Analysis of Mile per Gallon vs. Tranmission via Regression Models

Aliakbar Safilian[*]

January 21, 2019

Contents

1 Overview

In this report, we explore the relationship between a set of variables and *miles per gallon* (MPG). The dataset of interest in this report is **mtcars** from the dataset package.

The data includes the following variables:

- **cyl**: Number of cylinders
- **disp**: Displacement
- **hp**: Gross horsepower
- **drat**: Rear axle ratio
- **wt**: Weight
- **qsec**: 1/4 mile time
- **vs**: Engine shape, i.e., V-shaped or straight
- **am**: Automatic (0) or manual (1) transimission
- **gear**: Number of forward gears
- **carb**: Number of carburetors

We are particularly interested in the following two questions:

[*]Email: a.a.safilian@gmail.com

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

In Sect.2, we do some prelimanry analysis, including loading, transformation, and some summary and exploratory analysis. We show that, in general, we expect that manual transmission works better than automatic transmission with respect to fuel economy. We deeply investigate this in the subsequent sections, considering many other factors.

In Sect.3, for each numeric variable *var*, we build a linear model with **mpg** as the output and *var* as the regressor considering its interaction with the transmission type. We first select linear models which are worth considering. Then, we address our main questions using by these linear models.

In Sect.4, we consider multivariate regression models. Again, we select the best fitting models, and then study them to address our analysis questions.

The R scripts of Sect.2, Sect.3, and Sect.4 can be found in Appendix.A, Appendix.B, and Appendix.C. Appendix.D include the diagnosis plots for the fitting models.

We consider 0.05 as the significance rate in all statistical analyses in this report.

2 Preliminary Analysis

Let us first take a look at the structure of the data (we have transformed the variable **am** into its equaivalent factor variable and rename its levels):

```
'data.frame':   32 obs. of  11 variables:
 $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
 $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
 $ disp: num  160 160 108 258 360 ...
 $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
 $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
 $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
 $ qsec: num  16.5 17 18.6 19.4 17 ...
 $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
 $ am  : Factor w/ 2 levels "automatic","manual": 2 2 2 1 1 1 1 1 1 1 ...
 $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
 $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

As we see in the boxplot in Fig. 1, in general, the manual transformation is better than the automatic transmission in the sense of fuel economy. The mean of MPG for automatic transmission and manual transmission are 17.15 and 24.39, respectively.

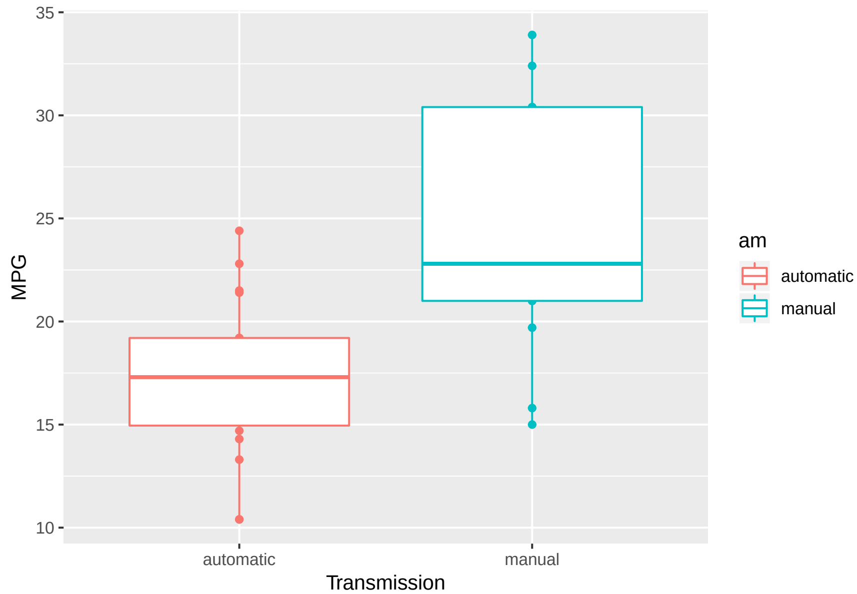

Figure 1: The Box Plot for MPG per Transmission Type

3 Single Variate Regression Models

In this section, for each numeric variable *var*, we build a linear model with MPG as the output and *var* as the regressor considering its interaction with the transmission type, i.e., the model $mpg \sim var * factor(am)$.

In Sect.3.1, we select the linear models which are worth considering. In Sect.3.2, Sect.3.3, and Sect.3.4, we address our main analysis question on the selected models.

3.1 Model Selection

We show that

Theorem 1 *The best three single variate models with mpg as the output are mpg* $\sim$ *wt*am, mpg* $\sim$ *disp*am, and mpg* $\sim$ *hp*am.* $\square$

The rest of this section is devoted to the proof of the theorem.

The correlation between the numeric variables in the dataset is shown in the chart in Fig. 2. As we see in the chart, the correlation between **mpg** and **qsec**, i.e., 0.42, is not that high. Therefore, we ignore the model fitting of **mpg** vs. **qsec**.

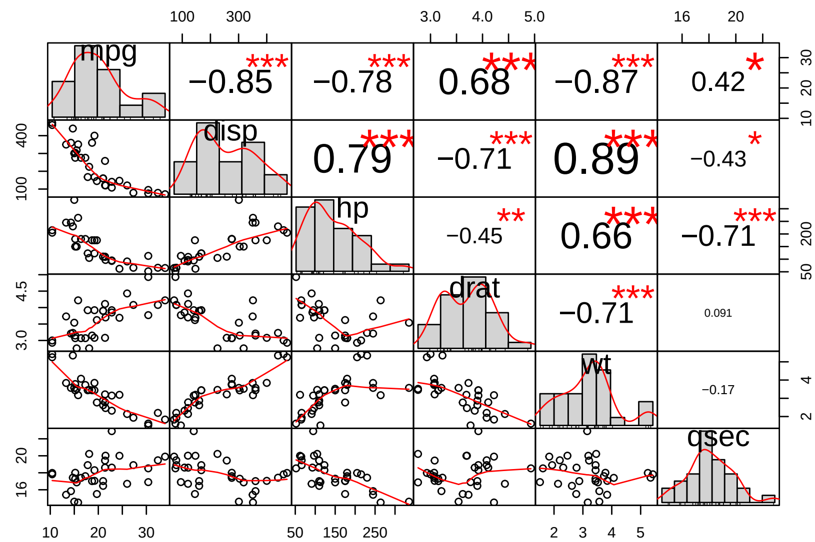

Figure 2: The Correlation Chart

Our models are as follow. Note that the weigth unit in the dataset is 1000lb; however, we change it to tonne.

