

Transmissions:

Which will save you money at the pump?

Executive Summary

We are here to answer the question “are manual transmissions better on gas than automatic transmissions”. We used a dataset from the 1974 Motor Trend US magazine, and ran some statistical tests and a regression analysis.

Although manual transmission cars in general do have higher MPG, yet, the MPG difference with the automatic ones may vary considerably, strongly affected by other characteristics like the cylinders of the engine or the total car weight.

Data Manipulation

```
data(mtcars)
auto <- split(mtcars, mtcars$am)[[1]]
man <- split(mtcars, mtcars$am)[[2]]
cars <- mtcars

for (i in 1:length(cars$am)) {
  if (cars$am[i] == 0) {
    cars$am[i] <- "Automatic"
  } else {
    cars$am[i] <- "Manual"
  }
}

car4 <- split(cars, cars$cyl)[[1]]
car6 <- split(cars, cars$cyl)[[2]]
car8 <- split(cars, cars$cyl)[[3]]
```

Exploratory analysis

Looking at some simple box plots allow us to make a good judgment that manual transmissions are easier on the wallet than automatic transmissions. This first box plot shows the overall difference between automatic and manual transmissions.

I then wanted to look at the relation between the different cylinders. The following boxplots are split by cylinders.

As we split it by cylinders we can still see the trend that manual transmissions are more likely to produce higher miles per gallons. The 8 cylinder data however is inconclusive. The medians look about the same but there were only 2 observations for manual transmissions.

Statistical Analysis

T test can make a more confirmed estimation of our visual analysis.

```
##
## welch Two Sample t-test
##
```

```
## data: mpg by am
## t = -3.767, df = 18.33, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -11.28 -3.21
## sample estimates:
## mean in group Automatic      mean in group Manual
##                17.15                24.39
```

The mean MPG for the manual transmission cars is **24.3923**, and the automatic transmission is **17.1474**.

```
##
## welch Two Sample t-test
##
## data: mpg by am
## t = -2.885, df = 8.999, p-value = 0.01802
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -9.232 -1.118
## sample estimates:
## mean in group Automatic      mean in group Manual
##                22.90                28.07
```

The mean MPG for the manual transmission cars is **28.075**, and the automatic transmission is **22.9**.

```
##
## welch Two Sample t-test
##
## data: mpg by am
## t = -1.561, df = 4.405, p-value = 0.1871
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -3.916 1.033
## sample estimates:
## mean in group Automatic      mean in group Manual
##                19.12                20.57
```

The mean MPG for the manual transmission cars is **20.5667**, and the automatic transmission is **19.125**.

```
##
## welch Two Sample t-test
##
## data: mpg by am
## t = -0.391, df = 10.2, p-value = 0.7039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -2.34 1.64
## sample estimates:
## mean in group Automatic      mean in group Manual
##                15.05                15.40
```

The mean MPG for the manual transmission cars is **15.4**, and the automatic transmission is **15.05**.

Regression Analysis.

Quantifying the difference between automatic and manual transmission MPG I used linear regression.

```
simp_fit <- lm(mpg ~ am, data = mtcars)
summary(simp_fit)
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -9.392 -3.092 -0.297 3.244 9.508
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.15      1.12    15.25 1.1e-15 ***
## am              7.24      1.76     4.11 0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.9 on 30 degrees of freedom
## Multiple R-squared:  0.36,    Adjusted R-squared:  0.338
## F-statistic: 16.9 on 1 and 30 DF,  p-value: 0.000285
```

This model does not offer us any additional information than that of the T test. The **Intercept** (17.1474) represents the mean MPG for automatic cars, while the **AM coefficient** (7.2449), is the difference between the MPG of manual and automatic transmission cars, which matched exactly the results the T test produced. We therefore conclude that the *MPG* is affected by additional variables. In order to determine these variables we shall examine the correlation between the MPG and the rest of the variables.

```
sort(cor(mtcars)[1, ])
##      wt      cyl      disp      hp      carb      qsec      gear      am      vs
## -0.8677 -0.8522 -0.8476 -0.7762 -0.5509  0.4187  0.4803  0.5998  0.6640
##      drat      mpg
##  0.6812  1.0000
```

We see that the top variables related to *MPG* are: Weight (wt), Number of Cylinders (cyl), Engine Displacement (disp) and Horse Power (hp), with considerable differences to the rest of the variables (keep in mind that in correlation, low negative numbers are as important as high positive ones). However, displacement is strongly related to the number of cylinders, so we shall include only the “wt”, “cyl” and “hp” variables in a new model that we shall create and compare to the simple *MPG~AM* model.

```
new_fit <- lm(mpg ~ wt + cyl + hp + am, data = mtcars)
best_fit <- anova(simp_fit, new_fit)
print(best_fit)
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ wt + cyl + hp + am
##   Res.Df RSS Df Sum of Sq    F Pr(>F)
## 1      30 721
## 2      27 170  3      551 29.2 1.3e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The comparison shows a very small **p-value** of 1.2742×10^{-8} , which allows us to reject the **Null Hypothesis** that the simple model may explain the *MPG* variable better, and thus opt for the multivariable model.

Residuals

Before concluding, we should examine the residuals plot.

```
par(mfrow = c(2, 2))
plot(new_fit)
```

Generally we cannot notice anything extremely peculiar. The *Normal Q-Q* plot shows that the residuals are normally distributed and the *Residuals vs Fitted* and *Scale Location* plots do not show any specific patterns. We may notice some outliers in the *Residuals vs Leverage* plot, but these are well within the Cook's Distance limits, so even if we removed them it would not change our model significantly.

In Conclusion

Lastly, we shall analyze the new multivariable model to see how *MPG* is affected by each variable.

```
summary(new_fit)
##
## Call:
## lm(formula = mpg ~ wt + cyl + hp + am, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.476 -1.847 -0.554  1.276  5.661
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   36.1465     3.1048   11.64 4.9e-12 ***
## wt           -2.6065     0.9198   -2.83  0.0086 **
## cyl           -0.7452     0.5828   -1.28  0.2119
## hp            -0.0250     0.0136   -1.83  0.0786 .
## am             1.4780     1.4411    1.03  0.3142
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.51 on 27 degrees of freedom
## Multiple R-squared:  0.849, Adjusted R-squared:  0.827
## F-statistic: 38 on 4 and 27 DF, p-value: 1.02e-10
```

First of all the *R-squared* is 0.849, indicating that a great part of the variance is explained by this model. According to the model, weight and number of cylinders have a negative effect on mileage. Specifically,

- The *MPG* of a car decreases by **-2.6065** for every **1000lbs** of extra weight.
- Each extra cylinder decreases the *MPG* by a factor of **-0.7452**.
- The car's HP does not affect consumption dramatically.
- Generally, cars with manual transmission, tend to produce **1.478** times more *miles per gallon*, compared with automatic transmission cars with the same weight, cylinders and horsepower.