

# MOTOR TRENDS DATA – REGRESSION MODELS PROJECT

## Executive Summary

The data (*mtcars*) that we are about to look at, was extracted from the 1974 Motor Trend US Magazine. It comprises fuel consumption along with 10 attributes and performances (**Appendix 2**) for 32 car models over the two years from 1973-74. We use regression analysis and data analysis to determine whether an automatic or manual transmission car is better for the MPG and quantify the difference. A simple regression model that looks at the MPG yield, ignoring all other variables other than the **am**, that is, the *transmission* of the car gave us an increased MPG of 7.245 if using a manual vs. automatic transmission. However the Adjusted R-squared value was only 0.3385, i.e. a poor fit. The null hypothesis ( $H_0$ ) that both automatic and manual vehicles yield the same MPG also failed.

Several multiple regression fitted models were tested and the conclusion was that a combination of factors/variables such as the displacement, the number of *cylinders*, the *weight* of the car, the *transmission* and the *horsepower* that were the best indicators of the MPG yields overall. The results showed that if the same car such as the Mazda RX4 was used, there was approximately 1.3 MPG increase when using the manual vs. automatic transmissions.

## EXPLORATORY DATA ANALYSIS

Firstly, a standard boxplot also created (Please see **Appendix 1**) to show whether a single variable such as *transmission* could show whether there was a difference in MPG yield between the two transmission types. This clearly indicated that there was a difference and the manual transmission had an edge. To ensure the conclusion was correct, a second test was performed as follows.

A simple regression model was used to determine whether a single variable like the transmission of the vehicles could show whether automatic transmission vehicles were better than manual ones or vice versa. The following was used:

```
amModel <- lm(mpg ~ am, data=mtcars); summary(amModel)
```

At face value, there appears to be a 7.245 increase in MPG yield in cars using manual transmissions over automatic ones. Yet the Adjusted R-squared fit was only 33.85%, indicating a poor fit and proving that a single variable such as *transmission* was not a good predictor of fuel efficiency.

## T-TESTING

Next a T-test and hypothesis testing was used to infer whether the MPG of automatic and manual vehicles was the same, and if other variables played a part in the MPG yield. We set the null hypothesis ( $H_0$ ) as both automatic and manual vehicles yielded the same MPG. The alternate Hypothesis ( $H_1$ ) being that there is a significant enough difference between the two types of transmission used with other variables too playing an important part. The following was used:

```
test <- t.test(mpg ~ am, data=mtcars); test$p.value  
[1] 0.001373638
```

The p-value was **0.0137**, therefore the null hypothesis  $H_0$  was rejected.

## SELECTING THE VARIABLES TO USE

```
all <- lm(mpg~., data=mtcars); summary(aov(all))
```

This showed us that the *cyl*, *disp*, *hp*, *drat*, *wt* and *am* were the most likely variables to use. We then tested how the *cyl*, *disp*, *hp*, *drat*, *wt* and *am* variables impact the MPG yield

## FITTING MULTIPLE REGRESSION MODELS

The last part of the testing involved fitting regression models using the 6 variables that could have an impact on the MPG yield. This is done by fitting multiple regression models:

- `model1 <- lm(mpg ~ cyl, data = mtcars)` Adjusted R-squared: 0.7171
- `model2 <- lm(mpg ~ wt, data = mtcars)` Adjusted R-squared: 0.7446
- `model3 <- lm(mpg ~ cyl + wt, data = mtcars)` Adjusted R-squared: **0.8185**
- `model4 <- lm(mpg ~ disp, data = mtcars)` Adjusted R-squared: 0.7090
- `model5 <- lm(mpg ~ disp + cyl, data = mtcars)` Adjusted R-squared: 0.7430
- `model6 <- lm(mpg ~ disp + cyl + wt, data = mtcars)` Adjusted R-squared: **0.8147**
- `model7 <- lm(disp + cyl + wt + am, data = mtcars)` Adjusted R-squared: 0.8079
- `model8 <- lm(disp + cyl + wt + am + hp, data = mtcars)` Adjusted R-squared: **0.8273**
- `model9 <- lm(disp + cyl + wt + am + drat, data = mtcars)` Adjusted R-squared: **0.8211**

