## Automatic or manual transmission better for mpg(miles per gallon)?

### Interested question

Is automatic or manual transimission better for mpg?

### Data description

Our used dataset is **mtcars** along with R software.
**am** is one column in **mtcars**, and it only has two values:

- 0 : automatic transimission
- 1 : manual transimission

### Exploratory data analysis

First, let's get some intuitive information from a boxplot between mpg and am. Figure 1 is the plot(in Appendix). And we can have an intuition that **manual transimission is better for miles per gallon**. Next, we will search the right model for mpg.

### Linear regression

We want to establish a linear model for mpg. First, let's try everything in.

```
fit1 <- lm(mpg ~ ., data = mtcars)
summary(fit1)$coef
```

```
##                Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
## am           2.52022689  2.05665055  1.2254035 0.23398971
## gear         0.65541302  1.49325996  0.4389142 0.66520643
## carb        -0.19941925  0.82875250 -0.2406258 0.81217871
```

p-values are all greater than 0.05 and the model may contain too many variables that have no relation with mpg. We filter out these variables one by one to get a reasonable model. Finally, the appropriate model is:

```
fit2 <- lm(mpg ~ wt + qsec + am - 1, data = mtcars)
summary(fit2)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am - 1, data = mtcars)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8820 -1.5401 -0.4246  1.6623  4.1711
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## wt    -3.1855     0.4828  -6.598 3.13e-07 ***
## qsec   1.5998     0.1021  15.665 1.09e-15 ***
## am     4.2995     1.0241   4.198 0.000233 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.497 on 29 degrees of freedom
## Multiple R-squared:  0.9871, Adjusted R-squared:  0.9858
## F-statistic:   741 on 3 and 29 DF,  p-value: < 2.2e-16
```

In our model, all three coefficients are significantly not equal to 0. And the model fits well, as the adjusted R squared is close to 1. We have quantified the mpg difference between automatic and manual transimission. The intepretation is: **holding all other variables constant, there is a 4.2995 rise in mpg from automatic to manual transimission, with a standard error of 1.0241.**

Next, let's do a residual plot. As figure 2 shows, the residuals are just like noise. Figure 3 is the QQ-plot of residuals. And we can see that residuals are almost normal. We are ensured that the model fits well.

### Summary

Manual transimission is better for mpg than automatic transimission. For people who want to save fuel using, manual transimission cars are preferred. And for car manufacturers, it's a good choice to optimize automatic transimission cars fuel use, to optimize the inner engine.

