# Do Manual Transmission Cars Have Better Gas Mileage?

Rhowell

2022-10-04

## Contents



*Created with Knitr*

### Executive Summary

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

`"Is an automatic or manual transmission better for MPG"`

`"Quantify the MPG difference between automatic and manual transmissions"`

We used a variety of comparison methods (t test, simple linear regression, and multivariate linear regression) to quantify the difference in mpg between automatic and manual transmissions.

We concluded that one can expect approximately 3 mpg advantage when one selects a manual transmission car vice an automatic transmission car.

### Data Description

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

A data frame with 32 observations on 11 (numeric) variables.

- [, 1] mpg Miles/(US) gallon
- [, 2] cyl Number of cylinders
- [, 3] disp Displacement (cu.in.)
- [, 4] hp Gross horsepower
- [, 5] drat Rear axle ratio
- [, 6] wt Weight (1000 lbs)
- [, 7] qsec 1/4 mile time
- [, 8] vs Engine (0 = V-shaped, 1 = straight)
- [, 9] am Transmission (0 = automatic, 1 = manual)
- [,10] gear Number of forward gears
- [,11] carb Number of carburetors

## Load Data

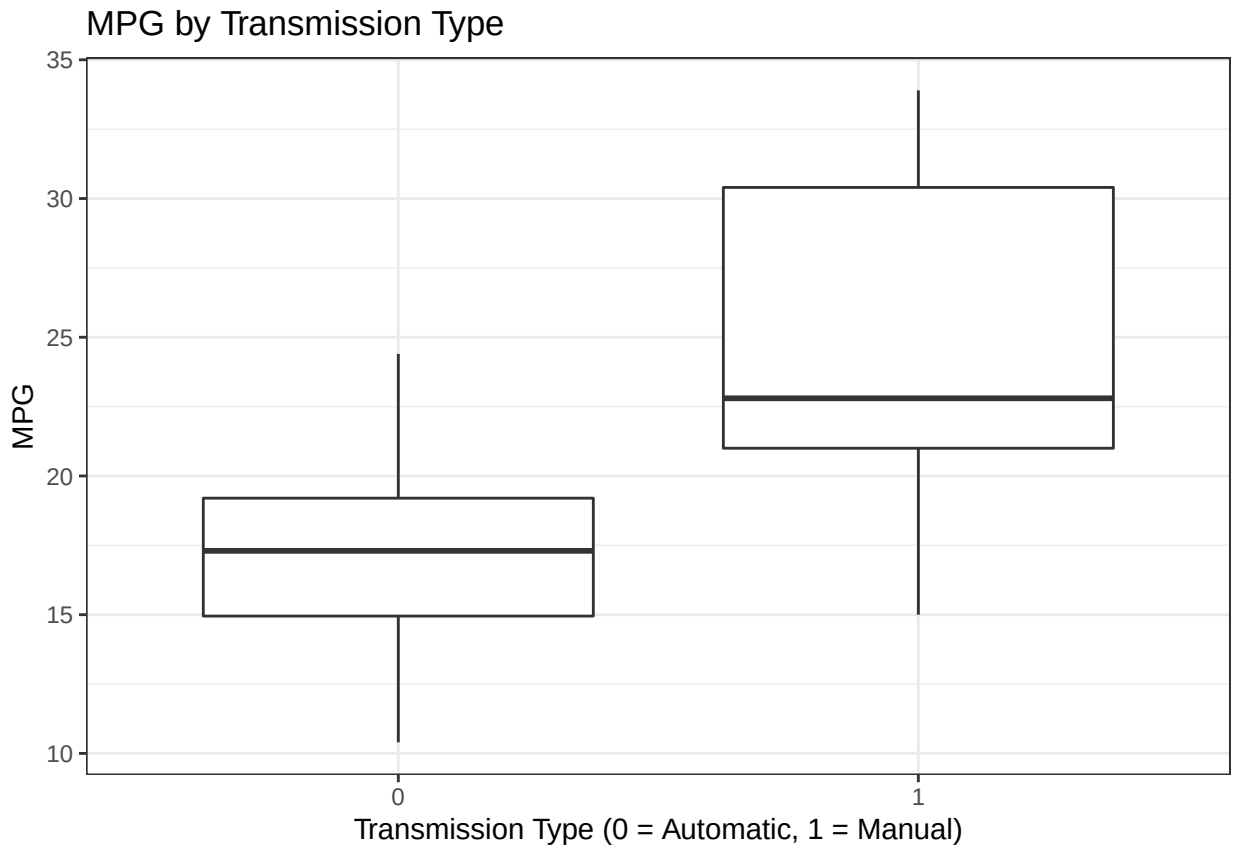
```
library(datasets)
data(mtcars)
str(mtcars)
```

```
## '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  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

## Data Analysis

How does the transmission type affect gas mileage?

