



Coordinación de
Educación Abierta y a Distancia
VICERRECTORADO ACADÉMICO



Análisis de Regresión Lineal Simple: Relación entre Peso y Rendimiento en el Dataset mtcars

Alisson Atupaña, Mario Camacho, Lenin Lopez

Universidad Nacional de Chimborazo

Facultad de Ingeniería

Ciencia de Datos e IA

Materia: Modelamiento

Docente: Estalin Mejia H.

Semestre: Tercero

Riobamba - Ecuador

16 de octubre de 2025

1 Introducción

1.1 Planteamiento del Problema

Existe una correlación entre el peso de un camión y en su rendimiento en el consumo de combustible constituye un aspecto esencial dentro de la industria automotriz moderna. Diversos investigadores sostienen que la cantidad de peso en la carga influye de manera directa en el rendimiento energético del vehículo, lo que quiere decir que entre mayor sea el peso, mayor será el consumo del combustible.

1.2 Preguntas de Investigación

1. ¿Existe una relación estadísticamente significativa entre el peso del vehículo y su rendimiento de combustible?
2. ¿Cuál sería el rendimiento de combustible estimado para un vehículo sin carga (peso teórico igual a cero)?

1.3 Objetivos

1.3.1 *Objetivo General*

Analizar la existencia y magnitud de la relación entre el peso del vehículo y su rendimiento de combustible mediante un modelo de regresión lineal simple haciendo uso del dataset `mtcars`.

1.3.2 *Objetivos Específicos*

1. Evaluar la normalidad de las variables mediante pruebas estadísticas apropiadas.
2. Calcular el coeficiente de correlación de Pearson entre el peso y el rendimiento.
3. Ajustar un modelo de regresión lineal simple y evaluar su significancia estadística.

4. Validar el cumplimiento de los supuestos del modelo de regresión.

1.4 Hipótesis

Hipótesis nula (H_0): No existe relación una relación lineal significativa entre el peso del vehículo y su rendimiento de combustible $H_0 : \beta_1 = 0$

Hipótesis alternativa (H_1): Existe una relación lineal significativa entre el peso del vehículo y su rendimiento de combustible $H_1 : \beta_1 \neq 0$

Donde β_1 representa el coeficiente de regresión (pendiente) el cual mide el efecto del peso sobre el rendimiento. Para la prueba de hipótesis se empleará un nivel de significancia de $\alpha = 0,05$ para el contraste de hipótesis.

2 Resultados

Se efectuó un análisis estadístico a travez de una metodología lineal desarrollada en seis etapas claves, que se se detallaran junto a los resultados obtenidos de dicho proceso estadístico.

2.1 Exploración del dataset

Consistió en la carga de todo el conjunto de datos empleados desde el dataset `mtcars`, que contiene información sobre 32 vehículos publicada por la revista *Motor Trend* de 1974, disponible en el entorno estadístico R (R Core Team, 2023). Dicho dataset contiene 11 variables de las cuales fueron seleccionadas las mas relevantes para dicho analisis:

- **mpg** (Miles Per Gallon): Hace referencia al rendimiento de combustible, y se ha asignada como la variable dependiente.
- **wt** (Weight): Hace referencia al peso del vehículo expresado en miles de libras, asignada como la variable independiente.

El beneficio de la exploracion reveló que el dataset no contiene valores faltantes (N/A) en ninguna de las dos variables de interés, lo que permitio proceder al calculo sin usar procedimientos de eliminacion o relleno de datos en las observaciones, la siguiente tabla 1 resume la estructura del dataset.

Cuadro 1: Estructura y Completitud del Dataset mtcars

Variable	Observaciones	Valores Perdidos	Completitud
Rendimiento (mpg)	32	0	100 %
Peso (wt)	32	0	100 %
Total	32	0	100 %

Nota. El dataset no requiere tratamiento de valores perdidos. Todas las observaciones están completas para ambas variables de estudio.

