Motor Trend

Tandin Dorji

2023-12-25

Executive Summary

This data analysis examines the relationship between the type of transmission (automatic or manual) of a car and its mileage (miles per gallon, MPG). The `mtcars` dataset available with R, that contains 32 observations on 11 variables is used to perform the analysis.

Preliminary data analysis indicates that cars with manual transmission have a significantly higher mpg than cars with automatic transmission. This is confirmed later on by the results of regression modelling. Manual cars have, on average, `7.24` mpg higher than auto cars in this dataset. While the result holds for this dataset, generalizing this would require further work to examine other factors that were not analysed.

Data Processing and Exploratory Data Analysis

In this section, a brief exploratory data analysis (EDA) will be performed after data cleaning and preparation.

As the data analysis is primarily interested in examining the relationship between miles per gallon (MPG) and transmission type (**am** with 0 = automatic, 1 = manual), we will focus on these two variables.

The exploratory analysis indicates that manual cars have a higher mpg than their auto counterparts (refer graphs in ANNEXURE 1). However, this will be verified in the following section via regression models.

Modelling

We will now formally model the relationship between the transmission type and mpg for cars via the regression model `lm(mpg ~ factor(am), data = cars)`.

Let us now examine the mean values for the two transmission types by removing the intercept from the model.

```
confint(lm(mpg ~ factor(am) - 1, data = cars))
```

```
##                    2.5 %   97.5 %
## factor(am)Auto   14.85062 19.44411
## factor(am)Manual 21.61568 27.16894
```

```
confint(lm(mpg ~ factor(am) - 1, data = cars), level = .99)
```

```
##                    0.5 %   99.5 %
## factor(am)Auto   14.05472 20.24002
## factor(am)Manual 20.65347 28.13114
```

It is clear from the above results that *transmission type* is a significant predictor of mileage with 33.85% of the variability in mpg accounted for by transmission type only. Furthermore, the upper limit of the mean mpg for automatic transmission cars are smaller than the lower limit of the mean mpg for manual cars. The results hold at a type-1 error rate (alpha or significance level) of 5% and 1%.

Model Selection

Although transmission type accounts for 33.85% variability in mpg, the other variables from the original dataset could explain more of the variability and shall be briefly explored in this section.

The analysis of variance of the initial model with just transmission type versus the model that includes weight of the car indicates that the weight of car is an important factor in determine the mpg. While we can build on this second model and examine the relevance of other factors, we will conclude the model comparison with just weight for this analysis.

Inference and Interpretation

Reverting to the initial model with just transmission type, let us interpret the model coefficients.

```
summary(lm.T)$coef
```

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

The model's intercept of 17.1473684 indicates the mean mpg for the base level, that is cars with automatic transmission. The second coefficient, or slope, indicates the expected, additional mpg for manual cars 7.2449393. Note that we are looking at the model with intercept in this interpretation.

Residual Diagnostics

So far, we have modelled the relationship between mpg and transmission type using a linear regression model but we have not paid any attention to the suitability of this relationship. We shall do this here by examining the residuals of the first linear and see its adherence to the assumptions for linear regression.

