Course project 4

Hazem Haffouz

2023-12-29

Load necessary libraries

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.3.2

## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode
```

Load the mtcars dataset

```
data(mtcars)
```

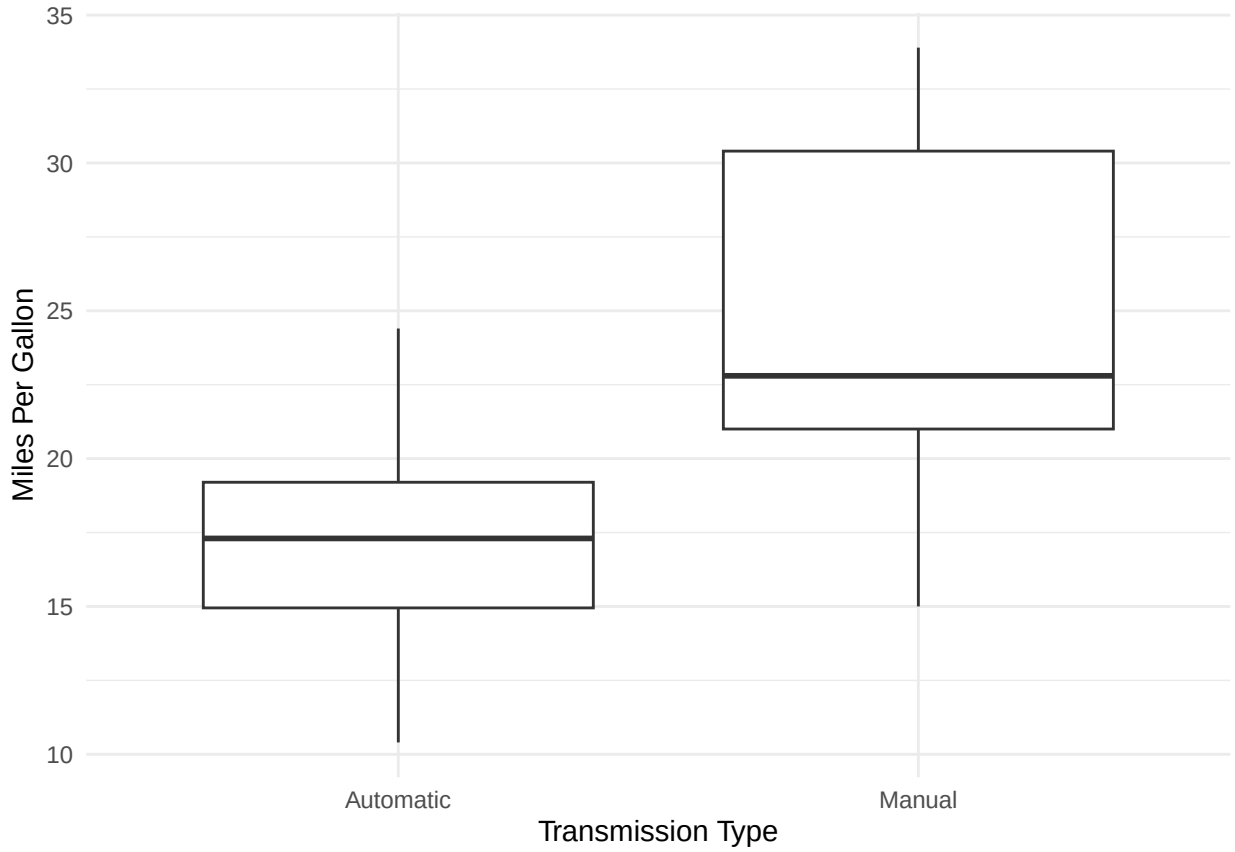
Modify the 'am' column to a factor for clarity in plots and models

```r
mtcars$am <- factor(mtcars$am, labels = c("Automatic", "Manual"))
```

1.Exploratory Data Analysis (EDA)

Boxplot of MPG by Transmission Type

```r
ggplot(mtcars, aes(x = am, y = mpg)) +
  geom_boxplot() +
  labs(x = "Transmission Type", y = "Miles Per Gallon") +
  theme_minimal()
```



2.Model Fitting

Simple model with transmission type

```r
simple_model <- lm(mpg ~ am, data = mtcars)
```

More complex model adjusting for other variables

```
complex_model <- lm(mpg ~ am + wt + hp + qsec, data = mtcars)
```

Compare models using AIC and BIC

```
aic_values <- AIC(simple_model, complex_model)[1, "AIC"]
bic_values <- BIC(simple_model, complex_model)[1, "BIC"]
```

Choose the model with the lowest AIC or BIC

```
chosen_model <- if(aic_values < bic_values) simple_model else complex_model
```

3.Coefficient Interpretation

Interpret the coefficient for transmission type

```
coef_summary <- summary(chosen_model)$coefficients
```

Print the coefficient summary

```
print(coef_summary)
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## amManual     7.244939   1.764422  4.106127 2.850207e-04
```

4.Quantify MPG Difference

Calculate the MPG difference between transmission types

```
mpg_difference <- coef(chosen_model)["amManual"]
```

5.Residual Diagnostics

Check residuals to ensure model assumptions are met

```r
library(car)

# Diagnostic Plots with the 'car' package
par(mfrow = c(2, 2))
# Use car::residualPlots to generate Residuals vs Fitted with better control
car::residualPlot(chosen_model)