2.2 Cálculo de Estadísticas Descriptivas

Se hizo énfasis al cálculo de medidas de tendencia central, como la dispersión y rango, que se realizó para observar las estadísticas de las dos variables. La Tabla 2 resume estos estadísticos descriptivos.

Cuadro 2: Estadísticas Descriptivas de las Variables de Estudio

Variable	Media	Mediana	DE	Mín	Máx	n
Rendimiento (mpg)	20.09	19.20	6.03	10.40	33.90	32
Peso (1000 lbs)	3.22	3.33	0.98	1.51	5.42	32

Nota. DE = Desviación Estándar; Mín = Mínimo; Máx = Máximo; n = Tamaño de muestra.

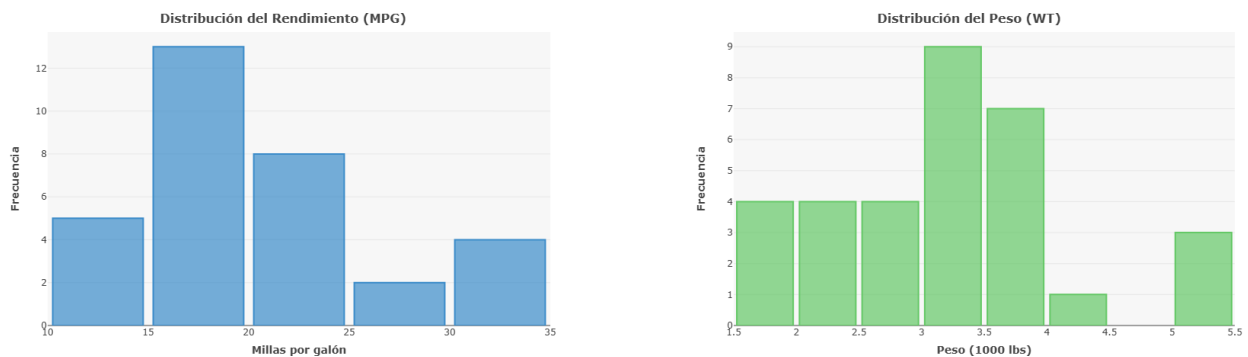


Figura 1: Histogramas de las variables de estudio

Por lo tanto se detallan los hallazgos mostrados en la tabla anterior:

- El rendimiento promedio de los vehículos es de 20.09 mpg con una desviación estándar de 6.03 mpg, lo que evidencia una variabilidad notable en la eficiencia del consumo de combustible.
- El peso promedio de los vehículos es de 3.22 miles de libras, que equivale a (3,220 lbs) con una desviación estándar de 0.98 miles de libras.
- El rendimiento varia entre 10.40 mpg (vehículo menos eficiente) y 33.90 mpg (vehículo más eficiente), reflejando una diferencia superior al triple.
- El rango de peso de los vehículos oscila entre 1.51 miles de libras (vehículo más ligero) y 5.42 miles de libras (vehículo más pesado).
- La escasa diferencia entre la media y la mediana en ambas variables sugiere que sus distribuciones son relativamente simétricas.

2.3 Proceso de Normalidad de las Variables

Consistió en evaluar si las variables se encuentran en una distribución normal, un requisito indispensable para verificar la validez de las inferencias en la regresión lineal. Además, se aplicó una prueba de Shapiro-Wilk Shapiro y Wilk (1965), cuya hipótesis nula define que los datos tienen un origen de una distribución normal.

Cuadro 3: Resultados de la Prueba de Normalidad de Shapiro-Wilk

Variable	Estadístico W	p-valor	Decisión ($\alpha = 0,05$)
Rendimiento (mpg)	0.9475	0.1229	No se rechaza H_0 (Normal)
Peso (1000 lbs)	0.9269	0.0473	Se rechaza H_0 (No normal)

Nota. Nivel de significancia: $\alpha = 0,05$. El estadístico W cercano a 1 indica mayor normalidad.