```
fit_wt <- lm(mpg ~ I(wt /2.20462) * factor(am), data = mtcars)
fit_disp <- lm(mpg ~ disp  * factor(am), data = mtcars)
fit_hp <- lm(mpg ~ hp * factor(am), data = mtcars)
fit_drat <- lm(mpg ~ drat * factor(am), data = mtcars)
```

Since the above models are non-nested models, we take advantage of the *coxtest* function from the *lmtest* package to test them. In Table. 1, each row and column belongs to a fitting model. For any $1 \leq i, j \leq 4$, the cell in the position $[i, j]$ represents the P-value of comparing m_i against n_j, where $(\forall t)$ m_t denotes the fitted model in row t, and n_t denotes the model represented in column j.

Table 1: P-values of Comparing Single Variate Models by coxtest

	fit_wt	fit_disp	fit_hp	fit_drat
fit_wt	0	0.01	0.00	0.38
fit_disp	0	0.00	0.03	0.11
fit_hp	0	0.01	0.00	0.15
fit_drat	0	0.00	0.00	0.00

Considering 0.05 as our significance rate, the results are as follow:

- All other fitting models are preferred over the model fit_drat.
- The models other than fit_drat have no preference over each other, though the best among all the models is fit_wt.

Therefore, we discard the model fit_drat.

One could find the corresponding diagnosis plots in Appendix. 1. As we see in the diagnostic plots, everything (including normality of errors and residuals) looks more less ok for the fitting models.

In the rest of this section, we investigate the fitting models excluding **fit_drat**.

3.2 MPG vs. Weight

The first model that we study is fit_wt, i.e., mpg ∼ wt*am. Let us take a look at its coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	31.42	3.02	10.40	0
I(wt/2.20462)	-8.35	1.73	-4.82	0
factor(am)manual	14.88	4.26	3.49	0
I(wt/2.20462):factor(am)manual	-11.68	3.19	-3.67	0

The estimated intercept in automatic transmission is about 31.42, and 14.88 is the estimated change in the intercept of the linear relationship between weigth and MPG going from automatic transmission to manual transmission. The estimated slope in automatic transmission is -8.35 while the estimated change in the slope switching from automatic to manual is -11.68. In other words:

- The estimated MPG for automatic and manual vehicles with 0 weight are 31.42 and 46.3, respectively.
- The expected change in MPG per 1 tonne change in weight for automatic and manual vehicles are -8.35 and -20.03, respectively.

Fig. 3 represents the corresponding plot, where the regression lines for automatic transmission and manaual transmission are shown in red and blue, respectively.

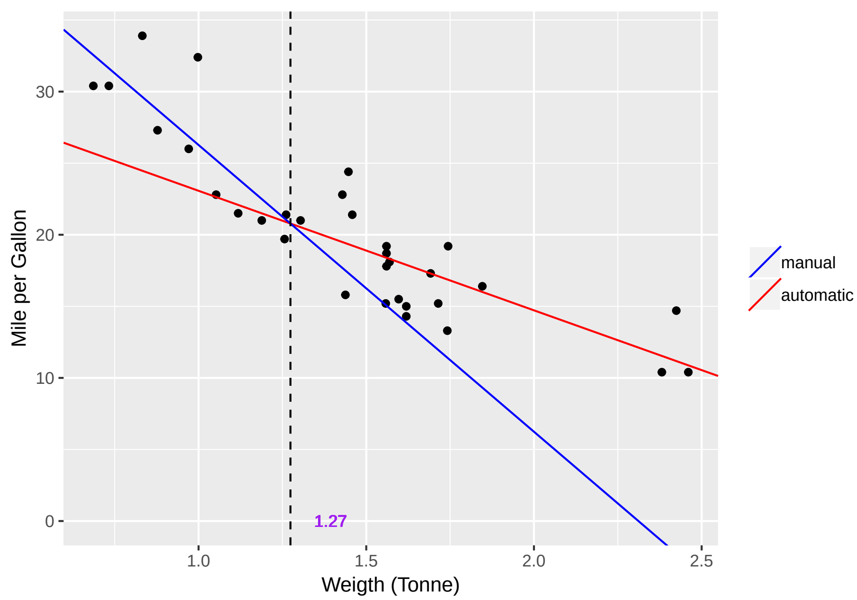

Figure 3: MPG vs. Weight

As it is clear heavier vehicle results in more fuel consumption. The regression lines meets at point **1.27** tonne. As seen, we predict that for vehicles with weight less (more, respectively) than 1.27 tonne, the manual (automatic, respectively) transmission is a better for fuel economy.

3.3 MPG vs. Displacement

The next model is fit_disp, i.e., mpg ∼ wt*am whose coefficients are as follow:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	25.16	1.93	13.07	0.00

```
disp                      -0.03      0.01   -4.44     0.00
factor(am)manual           7.71      2.50    3.08     0.00
disp:factor(am)manual     -0.03      0.01   -2.75     0.01
```

The estimated intercept in automatic transmission is about 25.16, while 7.71 is the estimated change in the intercept of the linear relationship between displacement and MPG going from automatic transmission to manual transmission. The estimated slope in automatic transmission is -0.03 while the estimated change in the slope switching from automatic to manual is -0.03. In other words:

- The estimated MPG for automatic transmission and manual transmission vehicles with 0 cu.in. displacement are 25.16 and 32.87, respectively.
- The expected change in MPG per 1 cu.in. change in displacement for automatic and manual vehicles are -0.03 and -0.06, respectively.

Fig. 4 the corresponding plot, where the regression lines for automatic transmission and manaual transmission are shown in blue and red, respectively.

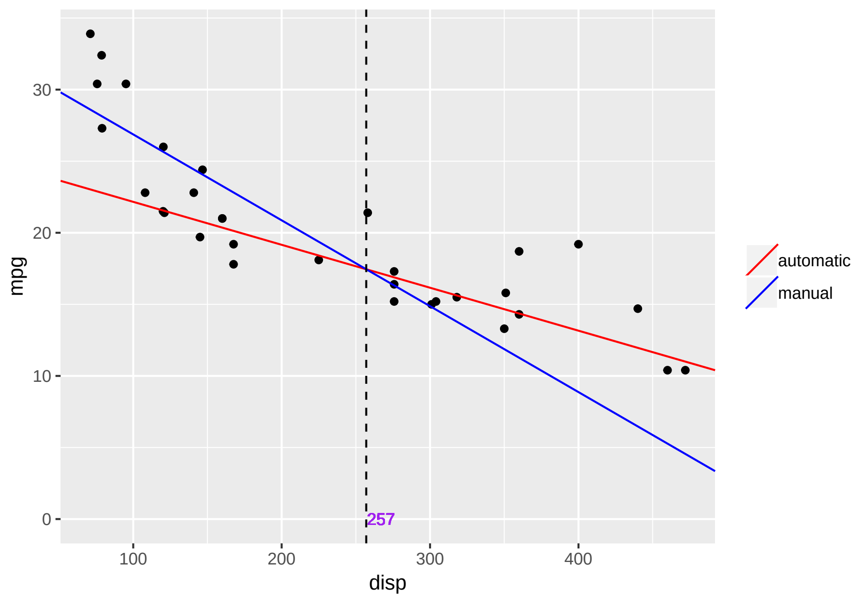

Figure 4: MPG vs. Displacement

As it is clear, higher displacement results in more fuel consumption. The regression lines meets at point **257** cu.in. As seen in the plot, we predict that for vehicles with displacement less (more, respectively) than 257 (cu.in.), the manual (automatic, respectively) transmission workes better w.r.t fuel economy.

3.4 MPG vs. Horsepower

The last model to study is fit_hp, i.e., mpg $\sim$ hp*am with the following coefficients:

