la validez de los datos de cada automóvil, se constató que, durante el período del estudio, la persona conductora había utilizado siempre el mismo automóvil.

Para obtener los datos de consumo de combustible propiamente dicho, se llevó a cabo el siguiente procedimiento:

- La persona conductora del automóvil registró el kilometraje indicado por el odómetro al momento que llenó el tanque de combustible una primera vez.
- La persona conductora efectuó un segundo registro la siguiente vez que llenó el tanque del auto.
- La distancia recorrida medida en kilómetros, durante el período transcurrido, se obtuvo de la diferencia entre ambos registros.
- Cuando un automóvil registró la distancia en millas, se realizó la correspondiente conversión a kilómetros (1 milla = 1.609 kilómetros).
- Se registró el tiempo transcurrido entre ambos registros, medido en días.

En el Cuadro 1 se presentan las variables explicativas o predictoras del estudio derivadas de los datos proporcionados por las personas conductoras. Se describe el tipo de variable (continua o categórica), el rango de valores y la unidad de medición.

Para medir la variable ‘peso adicional’, se preguntó a la persona conductora un estimado del peso adicional que transportó en el automóvil durante el período de observación, sin tomar en cuenta elementos básicos de un automóvil. Es decir, cuánto peso transportó en elementos tales como compras y electrodomésticos, entre otros. Para medir la variable ‘estado del tiempo’, se les preguntó a las personas conductoras en qué condiciones de estado de tiempo condujeron durante el período de observación. Por otra parte, para determinar el ‘tipo de carretera’, se consultó a las personas conductoras sobre la ubicación de las carreteras que frecuentaron durante el período de estudio, a saber, zona urbana (dentro de ciudades), autopistas, o zonas rurales.

Cuadro 1

Consumo promedio de combustibles: descripción de variables predictoras iniciales

Variable ^{1/}	Tipo	Valores posibles / Rango	Unidad de medición
Distancia (kilometraje)	Continua	0 - 50 000	kilómetro ^{2/}
Tiempo de conducción (tiempo)	Continua	1 – 50	Día
Edad del conductor (edad)	Continua	18 – 65	Año
Año del modelo (modelo)	Continua	1960 – 2023	Año
Peso adicional (peso)	Continua	0 – 1000	kilogramos
Número de acompañantes (acompañantes)	Continua	0 – 10	Persona
Uso de aire acondicionado (aire)	Categórica	0 = Siempre 1 = Con frecuencia 2 = Nunca	No aplica
Estado del tiempo (clima)	Categórica	0 = Muy caluroso/ Caluroso 1 = Fresco / Lluvioso 2 = Lluvioso/Caluroso	No aplica
Sexo del conductor (sexo)	Categórica	0 = Masculino 1 = Femenino	No aplica
Tipo de carretera (carretera)	Categórica	0 = Zona urbana/ Autopista 1 = Zona rural/ Otro	No aplica
Tipo de carro (carro)	Categórica	0 = Sedan 1 = Hatchback 2 = Pick-up 3 = Otro	No aplica

Tipo de combustible (combustible)	Catórica	0 = Gasolina 1 = Diesel	No aplica
--------------------------------------	----------	----------------------------	-----------

^{1/} Entre paréntesis se encuentran los nombres resumidos

^{2/} Los datos de distancia en millas, se convirtieron a kilómetros (1 milla = 1.60934 kilómetros)

Durante la limpieza y depuración de los datos se reagruparon algunas variables catóricas para las que se obtuvieron categorías con muy pocos o ningún valor.

El modelo inicial resultante es el siguiente:

$$\mu_{consumo|x_1,\dots,x_{12}} = \beta_0 + \sum_{i=1}^{12} \beta_i * X_i$$

Donde,

X_i es cada una de las 12 variables predictoras del modelo inicial, según el Cuadro 1.

β_i es cada uno de los coeficientes asociados con cada X_i

La evaluación de los supuestos se realizó mediante el análisis de gráficos estadísticos básicos como el gráfico cuantil-cuantil y gráficos de correlación y se complementó mediante pruebas estadísticas formales. Para la selección del modelo definitivo, se utilizó el Criterio de Información de Akaike (AIC) y se complementó la validez de los resultados llevando a cabo un segundo análisis de cumplimiento de los supuestos.