Dichos resultados dan a indicar que:

- **Variable rendimiento (mpg):** Con $W = 0,9475$ y $p = 0,1229 > 0,05$, se efectua que no se rechaza la hipótesis nula, por lo tanto permite concluir que el rendimiento sigue una distribución medianamente normal.
- **Variable peso (wt):** Con $W = 0,9269$ y $p = 0,0473 < 0,05$, da a entender que se rechaza la hipótesis nula, concluyendo en una ligera desviación de la normalidad.

2.4 Cálculo de la Correlación de Pearson

Este cálculo es importante porque permite determinar si existe una asociación lineal antes de proceder al ajuste del modelo de regresión. Por lo tanto, se le aplicó la prueba de correlación de Pearson con las siguientes hipótesis:

- H_0 : No existe correlación lineal entre peso y rendimiento ($\rho = 0$)
- H_1 : Existe correlación lineal entre peso y rendimiento ($\rho \neq 0$)

Una vez establecido esto, los resultados del anterior análisis de Pearson son:

Cuadro 4: Resultados del Análisis de Correlación de Pearson

Estadístico	Valor
Coefficiente de correlación (r)	$-0,868$
Estadístico t	$-9,559$
Grados de libertad	30
p-valor	$< 0,001$
Intervalo de confianza 95 %	$[-0,934, -0,744]$

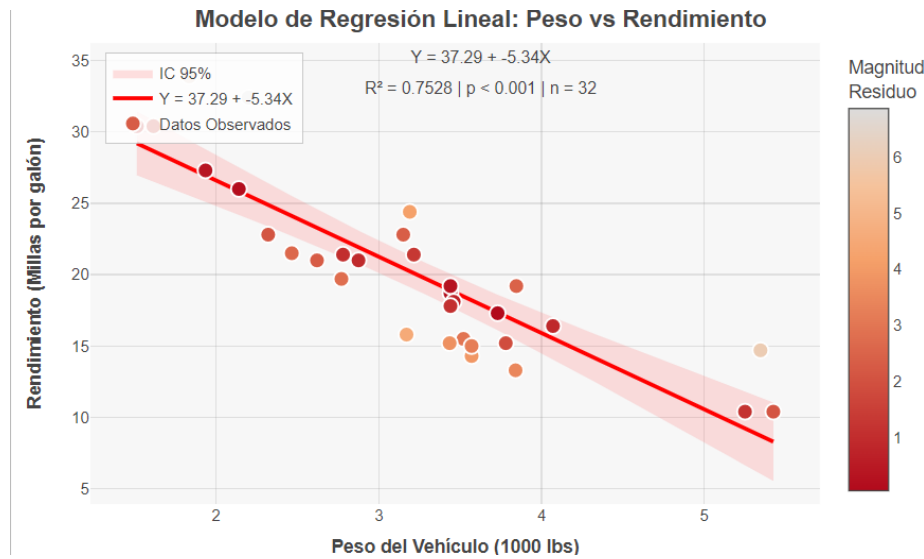


Figura 2: Gráfico de dispersión: Relación entre peso y rendimiento con línea de regresión ajustada

- El coeficiente con un valor de $r = -0,868$ indica una correlación lineal **negativa**, **ademas de fuerte** entre peso del camión y el rendimiento de combustible. Además, el signo negativo afirma que a mayor peso la carga, menor rendimiento en el combustible.
- El p-valor de $< 0,001$ da la capacidad de rechazar la hipótesis nula con un nivel de confianza superior al 99.9%, lo que confirma que la correlación es muy significativa.
- Con un intervalo de confianza al 95 % ($[-0,934, -0,744]$), da un refuerzo de que la correlación es negativa para la población.
- Finalmente, el estadístico da como $t = -9,559$ asignados a 30 grados de libertad indica que es altamente significativo, por lo que da a entender que la correlación observada es muy improbable bajo la hipótesis nula.