```
                    Estimate Std. Error t value Pr(>|t|)
(Intercept)            26.62       2.18   12.20     0.00
hp                     -0.06       0.01   -4.57     0.00
factor(am)manual        5.22       2.67    1.96     0.06
hp:factor(am)manual     0.00       0.02    0.02     0.98
```

The estimated intercept in automatic transmission is about 26.62. 5.22 is the estimated change in the intercept of the linear relationship between horsepower and MPG going from automatic transmission to manual transmission. The estimated slope in automatic transmission is -0.06 while the estimated change in the slope switching from automatic to manual is about 0. This shows that there is no signigicant interaction between **am** and **hp**. In other words:

- The expected MPG for automatic transmission and manual transmission vehicles with horsepower 0 are 26.62 and 31.84, respectively.

- The expected change in MPG per unit change in horsepower for both automatic and manual vehicles is about -0.06.

Note that the second bullet implies that the corresponding regression lines for automatic and manual would be parallel. Fig. 5 represents the corresponding plot, where the regression lines for automatic transmission and manaual transmission are shown in red and blue, respectively.

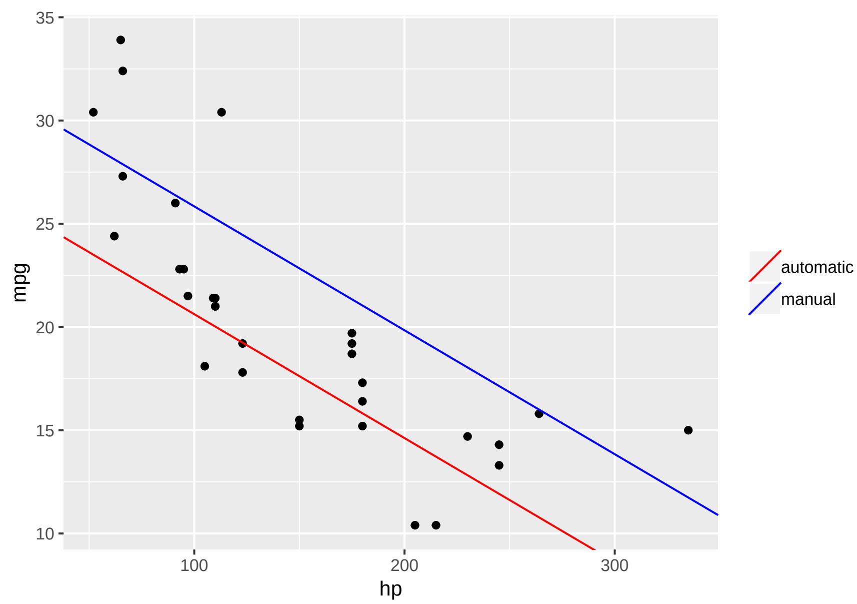

Figure 5: MPG vs. Horsepower

Clearly, higher horsepower results in worse fuel economy. As seen in the above plot, we predict that manual transmission is always better than automatic transmission with respect to fuel economy for a given horsepower.

4 Multivariate Regression Models

In this section, we consider more complicated models, i.e., multivariate regression models.

The structure of this section is as follows: In Sect.4.1, we select the multivariate linear regression models which are worth considering. In Sect.4.2 and Sect.4.3, we address our main analysis question on the selected models.

4.1 Model Selection

We show that:

Theorem 2 *The best two fitting linear models with mpg as the output are* $mpg \sim wt+hp$ *and* $mpg \sim wt+qsec$. $\square$

As we already saw, among the single variate models, the models worth to consider are **Model1**: mpg~wt, **Model2**: mpg~hp, and **Model3**: mpg~disp. Now, we want to see if adding some new regressors to these models make sense. We show that

Lemma 1 *The best models including wt as a regressor are mpg $\sim$ wt+hp and mpg $\sim$ wt+qsec.* □

Lemma 2 *The best model including hp as a must regressor is mpg $\sim$ hp+wt.* □

Lemma 3 *Considering disp as a must regressor in our models, the best linear model is mpg $\sim$ disp+wt.* □

Note that these three lemmas together prove our main theorem.

4.1.1 Proof of Lemma. 1

Let us first see if adding some regressors to mpg~wt makes senses. We take advantage of the *anova* function to address this question. The P-values for comparing the model $mpg \sim wt$ vs. $mpg \sim wt + var$, where $var \in \{disp, hp, drat, qsec\}$ are represented in Table. 2.

Table 2: P-values of Comparing 2-variate Models with wt as the Regressor

	mpg~wt+disp	mpg~wt+hp	mpg~wt+drat	mpg~wt+qsec
mpg~wt	0.06	0	0.33	0

Considering the significance rate 0.05, we see that only the two models "mpg ~ wt + hp" and "mpg ~ wt + qsec" are preferred over mpg ~ wt. Now, let us see which of these two models works better:

```
Cox test

Model 1: mpg ~ wt + hp
Model 2: mpg ~ wt + qsec
                Estimate Std. Error z value Pr(>|z|)
fitted(M1) ~ M2    -2.25       1.54   -1.46     0.14
fitted(M2) ~ M1    -2.30       1.53   -1.50     0.13
```

Therefore, considering 0.05 as the significance rate, none of them are preferred over the other.

Now, let us see if their combination, i.e., mpg ~ wt + hp + qsec, works better. In the following, we show that mpg ~ wt + hp + qsec is NOT preferred over mpg ~ wt + hp.

