Transmission Type vs Miles per Gallon Analysis

Marcus Vinicius Monteiro de Souza

28 de junho de 2017

Overview

In this project we'll analyze the Motor Trend Car Road Tests data set to answer the following questions:

1. Is an automatic or manual transmission better for MPG?
2. How can we quantify the MPG difference between automatic and manual transmissions?

Preparations

In this section, we'll install and load the required R packages:

```
if (!require('pacman')) {
  install.packages('pacman')
}
```

```
## Loading required package: pacman
```

```
pacman::p_load(GGally, lsr, plyr, tidyverse)
```

and load and preprocess the data for the analysis:

```
data("mtcars")

mtcars <- mtcars %>%
  mutate(cyl = as.factor(cyl), vs = as.factor(vs),
         am = as.factor(am), am = revalue(am, c('0' = 'automatic', '1' = 'manual')),
         gear = as.factor(gear), carb = as.factor(carb))
```

Exploring the data

The `mtcars` data has 32 observations of the following 11 variables:

Variable Name	Description
mpg	Miles/(US) gallon
cyl	Number of cylinders
disp	Displacement (cu.in.)
hp	Gross horsepower
drat	Rear axle ratio
wt	Weight (1000 lbs)
qsec	1/4 mile time
vs	V/S
am	Transmission (0 = automatic, 1 = manual)
gear	Number of forward gears
carb	Number of carburetors

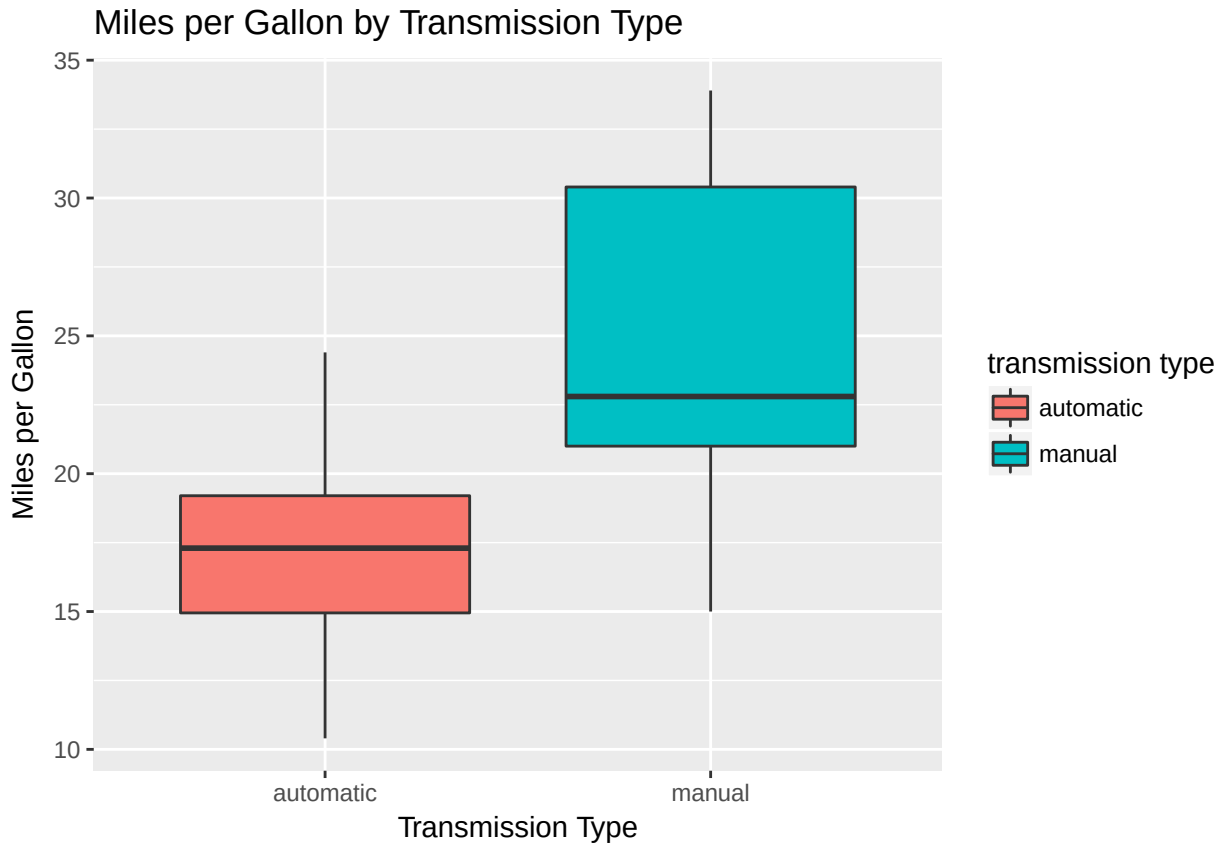
The observations look like this:

```
head(mtcars)
```

```
##    mpg cyl disp  hp drat    wt  qsec vs        am gear carb
## 1 21.0   6  160 110 3.90 2.620 16.46  0    manual    4    4
## 2 21.0   6  160 110 3.90 2.875 17.02  0    manual    4    4
## 3 22.8   4  108  93 3.85 2.320 18.61  1    manual    4    1
## 4 21.4   6  258 110 3.08 3.215 19.44  1 automatic    3    1
## 5 18.7   8  360 175 3.15 3.440 17.02  0 automatic    3    2
## 6 18.1   6  225 105 2.76 3.460 20.22  1 automatic    3    1
```