El procesamiento de los datos y la obtención de los resultados, se llevó a cabo utilizando el lenguaje de programación R versión 4.3.0 (R Core Team, 2023). Asimismo, se utilizaron las bibliotecas ("librerías") existentes en el lenguaje R: lattice (Sarkar, 2008) en la versión 0.21-8, car (Fox, Weisberg, 2019), versión 3.1-2, corrgram (Wright, 2021), versión 1.14 y ggplot2 (Wickham, 2016), versión 2.4.3.

RESULTADOS

Primeramente, se estudió la existencia de los valores extremos y valores de influencia presente que puedan afectar el modelo. Se estudiaron los valores extremos con los residuales estudentizados del modelo y los valores de influencia con las distancias de Cook. No se encontraron valores extremos con esta prueba, por lo que se procedió a analizar los supuestos del modelo.

En relación con las correlaciones entre las variables predictoras, se utilizó un diagrama de correlaciones entre las variables predictoras iniciales (ver Figura 1 en anexo). Se pudo asumir la no multicolinealidad ya que los coeficientes presentaron correlaciones bajas o medias.

En cuanto a la normalidad de los residuales condicionales a las variables predictoras, se utilizó el gráfico cuantil-cuantil (ver Figura 2 en anexo). Se observó que los residuales siguen una distribución normal. Esto se confirmó mediante una prueba Shapiro-Wilk. Se obtuvo una probabilidad $p = 0.7397$, lo cual indicó que se puede asumir normalidad.

Para probar el supuesto de linealidad se utilizó un gráfico de la variable respuesta vs las variables numéricas (ver Figura 3 en anexo). Las variables predictoras guardaron relación lineal con los residuales, lo cual hace presumir que se cumplió el supuesto de linealidad. En el caso de las variables Modelo y Kilometraje, se decidió asumir linealidad dado que hay una escasa cantidad de datos en los valores que corresponden a modelos de autos anteriores al 2005 y en los kilometrajes que superan los 2000 Km.

Se pudo asumir independencia en el modelo, ya que los datos recopilados para un automóvil fueron completamente independientes de los datos recopilados de otros automóviles. Además, los datos recogidos no fueron datos dependientes del tiempo o datos espaciales, que podrían presentar problemas. Cabe agregar que, al no utilizar muestreo por conglomerados, se descartaron de antemano posibles problemas de no independencia asociados con este tipo de datos.

Para escoger el modelo más adecuado, se analizó un modelo con interacción en el cual se estudiaron las distintas interacciones existentes entre las seis variables categóricas y las seis variables numéricas. Debido a que se obtuvieron muchas interacciones, se descartaron de antemano aquellas interacciones que no fueron de interés para el estudio. Con las interacciones que permanecieron, se realizó una prueba F para observar su significancia. En el modelo resultante, las interacciones restantes se dieron entre las variables *peso adicional vs tipo de combustible*, *número de personas vs estado del tiempo* y *kilometraje vs aire acondicionado* (ver Figura 4 en anexo).

Se pudo observar que, conforme aumentó el peso adicional que cargaba el automóvil, para los vehículos que utilizaron diésel, el consumo fue disminuyendo. En tanto, para los vehículos que utilizaron gasolina, conforme aumentó el peso adicional que el vehículo cargaba, el consumo de combustible fue aumentando.

De igual manera, conforme aumentó el kilometraje recorrido, para los vehículos que nunca utilizaron el aire acondicionado, el consumo de combustible fue aumentando más rápidamente en comparación a aquellos que siempre o frecuentemente utilizaron el aire acondicionado. Cuando el vehículo siempre utilizó el aire acondicionado, se pudo observar que el consumo de combustible fue aumentando más lentamente en comparación a las otras pendientes.

Para la interacción entre el 'número de personas' y el 'estado del tiempo', conforme aumentó el número de personas en el vehículo, para el estado del tiempo caluroso, el consumo de combustible fue aumentando más rápidamente que para las otras categorías. Para el 'estado del tiempo fresco/lluvioso', conforme aumentó el número de personas en el vehículo, el consumo de combustible fue disminuyendo. Para la categoría caluroso/lluvioso, el consumo de combustible fue aumentando a un ritmo más lento que la categoría caluroso, conforme aumentó el número de personas.

