Universiti Teknologi Malaysia

Johor , Skudai , Malaysia.

Probability And Statistical Data Analysis

Project 2

Group Member :

1. Hasan Adittya (A20EC4023)

2. DEIA MELAD MOHAMMED ELGHOUL (A20EC4013)

Introduction

We see many vehicles daily basis on roads of different size and weights. The fuel consumption also varies based on weight of the vehicle. Because more fuel is required to move higher weight. Therefore, it is quite important to know how millage is effected due to a change in weight of the vehicles.

My objective of this research is to compute effect of additional weight on millage of the vehicles. It is known that higher weight leads to fall in millage and vehicles with low weight can travel more millage.

For the above discussion, it is obvious that millage and weight have a relationship. The hypothesis of the study is in the same line of thoughts. I assume that a rise in weight of the vehicle will reduce the millage. Contrary to this, a fall in weight in weight will increase millage. Therefore, I construct my hypothesis as follow;

Null hypothesis: the effect of weight on millage is negative.

Alternative hypothesis: the effect of weight on millage is not negative.

From the study and analysis of the data, I expect negative correlation between millage and weight. The sign coefficient of my weight in regression analysis will be negative. I am expecting following equation;

$$Mpg = a - b*weight + error$$

Intercept will have a positive sign while coefficient of weight is expected to be negative.

Dataset

The dataset is taken from web source. It has different variables but I will use three variables. These are mpg (mile per gallon), weight (weight of the vehicles) and dummy variable for local and foreign made. The mpg is measured in miles while weight is measured in lb.

Data Analysis

In this part I will present statistical analysis of the data. It includes descriptive statistics, correlation analysis, regression analysis and anova. The descriptive analysis is used to explain the variables and understand their values like minimum, maximum, means etc. correlation analysis is used to measure strength of the relationship and understand the direction of the correlation between mpg and weight. The regression analysis is used to compute the effect of the weight on millage. The last analysis of variance (anova) is applied to see the difference between millage of local and foreign made vehicles.

Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Mile per gallon	74	21.297	5.786	12	41
Weight of vehicle	74	3019.459	777.194	1760	4840

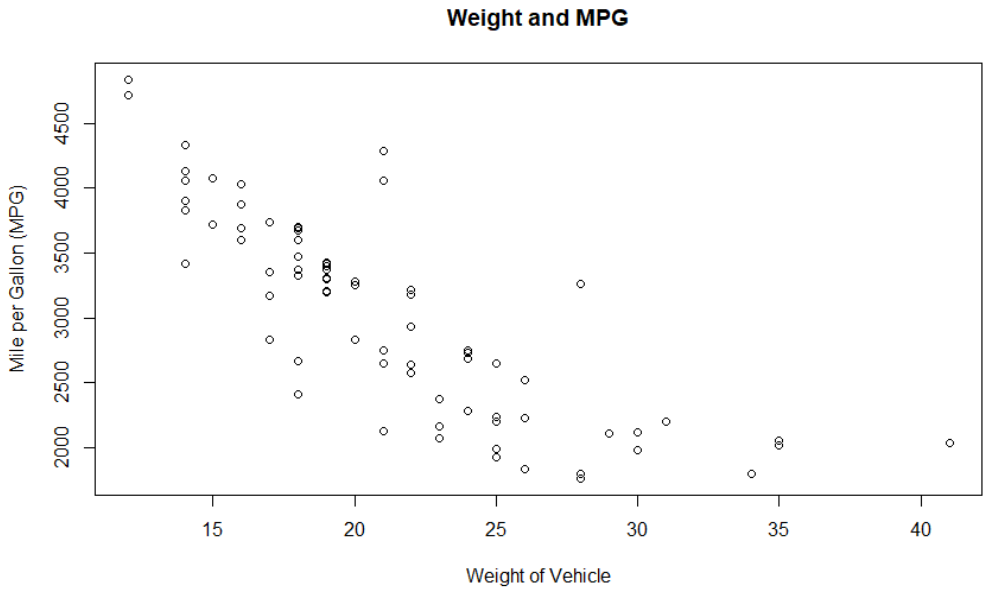
The above table presents summary stats for millage and weight. There are 74 observations in the data of the both variables. Means value of millage is 21.3 and it ranges from 12 to 41 while mean weight is 3019.5lb. Its minimum value is 1760 and highest value is 4840.

