

Mileage: Automatic or Manual

Kaung Myat Khant

2025-03-25

Executive Summary

The mtcars data is used to determine whether an automatic or manual transmission is better for MPG. Multivariate linear regression is applied to predict the average change in miles per gallon for being a manual transmission. Two models are fitted. One model include with transmission as the only regressor and the other model included transmission, engine shape, number of cylinder, engine displacement and weight of the automobile. Analysis of variance is used to test the model significance and variance inflation factor is used to assess multi-collinearity. Assumptions of linear regression are checked and outliers inflencing the models are identified using DFbetas, leverage and Cook's distance. The model showed that manual transmission has an average mileage that is 0.6 miles per gallon greater than the automatic transmission. But we can not conclude that manual transmission has greater mileage than automatic transmission regarding the inferential statistics.

Introduction

The 1974 *Motor Trend* US magazine surveyed 32 automobiles, which are produced in 1973-74. The survey included fuel consumption and 10 aspects of automobiles: number of cylinders, displacement in cubic inches, gross horsepower, rear axle ratio, weight(1000lbs per unit), quarter mile time, engine shape(V-shape or straight), gear transmission (automatic or manual), number of forward gears and number of carburetors. This data is used **to determine whether the automatic or manual transmission is better for mileage and to quantify their difference in miles per gallon.**

Method

All the variables in the data frame are numeric (double) format. Some variables need to be transformed into factor before the analysis. There is no missing values (NA) in the data. Number of cylinders, engine shape, transmission, number of forward gears and number of carburetors are factorized.

Some exploratory data analysis is carried out to see the realtionship between the mileage and the other aspects of automobiles.

To select a model to explain the mileage difference between the automatic and manual transmission, the transmission variable alone is first fitted with the mileage. Next, other covariates identified from the exploratory data analysis are added to the model. These includes engine shape, number of cylinder, engine displacement and weight of the automobile. Then, analysis of variance is performed to test the bias introduced by under-fitting. The model with five regressors is a better fit than the transmission alone to predict the mileage.

However, variance inflation factors for each variables are assessed for multi-collinearity leading to over-fitting. Engine displacement has a high variance inflation with nearly 13 times. So, the engine displacement variable is omitted in the another model and the new model is compared with the previous one using `anova()`

function. The test results showed there is no actual difference between including the displacement and excluding it. Therefore, the model with transmission, engine shape, number of cylinder and weight of the automobile.

After that, the assumptions of the regression models are checked and the influential points are identified (**Annex**). The QQ-Residual plots showed that the data are residual are normally distributed. The residual vs fitted plot showed that residuals are spread symmetrically around zero. There is neither increasing nor decreasing of the residual values along with increasing fitted values. Thus, there seems to be homogeneity of variance.

The plots showed that Toyota Corolla and Toyota Corona are outliers. DF-betas, leverage values and Cook's distance are used to assess the outliers and influential points. And the values showed these car models did not have much leverage and influence on the fitted line.

```
fit1 <- lm(mpg ~ am, data = mtcars)
summary(fit1)
fit5 <- lm(mpg ~ am + vs + cyl + disp + wt, data = mtcars)
summary(fit5)
anova(fit1, fit5)
```

```
car::vif(fit5)
fit4 <- lm(mpg ~ am + vs + cyl + wt, data = mtcars)
summary(fit4)
anova(fit4, fit5)
car::vif(fit4)
```