2.5 Ajuste del Modelo de Regresión Lineal Simple

Se ajustó el modelo de regresión lineal simple usando un método de mínimos cuadrados ordinarios, que permite estimar los coeficientes β_0 (intercepto) y β_1 (pendiente) minimizando

la suma de los cuadrados de los residuos. Por lo tanto, el modelo ajustado es expresado mediante la siguiente ecuación:

$$\widehat{\text{mpg}} = 37,285 - 5,344 \times \text{wt} \quad (1)$$

2.5.1 Estimación de los Coeficientes

En la siguiente tabla se presenta los coeficientes que se han estimados en el modelo junto a sus errores estandares, así como estadísticos de pruebas y valores p.

Cuadro 5: Coeficientes del Modelo de Regresión Lineal Simple

Coeficiente	Estimación	EE	Estadístico t	p-valor
β_0 (Intercepto)	37.285	1.878	19.858	< 0,001
β_1 (Peso)	-5,344	0.559	-9,559	< 0,001

Nota. EE = Error Estándar. Los estadísticos t se calcularon como: $t = \frac{\hat{\beta}_i}{EE(\hat{\beta}_i)}$.

1. Intercepto ($\beta_0 = 37,285$ mpg):

- Es el valor predicho al rendimiento cuando el peso es cero ($wt = 0$).
- Por lo tanto es estadísticamente significativo ($t = 19,858$, $p < 0,001$).

2. Pendiente ($\beta_1 = -5,344$ mpg por 1000 lbs):

- Estima el cambio esperado en el rendimiento por cada aumento de en las unidades de peso.
- El signo negativo confirma la relación inversa: a mayor peso, menor rendimiento.
- Por lo tanto es altamente significativo ($t = -9,559$, $p < 0,001$), lo que permite rechazar la hipótesis nula $H_0 : \beta_1 = 0$.
- Finalmente el error estándar de 0.559 da a entender que tiene una alta precisión en la estimación del coeficiente.

2.5.2 Evaluación de la Bondad de Ajuste

Para probar la calidad del modelo ya ajustado se calculo multiples metricas de bondad de ajuste:

Cuadro 6: Metricas de Bondad de Ajuste del Modelo

Métrica	Valor
Coefficiente de determinación (R^2)	0.7528
Error estándar residual (RSE)	3.046 mpg
Estadístico F	91.375
Grados de libertad	(1, 30)
p-valor (prueba F)	< 0,001

- $R^2 = 0,7528$: El coeficiente de determinación nos muestra que el 75.28 % de la variabilidad total observada en el rendimiento del combustible puede atribuirse al peso del vehículo. Este resultado refleja un nivel de ajuste sobresaliente para un modelo de regresión lineal simple, lo que confirma que el peso es un predictor muy relevante del rendimiento.
- **Error estándar residual = 3.046 mpg**: Nos indica que las diferencias promedio entre los valores reales y los valores predichos por el modelo son de aproximadamente 3.05 millas por galón. Es decir, las predicciones realizadas presentan una precisión adecuada.
- **Prueba F global**: Con $F_{1,30} = 91,375$ y $p < 0,001$, se rechaza la hipótesis nula, la cual plantea que todos los coeficientes de regresión (excepto el intercepto) son cero. Esto confirma que el modelo en su conjunto es estadísticamente significativo y tiene capacidad predictiva real.

2.5.3 Normalidad de los Residuos

Método de verificación: Prueba de Shapiro-Wilk aplicada a los residuos y análisis del gráfico Q-Q (cuantiles teóricos versus observados).

Cuadro 7: Prueba de Normalidad de los Residuos (Shapiro-Wilk)

Estadístico	Valor
Estadístico W	0.946
p-valor	0.279

Resultado: Dado que el estadístico obtenido fue $W = 0,946$ y $p = 0,279 > 0,05$, no se rechaza la hipótesis nula de normalidad. Además, el gráfico Q-Q mostró que los residuos se ajustan razonablemente bien con la línea diagonal teórica, con desviaciones mínimas en los extremos.

Conclusión: El supuesto de normalidad de los residuos se cumple satisfactoriamente.