Matrix of correlations

Variables	(1)	(2)
(1) mpg	1.000	
(2) weight	-0.807	1.000

Above table presents correlation between millage and weight. It has a negative sign which indicates that there is negative relationship between millage and weight. A rise in weight leads to lower millage and vice versa. The value of the correlation coefficient is 0.807. This means that the there is strong relationship between the two variables.

Scatter Plot



To further expand the analysis, I plotted millage and weight in a scatter plot. Weight is taken on x-axis and millage is plotted on y-axis. The scatter plot indicates negative relation between millage and weight because millage falls when weight rises.

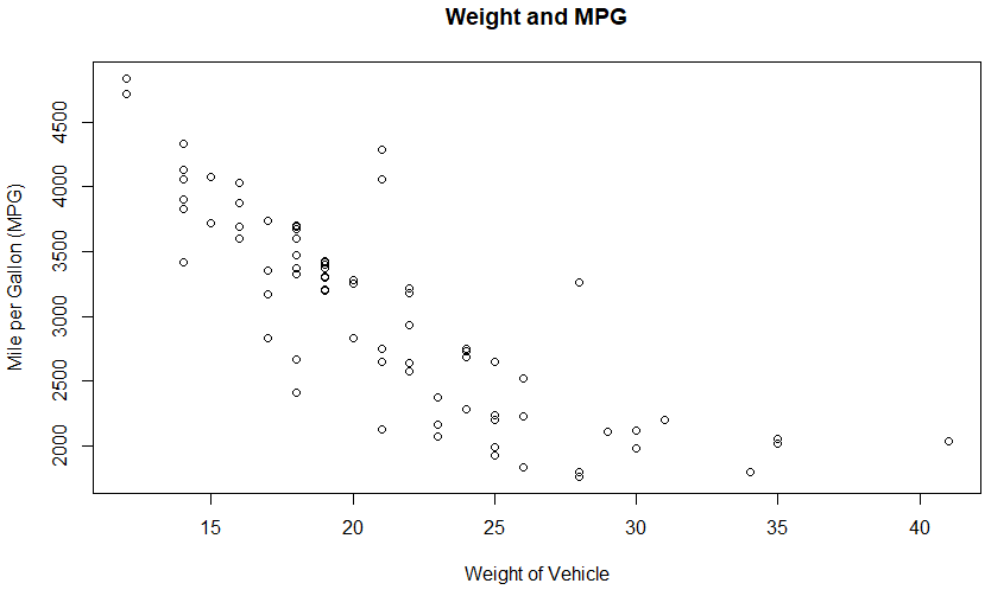
Linear regression

Millage	Coef.	St.Err.	t-value	p-value	Sig
Weight	-.006	.001	-11.60	0.000	***
Constant	39.44	1.614	24.44	0.000	***

R-squared	0.652	Number of obs	74.000
F-test	134.618	Prob > F	0.000

**** p<.01, ** p<.05, * p<.1*

The above table presents regression analysis. The dependent variable is millage and independent variable is weight. This is a simple linear regression model with one independent variable and one dependent variable.



Anova Test

The coefficient of weight has a negative sign as it was expected. This means that with additional weight and the millage will fall. The value of coefficient is 0.006. this means that additional one lb weight reduces millage by 0.006 miles. In other words, if we increase weight by 1000 lbs, the millage will fall by 6 miles. This coefficient is statistically significant at 0.05 level of significant.

The r-squared of the model is 0.65 which is good. This indicates that 65% of the variation in the millage are explained by weight in this model.

Anova results

	Df	Sum sq	Mean sq	F value	Pr
FOREIGN	1	378.2	378.2	13.18	0.0005***
RESIDUALS	72	2065.2	28.7		

In above table I presented results of analysis of variance. This is conducted to test the difference of millage between local and foreign made. The p-value of the test is 0.0005. It is below 0.05 level of significance. Therefore, I conclude that there is a difference of millage.