Results

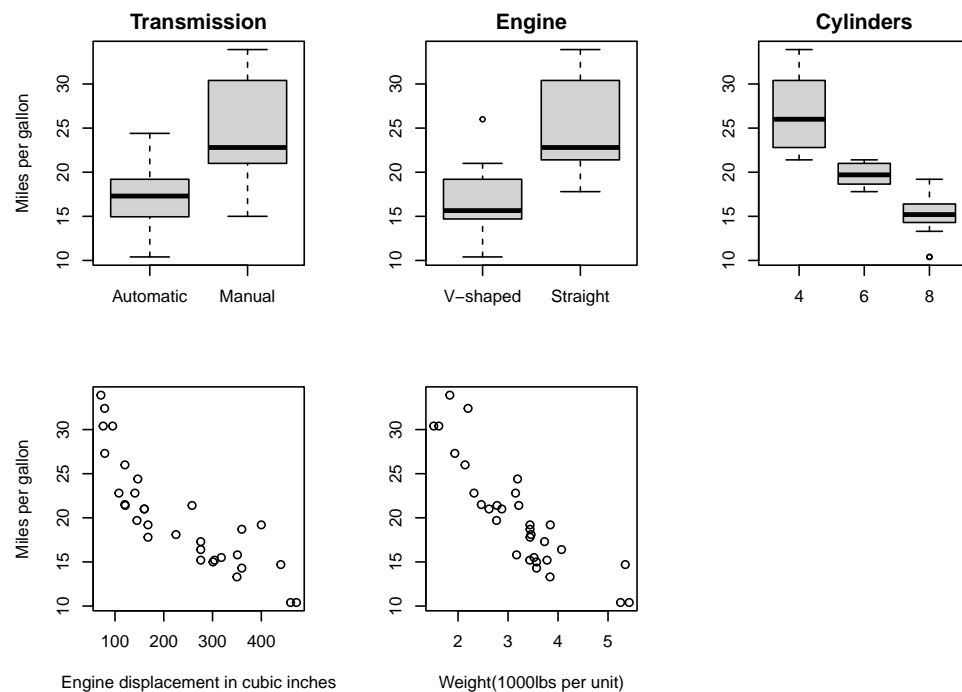


Figure 1: Factors influencing mileage

Table 1: Multivariate regression to predict the mileage from transmission of automobile

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	32.181	3.729	8.631	0.000	24.517	39.845
amManual	0.623	1.501	0.415	0.681	-2.463	3.709
vsStraight	1.242	1.904	0.652	0.520	-2.672	5.156
cyl6	-3.733	1.638	-2.280	0.031	-7.099	-0.367
cyl8	-4.748	2.657	-1.787	0.086	-10.210	0.714
wt	-3.106	0.920	-3.375	0.002	-4.998	-1.214

The table showed that the model is a good fit and can explain 84% of variation in the miles per gallon. And the automobiles with manual transmission had a little bit higher mileage than the automobiles with automatic transmission. The average mileage of automobiles with manual transmission is 0.62 mpg greater than that of the automobiles with automatic transmission, holding the engine shape, number of cylinders and weight of the automobiles constant. However, the t-values suggested that there is weak evidence to reject the null hypothesis of the average difference of zero (P-value 0.6814).

Conclusion

The fitted model to predict the miles per gallon using transmission, engine shape, number of cylinders and weight of the automobiles is a good fit. However, we cannot conclude that manual transmission has higher mileage than automatic transmission since there is no significant difference in average miles per gallon.

Annex: Diagnostic plots

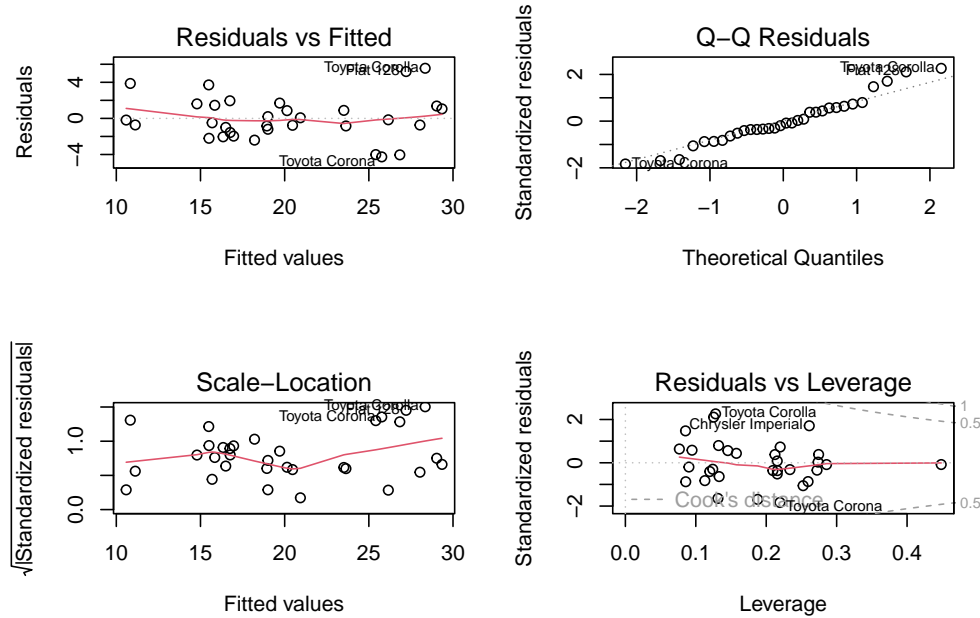


Figure 2: Regression Diagnostic plots

Dfbetas

```
##              (Intercept)  amManual vsStraight      cyl6      cyl8
## Toyota Corolla  0.03840397 0.2867938  0.3430191 -0.08102151 0.1973432
## Toyota Corona  -0.67009594 0.7677764  0.2842099  0.49726000 0.4529384
##              wt
## Toyota Corolla -0.1980967
## Toyota Corona  0.3071057
```

Toyota Corona has high df-betas.

Leverage

```
## Toyota Corolla  Toyota Corona
##      0.1280776      0.2194430
```

But hat values are not very high.

Cook's distance

```
## Toyota Corolla  Toyota Corona
##      0.1250681      0.1578109
```

Cook's distances are also small.