Regression Models Course Project

Jessica Dyer

1/23/2021

1 Executive summary

---

2 Introduction

Motor Trend, a magazine about the automobile industry are interested in exploring the relationship between a set of variables and miles per gallon (MPG). 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

Research shows that there are a number of factors that affect fuel economy in vehicles including: weight, horsepower, transimission type and engine displacement. The goal of this analysis is to understand what variables have the largest impact on MPG.

3 Methods

We will be using the mtcars data set provided with the R statistical package. The data are from a 1974 edition of Motor Trend magazine and is comprised of fuel consumption and 10 aspects of automobile design and performance for 32 automobiles, all of them 1973-74 models.

- The 'mtcars' dataset has 32 observations and 11 variables.
- The data dictionary can be found by typing this into the console: **?mtcars**

We used nested likelihood ratio tests (ANOVA) to help find the model that best explains variation in MPG in this data. We conducted linear regression using the model derived from the step-wise model search. The final model did not include transmission type. However, because the client is interested in transmission type, we developed two models, one best fit and one best fit including transmission type. After finding the best fitting model with the available data, we then evaluated the relationship between weight and transmission type. We generated a box plot and compared mean weights of cars with a manual transmission versus automatic transmission using a t-test.

We then produced a series of four diagnostic plots: 1) residuals vs fitted, 2) scale-location, 3) normal Q-Q, and 4) residuals vs leverage and examined them to evaluate the fit and residuals of the linear model.

An additional diagnostic step included calculating the variance inflation factors (VIF) associated with our model to examine the increase in the variance for the ith regressor compared to the ideal setting where it is orthogonal to the other regressors.

Lastly, we ran the model, which included transmission type to evaluate the relationship between car weight, number of cylinders, and transmission type to MPG.

4 Results

The results of the likelihood ratio tests suggested that the model that captured the majority of the variability in the data is: `mpg ~ wt + as.factor(cyl)`. With this model, 81% of the variation in MPG is explained by both the weight of the vehicle and the number of cylinders. However, only 7.77% of the variation in MPG is explained by the number of cylinders, meaning the majority of the variation is explained by the weight of the vehicle. We also added transmission type into the model and found that only 0.66% of the variation in MPG is explained by transmission type.

In this dataset, cars with a manual transmission are heavier than cars with an automatic transmission. Manual transmission cars have a mean weight of 3768.89 pounds and automatic transmission cars have a mean weight of 2411 pounds, a statistically significant difference ($p = 0$). See Figure 3.

The four diagnostic plots show. . .

Linear regression results show that 1000 pound increase in weight, miles per gallon decrease by 3.15 (p=0.002). Additionally, a car with six cylinders gets 4.26 MPG less (p=0.006) and a car with eight cylinders gets 6.08 MPG less (p=0.001) when compared to cars with four cylinders. There is no statistically significant difference in MPG between automatic and manual transmissions when weight and number of cylinders are held constant (p=0.9). See Figure 1.

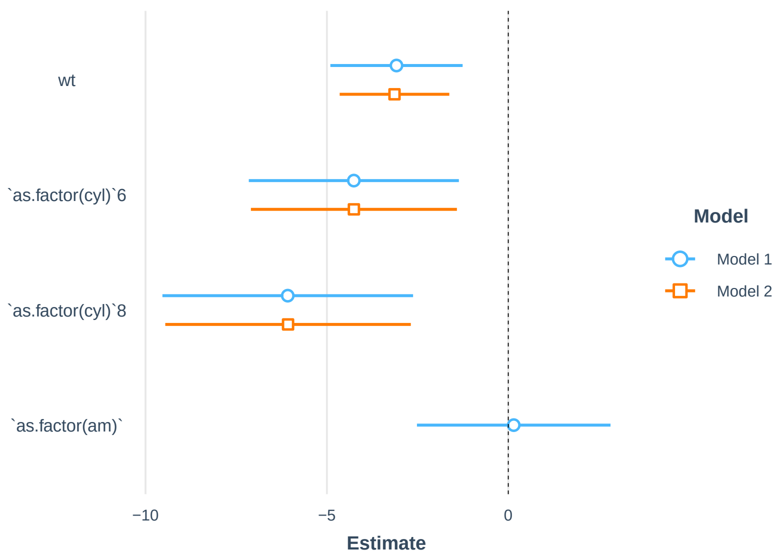

Figure 1: Linear regression results.