The results of the above fitted models indicated the following:

The strongest fit from the above models were model3, model6, model7, model8 and model9. The two models had the strongest fit thereby indicating a strong correlation to the MPG however, were models **8** and **9**. However adding drat, that is, rear axle ratio did not add that much value to the argument as it pulled the Adjusted R-squared down. When hp was added (but drat excluded), the model had the strongest explanation of MPG yields.

Therefore, the overall results lead us to the conclusion that all other factors remaining the same, the displacement, the number of cylinders, the weight of the car, the transmission and the horsepower combined were the best indicators of the MPG yield of a vehicle, not the type of transmission type on its own.

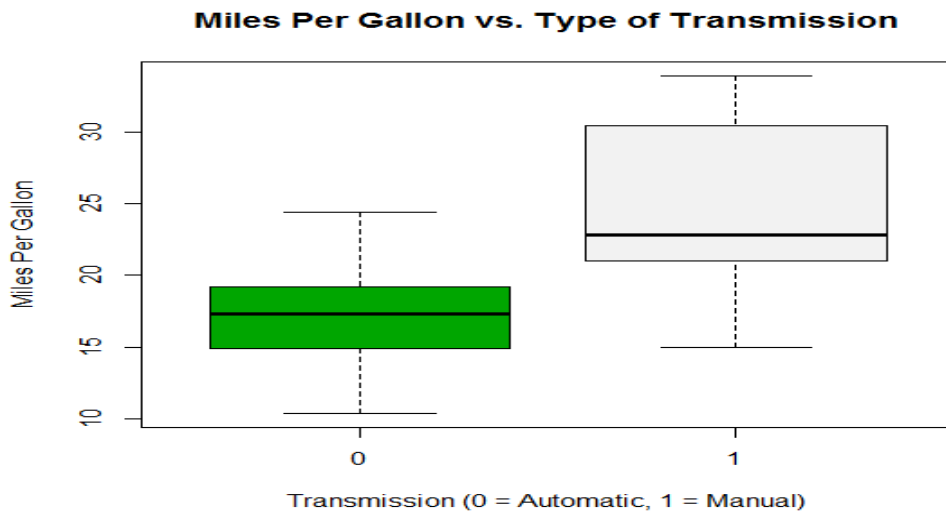
If a standard Mazda RX4 is used as an example, using model8, using the data in **Appendix 3**,

- $21 + 0.1225708(160) + (-1.106379846) + (-3.302623012.620) + (1.556491631) + (-0.0276002110)$

gives us 22.3 vs. 21.0 MPG using a manual transmission over an automatic one.

## Appendix 1

Boxplot of MPG vs Transmission



```
boxplot(mpg ~ am, xlab="Transmission (0 = Automatic, 1 = Manual)",  
ylab="Miles Per Gallon", col =terrain.colors(2))  
title(main="Miles Per Gallon vs. Type of Transmission")
```

## Appendix 2

List of cars and attributes

	mpg	cyl	displacement	horsepower	drat	weight	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.82	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

```
data("mtcars"); mtcars
```

## Appendix 3

### Model 8 evaluation

```
model8 <- lm(mpg ~ disp + cyl + wt + am + hp, data = mtcars)
summary(model8)$coef
```

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	38.20279869	3.66909647	10.412045	9.084987e-11
disp	0.01225708	0.01170645	1.047036	3.047194e-01
cyl	-1.10637984	0.67635506	-1.635797	1.139322e-01
wt	-3.30262301	1.13364263	-2.913284	7.256888e-03
am	1.55649163	1.44053603	1.080495	2.898430e-01
hp	-0.02796002	0.01392172	-2.008374	5.509659e-02

```
model8 <- lm(mpg ~ disp + cyl + wt + am + hp, data = mtcars)
summary(model8)$coef
```