En las tres interacciones, la prueba F presentó una probabilidad menor a 0.05, por lo que se asumió la existencia de interacción entre las variables anteriormente mencionadas. Por tanto, el modelo preliminar con interacción es el siguiente:

$$\hat{Y}_i = \beta_0 + \sum_i \beta_i * X_i$$

Donde,

$i \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 6-10, 3-8, 3-9, 5-12, 5-13\}$

y

X_1 = Edad

X_2 = Modelo

X_3 = Kilometraje

X_4 = Días

X_5 = Número_personas

X_6 = Peso_adicional

X_7 = SexoMasculino

X_8 = Aire_acondicionadoConFrecuencia

X_9 = Aire_acondicionadoNunca

X_{10} = Tipo_combustibleGasolina

X_{11} = Tipo_carreteraRural/Otro

X_{12} = Estado_tiempoFresco/Lluvioso

X_{13} = Estado_tiempoCaliente/Lluvioso

X_{14} = Tipo_carroOtro

X_{15} = Tipo_carroPick-up

X_{16} = Tipo_carroSedan

X_{6-10} = Peso_adicional:Tipo_combustibleGasolina

X_{3-8} = Kilometraje:Aire_acondicionadoCon frecuencia

X_{3-9} = Kilometraje:Aire_acondicionadoNunca

X_{5-12} = Num_personas:Estado_tiempoFresco/Lluvioso

X_{5-13} = Num_personas:Estado_tiempoCaluroso/Lluvioso

Este modelo cumplió con los supuestos de normalidad, no multicolinealidad e independencia. Para comprobar el cumplimiento del supuesto de homocedasticidad, se realizó la prueba Breusch-Pagan, la cual arrojó una probabilidad de 0.2646, por lo que se pudo asumir que los residuales del modelo sí cumplieron con el supuesto de homocedasticidad.

Para efectuar la reducción del modelo, se llevaron a cabo dos procesos de selección de variables utilizando el método por pasos: un proceso hacia atrás y otro proceso hacia adelante, utilizando el Criterio de Información de Akaike (AIC, por sus siglas en inglés). Durante este proceso, se encontró que el R cuadrado ajustado del modelo hacia atrás fue mayor (0.5442) que el R cuadrado ajustado del modelo hacia adelante. Por tanto, se escogió el modelo del proceso hacia atrás.

El Cuadro 2 presenta las variables predictoras seleccionadas junto con sus respectivos coeficientes:

Cuadro 2

Variables predictoras seleccionadas

X_i	Variable predictor	Coefficiente beta	Límite inferior	Límite superior
	(intercepto)	0.032	-2007.299	-95.465
X_1	Modelo	0.025	0.071	1.019
X_2	Kilometraje	0.072	-0.001	0.012
X_3	Días	0.114	-1.088	0.122
X_4	Num_personas	0.176	-4.950	25.993
X_5	Peso_adicional	0.112	-0.562	0.061
X_6	Aire_acondicionadoCon frecuencia	0.478	-15.173	7.255
X_7	Aire_acondicionadoNunca	0.117	-24.534	2.844
X_8	Tipo_combustibleGasolina	0.001	-26.233	-8.033
X_9	Estado_tiempoFresco/Lluvioso	0.294	-3.927	12.564
X_{10}	Estado_tiempoCaluroso/Lluvioso	0.413	-12.906	5.431
X_{5-8}	Peso_adicional:Tipo_combustibleGasolina	0.017	0.112	0.873
X_{2-6}	Kilometraje:Aire_acondicionadoCon frecuencia	0.807	-0.014	0.011
X_{2-7}	Kilometraje:Aire_acondicionadoNunca	0.053	0.000	0.047
X_{4-9}	Num_personas:Estado_tiempoFresco/Lluvioso	0.123	-28.391	3.553
X_{4-10}	Num_personas:Estado_tiempoCaluroso/Lluvioso	0.325	-23.520	8.023

La ecuación de regresión resultante, de acuerdo con los datos del Cuadro 2, es la siguiente:

$$\hat{Y} = 0.032 + 0.025 * X_1 + 0.072 * X_2 + 0.114 * X_3 + 0.176 * X_4 + 0.112 * X_5 + 0.478 * X_6 + 0.117 * X_7 + 0.001 * X_8 + 0.294 * X_9 + 0.413 * X_{10} + 0.017 * X_{5-8} + 0.807 * X_{2-6} + 0.053 * X_{2-7} + 0.123 * X_{4-9} + 0.325 * X_{4-10}$$

Para el modelo seleccionado se obtuvo un $R^2 = 0.69$. Eso indica que es posible explicar un 69% de la variabilidad en el consumo de combustible con este modelo de regresión.

