Regression Models Course Project

Sebastien Plat

Executive Summary

This document aims to answer the following two questions relative to automotive fuel consumption:

- "Is an automatic or manual transmission better for MPG"
- "Quantify the MPG difference between automatic and manual transmissions"

The analysis will based on data extracted from the 1974 Motor Trend US magazine. This data comprises fuel consumption and 10 aspects of automobile design & performance for 32 automobiles (1973-74 models).

We will see that for the available data, **MPG is higher with manual transmission**. But the sample size is **very small** and crucial parameters like Weight and Horse Power are **very different** (manual transmission models are lighter and with less HP), so the results are probably biased. Additionaly, predictions made using the model show that the **uncertainty is too great to draw reliable conclusions**.

Exploratory Analysis

We start the analysis by trying to find interesting relations between mpg and other available parameters. As shown Fig. 1:

- mpg seems highly correlated with weigth, HP and quarter mile time
- manual transmission cars have a higher average mpg
- but on average, they are lighter and with less HP

The last two points are illustrated Table 1:

Table 1: Comparizon Auto vs Manual Transmissions

am	avgMPG	sdMPG	avgWeight	sdWeight	avgHP	sdHP
auto	17.1	3.8	3.8	0.8	160.3	53.9
manual	24.4	6.2	2.4	0.6	126.8	84.1

The fact the the two sets of cars are so different could be problematic as it could induce a bias. But let's build some linear models based on point 1:

```
fit1 <- lm (mpg ~ am * wt, data = mtcars)
fit2 <- lm (mpg ~ am * (wt + hp), data = mtcars)
fit3 <- lm (mpg ~ am * (wt + hp + qsec), data = mtcars)
```

Model Fit

The ANOVA table shows that **fit2 and fit3 improve the model** (the hypothesis of normal residuals is validated by Shapiro-Wilk tests in all three cases).

Table 2: Analysis of Variance Table

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
28	188	NA	NA	NA	NA
26	135.9	2	52.1	6	0.008
24	104.7	2	31.2	3.6	0.044

We are not sure how to explain the relationship between quarter mile time and fuel consumption, so we will continue with the **fit2 model**, with $R^2 = 0.8793$.

Fig. 2 shows that:

- errors are **symetrically distributed** around mean 0 and without heteroskedasticity
- errors are **roughly normal**, as confirmed by the Shapiro-Wilk test
- there are some **outliers**, mainly the **Maserati Bora**: the only manual car with more than 300 HP

Estimates

Let's compare the predicted fuel consumption for auto VS manual in two cases:

- test1: average weight & HP based on auto samples
- test2: average weight & HP based on manual samples

test	hp	wt	am	fit	lwr	upr	predict.se
test1	160.3	3.8	auto	17.1	12.3	21.9	2.3
test1	160.3	3.8	manual	13.4	6.7	20	3.2
test2	126.8	2.4	auto	21.1	15.8	26.3	2.6
test2	126.8	2.4	manual	24.5	19.6	29.4	2.4

Note: the 95% Confidence Intervals are built with a t-distribution of 26 degrees of freedom (see SO question on prediction intervals and on r calculations).

The 95% Confidence Intervals overlap significantly (as shown Fig. 3). t-tests with `qt(0.975,52)` fail to reject the null hypothesis that their means are statistically different (more details here and here):

	lwr	upr	CI with 0
test1	-4.3	11.8	Yes
test2	-10.4	3.6	Yes

These results show that there are **no reliable answers** to the questions as the **uncertainty is too large** with such a small sample of cars (an illustration of this uncertainty is shown Fig. 4 for the model fit1 which is easier to plot - only 2D).

