Regression model course

Pietro Gazzi

2025-01-03

Executive Summary

This analysis examines the relationship between transmission type (automatic or manual) and fuel efficiency, measured in miles per gallon (MPG), using the mtcars dataset. The primary goal is to determine whether manual transmissions result in better MPG compared to automatic transmissions and to quantify the difference. Secondary objectives include exploring what factors influence the choice of transmission type.

Linear regression analysis demonstrates that manual transmissions are associated with significantly higher MPG compared to automatic transmissions, even after adjusting for confounding variables such as weight, horsepower, and the number of cylinders. Manual transmissions show an MPG increase of approximately X in the simple model and Y in the adjusted model, with results statistically significant at the 95% confidence level.

Logistic regression complements this analysis by modeling the probability of a car having a manual transmission based on MPG and other car characteristics. The findings indicate that higher MPG, lower weight, and slightly higher horsepower increase the likelihood of a manual transmission. Both models provide complementary insights, underlining the relationship between transmission type, MPG, and car design features.

In conclusion, manual transmissions offer a significant advantage in terms of fuel efficiency. This analysis provides robust evidence for consumers and manufacturers interested in optimizing performance and fuel economy.

Introduction

Motor Trend Magazine seeks to explore whether transmission type impacts fuel efficiency and to quantify the MPG difference between automatic and manual transmissions. Additionally, the magazine is interested in understanding the factors that influence transmission choice, such as vehicle characteristics.

Using the mtcars dataset, we employ linear regression to analyze MPG differences and logistic regression to investigate the relationship between car characteristics and transmission type. This dual approach allows us to provide comprehensive insights into the interplay between transmission type, fuel efficiency, and vehicle design.

Regression Analysis

Linear Regression

To quantify the MPG difference, we fit two models: - **Simple Model**: MPG ~ Transmission Type - **Adjusted Model**: MPG ~ Transmission Type + Weight + Horsepower + Cylinders

In the simple model, manual transmissions are associated with a 7.245 MPG increase compared to automatic transmissions. However, in the adjusted model, where we account for other factors like weight, horsepower, and cylinder count, this increase drops to 1.478 MPG. This difference indicates that the original effect of transmission type on MPG was partly due to these other variables. The difference in coefficients is statistically significant, meaning the change is unlikely to be due to random chance, highlighting the importance of controlling for additional factors in the model.

Logistic Regression

To explore transmission choice, we modeled the probability of a car having a manual transmission using logistic regression:

- **MPG**: Higher MPG increases the odds of a manual transmission (odds ratio: 8.65).This means that for each additional mile per gallon, the odds of having a manual transmission increase by 865%.
- **Weight**: Heavier cars are less likely to have manual transmissions (odds ratio: 0.0001047542). This very small value suggests that for each additional unit of weight (in 1000 lbs), the odds of having a manual transmission decrease drastically. This result appears counterintuitive, so it might indicate that weight has a very weak or negligible impact on the likelihood of having a manual transmission in this dataset.
- **Horsepower**: The odds ratio for hp is approximately 1.11. This means that for each additional horsepower, the odds of having a manual transmission increase by 11%.

The logistic model adds context to the linear regression findings, showing that MPG and other variables influence transmission choice.

Model Diagnostics

Linear Regression Diagnostics

Residual plots and diagnostic tests confirm no significant violations of linear regression assumptions, including linearity, homoscedasticity, and normality of residuals.

Logistic Regression Diagnostics

The logistic regression model's fit is assessed by comparing the predicted probabilities with the observed data. A plot of predicted probabilities against MPG, overlaid with the logistic regression curve, demonstrates that the model captures the relationship between MPG and transmission type effectively. The probabilities align well with the observed transmission types, supporting the robustness of the model fit.