Se observó que los vehículos más antiguos tuvieron, en promedio, un mayor consumo de combustible. Se puede esperar con un 95% de confianza que, por cada incremento de 6 años en el modelo del vehículo, el consumo de combustible aumente entre 0.427 y 6.113 litros, manteniendo constantes las demás variables.

Se observó además, que el tipo de combustible afecta el consumo promedio de litros de combustible. Se espera con un 95% de confianza que el consumo de combustible se reduzca entre 8.033 y 26.233 litros cuando el vehículo utiliza gasolina en comparación a un vehículo de diésel, manteniendo constantes las demás variables.

En cuanto a la interacción Peso adicional: Tipo de combustible, se espera con un 95% de confianza que con un aumento de 1Kg en el peso adicional, el consumo de combustible para los

vehículos de gasolina aumente más que en los vehículos de diesel. Este aumento podría variar entre 0.112 y 0.873 litros.

Los restantes coeficientes del modelo no tuvieron interpretación, debido a que el valor cero se encontraba dentro del intervalo de confianza.

Finalmente, el modelo desarrollado permite predecir el consumo de combustible según los valores dados para cada uno de los predictores. Por ejemplo, para un vehículo de gasolina modelo 2014, kilometraje recorrido de 577 Km, durante un período de 5 días, con una persona adicional, con un peso adicional de 17 Kg, usando el aire acondicionado siempre, y bajo condiciones de estado del tiempo “Caluroso/Lluvioso”, se espera, con un 95% de confianza, que el consumo de combustible esté entre 25.9 y 40.7 Litros.

CONCLUSIONES

Fue posible obtener un modelo de regresión con 15 predictores que permite explicar un 69% de la variabilidad en el consumo de combustible.

Se observó que los vehículos de años más recientes consumieron en promedio menos combustible. Esto concuerda con lo indicado por McCarthy y Tay (1998), quienes afirman que los nuevos automóviles deben adherirse a ciertos lineamientos de consumo de combustible como el Corporate Average Fuel Economy (CAFE), establecido en los Estados Unidos. Sprei, Karlsson y Holmberg (2008), también afirman que el consumo de combustibles en automóviles ha disminuido con la ayuda de los avances tecnológicos.

En promedio, los vehículos de gasolina consumieron menos combustible que los vehículos de diesel. Esto contradice a Sullivan, Baker, Boyer, Hammerle, Kenneym, Muniz y Wallington (2004), quienes afirman que debido al diseño de los motores diesel, estos consumen menos combustibles en comparación con los motores que utilizan gasolina.

Se encontró que aunque el uso del aire acondicionado sí aumentó el gasto en combustible, es menor en comparación a que si no se usará el aire acondicionado. Esto contradice a lo indicado en los estudios del Departamento de Energías de Estados Unidos (2008), donde se afirma que el aire acondicionado reduce el rendimiento en Km/litro hasta en un 20 %. Cabe recalcar que esto se podría deber al tamaño de muestra recolectado. Se necesitaría un tamaño de muestra mayor para llevar a cabo un análisis más específico.

Se observó que para los automóviles que conducían en un estado de tiempo *caluroso*, el consumo de combustible aumentó más aceleradamente en comparación con aquellos automóviles que conducían durante los estados de tiempo *lluviosos* o *calurosos/lluviosos*. Esto concuerda con lo encontrado por el Departamento de Energía de Estados Unidos (2008), quienes afirman que los autos consumen más combustible en climas calientes.

De acuerdo con el criterio de la Dra. Leonora De Lemos se recomienda, para estudios futuros, la utilización de sistemas de diagnóstico a bordo (Ingals, 1996), los cuales permiten

recolectar, entre otros datos, el kilometraje recorrido y el consumo de combustible de una forma más exacta y detallada para las variables estudiadas.

