

Regression Models Course Project Report

Martin Hediger

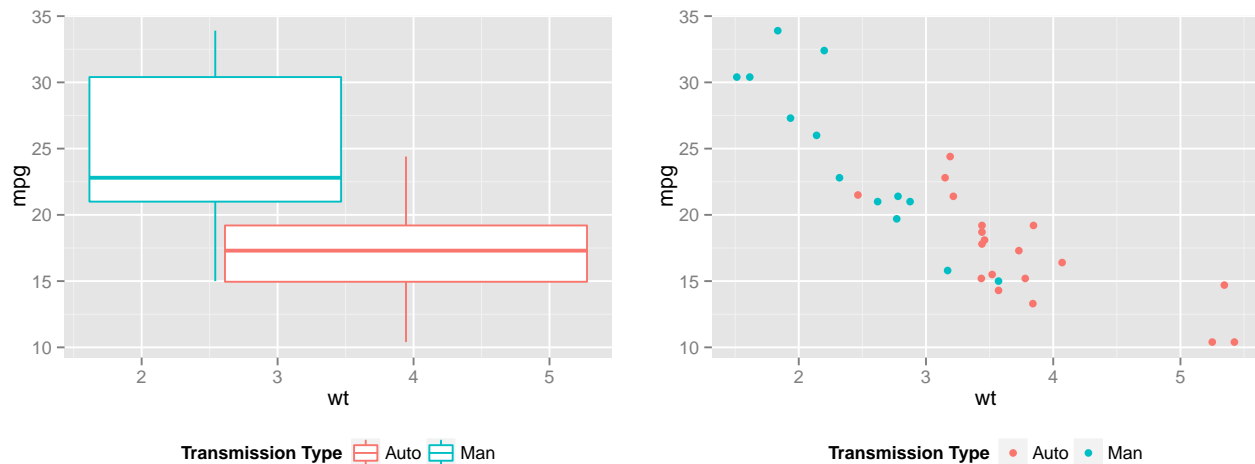
Executive Summary

This short report analyses if 1) automatic or manual transmission results in larger miles-per-gallon (mpg) and 2) quantifies the effect of manual or automatic transmission on mpg in the cars from the `mtcars` dataset. Automatic transmission vehicles appear to be less efficient than manual transmission vehicles (Fig. “box-scat” A). A linear relationship between miles-per-gallon range and car weight for both automatic and manual transmission vehicles is found (Fig. “box-scat” B). Inference of a linear model of the form $mpg_i = \beta_0 + \gamma \cdot (am_i \cdot wt) + \beta_1 \cdot wt_i + e_i$ for the dummy variable am_i (0: auto, 1: man) indicates that the dependence of mpg on wt is not different for the cars with automatic and manual transmission.

Exploratory Analysis

As an initial test, dependence of `mpg` on `wt` is analysed.

box-scat A: Boxplot, B: Scatterplot of `mpg` against `wt`.



[Source code for the panel plot.](#)

According to the boxplot, automatic cars have lower MPG (and possibly lower variance in the data). Importantly, the relationships appear linear and no outliers which could affect correlation values are identified. The only aspect which is slightly problematic is the limited dataset size ($n=32$).

It is noted that apparently most cars with automatic transmission also are heavier which possibly confounds the observation, this would be subject to further research.

Results

Question 1: Which transmission type has higher MPG?

Based on the exploratory analysis above, it is found that on average the difference between $MPG(auto)$ and $MPG(man)$ is around 7.2449 miles-per-gallon ($= \text{mean}(\text{mtcars}[\text{mtcars}\$am==1, "mpg"]) - \text{mean}(\text{mtcars}[\text{mtcars}\$am==0, "mpg"])$).

Question 2: Quantification of MPG-difference between transmission types

MODEL 1

In the full model `lm(mpg ~ . + factor(am), data=mtcars)` there is no strong evidence against the null hypothesis for any variable, i.e. that $H_{0,i} \neq 0$ where $i \in \{cyl, disp, hp, drat, wt, qsec, vs, am, gear, carb\}$, that is all p-values are larger than any accepted significance level.

