Regression Models Course Project

Andrea Eoli [19-01-2023]

Executive summary

We try to have a look at the difference in gasoline compumption between cars with automatic transmission and those with a manual one. Although automatic cars are more efficient (higher number of miles per gallon, or MPG), the type of transmission is not the most influetial factors that affect the MPG. In fact, when fitting a model that can explain MPG efficiency, the number of cylinders, the horsepower, and the car weight are the only statistically significant covariates and are negatively correlated with MPG.

Analysis

```
data(mtcars)

# correlation data for later
carscor <- cor(mtcars$mpg,mtcars[,-1])

# Factorize discrete variables
mtcars$mpg <- as.numeric(mtcars$mpg)
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$am <- factor(mtcars$am)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
```

Q1: "Is an automatic or manual transmission better for MPG"

Fig.1 shows that automatic cars have a higher mean MPG. This is confirmed by the t test (p-value = 0.0014). We can therefore conclude that automatic cars are better for MPG.

```
t.test(mtcars$mpg~mtcars$am)$p.val
```

```
## [1] 0.001373638
```

Q2: "Quantify the MPG difference between automatic and manual transmissions"

```
fit_am <- lm(mpg ~ am, data = mtcars)
round(summary(fit_am)$r.squared*100, digits=2)
```

```
## [1] 35.98
```

First, we fit a linear model between "mpg" and the transmission type. However, as only the 36% of MPG variation is explained by "am", we need to include more covariates. Let's investigate their correlations.

```
carscor
```

```
##            cyl       disp         hp      drat         wt     qsec        vs
## [1,] -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684 0.6640389
##             am      gear       carb
## [1,] 0.5998324 0.4802848 -0.5509251
```