```
ggplot(data = mtcars, aes(as.factor(am), mpg)) +
  geom_boxplot() +
  theme_bw() +
  xlab("Transmission Type (0 = Automatic, 1 = Manual)") +
  ylab("MPG") +
  labs(title="MPG by Transmission Type")
```




Visually, it appears that manual transmissions do have better gas mileage than automatics.

### Significant Statistics

Let's next test statistically if there is a difference between automatic and manual transmissions. We will perform a hypothesis test, with the automatic transmission as the null, and manual if we reject the null.

```
t.test(mpg ~ am, data = mtcars, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  mpg by am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group 0 mean in group 1
##        17.14737        24.39231
```

Since the p-value is 0.0013 (less than $p < 0.005$), and the confidence interval does not contain 0, we can reject the null hypothesis, restating, we can conclude that there is statistically significant difference between manual and automatic transmissions. This supports our visual estimation above.

## Linear Regression

If we continue to assess with different variables, more patterns may emerge. Whereas testing for differences in just the transmission did seem significant, we can take into account confounding variables.

Beginning with a simple linear regression:

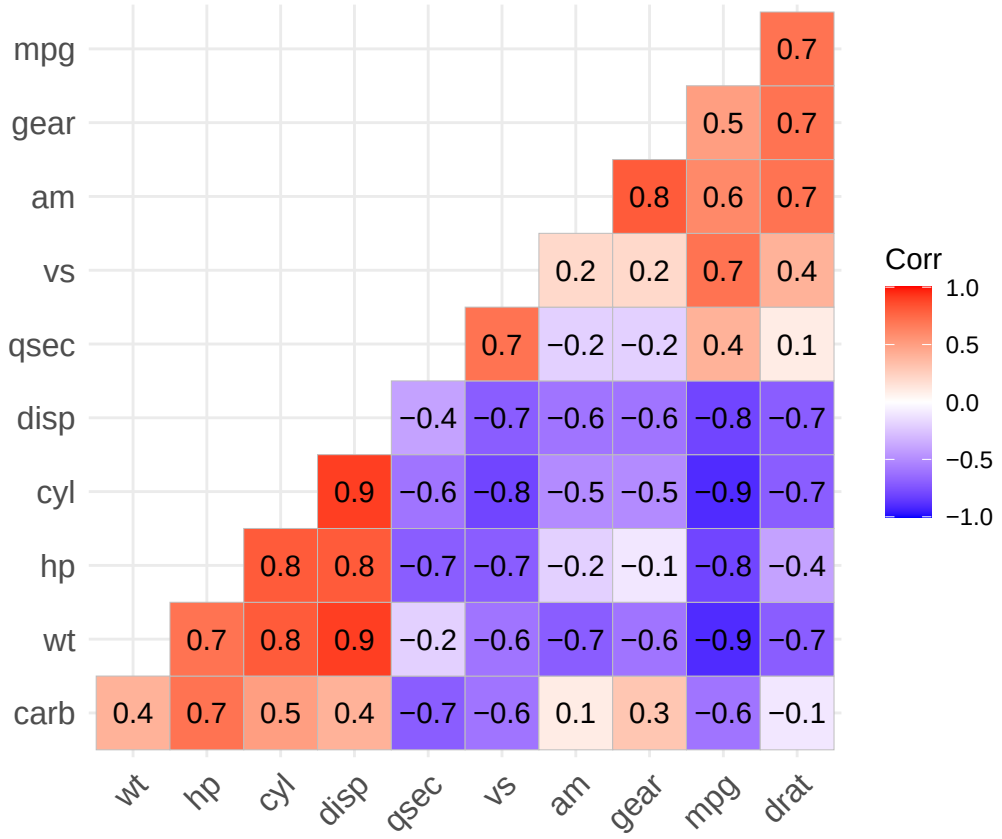
```
summary(lm(mpg ~ as.factor(am), mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ as.factor(am), data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      17.147      1.125  15.247 1.13e-15 ***
## as.factor(am)1    7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