Appendix

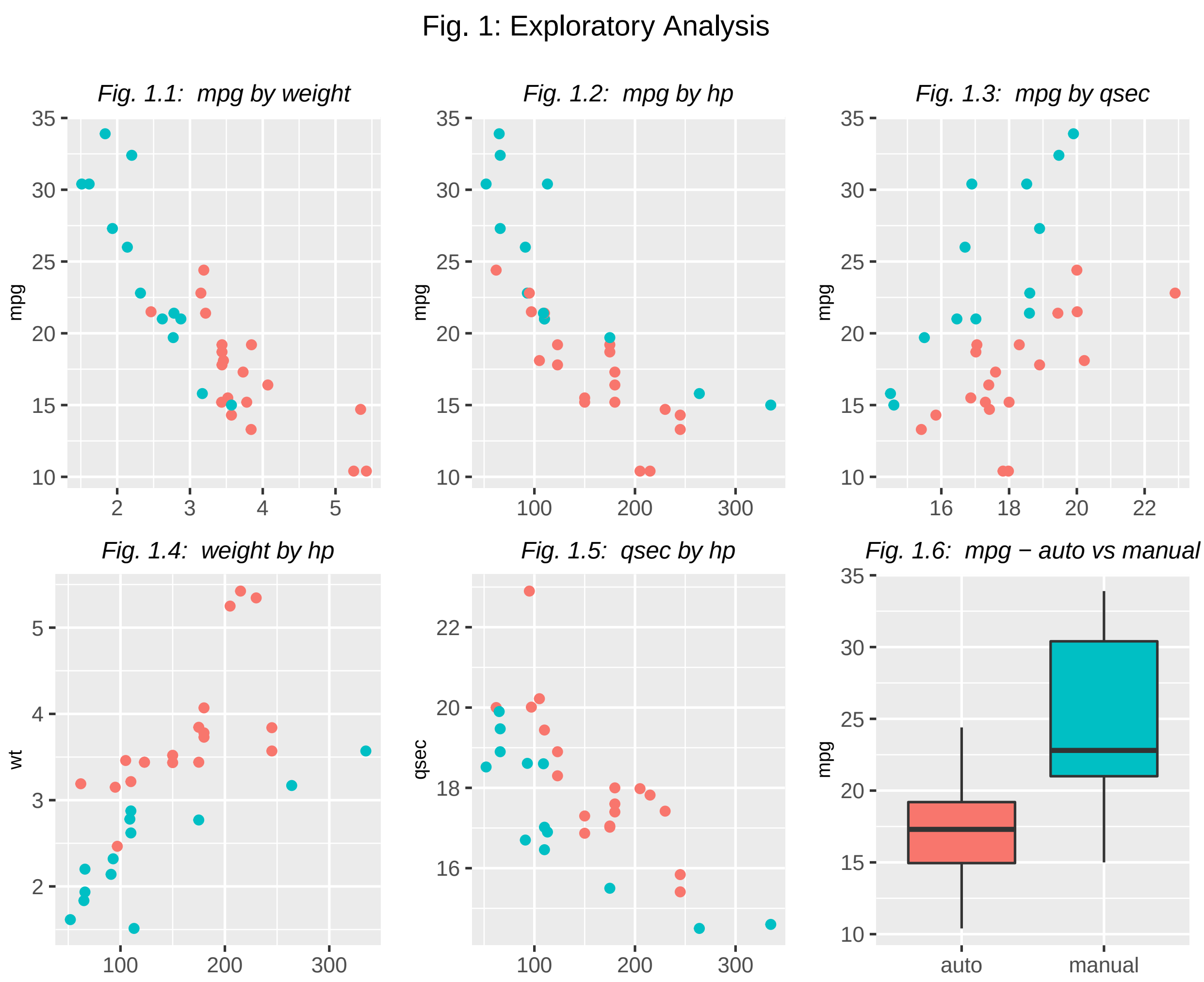


Fig. 1: Exploratory Analysis
Fig. 1.1: mpg by weight
Fig. 1.2: mpg by hp
Fig. 1.3: mpg by qsec
Fig. 1.4: weight by hp
Fig. 1.5: qsec by hp
Fig. 1.6: mpg − auto vs manual
mpg
wt
qsec
auto
manual