BIBLIOGRAFÍA

- El Financiero (5 de junio de 2023). Estos son los cantones que consumen más diésel y gasolina. El Financiero. <https://www.elfinancierocr.com/economia-y-politica/estos-son-los-cantones-que-consumen-mas-diesel-y/ZBYKEAZWBZAV5BNTWCU4KCBOQQ/story/#:~:text=%E2%80%8C&text=Seg%C3%BAAn%20datos%20recientes%20de%20la,la%20pandemia%20de%20covid%2D19>.
- Fox J, Weisberg S (2019). *An R Companion to Applied Regression*, Third edition. Sage, Thousand Oaks CA. <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>.
- H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- Ingals, M (1996). On-board diagnostics -Second generation (OBD-II) system criteria for alternate-fueled vehicles. Southwest Research Institute. <https://nepis.epa.gov/Exe/ZyNET.exe/94004KUR.txt?ZyActionD=ZyDocument&Client=EPA&Index=1995%20Thru%201999&Docs=&Query=&Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&UseQField=&IntQFieldOp=0&ExtQFieldOp=0&XmlQuery=&File=D%3A%5CZYFILES%5CINDEX%20DATA%5C95THRU99%5CTXT%5C0000040%5C94004KUR.txt&User=ANONYMOUS&Password=anonymous&SortMethod=h%7C-&MaximumDocuments=1&FuzzyDegree=0&ImageQuality=r75g8/r75g8/x150y150g16/i425&Display=hpfr&DefSeekPage=x&SearchBack=ZyActionL&Back=ZyActionS&BackDesc=Results%20page&MaximumPages=1&ZyEntry=8>
- MacCarthy, P.S., Tay, R.S. (2008). New vehicle consumption and fuel efficiency: a nested logit approach. *Transportation Research Part E: Transport and Environment*. 34(1), 39-51. [https://doi.org/10.1016/S1366-5545\(97\)00042-2](https://doi.org/10.1016/S1366-5545(97)00042-2)
- Organización de las Naciones Unidas. (2023). ¿Qué es el cambio climático?. Organización de las Naciones Unidas. <https://www.un.org/es/climatechange/what-is-climate-change#:~:text=La%20quemado%20combustibles%20f%C3%B3siles%20genera%20emisiones%20de%20gases%20de,de%20carbono%20y%20el%20metano>.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Sarkar D (2008). *Lattice: Multivariate Data Visualization with R*. Springer, New York. ISBN 978-0-387-75968-5, <http://lmdvr.r-forge.r-project.org>.
- Sprei, F., Karlsson, S., Holmberg, J. (2008). Better performance or lower fuel consumption: Technological development in the Swedish new car fleet 1975-2002. *Transportation Research Part D: Transport and Environment*, 13(2), 75-85. <https://doi.org/10.1016/j.trd.2007.11.003>

Sullivan, J.L., Baker, R.E., Boyer, B.A., Hammerle, R.H., Kenney, T.E., Muniz, L., Wallington, T.J. (2004). CO2 Emission Benefit of Diesel (versus Gasoline) Powered Vehicles. *Environmental Science & Technology*.38 (12), 3217-3223 DOI: 10.1021/es034928d

US Department of Energy (2008). Fuel economy guide. Available at: www.fueleconomy.gov

Wright K (2021). `_corrgram: Plot a Correlogram`. R package version 1.14, <https://CRAN.R-project.org/package=corrgram>.