```
Analysis of Variance Table

Model 1: mpg ~ wt + hp
Model 2: mpg ~ wt + hp + qsec
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     29 195.05
2     28 186.06  1    8.9885 1.3527 0.2546
```

Now, we show that mpg ~ wt + hp + qsec is NOT preferred over mpg ~ wt + qsec:

```
Analysis of Variance Table

Model 1: mpg ~ wt + qsec
Model 2: mpg ~ wt + hp + qsec
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     29 195.46
2     28 186.06  1    9.4043 1.4153 0.2442
```

Therefore, we showed that the best models including **wt** as a must regressor is *mpg ~ wt + hp* or *mpg ~ wt + qsec*. Lemma. 1 was proven.

4.1.2 Proof of Lemma. 2

Let us now see if adding some regressors to mpg ∼ hp makes our fitting model any better. Again, we apply the *anova* function to address this question. The P-value for comparing the model mpg~hp vs. mpg~hp+*var*, where $var \in \{disp, wt, drat, qsec\}$ are represented in Table. 3.

Table 3: P-values of Comparing 2-variate Models with hp as the Regressor

	mpg~hp+disp	mpg~hp+wt	mpg~hp+drat	mpg~hp+qsec
mpg~hp	0	0	0	0.11

Considering the significance rate 0.05, all the models excluding mpg~hp+qsec are preferred over mpg~hp. Now, using the *coxtest* command, we want to see which of them works the best:

The following tables show that *mpg~hp+wt* has preference over mpg~hp+disp and mpg~hp+drat:

```
Cox test

Model 1: mpg ~ hp + disp
Model 2: mpg ~ hp + wt
                Estimate Std. Error z value  Pr(>|z|)
fitted(M1) ~ M2  -9.0487     1.7108 -5.2892 1.229e-07 ***
fitted(M2) ~ M1  -0.2062     2.1047 -0.0980     0.922
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Cox test

Model 1: mpg ~ hp + wt
Model 2: mpg ~ hp + drat
                Estimate Std. Error z value  Pr(>|z|)
fitted(M1) ~ M2  -2.9554     1.7423 -1.6963   0.08983 .
fitted(M2) ~ M1 -10.9891     1.5147 -7.2548 4.023e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Therefore, we showed that the best model including **hp** as a must regressor is *mpg~hp+wt*. Lemma. 2 is proven.

4.1.3 Proof of Lemma. 3

The last model to be considered is mpg~disp. Applying the *anova* function, the P-values for comparing the model mpg~disp vs. mpg~disp+*var*, where $var \in \{hp, wt, drat, qsec\}$ are represented in Table. 4.

As we see in the table, *mpg~disp+wt* is the only model which is preferred over mpg~disp. In other words, considering disp as a must regressor in our models, the best linear model is *mpg~disp+wt*. Lemma. 3 is proven.

Table 4: P-values of Comparing 2-variate Models with disp as the Regressor

	mpg~disp+hp	mpg~disp+wt	mpg~disp+drat	mpg~dispp+qsec
mpg~disp	0.07	0.01	0.25	0.57

4.2 MPG vs. Weight plus Horsepower

In this section, we study the model with **wt** and **hp** as regressors. Let's first consider the full interaction between transmission type and the regressors, i.e., the following model:

```
fit_wt_hp <- lm(mpg ~ (I(wt/2.20462) + hp) * factor(am), data = mtcars)
```

The coefficients of the model are as follow:

```
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                      30.70       2.68   11.48     0.00
I(wt/2.20462)                    -4.09       2.08   -1.96     0.06
hp                               -0.04       0.01   -3.00     0.01
factor(am)manual                 13.74       4.22    3.25     0.00
I(wt/2.20462):factor(am)manual  -12.72       4.57   -2.78     0.01
hp:factor(am)manual               0.03       0.02    1.45     0.16
```

As we see, the P-value for the interaction of **hp** and **am** is not significant. Moreover, the P-value for the coefficient **wt** is a little higher than the sigficance rate (0.05). Therefore, we modify the model as follows:

```
fit_wt_hp <- lm(mpg ~ (I(wt/2.20462) * factor(am)) + hp, data = mtcars)
```

The coefficients of the model are as follow:

```
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                      30.95       2.72   11.36     0.00
I(wt/2.20462)                    -5.55       1.86   -2.98     0.01
factor(am)manual                 11.55       4.02    2.87     0.01
hp                               -0.03       0.01   -2.75     0.01
I(wt/2.20462):factor(am)manual   -7.89       3.18   -2.48     0.02
```

The estimated intercept in automatic transmission is about 30.95, and 11.55 is the estimated change in the intercept of the linear relationship going from automatic transmission to manual transmission. The estimated coefficient of weigth in automatic transmission is -5.55 while the estimated change in the weigth coefficient switching from automatic to manual is -7.89. The estimated coefficient of horsepower is -0.03.

In other words:

- The estimated MPG is 30.95 and 42.5 for the automatic transmission and the manual transmission vehicle with weight 0 and horsepower 0, respectively.
- The expected change in MPG for an automatic transmission and manual transmission per tonne change in weight are -5.55 and -13.44, respectively, by holding the horsepwer constant.
- The expected change in MPG for both automatic and manual vehicle per unit change in horsepower is -0.03, by holding weight constant.

We have represented the diagnosis plots of this model in Fig. 9.

4.3 MPG vs. Weight plus 1/4-Mile-Time

In this section, we study the model with **wt** and **qsec** as regressors. We first consider the full interaction between transmission type and the regressors, i.e., the following model:

```
fit_wt_qsec <- lm(mpg ~ (I(wt/2.20462) + qsec) * factor(am), data = mtcars)
```

The coefficients of the model are as follow:

```
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                      11.25       6.99    1.61     0.12
```

```
I(wt/2.20462)                     -6.61       1.52   -4.34     0.00
qsec                               0.95       0.31    3.08     0.00
factor(am)manual                   8.93      12.67    0.70     0.49
I(wt/2.20462):factor(am)manual    -8.29       3.34   -2.48     0.02
qsec:factor(am)manual              0.24       0.56    0.42     0.68
```

As we see, the P-value for the interaction of **qsec** and **am** is not significant. Therefore, we modify the model as follows:

```
fit_wt_qsec <- lm(mpg ~ (I(wt/2.20462) * factor(am)) + qsec, data = mtcars)
```

The coefficients of the new model are as follow:

```
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                        9.72       5.90    1.65     0.11
I(wt/2.20462)                     -6.47       1.47   -4.41     0.00
factor(am)manual                  14.08       3.44    4.10     0.00
qsec                               1.02       0.25    4.04     0.00
I(wt/2.20462):factor(am)manual    -9.13       2.64   -3.46     0.00
```

Since the P-value associated with intercept is high, our estimated intercept in automatic transmission is 0. 14.08 is the estimated change in the intercept of the linear relationship going from automatic transmission to manual transmission. The estimated coefficient of weigth in automatic transmission is -6.47 while the estimated change in the weigth coefficient switching from automatic to manual is -9.13. The estimated coefficient of qsec is 1.02.