Refer ANNEXURE 2 for residual diagnostic plots

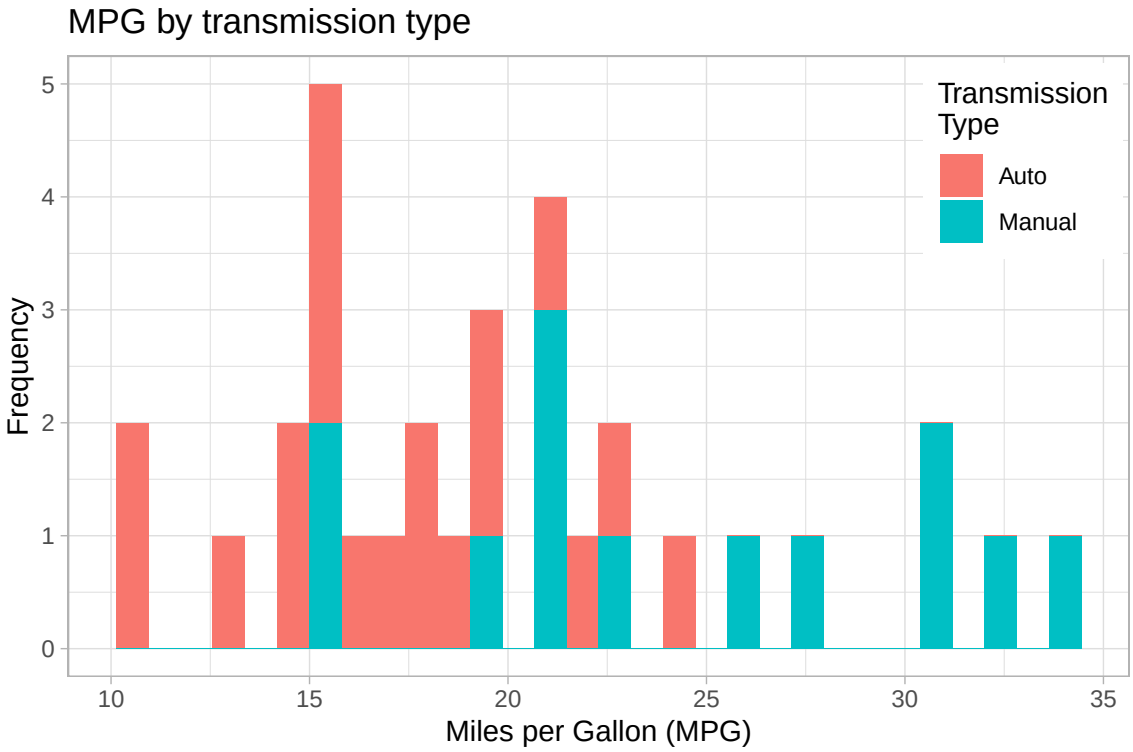
While the residual plots are a little difficult to interpret as the predictor is a categorical variable with two levels, thus making it a binary predictor, the plots nevertheless indicate that the model adheres to the assumptions for linear regression. Plots 1 and 3 (bottom left) indicate that residuals do not vary much with the fitted values, and consequently the predictors. While the standard error of predicted mpg for manual cars are higher, it is not significantly concerning. Lastly, the two plots in the second column indicate that the residuals are normally distributed and there are no influential observations to distort the model. Thus, this model is a pretty good one to begin with that conforms to the assumptions required for linear regression and models the relationship between mpg and transmission type reasonably well.

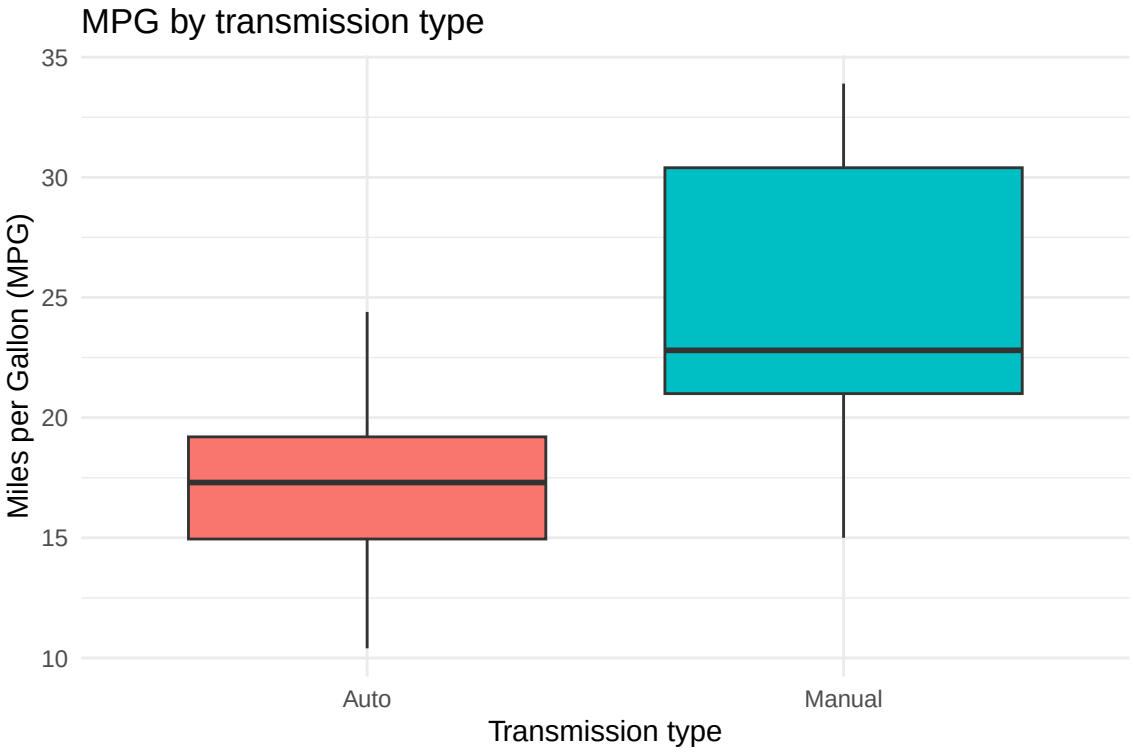
Conclusion - answer questions

In conclusion, the analysis indicates that manual transmission is better for MPG, with a gain of 7.2449 mpg on average, compared to their auto counterparts.

ANNEXURE 1

Exploratory data visualization





ANNEXURE 2

Diagnostic plots