

Peer-graded Assignment: Regression Models Project

Nikhil Kalyankar

June 25, 2017

Executive Summary

Miles per gallon (*mpg*) is an important criteria for assessing and rating automobiles. It depends of various factors: fuel quality, weight, number of cylinders, engine efficiency, transmission efficiency, driver ability, road conditions, etc. The analysis presented in this report show that *manual transmission* is better than *automatic transmission* for *mpg*. However, it is not possible to quantify the relationship.

Exploratory Analysis

- **Describe the dataset:** The **mtcars** dataset of contains information on various cars set on 11 variables namely: **mpg** (Miles per gallon), **cyl** (Number of cylinders), **disp** (Displacement cubic inches), **hp** (Gross horsepower), **drat** Rear axle ration, **wt** (Weight 1000lbs), **qsec** (1/4 mile time), **vs** (V/S), **am** (Transmission; 0 = automatic; 1 = manual), **gear** (Number of forward gears) and **carb** (Number of carburetors).
- **Identify the key variables:** This report identifies the **mpg** and **am** variables as the key for the purpose of this report. Other variables particularly *cyl* and *gear* would be considered for additional analysis.
- **Explore the dataset:** Firstly, the mean **mpg** when grouped by *automatic* and *manual* transmission is **17.14737** and **24.39231** respectively. Secondly, the average **mpg** when grouped by *cyl* and *am* shows that the *mpg* values are **higher** for *manual* transmission (Note: The values are very close in case of 8 cylinder automobiles). The other variables related to *cyl* i.e., *wt*, *disp* and *hp* show similar results. Finally, the average **mpg** when grouped by *gear* and *am* shows that only **4 gear** automobiles have both types of transmission and again the *manual* transmission automobiles give **higher mpg**. We can consider that on an average **manual** transmission is **better** than **automatic** transmission.

Regression Analysis

The regression analysis with outcome variable **mpg** and predictor variable **am** by excluding the intercept term gives us the coefficients for automatic and manual as **17.14737** and **24.39231** respectively. This show that manual transmission gives higher *mpg*. Note: these values are same as the mean *mpg* when grouped by *am*. The next analysis with **wt** as an addtional predictor variable changes the *am* coefficients to **37.321551** and **37.297936** respectively. It can be inferred that *wt* has significant impact on *mpg* that removes the effect of the type of transmission. The regression analysis ran on all variables by excluding the intercept term: $\text{lm}(\text{mpg} \sim . - 1, \text{data} = \text{mtcars})$ show that when all variables are considered together they do not have statistically significant impact. S

Is an automatic or manual transmission better for MPG?

A manual transmission is surely better than an automatic transmission for MPG.

Quantifying the MPG difference between automatic and manual transmissions.

It is impossible to quantify the MPG difference between automatic and manual transmission as we have only two possible outcomes as the predictor variable can only provide two values. Additional variables are required for quantification.

Appendix

1.1. Dataset Overview

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

1.2. Dataset Summary

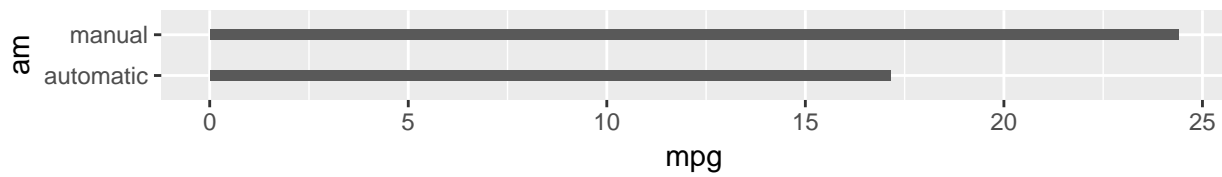
mpg	Min. :10.40	1st Qu.:15.43	Median :19.20	Mean :20.09	3rd Qu.:22.80	Max. :33.90
cyl	Min. :4.000	1st Qu.:4.000	Median :6.000	Mean :6.188	3rd Qu.:8.000	Max. :8.000
disp	Min. : 71.1	1st Qu.:120.8	Median :196.3	Mean :230.7	3rd Qu.:326.0	Max. :472.0
hp	Min. : 52.0	1st Qu.: 96.5	Median :123.0	Mean :146.7	3rd Qu.:180.0	Max. :335.0
drat	Min. :2.760	1st Qu.:3.080	Median :3.695	Mean :3.597	3rd Qu.:3.920	Max. :4.930
wt	Min. :1.513	1st Qu.:2.581	Median :3.325	Mean :3.217	3rd Qu.:3.610	Max. :5.424
qsec	Min. :14.50	1st Qu.:16.89	Median :17.71	Mean :17.85	3rd Qu.:18.90	Max. :22.90
vs	Min. :0.0000	1st Qu.:0.0000	Median :0.0000	Mean :0.4375	3rd Qu.:1.0000	Max. :1.0000
am	Min. :0.0000	1st Qu.:0.0000	Median :0.0000	Mean :0.4062	3rd Qu.:1.0000	Max. :1.0000
gear	Min. :3.000	1st Qu.:3.000	Median :4.000	Mean :3.688	3rd Qu.:4.000	Max. :5.000
carb	Min. :1.000	1st Qu.:2.000	Median :2.000	Mean :2.812	3rd Qu.:4.000	Max. :8.000