We see the $R^2$ value is 0.338, so this one variable linear regression model only accounts for 34% of the variance in fuel consumption.

The next question we want to ask, is what other variables correlate to gas mileage?

```
correlation_matrix <- round(cor(mtcars),1)
corrp.mat <- cor_pmat(mtcars)
ggcorrplot(correlation_matrix, hc.order =TRUE,
           type ="lower", lab =TRUE)
```




It appears that number of gears, transmission, engine shape (vs), and quarter mile time (qsec) all correlate with mpg.

## Multiple Linear Regression

Can we fit a better model with multiple variables? We'll create a stepwise regression using the `stepAIC()` from the `MASS` package to find the variables which result in the best fit.

```
initialModel <- lm(mpg ~ ., data = mtcars)
stepReg <- stepAIC(initialModel, direction = "both")
# note: results are hidden
```

```
print(stepReg$anova)
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##
## Final Model:
## mpg ~ wt + qsec + am
##
##
##     Step Df   Deviance Resid. Df Resid. Dev      AIC
```

```
## 1                            21   147.4944 70.89774
## 2  - cyl  1 0.07987121        22   147.5743 68.91507
## 3   - vs  1 0.26852280        23   147.8428 66.97324
## 4 - carb  1 0.68546077        24   148.5283 65.12126
## 5 - gear  1 1.56497053        25   150.0933 63.45667
## 6 - drat  1 3.34455117        26   153.4378 62.16190
## 7 - disp  1 6.62865369        27   160.0665 61.51530
## 8   - hp  1 9.21946935        28   169.2859 61.30730
```

The model with the best fit accounts for weight and quarter mile time in addition to transmission.