# Use car::qqPlot to generate a QQ plot with better label management
car::qqPlot(chosen_model, main="Q-Q Plot")
```

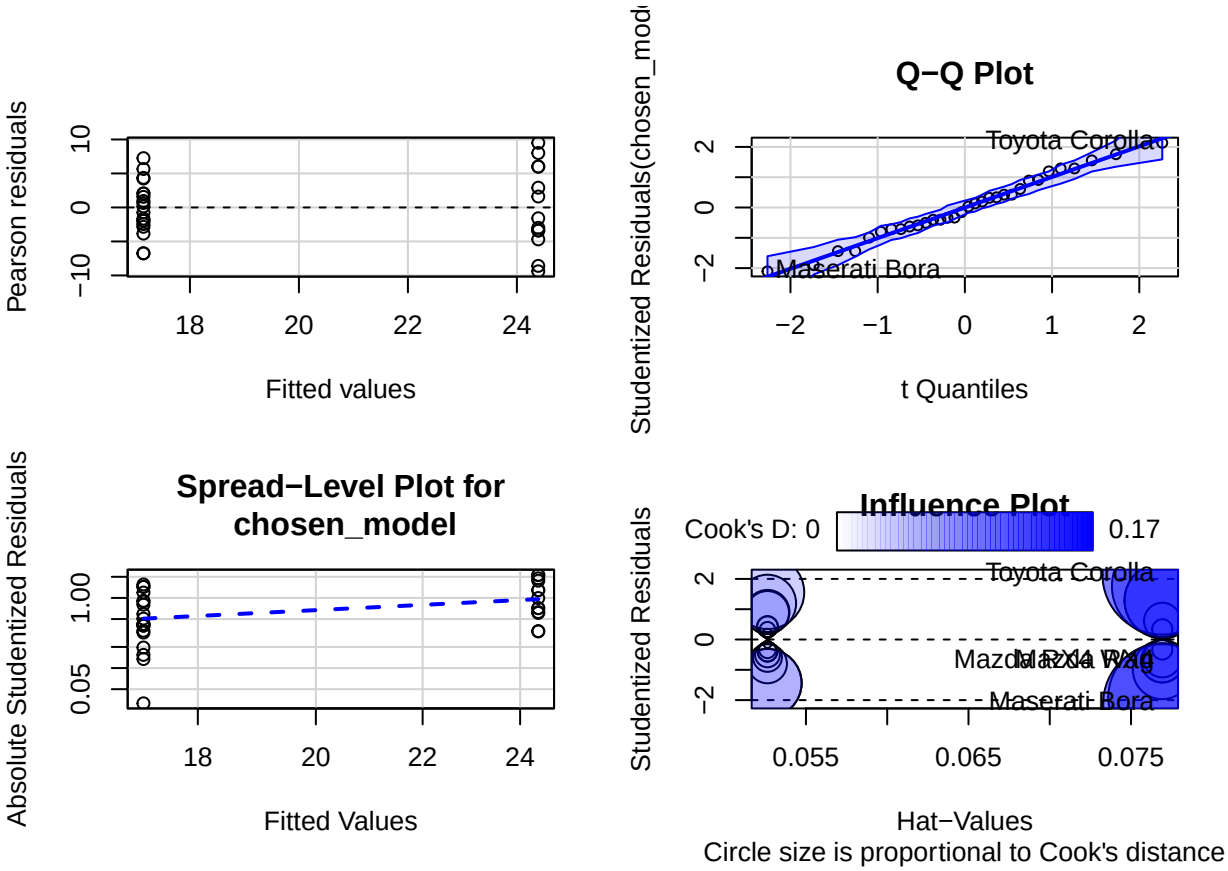
```
## Toyota Corolla  Maserati Bora
##             20             31
```

```r
# Use car::spreadLevelPlot to check for homoscedasticity
car::spreadLevelPlot(chosen_model)
```

```
##
## Suggested power transformation:  -0.814442
```

```r
# Influence Plot
car::influencePlot(chosen_model, id.method="identify", main="Influence Plot", sub="Circle size is propo
```

```
## Warning in plot.window(...): "id.method" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "id.method" is not a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "id.method" is not
## a graphical parameter

## Warning in axis(side = side, at = at, labels = labels, ...): "id.method" is not
## a graphical parameter

## Warning in box(...): "id.method" is not a graphical parameter

## Warning in title(...): "id.method" is not a graphical parameter

## Warning in plot.xy(xy.coords(x, y), type = type, ...): "id.method" is not a
## graphical parameter
```




```
##                    StudRes        Hat      CookD
## Mazda RX4        -0.714376 0.07692308 0.02161671
## Mazda RX4 Wag    -0.714376 0.07692308 0.02161671
## Toyota Corolla    2.135121 0.07692308 0.16980457
## Maserati Bora    -2.105221 0.07692308 0.16570811
```

```
# Reset to default parameters after plotting
par(mfrow = c(1, 1))
```

6.Uncertainty Quantification and Inference

Use confidence intervals to quantify the uncertainty around the MPG difference

```
mpg_diff_confint <- confint(chosen_model, "amManual")
```

Print results to the console

```
cat("AIC Comparison:\n", aic_values, "\n\n")
```

```
## AIC Comparison:
##  196.4844
```

```r
cat("BIC Comparison:\n", bic_values, "\n\n")
```

```
## BIC Comparison:
##  200.8816
```

```r
cat("Coefficient for Manual Transmission:\n", mpg_difference, "\n\n")
```

```
## Coefficient for Manual Transmission:
##  7.244939
```

```r
cat("95% Confidence Interval for MPG Difference:\n", mpg_diff_confint, "\n\n")
```

```
## 95% Confidence Interval for MPG Difference:
##  3.64151 10.84837
```

Reset par to default

```r
par(mfrow = c(1, 1))
```