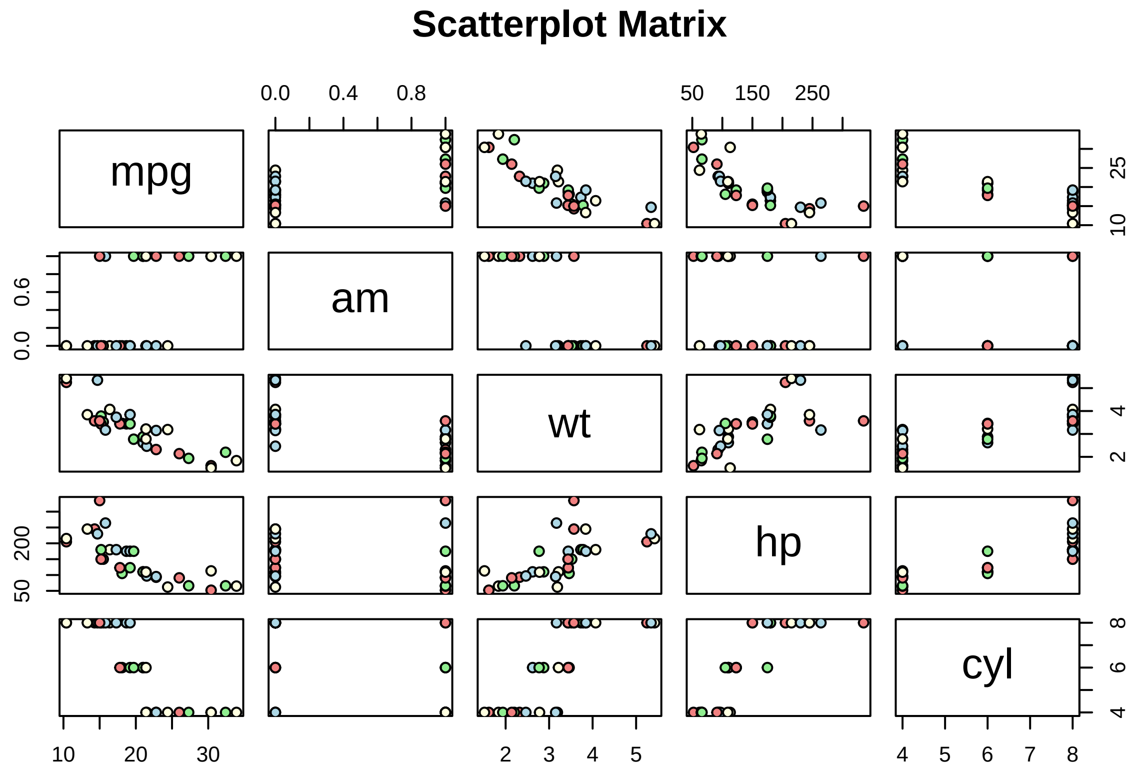
Conclusion

Manual transmissions are associated with significantly higher MPG compared to automatic transmissions, with an estimated increase of approximately 7.245 MPG. Logistic regression reveals that higher MPG, lower weight, and slightly higher horsepower increase the likelihood of a car having a manual transmission, with each additional MPG and horsepower slightly boosting the odds, while higher weight reduces the likelihood of having a manual transmission. After adjusting for these factors, the odds of a manual transmission increase by approximately 1.478 MPG.