Theory suggests that the smallest reasonable model should be employed, so based on the collectively large p-values for the different variables in the full model, the variables **wt**, **am**, **qsec** and **hp** are considered for a reduced model because of all variables, these have lowest p-values.

MODEL 2

It is tested if miles-per-gallon are sufficiently explained by **wt**, **am**, **qsec** and **hp**.

A residual against fitted values plot (appendix) shows uniform distribution and a Cook's distance plot confirms that no data point has substantially stronger influence on the fit than the other points.

MODEL 3

It is found that the evidence against **hp** being different from zero is very weak (p-value 0.223). Therefore, in the last model only **wt**, **am** and **qsec** remain.

There is strong evidence that the coefficient of **wt** is different from zero (p-value $\ll 0.001$).

Therefore, we perform inference using a Dummy variable in the model $mpg = \hat{\beta}_0 + \hat{\beta}_1 \cdot wt + \hat{\gamma} \cdot (am_i \cdot wt)$ under the assumption that both (auto and manual cars) have the same zero-weight MPG.

```
fit_4 <- summary(lm(mpg ~ wt + I(am*wt), data=mtcars))
fit_4$coefficients
```

##		Estimate	Std. Error	t value	Pr(> t)
##	(Intercept)	38.8798	2.5094	15.4937	1.453e-15
##	wt	-5.6895	0.6654	-8.5506	2.029e-09
##	I(am * wt)	-0.4947	0.5156	-0.9595	3.452e-01

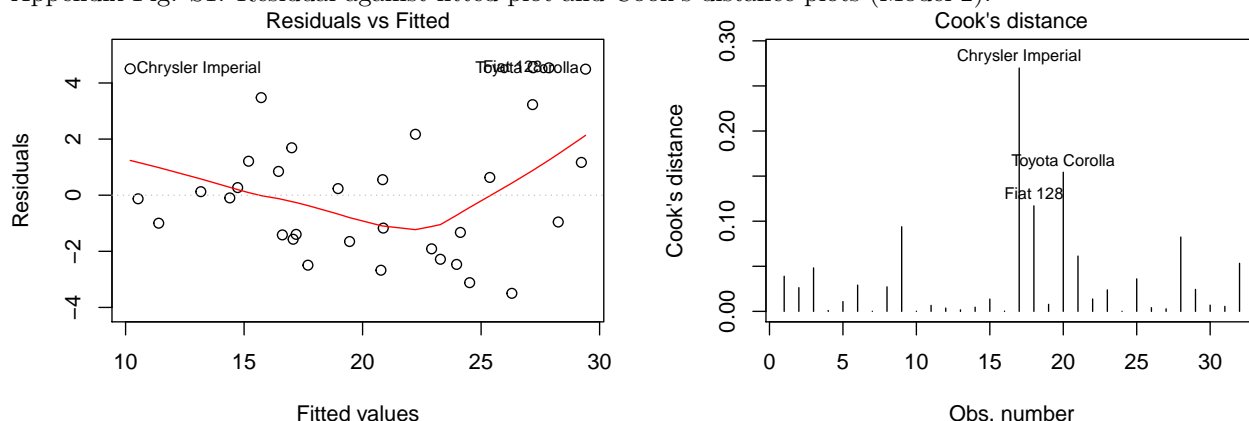
The coefficients are interpreted as follows. The **wt** dependence of automatic cars (**am** = 0) is such that for every unit in **wt**, the MPG decreases by 5.7 units. However, given that the p-value for the dummy variable **I(am*wt)** is around 0.34, it is not plausible to believe that the **wt** dependence of manual cars is different from automatic cars. This is illustrated in a figure in the appendix where a single linear model can explain the MPG dependence of both manual and automatic cars.

Conclusion

MPG is higher for manual cars and the **wt** dependence of MPG is not different between automatic and manual cars.

Appendix

Appendix Fig. S1: Residual against fitted plot and Cook's distance plots (Model 2).



Appendix Fig. S2: Dummy Variable Analysis: Does MPG dependence on `wt` differ for automatic and manual cars?