ANEXOS

Figura 1

Diagrama de correlaciones entre las variables predictoras

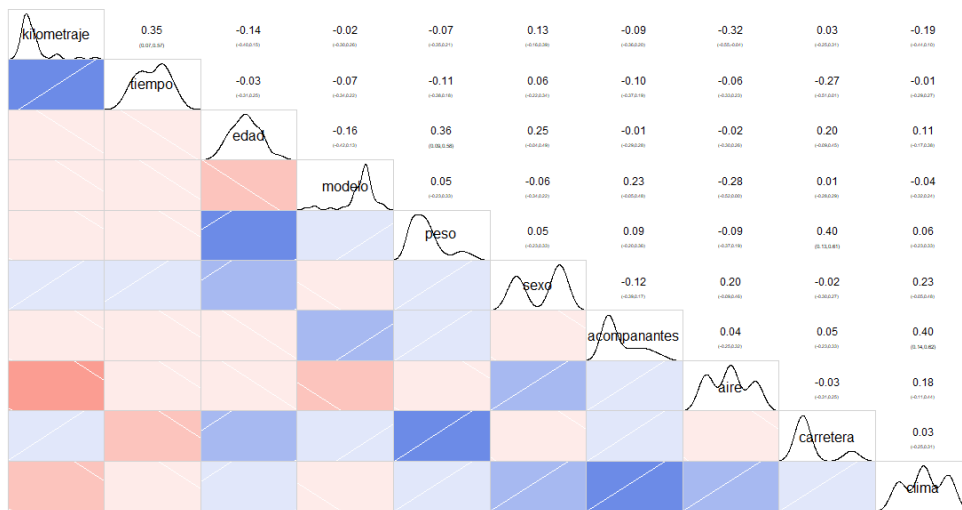


Figura 2

Gráfico cuantil-cuantil para los residuales del modelo propuesto

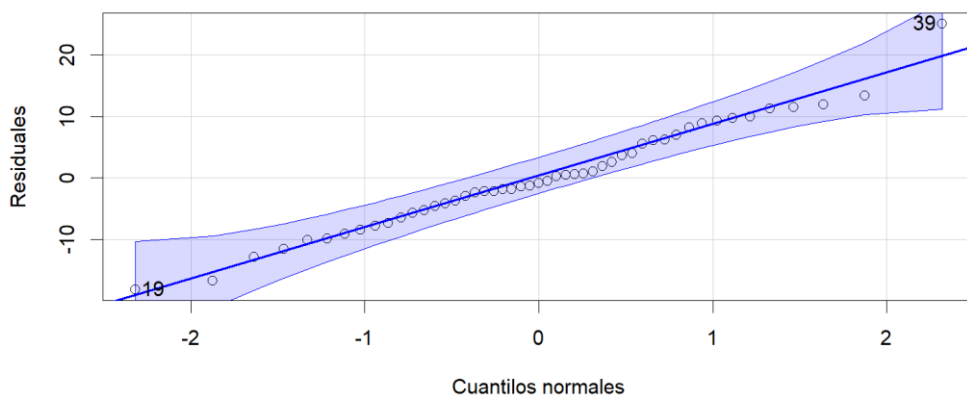


Figura 3

Verificación del supuesto de linealidad de la variable de respuesta versus las variables numéricas

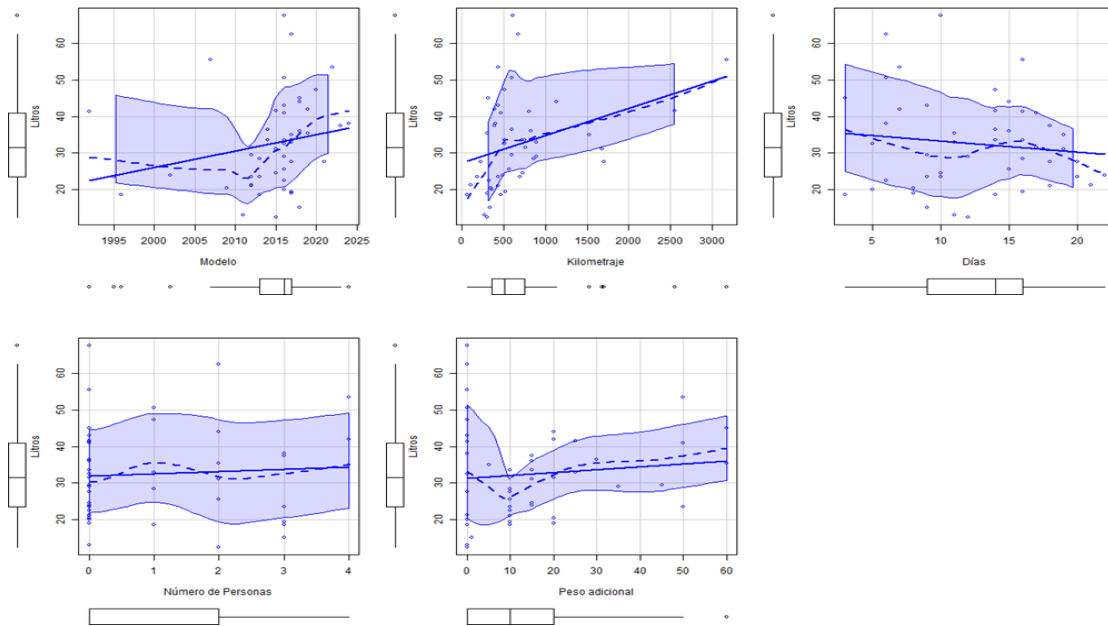


Figura 4

Interacciones entre variables predictoras