In other words:

- The estimated MPG is 0 for an automatic transmission vehicle with weight 0 and qsect 0.
- The estimated MPG is 14.08 for a manual transmission vehicle with weight 0 and qsect 0.
- The expected change in MPG for an automatic transmission vehicle per tonne change in weight is -6.47, by holding the qsec constant.
- The expected change in MPG for an automatic manual vehicle per tonne change in weight is -15.6, by holding the qsec constant.
- The expected change in MPG for both automatic and manual vehicle per unit change in qsec is 1.02, by holding weight constant

The diagnosis plot of this models can be found in Fig. 10.

Appendix A: R Scripts of Sect. 2

Loading and transforming the data:

```
library(datasets)
data("mtcars")
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <- c("automatic", "manual")
str(mtcars)
```

The boxplot for MPG per each transmission type:

```
library(ggplot2)
qplot(am, mpg, data = mtcars, colour = am) + geom_boxplot() +
        xlab("Transmission") + ylab("MPG")
```

Appendix B: R Scripts of Sect. 3

The correlation chart:

```
library(corrplot)
library(PerformanceAnalytics)
chart.Correlation(mtcars[, c(1, 3:7)], histogram=TRUE, pch=19)
```

Code of Table. 1 (P-values of Comparing Single Variate Models)

```
library(lmtest)
wt_disp <- round(coxtest(fit_wt, fit_disp)$`Pr(>|z|)`, 2)
wt_hp <- round(coxtest(fit_wt, fit_hp)$`Pr(>|z|)`, 2)
wt_drat <- round(coxtest(fit_wt, fit_drat)$`Pr(>|z|)`, 2)
disp_hp <- round(coxtest(fit_disp, fit_hp)$`Pr(>|z|)`, 2)
disp_drat <- round(coxtest(fit_disp, fit_drat)$`Pr(>|z|)`, 2)
hp_drat <- round(coxtest(fit_hp, fit_drat)$`Pr(>|z|)`, 2)
wt_c <- c(0, wt_disp[2], wt_hp[2], wt_drat[2])
disp_c <- c(wt_disp[1], 0, disp_hp[2], disp_drat[2])
hp_c <- c(wt_hp[1], disp_hp[1], 0, hp_drat[2])
drat_c <- c(wt_drat[1], disp_drat[1], hp_drat[1], 0)
tests_pval <- as.data.frame(cbind(wt_c, disp_c, hp_c, drat_c))
row.names(tests_pval) <- c("fit_wt", "fit_disp", "fit_hp", "fit_drat")
colnames(tests_pval) <- paste("   ", row.names(tests_pval), sep = "")

library(knitr)
library(kableExtra)
kable(tests_pval, caption = "P-values of Comparing Single Variate Models by coxtest") %>%
        kable_styling(latex_options = "hold_position")
```

The coefficients of fit_wt:

```
cff_wt <- round(summary(fit_wt)$coefficient, 2)
cff_wt
```

The corresponding plot for fit_wt

```
line <- data.frame(intercept = c( cff_wt[1, 1], cff_wt[1, 1] + cff_wt[3, 1]),
                   slope = c( cff_wt[2, 1], cff_wt[2, 1] + cff_wt[4, 1]),
                   row.names = c("automatic", "manual") )
x_com_wt <- (line[1,1] - line[2,1]) / (line[2,2] - line[1, 2])
qplot(wt/2.20462, mpg, data = mtcars) +
        xlab("Weigth (Tonne)") + ylab("Mile per Gallon") +
        geom_abline(aes(intercept=intercept, slope=slope,
                        colour=c("red", "blue")), data=line) +
        theme(legend.title=element_blank()) +
        scale_color_manual(labels = c("manual", "automatic"), values = c("blue", "red")) +
        geom_vline(xintercept = x_com_wt, linetype = "dashed") +
        geom_text(aes(x_com_wt+0.12,0,label = round(x_com_wt, 2)), size = 3,
                  color = "purple")
```

The coefficients of fit_disp:

```
cff_disp <- round(summary(fit_disp)$coefficient, 2)
cff_disp
```

The corresponding plot for fit_disp:

```r
line <- data.frame(intercept = c( cff_disp[1, 1], cff_disp[1, 1] + cff_disp[3, 1]),
                   slope = c( cff_disp[2, 1], cff_disp[2, 1] + cff_disp[4, 1]),
                   row.names = c("automatic", "manual") )
x_com_disp <- (line[1,1] - line[2,1]) / (line[2,2] - line[1, 2])
qplot(disp, mpg, data = mtcars) +
        geom_abline(aes(intercept=intercept, slope=slope,
                        colour=c("blue", "red")), data=line) +
        theme(legend.title=element_blank()) +
        scale_color_manual(labels = c("automatic", "manual"), values = c("red", "blue")) +
        geom_vline(xintercept = x_com_disp, linetype = "dashed") +
        geom_text(aes(x_com_disp+10,0,label = round(x_com_disp, 2)), size = 3,
                  color = "purple")
```

The coefficients of fit_hp:

```r
cff_hp <- round(summary(fit_hp)$coefficient, 2)
cff_hp
```

The corresponding plot for fit_hp:

```r
line <- data.frame(intercept = c( cff_hp[1, 1], cff_hp[1, 1] + cff_hp[3, 1]),
                   slope = c( cff_hp[2, 1], cff_hp[2, 1] + cff_hp[4, 1]),
                   row.names = c("automatic", "manual") )

qplot(hp, mpg, data = mtcars) +
        geom_abline(aes(intercept=intercept, slope=slope,
                        colour=c("blue", "red")), data=line) +
        theme(legend.title=element_blank()) +
        scale_color_manual(labels = c("automatic", "manual"), values = c("red", "blue"))
```

Appendix C: R Scripts of Sect. 4

The script for Table. 2

```r
fit11 <- lm(mpg ~ wt, data = mtcars)
fit12 <- lm(mpg ~ wt + disp, data = mtcars)
fit13 <- lm(mpg ~ wt + hp, data = mtcars)
fit14 <- lm(mpg ~ wt + drat, data = mtcars)
fit15 <- lm(mpg ~ wt + qsec, data = mtcars)

disp_val <- round(anova(fit11, fit12)$`Pr(>F)`[2], 2)
hp_val <- round(anova(fit11, fit13)$`Pr(>F)`[2], 2)
drat_val <- round(anova(fit11, fit14)$`Pr(>F)`[2], 2)
qsec_val <- round(anova(fit11, fit15)$`Pr(>F)`[2], 2)


tests_pval <- as.data.frame(cbind(disp_val, hp_val, drat_val, qsec_val))
row.names(tests_pval) <- c("mpg~wt")
colnames(tests_pval) <- c("mpg~wt+disp", "  mpg~wt+hp", "  mpg~wt+drat", "  mpg~wt+qsec")

kable(tests_pval,
      caption = "P-values of Comparing 2-variate Models with wt as the Regressor") %>%
        kable_styling(latex_options = "hold_position")
```