Conclusion

This study was aimed at computing the effect of weight on millage. The regression analysis confirms our hypothesis. The coefficient of weight has a negative sign. Therefore, I conclude that additional weigh leads to falls in millage. On the other side I tested difference between millage of local and foreign made. The anova test significantly confirmed that there is a difference between the two type of vehicles.

Appendix A

Dataset

Make	Mpg	weight	foreign	Make	mpg	weight	foreign
AMC Concord	22	2930	Domestic	Olds Cutl Supr	19	3310	Domestic
AMC Pacer	17	3350	Domestic	Olds Cutlass	19	3300	Domestic
AMC Spirit	22	2640	Domestic	Olds Delta 88	18	3690	Domestic
Buick Century	20	3250	Domestic	Olds Omega	19	3370	Domestic
Buick Electra	15	4080	Domestic	Olds Starfire	24	2730	Domestic
Buick LeSabre	18	3670	Domestic	Olds Toronado	16	4030	Domestic
Buick Opel	26	2230	Domestic	Plym. Arrow	28	3260	Domestic
Buick Regal	20	3280	Domestic	Plym. Champ	34	1800	Domestic
Buick Riviera	16	3880	Domestic	Plym. Horizon	25	2200	Domestic
Cad. Skylark	19	3400	Domestic	Plym. Sapporo	26	2520	Domestic
Cad. Deville	14	4330	Domestic	Plym. Volare	18	3330	Domestic
Cad. Eldorado	14	3900	Domestic	Pont. Catalina	18	3700	Domestic
Cad. Seville	21	4290	Domestic	Pont. Firebird	18	3470	Domestic
Chev. Chevette	29	2110	Domestic	Pont. Grand Prix	19	3210	Domestic
Chev. Impala	16	3690	Domestic	Pont. Le Mans	19	3200	Domestic
Chev. Malibu	22	3180	Domestic	Pont. Phoenix	19	3420	Domestic

Chev. Monte Carlo	22	3220	Domestic	Pont. Sunbird	24	2690	Domestic
Chev. Monza	24	2750	Domestic	Audi 5000	17	2830	Foreign
Chev. Nova	19	3430	Domestic	Audi Fox	23	2070	Foreign
Dodge Colt	30	2120	Domestic	BMW 320i	25	2650	Foreign
Dodge Diplomat	18	3600	Domestic	Datsun 200	23	2370	Foreign
Dodge Magnum	16	3600	Domestic	Datsun 210	35	2020	Foreign
Dodge St. Regis	17	3740	Domestic	Datsun 510	24	2280	Foreign
Ford Fiesta	28	1800	Domestic	Datsun 810	21	2750	Foreign
Ford Mustang	21	2650	Domestic	Fiat Strada	21	2130	Foreign
Linc. Continental	12	4840	Domestic	Honda Accord	25	2240	Foreign
Linc. Mark V	12	4720	Domestic	Honda Civic	28	1760	Foreign
Linc. Versailles	14	3830	Domestic	Mazda GLC	30	1980	Foreign
Merc. Bobcat	22	2580	Domestic	Peugeot 604	14	3420	Foreign
Merc. Cougar	14	4060	Domestic	Renault Le Car	26	1830	Foreign
Merc. Marquis	15	3720	Domestic	Subaru	35	2050	Foreign
Merc. Monarch	18	3370	Domestic	Toyota Celica	18	2410	Foreign
Merc. XR-7	14	4130	Domestic	Toyota Corolla	31	2200	Foreign
Merc. Zephyr	20	2830	Domestic	Toyota Corona	18	2670	Foreign

				VW			
Olds 98	21	4060	Domestic	Dasher	23	2160	Foreign
VW				VW			
Scirocco	25	1990	Foreign	Diesel	41	2040	Foreign
				VW			
Volvo 260	17	3170	Foreign	Rabbit	25	1930	Foreign