Fig. 2: Fit2 Model Diagnostics

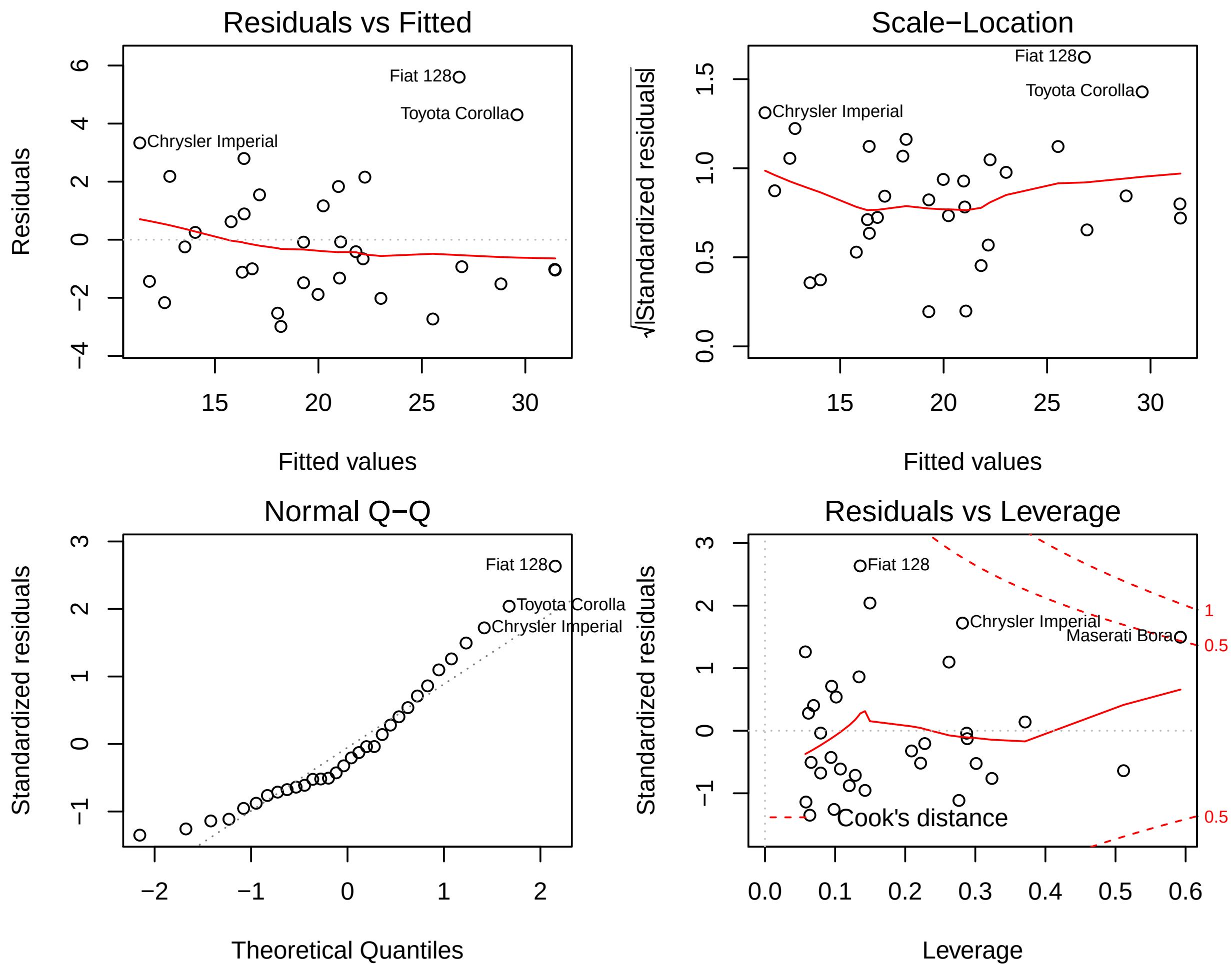
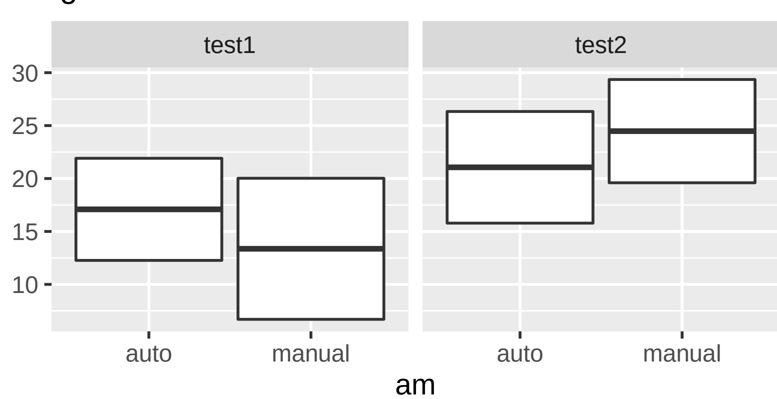