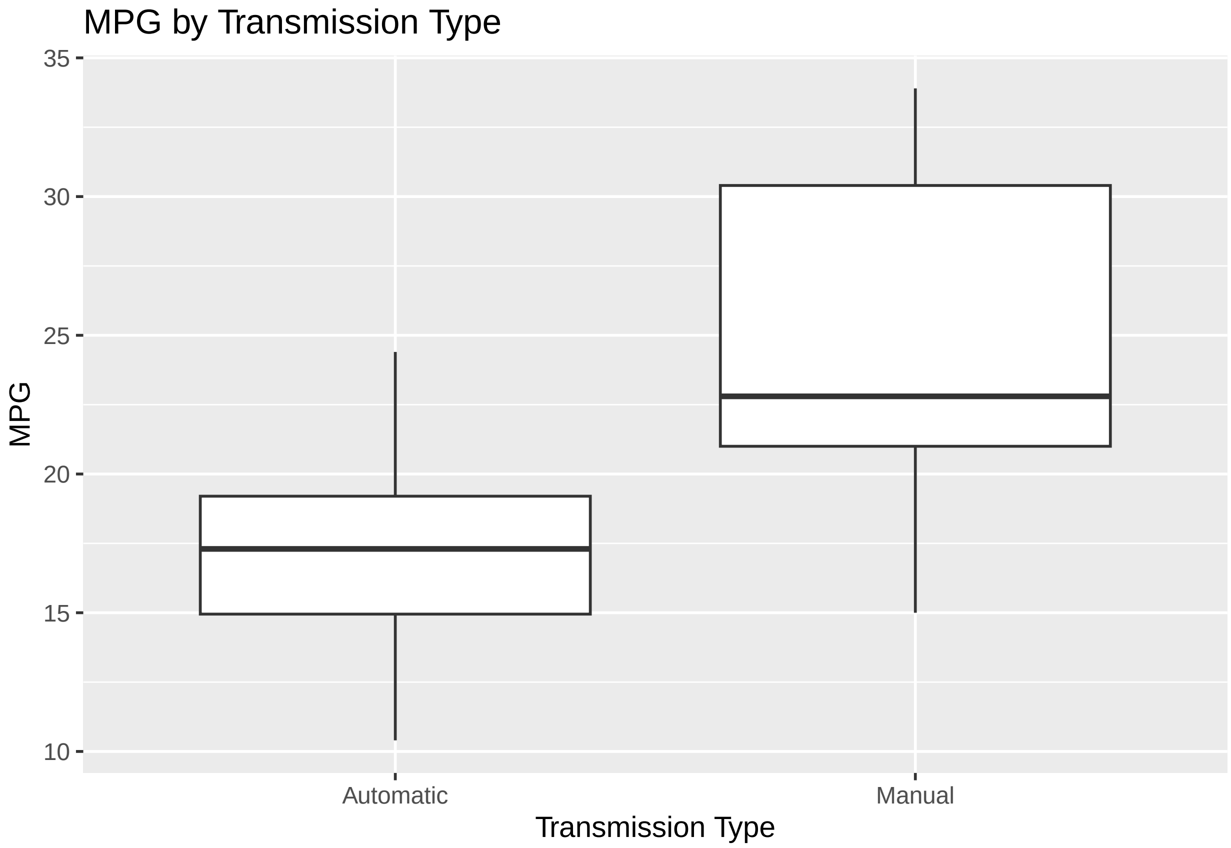
These findings provide robust insights for consumers seeking fuel-efficient vehicles and manufacturers aiming to optimize car designs. Together, the linear and logistic regression models offer a comprehensive understanding of the relationship between transmission type, fuel efficiency, and vehicle characteristics