Comparing two models mpg~wt+hp and mpg~wt+qsec:

```
round(coxtest(fit13, fit15), 2)
```

Comparing two models mpg~wt+hp+qsec and mpg~wt+hp:

```
fit <- lm(mpg ~ wt + hp + qsec, data = mtcars)
anova(fit13, fit)
```

Comparing two models mpg~wt+hp+qsec and mpg~wt+qsec:

```
anova(fit15, fit)
```

The script fot Table. 3:

```
fit21 <- lm(mpg ~ hp, data = mtcars)
fit22 <- lm(mpg ~ hp + disp, data = mtcars)
fit23 <- lm(mpg ~ hp + wt, data = mtcars)
fit24 <- lm(mpg ~ hp + drat, data = mtcars)
fit25 <- lm(mpg ~ hp + qsec, data = mtcars)

disp_val <- round(anova(fit21, fit22)$`Pr(>F)`[2], 2)
wt_val <- round(anova(fit21, fit23)$`Pr(>F)`[2], 2)
drat_val <- round(anova(fit21, fit24)$`Pr(>F)`[2], 2)
qsec_val <- round(anova(fit21, fit25)$`Pr(>F)`[2], 2)


tests_pval <- as.data.frame(cbind(disp_val, wt_val, drat_val, qsec_val))
row.names(tests_pval) <- c("mpg~hp")
colnames(tests_pval) <- c("mpg~hp+disp", "  mpg~hp+wt", "  mpg~hp+drat", "  mpg~hp+qsec")

kable(tests_pval,
      caption = "P-values of Comparing 2-variate Models with hp as the Regressor") %>%
        kable_styling(latex_options = "hold_position")
```

Comparing mpg~hp+wt vs. mpg~hp+disp and mpg~hp+drat:

```
coxtest(fit22, fit23)
coxtest(fit23, fit24)
```

The script for Table. 4.

```
fit31 <- lm(mpg ~ disp, data = mtcars)
fit32 <- lm(mpg ~ disp + hp, data = mtcars)
fit33 <- lm(mpg ~ disp + wt, data = mtcars)
fit34 <- lm(mpg ~ disp + drat, data = mtcars)
fit35 <- lm(mpg ~ disp + qsec, data = mtcars)

hp_val <- round(anova(fit31, fit32)$`Pr(>F)`[2], 2)
wt_val <- round(anova(fit31, fit33)$`Pr(>F)`[2], 2)
drat_val <- round(anova(fit31, fit34)$`Pr(>F)`[2], 2)
qsec_val <- round(anova(fit31, fit35)$`Pr(>F)`[2], 2)


tests_pval <- as.data.frame(cbind(hp_val, wt_val, drat_val, qsec_val))
row.names(tests_pval) <- c("mpg~disp")
colnames(tests_pval) <-
        c("mpg~disp+hp", "  mpg~disp+wt", "  mpg~disp+drat", "  mpg~dispp+qsec")
```

```r
kable(tests_pval,
      caption = "P-values of Comparing 2-variate Models with disp as the Regressor") %>%
        kable_styling(latex_options = "hold_position")
```

The coefficients of the fit_wt_hp:

```r
cff_wt_hp <- round(summary(fit_wt_hp)$coefficient, 2)
cff_wt_hp
```

The coefficients of the new fit_wt_hp:

```r
cff_wt_hp <- round(summary(fit_wt_hp)$coefficient, 2)
cff_wt_hp
```

The coefficients of the model fit_qsec:

```r
cff_wt_qsec <- round(summary(fit_wt_qsec)$coefficient, 2)
cff_wt_qsec
```

The coefficients of the new model fit_qsec:

```r
cff_wt_qsec <- round(summary(fit_wt_qsec)$coefficient, 2)
cff_wt_qsec
```

Appendix D: The Diagnosis Plots

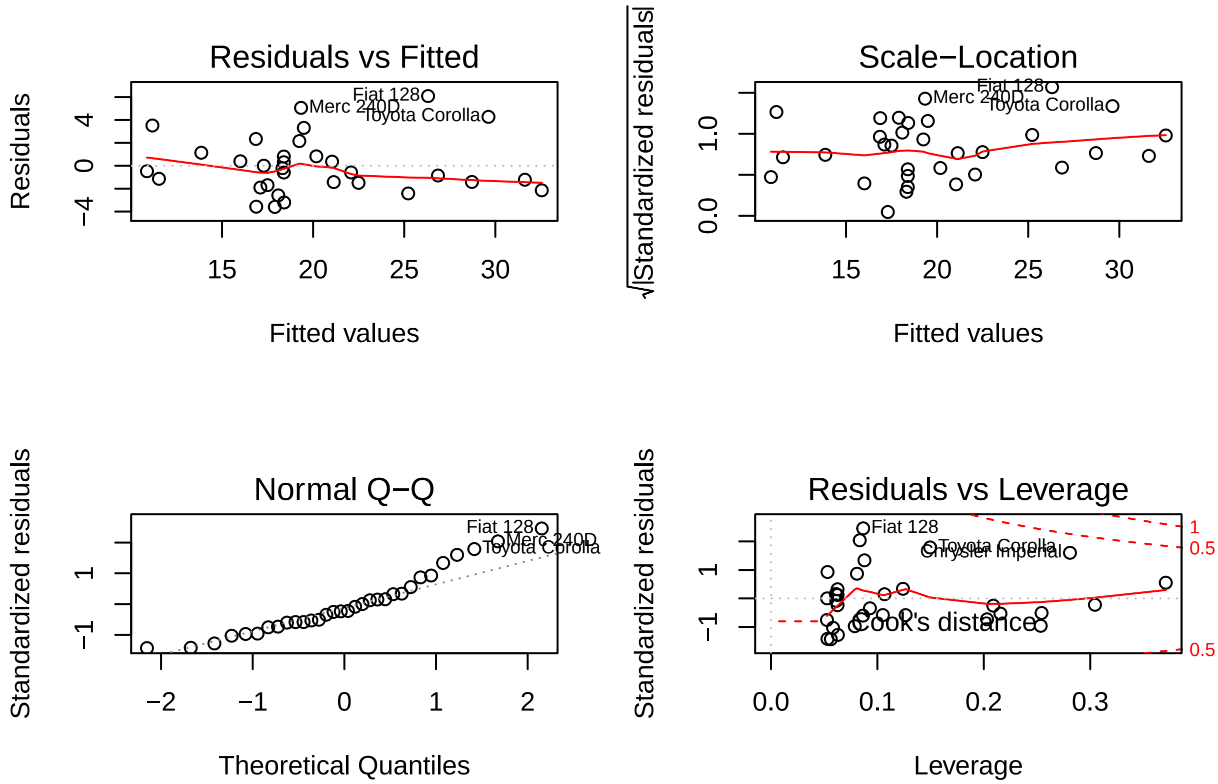

Figure 6: Diagnosis Plots for mpg vs. wt*factor(am)