1.3. Summary of factorized *am*

	automatic	manual
am	19	13

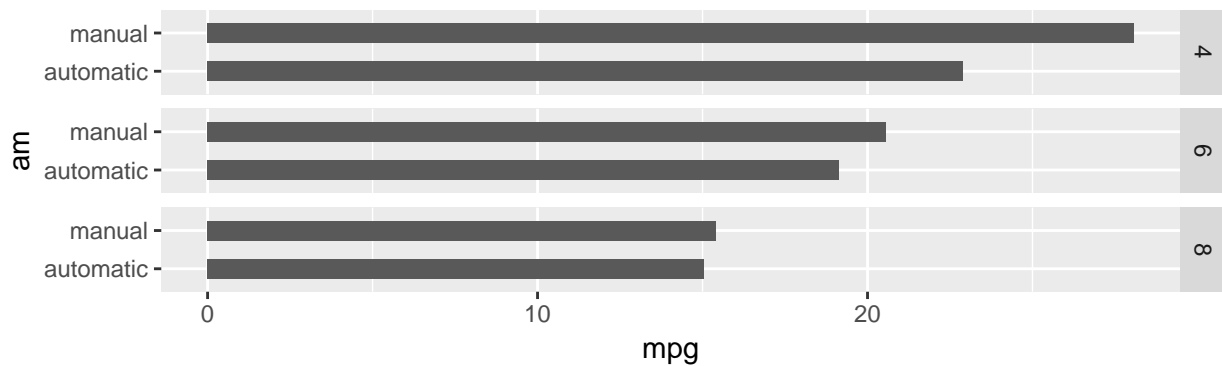
2. Average *mpg* grouped by *am*

am	mpg
automatic	17.14737
manual	24.39231



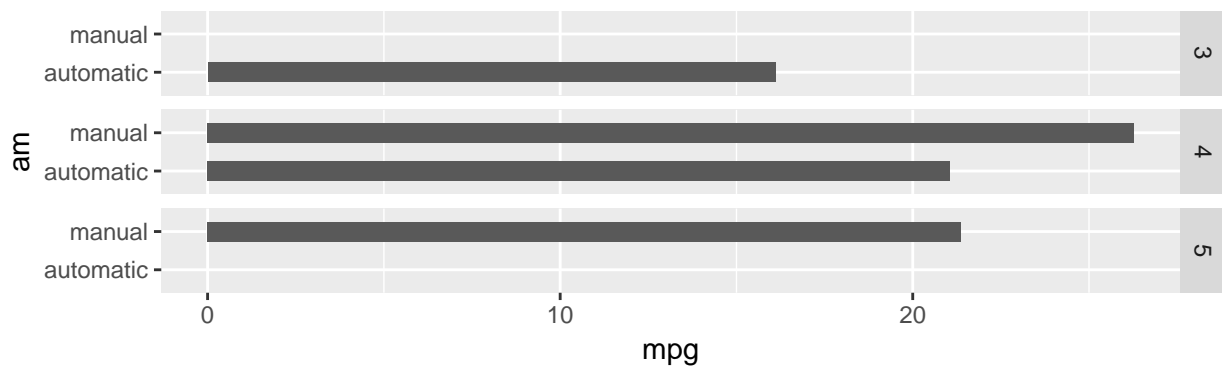
3. Average *mpg* grouped by *cyl* and *am*

am	cyl	mpg
automatic	4	22.90000
automatic	6	19.12500
automatic	8	15.05000
manual	4	28.07500
manual	6	20.56667
manual	8	15.40000



4. Average *mpg* grouped by *gear* and *am*

am	gear	mpg
automatic	3	16.10667
automatic	4	21.05000
manual	4	26.27500
manual	5	21.38000



