Fuel efficiency of manual and automatic transmissions

Andrea Schioppa

Summary

We use the dataset `mtcars` to compare the fuel efficiency of automatic and manual trasmissions. We fit a linear model predicting the miles per US gallon (`mpg`) using the transmission (`am=0` for automatic, `am=1` for manual), the weight `wt` (in 1000lbs) and the gross horsepower `hp` as predictors; the choice of the model is motivated by analysis of variance.

We find that manual transmission leads to an estimated increase of `mpg` of 8.42 with a 95% confidence interval [1.96,14.88].

However, we also find that with increasing weight automatic cars lose less efficiency (see interpretation of the coefficient `am:wt`). A limitation of the model is that manual cars tend to weigh less than automatic ones, so comparing efficiency for a given weight seems reasonable in a range of about $[2500, 3500]$ lbs.

Analysis

The data set `mtcars` consists of 32 observations of 11 variables. In Figure 1 we make some exploratory plots to compare `mpg` with the other variables. Besides using `am` (transmission type), we decide to try as continuous predictors `wt`, `hp`, `disp`, and as categorical predictors `cyl`, `vs`, `gear`.

To select the model we first tried

```
fit1 <- lm(mpg ~ am + wt, data=mtcars); fit2 <- update(fit1, mpg ~ am*wt)
fit3 <- update(fit2, . ~ .+ hp); fit4 <- update(fit3, . ~ . + am*hp);
anova(fit1,fit2,fit3,fit4)$"Pr(>F)"
```

```
## [1]           NA 0.0003103491 0.0093689877 0.1598330395
```

because of the p-values we decide to use the model `mpg~am*wt+hp`. We have also tried to include any of `disp`, `vs`, `cyl` and `gear`: `anova` suggests that adding any of these predictors is not significant (we omit the code for reasons of space).

We fit the model with `fit <- lm(mpg ~ am * wt + hp, data = mtcars)`; the diagnostic graphs in Figures 2,3 indicate "outliers", Maserati Bora(31), Chrysler Imperial(17), Toyota Corolla(20), Fiat 128(18) that we decide to remove after inspecting `dfbetas(fit)`, `hatvalues(fit)`.

We refit the model and observe an improvement through the quantile-quantile plot Figure 4.

```
newmtcars <- mtcars[-c(17, 20, 31, 18),]
fit.new <- lm(mpg ~ am * wt + hp, data = newmtcars)
cf.new <- coef(fit.new)
sfit.new <- summary(fit.new)
```

We show a summary of the coefficients and p-values.

	Estimate	Std. Error	p-value
(Intercept)	33.22	2.09e+00	6.61e-14
am	8.42	3.12e+00	1.29e-02
wt	-3.05	6.63e-01	1.26e-04
hp	-0.03	7.52e-03	5.72e-04
am:wt	-3.06	1.12e+00	1.19e-02

The coefficient `am` is interpreted as the expected increase in `mpg` when considering a manual vs. an automatic transmission. The coefficient `wt` is interpreted as the change in `mpg` when increasing the weight by 1 unit (=1000 lbs) for an automatic transmission. Finally, `am:wt` quantifies the change in `wt` if we consider a manual transmission.

The 95% confidence interval for `am` is:

```
interval_am <- sfit.new$coefficients[2,1] + c(-1,1) * sfit.new$coefficients[2,2] *
    qt(0.975, df = sfit.new$df[2])
```

```
interval_am
```

```
## [1]  1.95903 14.88342
```

We similarly compute the 95% confidence interval for `wt`:

```
## [1] -4.424002 -1.679864
```