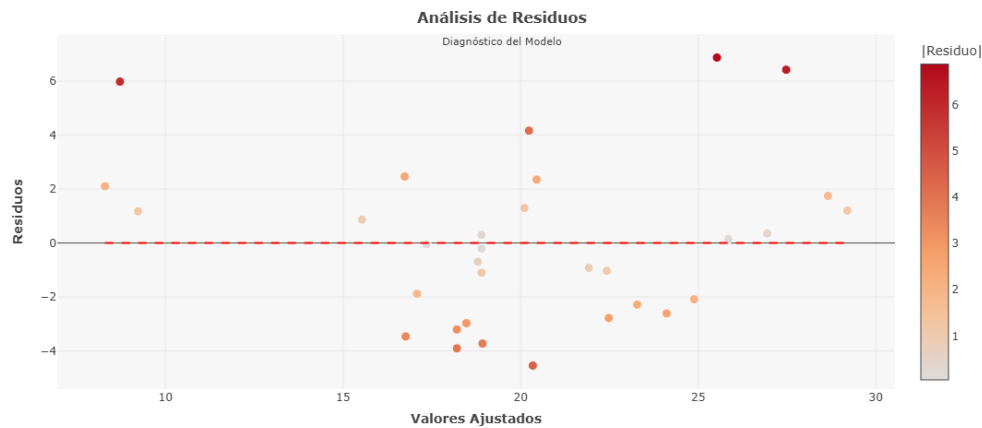


Figura 3: Gráficos de diagnóstico del modelo: Residuos vs. Valores ajustados

2.6 Generación de Predicciones

El séptimo y último paso consistió en aplicar el modelo de regresión validado para estimar el rendimiento del combustible para diferentes valores de peso. Este procedimiento permitió responder a las preguntas prácticas de la investigación y demostrar la utilidad predicha del modelo obtenido.

2.6.1 Metodología de Predicción

Las predicciones se generaron utilizando la ecuación del modelo (Ecuación 1):

$$\widehat{\text{mpg}} = 37,285 - 5,344 \times \text{wt}$$

Se seleccionaron valores de peso que abarcan tanto el rango observado en los datos (1.51 a 5.42 miles de libras) además de incluir una extrapolación al caso teórico de peso cero, con el propósito de responder directamente a una de las preguntas de investigación formuladas.

2.6.2 Resultados de las Predicciones

Cuadro 8: Predicciones de Rendimiento según Peso del Vehículo

Peso (1000 lbs)	Rendimiento Predicho (mpg)	Tipo de Predicción
0.0*	37.29	Extrapolación
1.5	29.27	Interpolación
2.0	26.60	Interpolación
2.5	23.92	Interpolación
3.0	21.25	Interpolación
3.5	18.58	Interpolación
4.0	15.91	Interpolación
4.5	13.24	Interpolación
5.0	10.56	Interpolación

Nota. *El valor de peso = 0 está fuera del rango observado (1.51-5.42 miles de libras) y constituye una extrapolación teórica. Las predicciones de interpolación son más confiables.

2.6.3 Interpretación de las Predicciones

1. **Predicción para peso cero :** Un vehículo hipotético sin carga presentaría un rendimiento predicho de 37.29 mpg. No obstante, esta predicción requiere interpretación cautelosa debido a tres factores:

- Es una extrapolación fuera del rango de datos observados

- No existe un vehículo real con un peso en cero
- La relación lineal podría no mantenerse en valores extremos

Sin embargo, este valor responde teóricamente a la pregunta de investigación planteada y representa el intercepto del modelo.

2. **Tendencia sistemática:** Las predicciones muestran una disminución consistente del rendimiento a medida que aumenta el peso del vehículo, confirmando la relación inversa identificada en el análisis estadístico .

3 Discusion

3.1 Interpretación de los Hallazgos

El presente análisis se llevo a acabo con el único propósito de evaluar de manera estadística y objetiva la relación existente entre el peso de un vehículo y su rendimiento en el consumo de combustible, las variables fueron consideradas fundamentales dentro del ámbito de la ingeniería automotriz y la eficiencia energética.