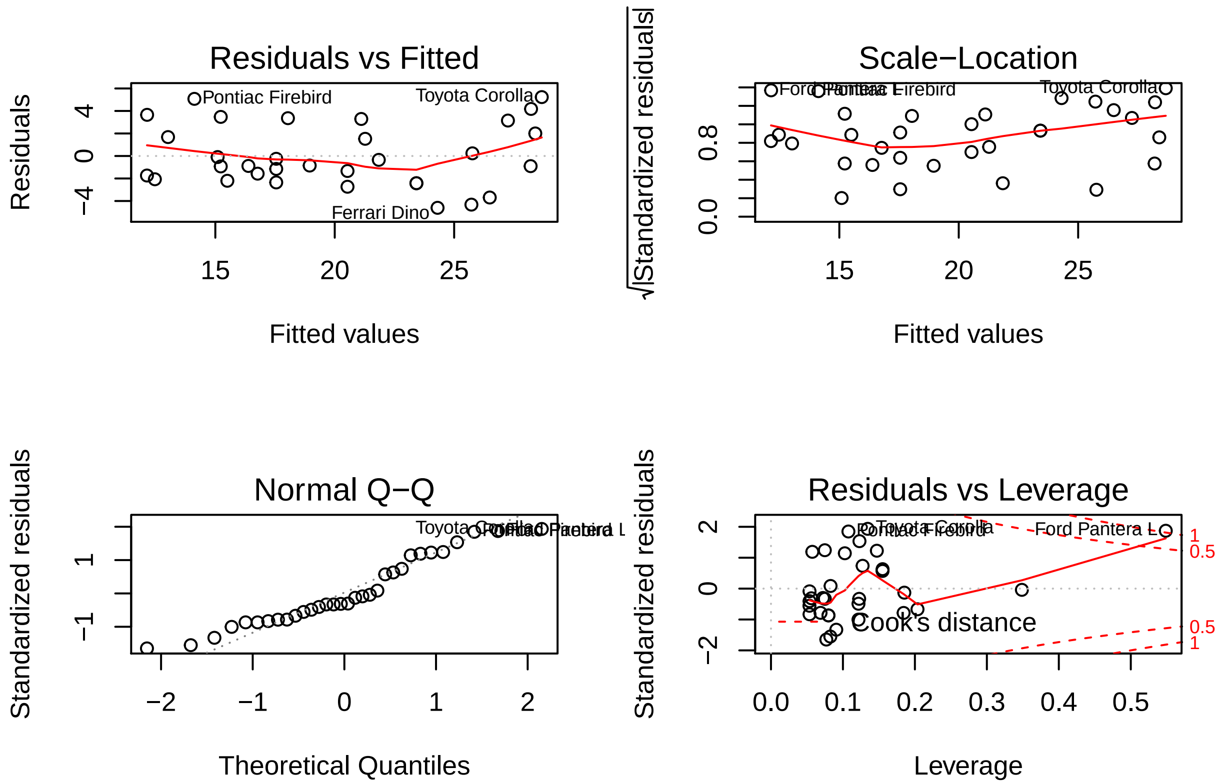

Figure 7: Diagnosis Plots for mpg vs. disp*factor(am)

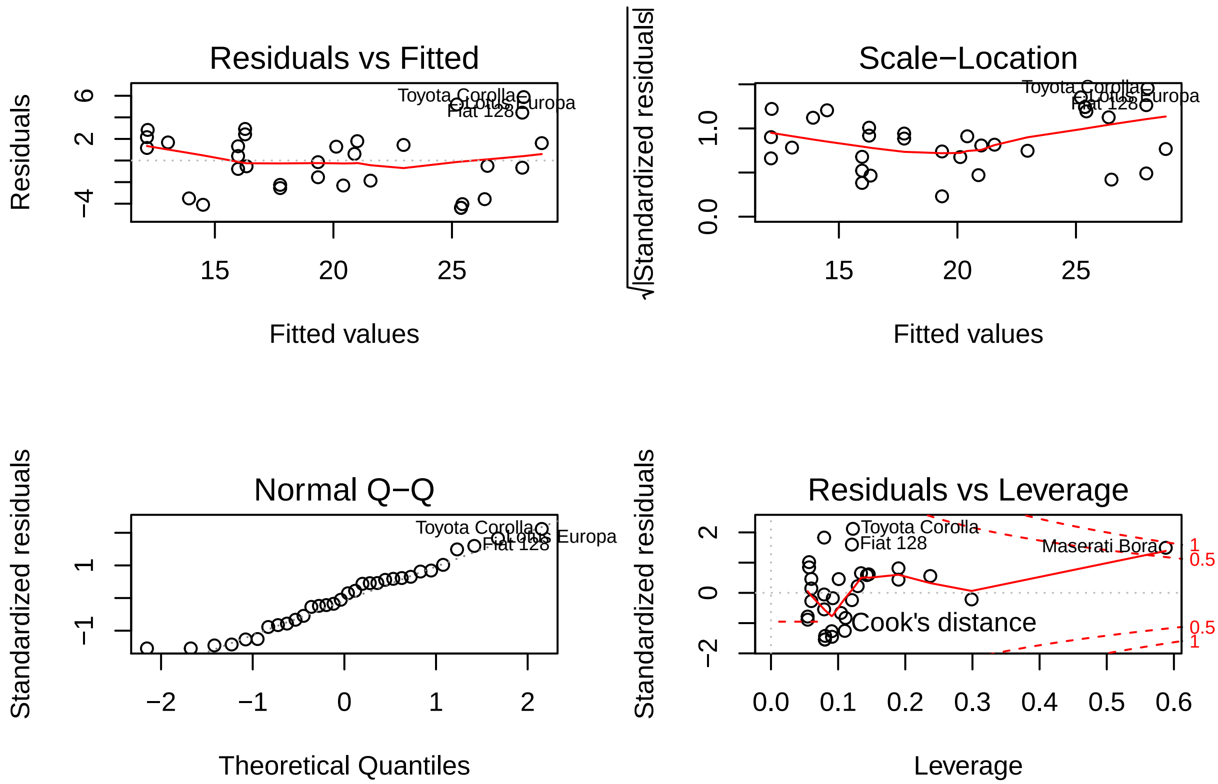

Figure 8: Diagnosis Plots for mpg vs. hp*factor(am)