and `am:wt`:

```
## [1] -5.3734631 -0.7424189
```

Appendix: Figures

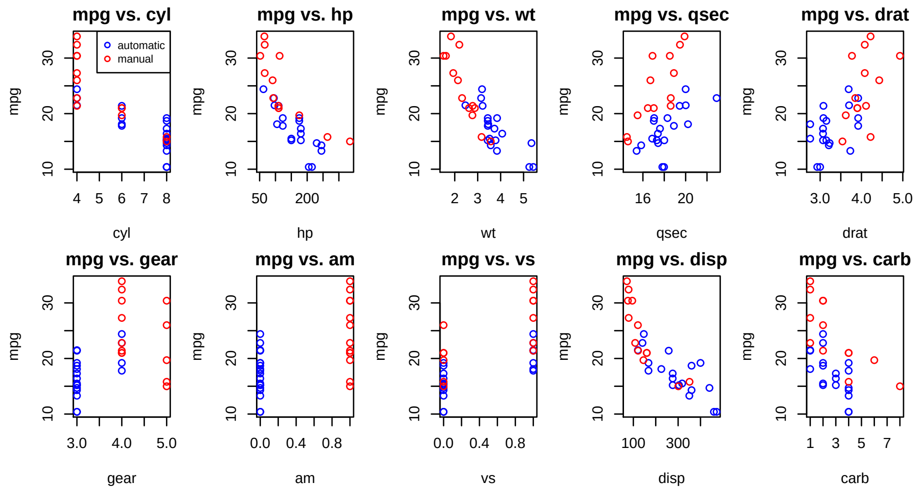

Figure 1: Exploratory plot 1

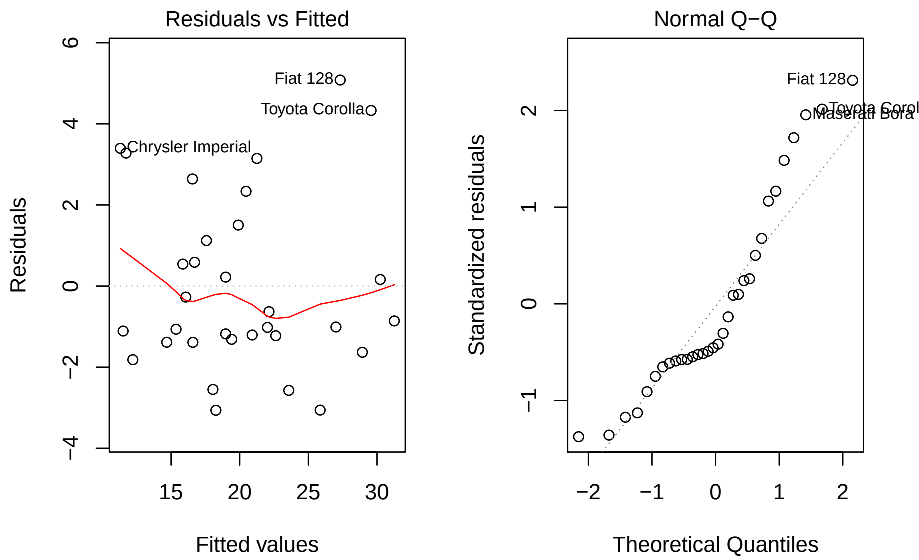

Figure 2: Diagnostic plot 1

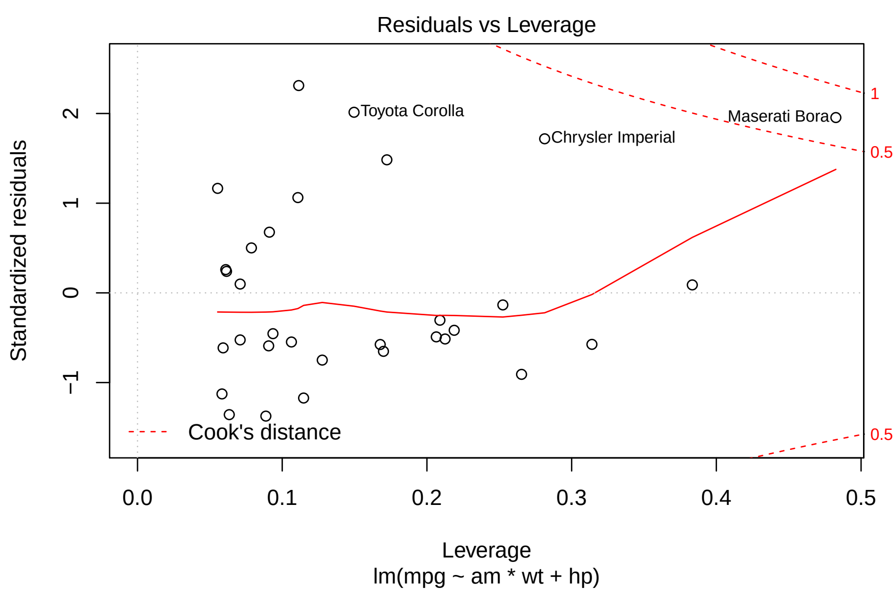

Figure 3: Diagnostic plot 2

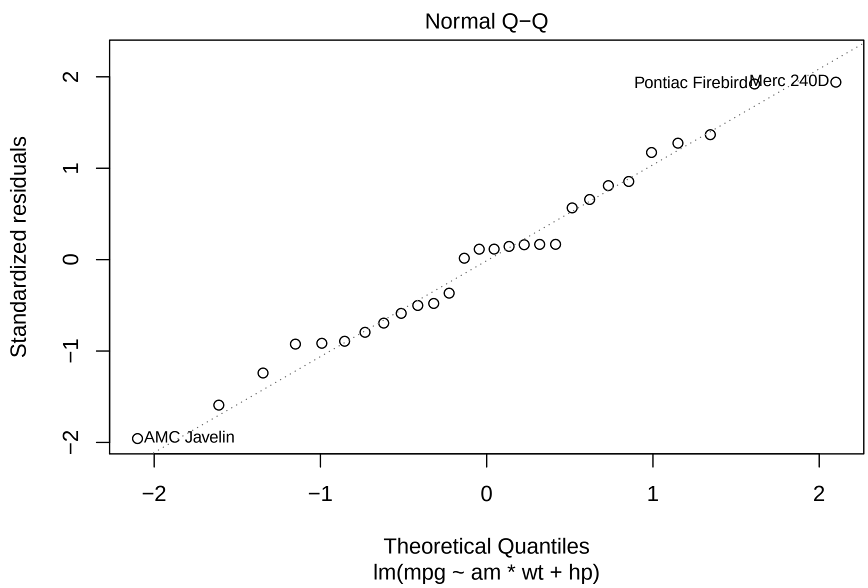

Figure 4: QQ-plot after removing “outliers”