El peso se eligió por su influencia directa en la resistencia al movimiento y el consumo, mientras que el rendimiento (millas por galón) el cual nos permitió medir de forma cuantitativa la eficiencia del vehículo.

Los resultados del análisis proporcionan evidencia sólida que respalda la hipótesis de que existe una relación lineal significativa entre el peso del vehículo y su rendimiento de combustible. Los principales hallazgos se interpretan de la siguiente manera:

1. **Significancia estadística:** El p-valor inferior a 0.001 permite rechazar la hipótesis nula ($H_0 : \beta_1 = 0$) con un alto nivel de confianza, confirmando que el peso del vehículo es un predictor estadísticamente significativo del rendimiento de combustible.
2. **Magnitud de la relación:** El coeficiente de correlación de Pearson ($r = -0,868$) indica una asociación negativa fuerte entre ambas variables. Esto significa que conforme

aumenta el peso, el rendimiento disminuye de manera consistente y pronunciada.

3. **Capacidad explicativa:** El coeficiente de determinación ($R^2 = 0,7528$) revela que el 75.28 % de la variabilidad observada en el rendimiento puede ser explicada únicamente por el peso del vehículo, lo que constituye un ajuste muy satisfactorio para un modelo de regresión simple.
4. **Interpretación práctica del modelo:** El coeficiente de regresión ($\beta_1 = -5,344$) indica que por cada incremento de 1000 libras en el peso del vehículo, se espera una disminución promedio de 5.34 millas por galón en el rendimiento. Esta cuantificación precisa del efecto tiene implicaciones directas para la eficiencia energética y económica del transporte.

En conjunto, este procedimiento permitió verificar el impacto del peso en la eficiencia del combustible y proporcionar una base analítica útil para futuros estudios sobre transporte eficiente.

3.2 Respuestas a las Preguntas de Investigación

Con base en los resultados obtenidos, se responden las preguntas planteadas inicialmente:

Pregunta 1: ¿Existe una relación estadísticamente significativa entre el peso del vehículo y su rendimiento de combustible?

Respuesta: Sí, existe una relación estadísticamente significativa y negativa. El análisis demuestra de manera concluyente que el peso del vehículo incide negativamente en su rendimiento de combustible ($\beta_1 = -5,344$, $t = -9,559$, $p < 0,001$). La correlación negativa fuerte ($r = -0,868$) confirma que vehículos más pesados presentan sistemáticamente un menor rendimiento de combustible.

Pregunta 2: ¿Cuál sería el rendimiento de combustible predicho para un vehículo sin carga?

Respuesta: Según el modelo ajustado, un vehículo hipotético sin carga (peso = 0) tendría un rendimiento teórico de 37.29 millas por galón. No obstante, esta predicción debe interpretarse con precaución, ya que constituye una extrapolación más allá del rango de valores observados en el dataset (peso mínimo: 1.51 miles de libras). En términos prácticos, esta estimación representa el intercepto del modelo y no necesariamente refleja una condición real alcanzable.

4 Referencias

- Anscombe, F. J. (1973). Graphs in Statistical Analysis. *The American Statistician*, 27(1), 17-21.
- Breusch, T. S., & Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica*, 47(5), 1287-1294.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (3rd). Lawrence Erlbaum Associates.
- Fox, J. (2016). *Applied Regression Analysis and Generalized Linear Models* (3rd). Sage Publications.
- Henderson, H. V., & Velleman, P. F. (1981). Building multiple regression models interactively. *Biometrics*, 37(2), 391-411.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. Springer.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models* (5th). McGraw-Hill/Irwin.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). *Introduction to Linear Regression Analysis* (5th). John Wiley & Sons.
- Pearson, K. (1896). Mathematical Contributions to the Theory of Evolution III: Regression, Heredity and Panmixia. *Philosophical Transactions of the Royal Society of London*, 187, 253-318.
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>
- Rawlings, J. O., Pantula, S. G., & Dickey, D. A. (1998). *Applied Regression Analysis: A Research Tool* (2nd). Springer-Verlag.
- Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3/4), 591-611.