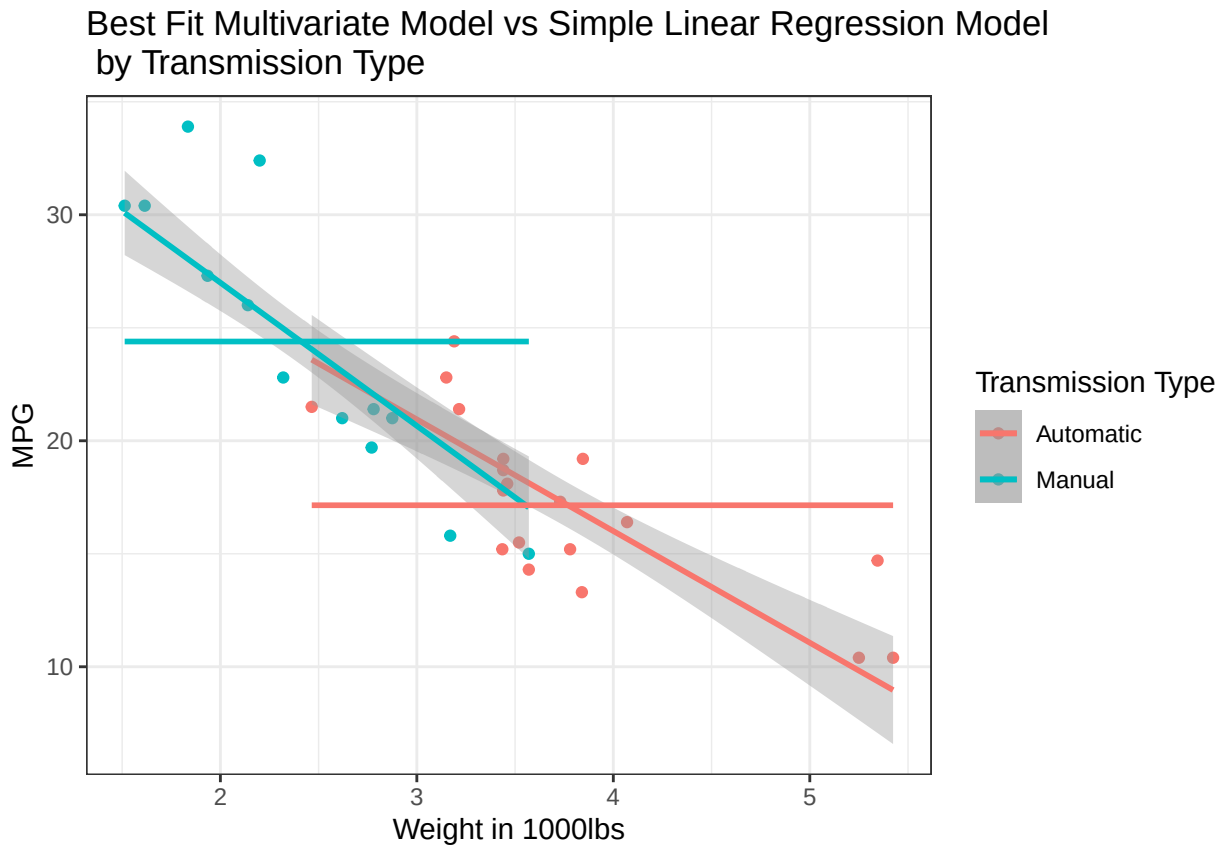
```
summary(bestModelFit <- lm(mpg ~ wt + qsec + as.factor(am), data = mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + as.factor(am), data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      9.6178     6.9596   1.382 0.177915
## wt              -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec             1.2259     0.2887   4.247 0.000216 ***
## as.factor(am)1   2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

Now our $R^2$ value is .833, so our adjusted model accounts for 83% of variance in gas mileage.

```
ggplot(mtcars, aes(wt, mpg, color = as.factor(am))) +
  geom_point() +
  geom_smooth(aes(y = predict(bestModelFit, mtcars)),method = lm) +
  geom_smooth(aes(y = predict(lm(mpg ~ as.factor(am), mtcars))), method = lm) +
  theme_bw() +
  labs(title = "Best Fit Multivariate Model vs Simple Linear Regression Model \n by Transmission Type")
  xlab("Weight in 1000lbs") +
  ylab("MPG") +
  scale_color_discrete(name = "Transmission Type", labels = c("Automatic", "Manual"))
```