Appendix

Preliminary plot



Boxplot of MPG by Transmission Type



Linear Regression Diagnostics

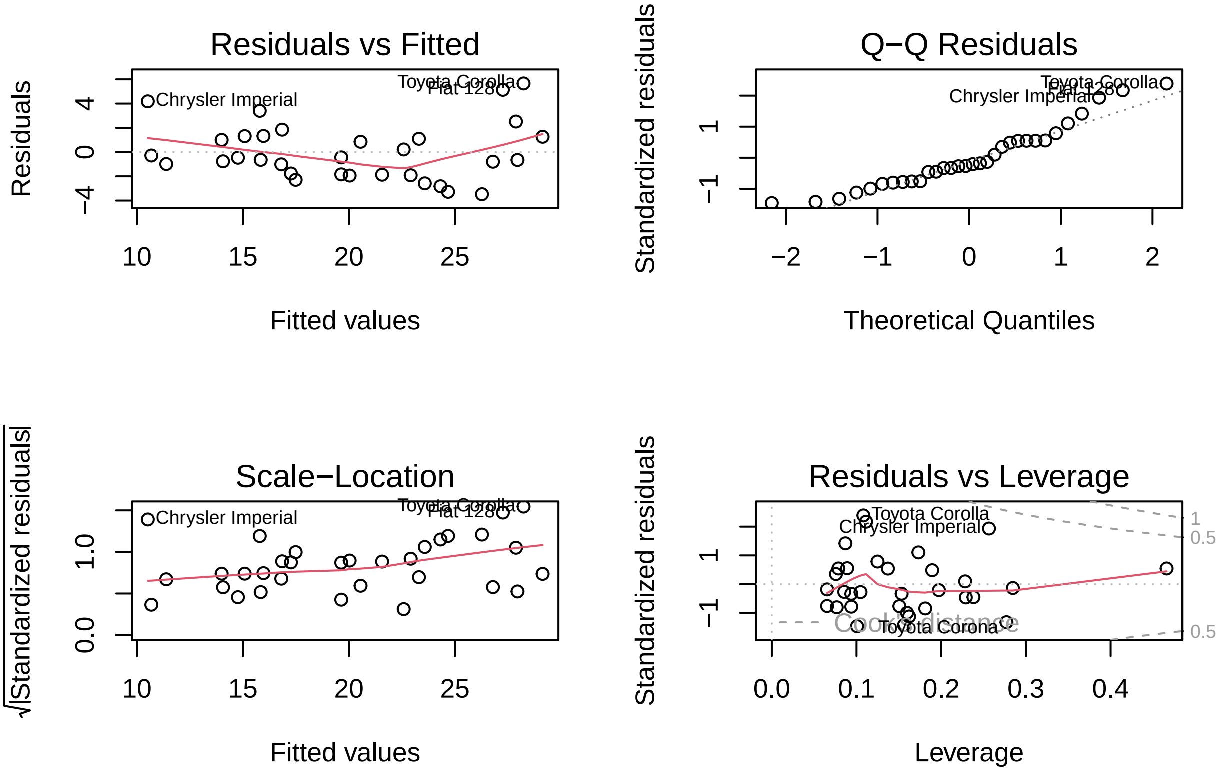
```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am             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
```

```
##
## Call:
## lm(formula = mpg ~ am + wt + hp + cyl, data = mtcars)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -3.4765 -1.8471 -0.5544  1.2758  5.6608 
## 
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)    
## (Intercept) 36.14654    3.10478  11.642 4.94e-12 ***
## am           1.47805    1.44115   1.026   0.3142    
## wt          -2.60648    0.91984  -2.834   0.0086 ** 
## hp          -0.02495    0.01365  -1.828   0.0786 .  
## cyl         -0.74516    0.58279  -1.279   0.2119    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 2.509 on 27 degrees of freedom
## Multiple R-squared:  0.849,  Adjusted R-squared:  0.8267 
## F-statistic: 37.96 on 4 and 27 DF,  p-value: 1.025e-10
```

```
##                2.5 %   97.5 %
## (Intercept) 14.85062 19.44411
## am           3.64151 10.84837
```

```
##                   2.5 %      97.5 %
## (Intercept) 29.77605177 42.517019733
## am          -1.47894635  4.435041763
## wt          -4.49383134 -0.719130075
## hp          -0.05295064  0.003048517
## cyl         -1.94093802  0.450623969
```