Let's see a raw data box plot of `mpg` by transmission type:

```
mtcars %>%
 ggplot(aes(am, mpg)) +
 geom_boxplot(aes(fill = am)) +
 labs(title = 'Miles per Gallon by Transmission Type', x = 'Transmission Type',
      y = 'Miles per Gallon', fill = 'transmission type')
```

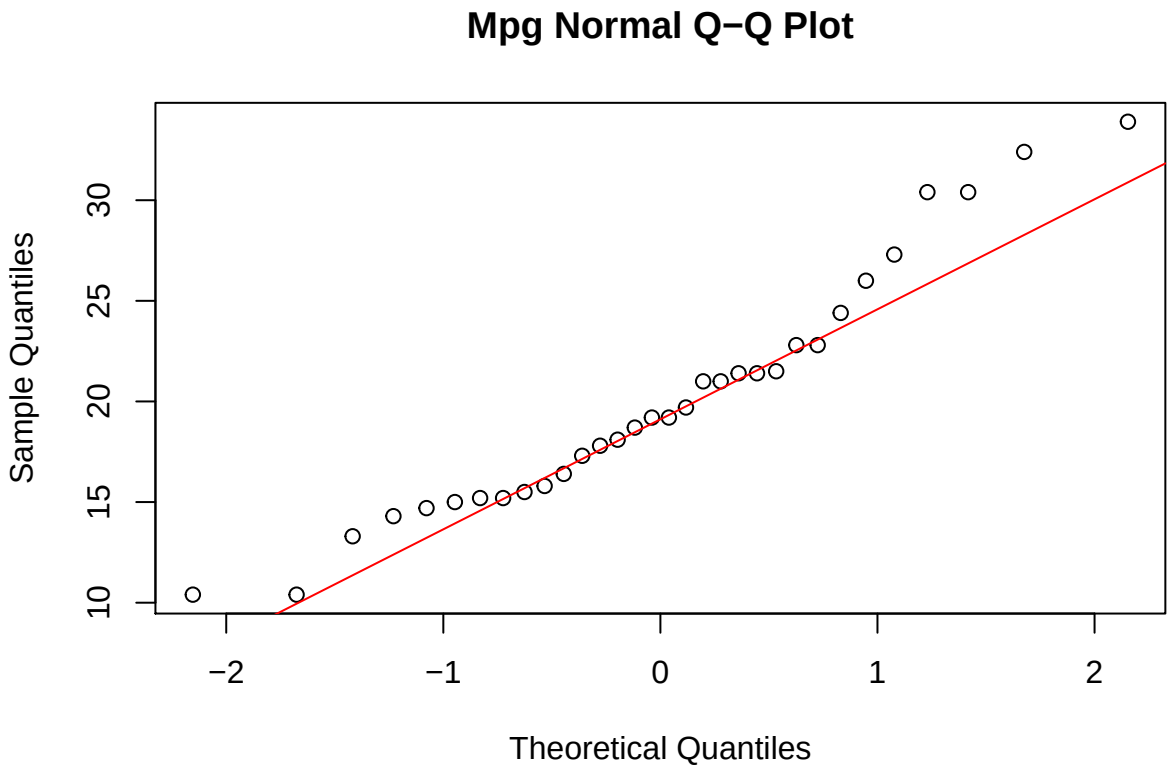


Apparently, manual transmission cars do more miles per gallon than automatic transmission ones... but is it really? Let's analyze it.

Is an automatic or manual transmission better for MPG?

To answer this question, we'll statistically compare the means of `mpg`s of both automatic and manual transmission populations. First, we'll check if `mpg` can be considered as normally distributed:

```r
qqnorm(mtcars$mpg, main = 'Mpg Normal Q-Q Plot')
qqline(mtcars$mpg, col = 2)
```



mpg's distribution appears normal. Now, because each observation was taken independently, let's do a unpaired, or independent samples test:

```r
independentSamplesTTest(mpg ~ am, mtcars)
```

```
##
##    Welch's independent samples t-test
##
## Outcome variable:   mpg
## Grouping variable:  am
##
## Descriptive statistics:
##             automatic manual
##    mean        17.147 24.392
##    std dev.     3.834  6.167
##
## Hypotheses:
##    null:        population means equal for both groups
##    alternative: different population means in each group
##
## Test results:
##    t-statistic:  -3.767
##    degrees of freedom:  18.332
##    p-value:  0.001
```

```
##
## Other information:
##    two-sided 95% confidence interval:  [-11.28, -3.21]
##    estimated effect size (Cohen's d):  1.411
```

The conclusion is: the mean `mpg` of automatic transmission cars was *17.147 (std dev = 3.834)*, whereas the mean `mpg` of manual transmission ones was *24.392 (std dev = 6.167)*. A Welch's independent samples t-tests showed that this *7.245* miles per gallon difference between the populations is was **significant** *(t(18.332) = -3.767, p = .001, CI95 = [-11.28, -3.21], d = 1.411)*, rejecting the null hypothesis that population means equal for both groups and suggesting that **manual transmission cars are better for `mpg`**.