Fig. 3: 95% Confidence Interval for MPG Prediction



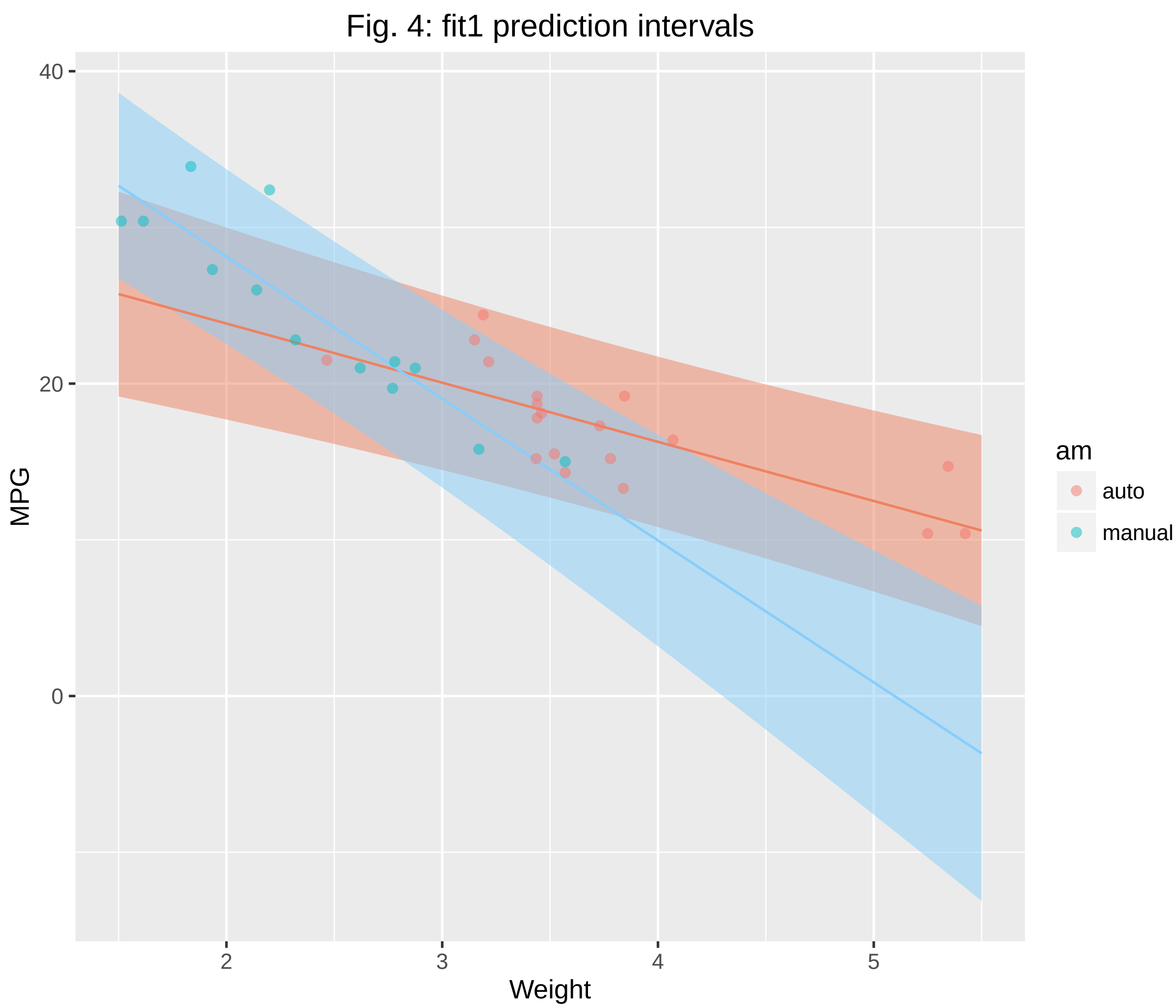


Fig. 4: fit1 prediction intervals
40
20
0
MPG
2
3
4
5
Weight
am
auto
manual