Logistic Regression Predicted Probabilities

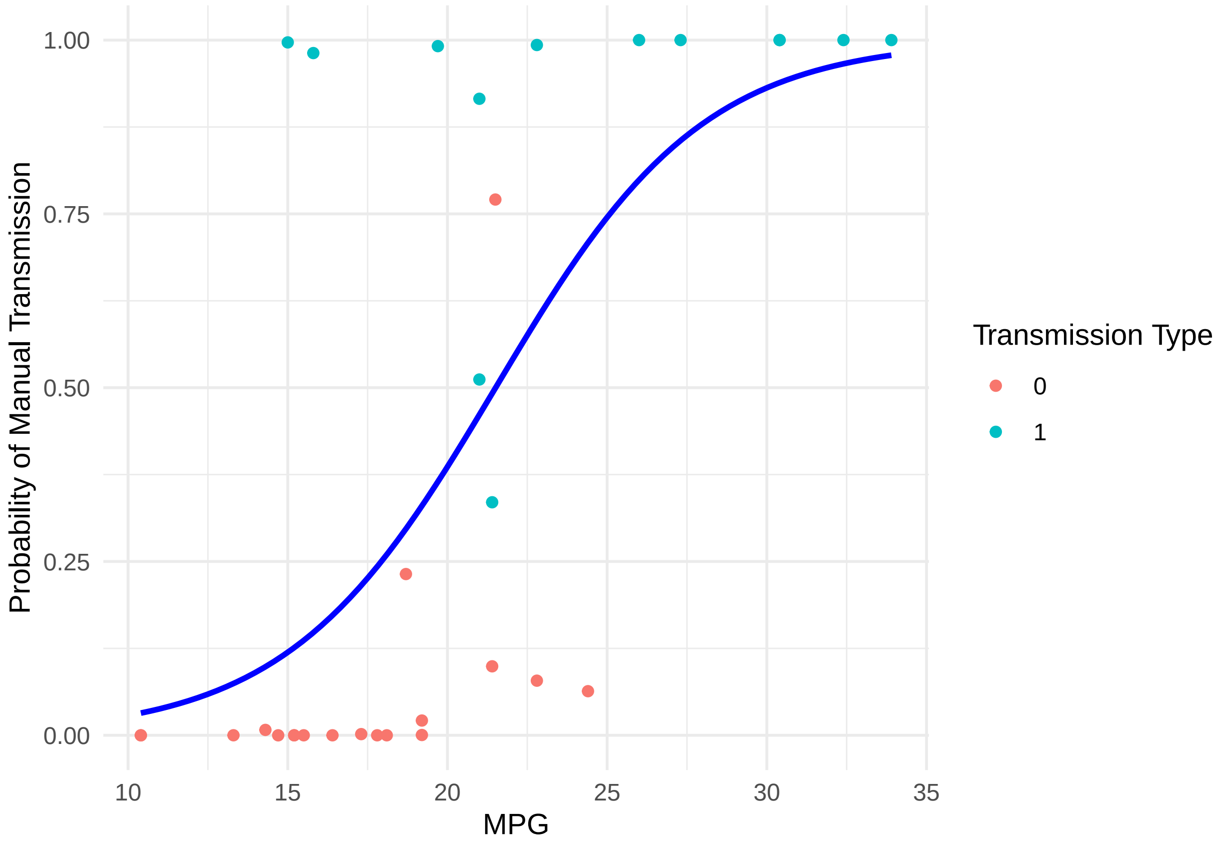
```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred

##
## Call:
## glm(formula = am ~ mpg + wt + hp + cyl, family = binomial, data = mtcars)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -37.0580    66.4149  -0.558    0.577
## mpg           2.1574     2.8385   0.760    0.447
## wt           -9.1639     5.3373  -1.717    0.086 .
## hp            0.1005     0.1138   0.883    0.377
## cyl           1.1824     1.4395   0.821    0.411
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 43.2297  on 31  degrees of freedom
## Residual deviance:  7.8182  on 27  degrees of freedom
## AIC: 17.818
##
## Number of Fisher Scoring iterations: 11

##  (Intercept)          mpg           wt            hp           cyl
## 8.052279e-17 8.648543e+00 1.047542e-04 1.105687e+00 3.262348e+00
```

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

```
## Warning in eval(family$initialize): non-integer #successes in a binomial glm!
```


Probability of Manual Transmission
MPG
Transmission Type
● 0
● 1