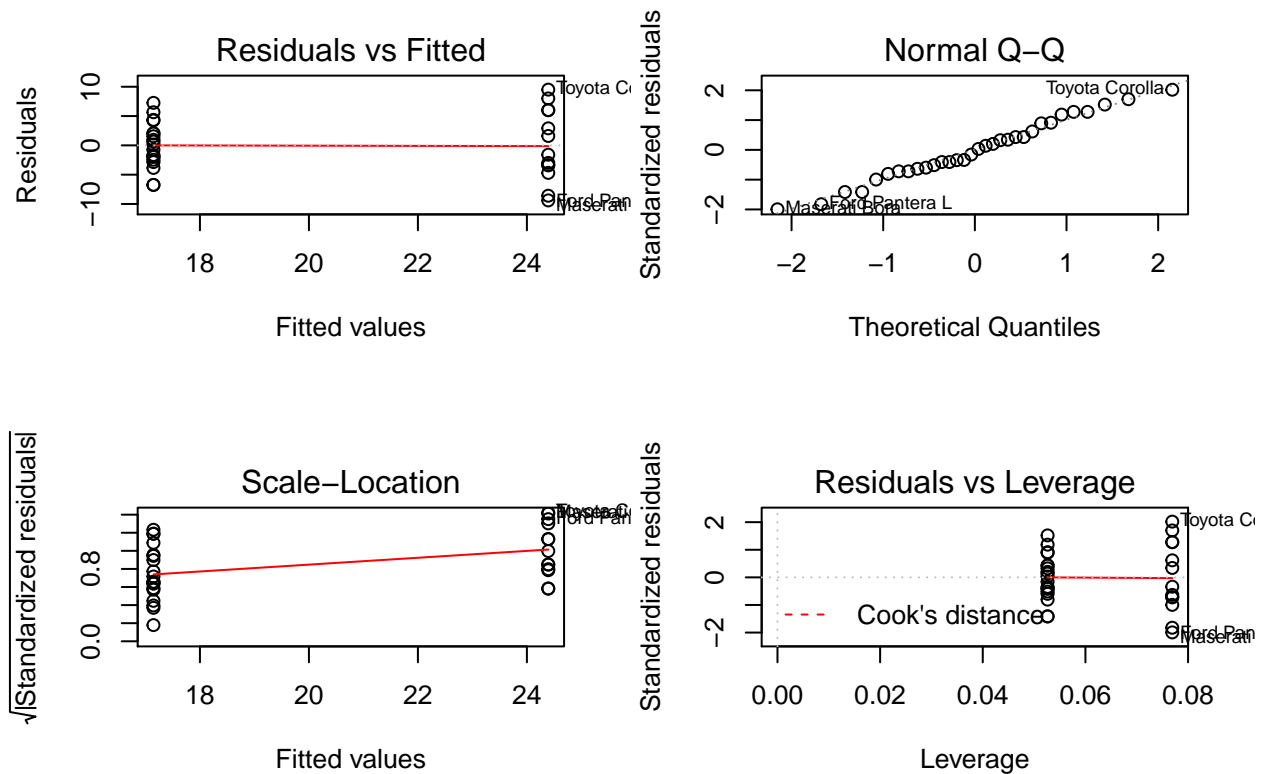
5.1. `lm(mpg ~ am - 1, data = mtcars)`

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Residuals	-9.392308	-3.092308	-0.2973684	0	3.243927	9.507692

	Estimate	Std. Error	t value	Pr(> t)
amautomatic	17.14737	1.124602	15.24749	0
ammanual	24.39231	1.359578	17.94109	0

value	numdf	dendf
277.1843	2	30

label	R-squared	Adjusted R-squared	Sigma
value	0.9486625	0.9452400	4.9020288



5.2. Prediction

Newdata	Predicted_Value
automatic	17.14737
manual	24.39231

6. `lm(mpg ~ am + wt - 1, data = mtcars)`

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Residuals	-4.529524	-2.361895	-0.1317223	0	1.402517	6.878249

	Estimate	Std. Error	t value	Pr(> t)
amautomatic	37.321551	3.0546385	12.217993	0e+00
ammanual	37.297936	2.0856607	17.883032	0e+00
wt	-5.352811	0.7882438	-6.790807	2e-07

value	numdf	dendf
478.0542	3	29

label	R-squared	Adjusted R-squared	Sigma
value	0.9801799	0.9781296	3.0979398

7. `lm(mpg ~ . - 1, data = mtcars)`

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Residuals	-3.450644	-1.604402	-0.1196051	0	1.219268	4.627094

	Estimate	Std. Error	t value	Pr(> t)
cyl	-0.1114405	1.0450234	-0.1066392	0.9160874
disp	0.0133352	0.0178575	0.7467585	0.4634887
hp	-0.0214821	0.0217686	-0.9868407	0.3349553
drat	0.7871110	1.6353731	0.4813036	0.6352779
wt	-3.7153039	1.8944143	-1.9611887	0.0632522
qsec	0.8210407	0.7308448	1.1234133	0.2739413
vs	0.3177628	2.1045086	0.1509915	0.8814235
amautomatic	12.3033742	18.7178844	0.6573058	0.5181244
ammanual	14.8236010	18.3526517	0.8077090	0.4283130
gear	0.6554130	1.4932600	0.4389142	0.6652064
carb	-0.1994193	0.8287525	-0.2406258	0.8121787

value	numdf	dendf
179.8472	11	21

label	R-squared	Adjusted R-squared	Sigma
value	0.9894964	0.9839946	2.6501970