"Cyl", "disp", "hp", and "wt" are negatively correlated with MPG. This is also visible in Fig.2. We then use step() to test all covariates and find the best model.

```
fit_all <- lm(mpg ~ ., data = mtcars)
step_fit <- step(fit_all, direction = "both", trace = 0)

shapiro.test(summary(step_fit)$residuals)$p.val
```

```
## [1] 0.4478562
```

```
summary(step_fit)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## am1          1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

The function automatically keeps the model with the lowest AIC ("mpg ~ cyl + hp + wt + am", AIC=61.65), which corresponds to the best fit. The model explains the 86.6% of MPG variation. Residuals are normally distributed (p-value = 0.4479) and model assumptions are met (Fig.5). 6 cylinders (-3.03), horsepower(-0.03), and weight (-2.49) are the variables with the biggest impact on MPG. Although we were expecting automatic cars to be more efficient, manual transmission has a positive coefficient (1.8). However, it is not significant (Pr = 0.21) and we cannot therefore conclude that, according to this model, manual cars are more efficient than automatic ones. This contraddiction can be seen in Fig.2-4: automatic cars are correlated with heavier vehicles, higher cylinder number, and higher horsepower, which decrease their MPG.

Appendix

Fig.1 − MPG by transmission type

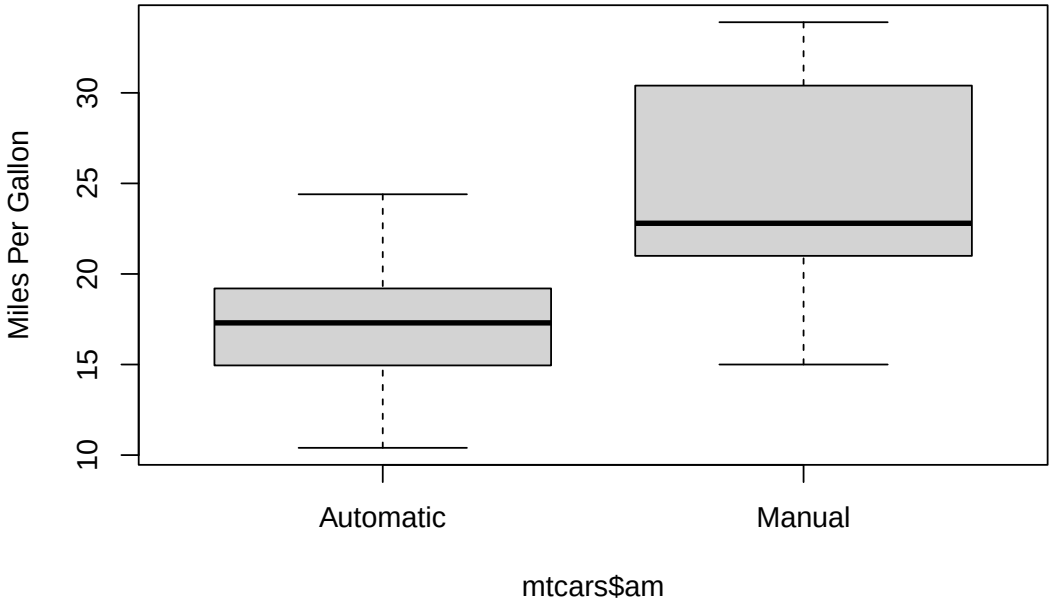

Fig.2 − Correlations

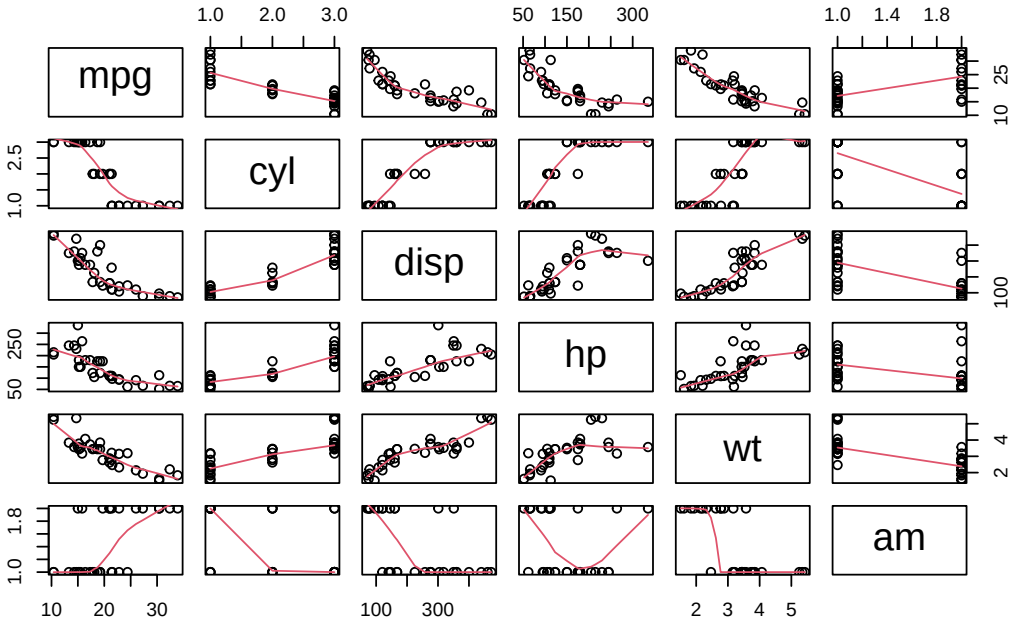

Fig.3 − MPG by transmission type and cylinders

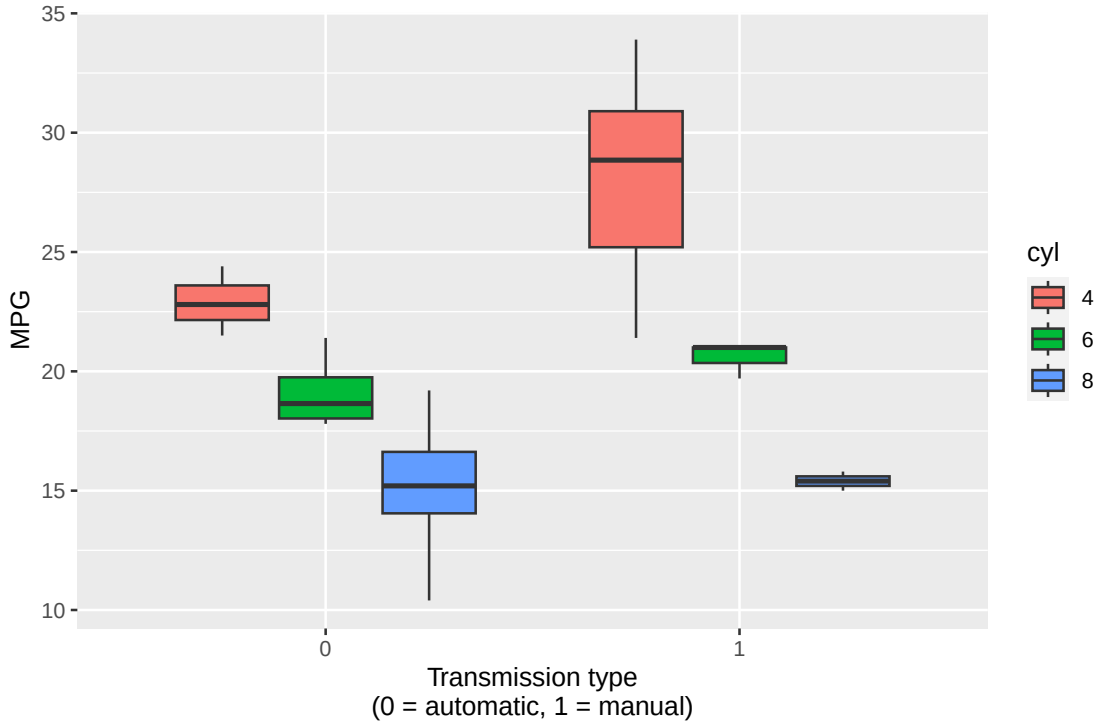

Fig.4 − MPG by car weight and cylinders

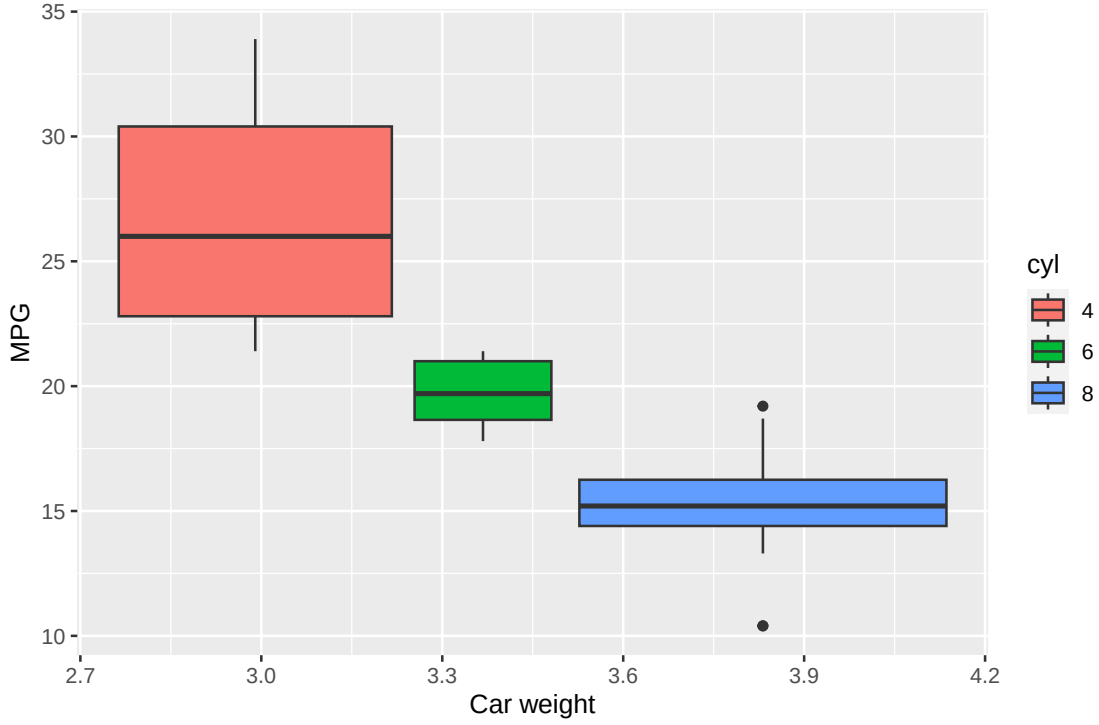
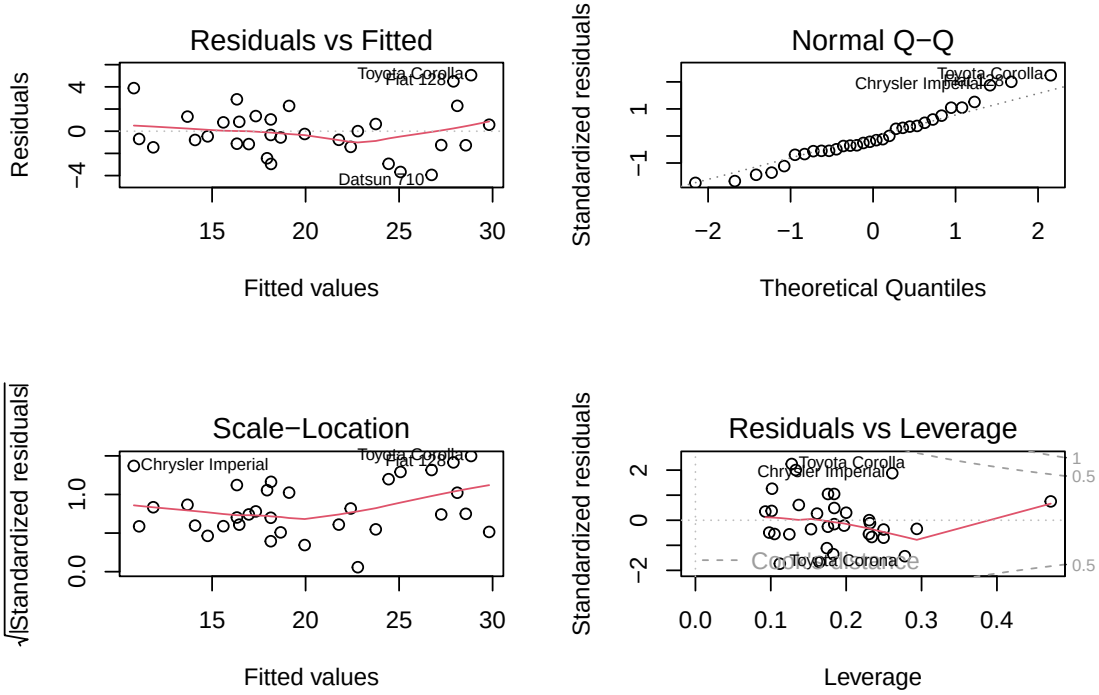
Fig.5 − Residuals plots