How can we quantify the MPG difference between automatic and manual transmissions?

To answer this question we'll fit a multiple regression model for `mpg`, check the residuals and then draw our conclusion.

Model Selection

```
mpg.model <- lm(mpg ~ ., mtcars)
# Select a formula-based model by AIC
mpg.model <- step(mpg.model, trace = FALSE)
summary(mpg.model)
```

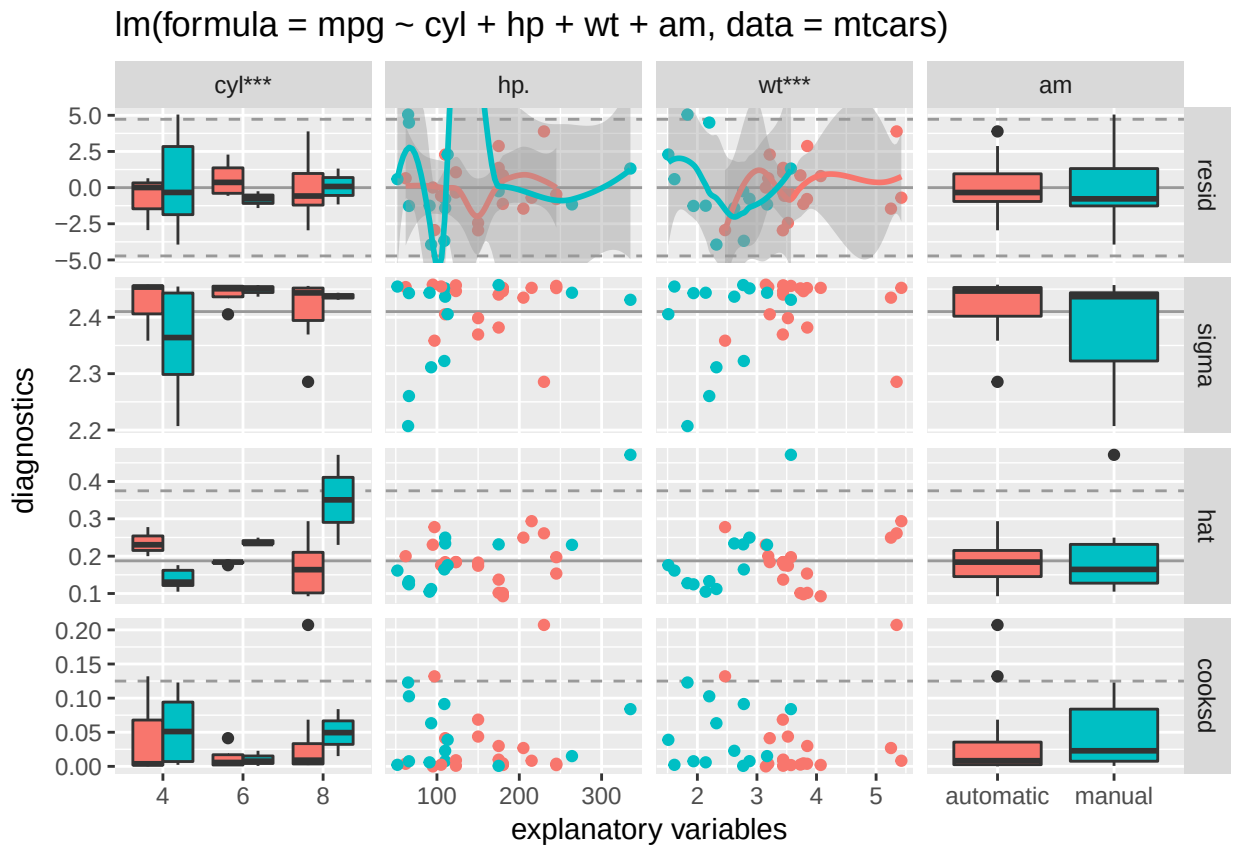
```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## 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 **
## ammanual     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
```

Checking the residuals

```
ggnostic(mpg.model, aes(color = am))
```

```
## Loading required package: broom
```

```
## `geom_smooth()` using method = 'loess'
## `geom_smooth()` using method = 'loess'
```



We can see that the residuals for `am` are within the lines, appearing to be normally distributed. We can now draw our conclusion.

Conclusion

Our final model has Adjusted R-squared value of 0.8401, meaning that it explains 84% of the variation in `mpg`. According to it, **manual transmission cars can run 1.80921 miles per gallon more than automatic transmission ones**.