---

## Appendix

Figure 1: am VS mpg boxplot

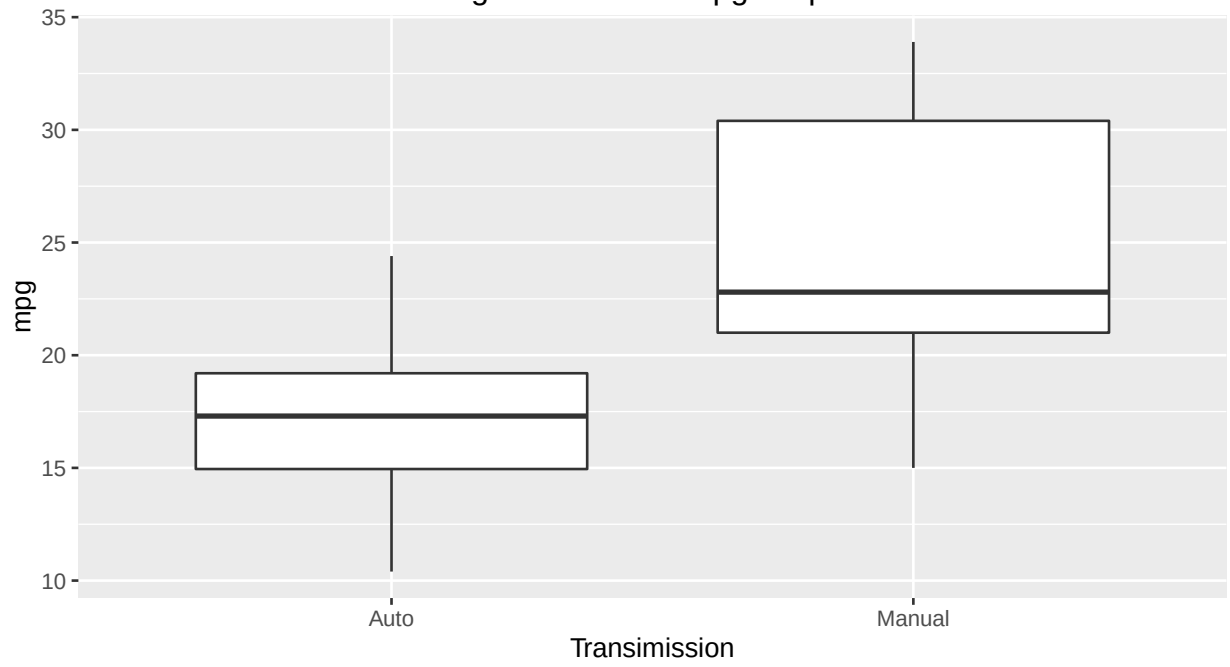


Figure 2: residual plot

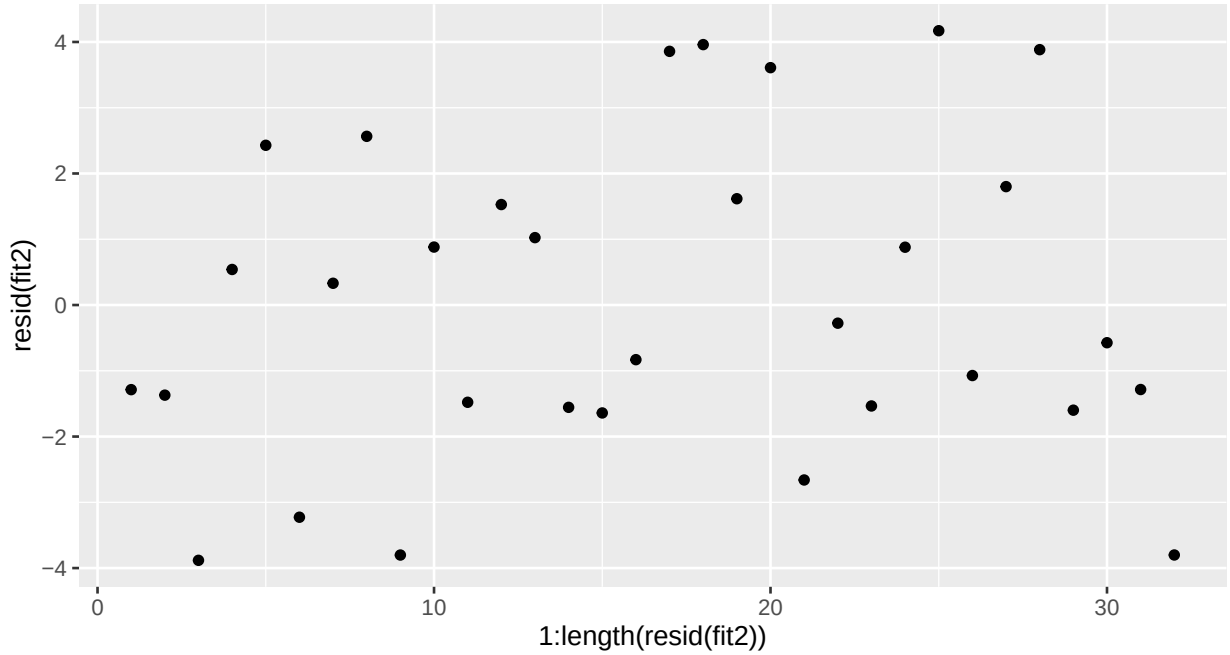

**Figure 3: Normal QQ−plot of residuals**