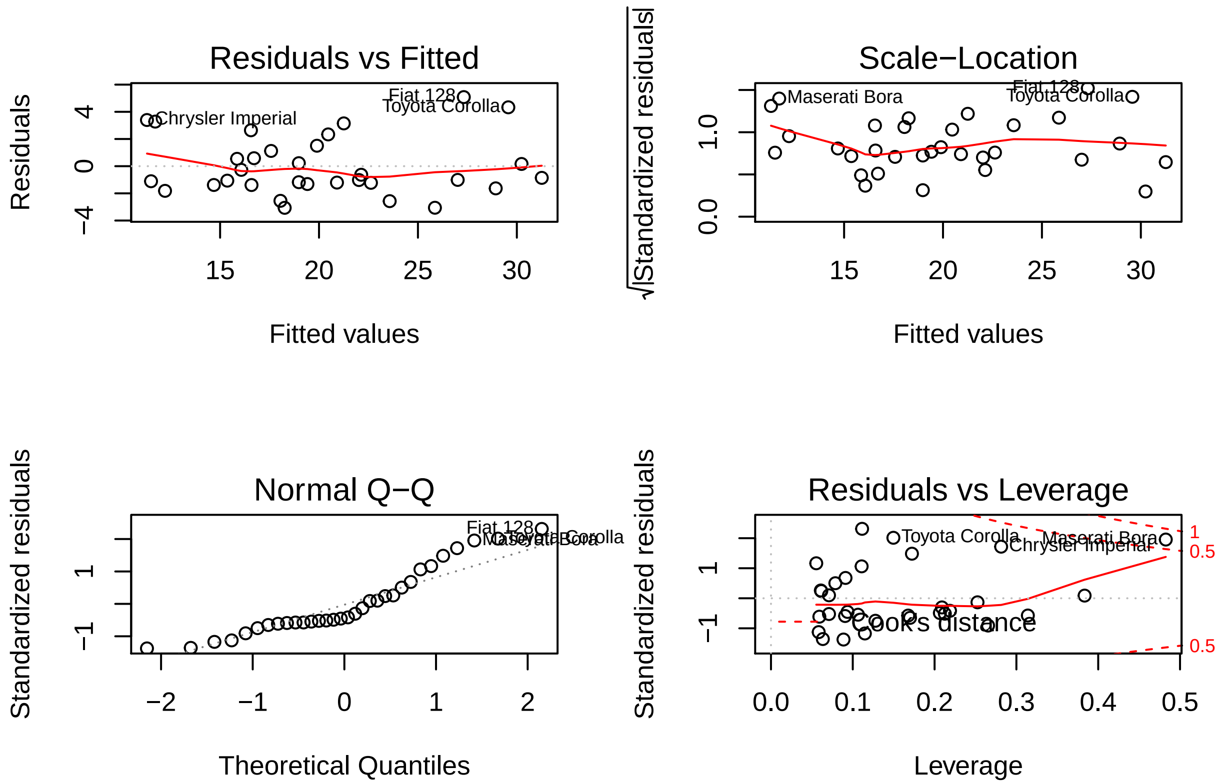

Figure 9: Diagnosis Plots for mpg vs. (wt + hp) * factor(am)

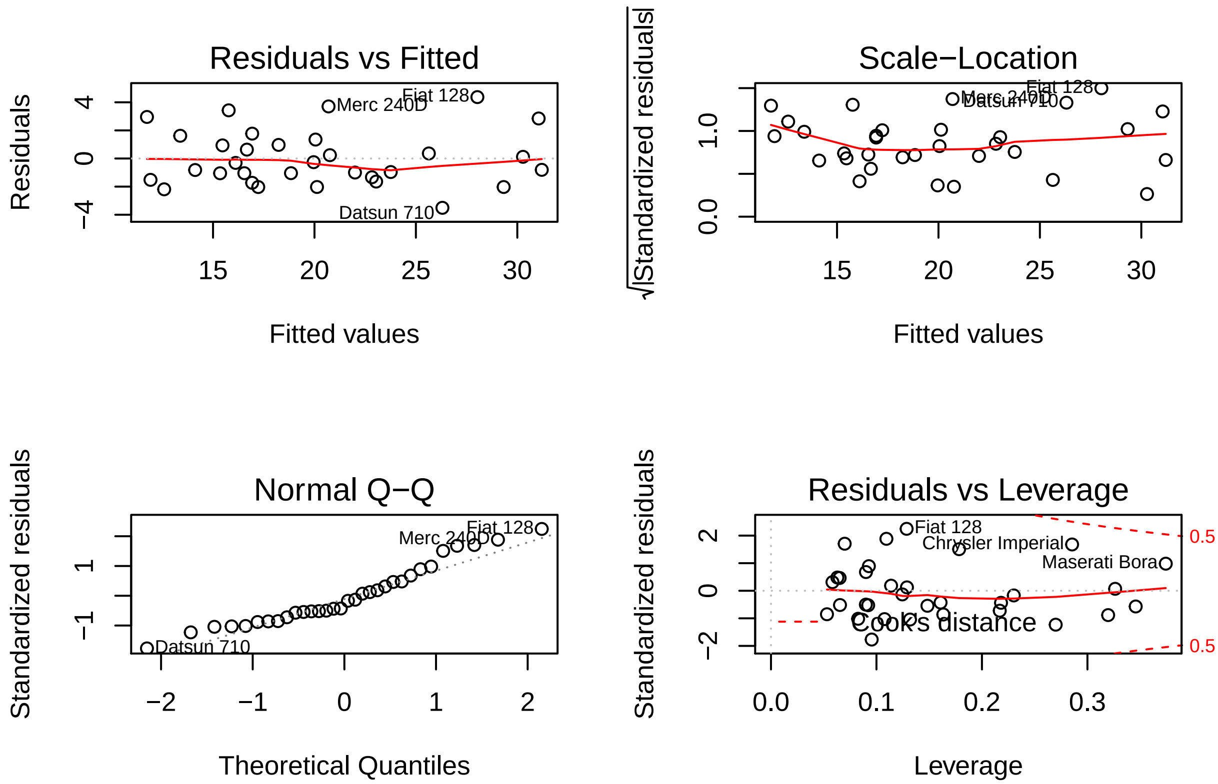

Figure 10: Diagnosis Plots for mpg vs. (wt + qsec) * factor(am)