```
## `geom_smooth()` using formula 'y ~ x'
## `geom_smooth()` using formula 'y ~ x'
```




## Residuals

The plots of the residuals support the conclusion that accounting for multiple variables accounts for more of the variance.

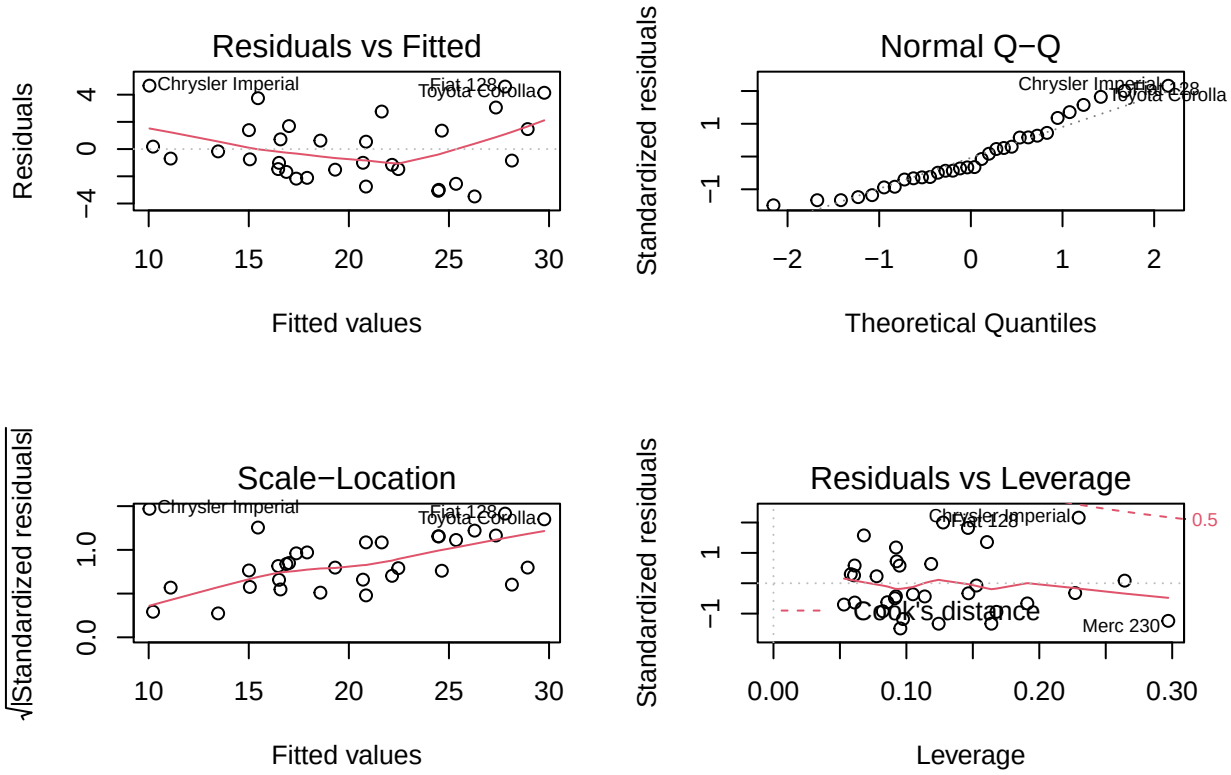
The points in the *Residuals vs. Fitted* plot appear to be random which indicate the data are independent. The plot also suggests three potential outliers for the Chrysler Imperial, Fiat 128, and Toyota Corolla.

The points of the *Normal Q-Q* plot follow the line indicating that the residuals are normally distributed.

The points on the *Scale-Location* plot appear to be spread equally along the line with random points spread out allowing us to conclude equal variance (homoscedasticity).

The *Residuals vs. Leverage* plot doesn't show any influential points. All points are within the 0.05 lines which conclude there are no outliers.

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




## Conclusion

1. **Is an automatic or manual transmission better for MPG?**

Yes, one can expect better gas mileage with a manual transmission vice an automatic transmission. There were a few confounding variables, such as weight and quarter mile time in addition to transmission type that better predict variance in gas mileage.

2. **Quantify the MPG difference between automatic and manual transmissions.**

Both the t test and the simple linear regression showed that there is a 7.25 mpg premium when selecting a manual transmissions vice automatic transmissions.

However, by using the best fitting model (holding weight and quarter mile time constant), the premium dropped to 2.94 mpg when selecting a manual transmission vice automatic transmission.