5 Discussion

We know that heavier cars get lower gas mileage. I think that the transmission variable is co-linear with weight and actually measuring the same thing as weight. Much of the difference seen when looking at MPG by transmission type is actually due to the weight of the vehicle.

6 Conclusion

7 Appendix

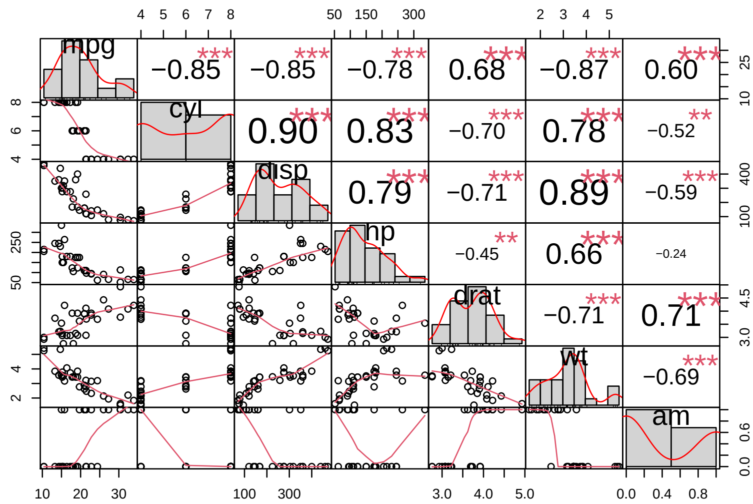

Figure 2: Correlation between various indicators and MPG

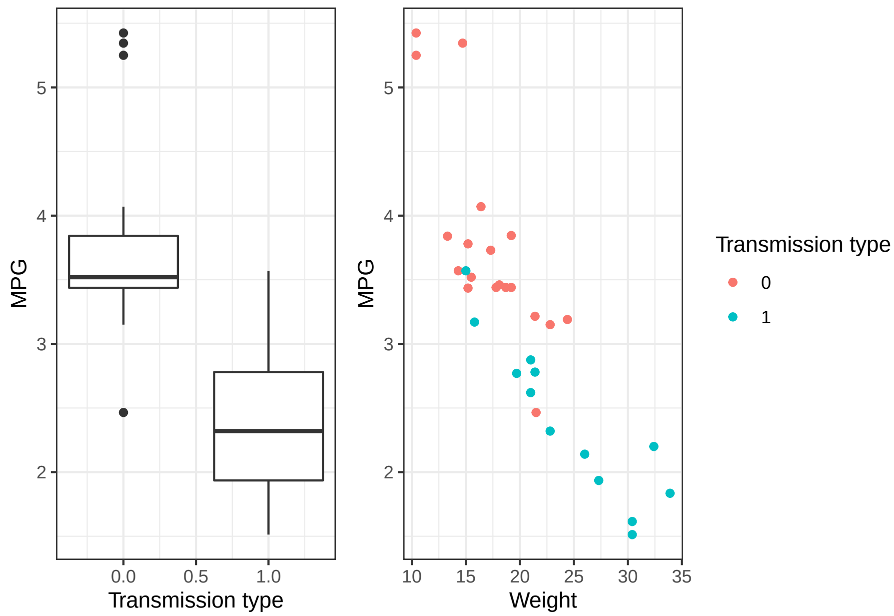

Figure 3: Relationship between MPG and transmission type

```
## Start:  AIC=80
## mpg ~ wt + disp + drat + as.factor(cyl) + as.factor(am)
##
##                   Df Sum of Sq    RSS    AIC
## - disp             1     0.075 182.75 76.551
## - drat             1     0.195 182.87 76.572
## - as.factor(am)    1     0.211 182.88 76.574
## <none>                         182.68 80.003
## - as.factor(cyl)   2    60.831 243.51 82.270
## - wt               1    47.657 230.33 83.956
##
## Step:  AIC=76.55
## mpg ~ wt + drat + as.factor(cyl) + as.factor(am)
##
##                   Df Sum of Sq    RSS    AIC
## - drat             1     0.219 182.97 73.123
## - as.factor(am)    1     0.236 182.99 73.126
## <none>                         182.75 76.551
## - as.factor(cyl)   2    84.238 266.99 81.750
## - wt               1    81.574 264.32 84.895
##
## Step:  AIC=73.12
## mpg ~ wt + as.factor(cyl) + as.factor(am)
##
##                   Df Sum of Sq    RSS    AIC
## - as.factor(am)    1     0.090 183.06 69.673
## <none>                         182.97 73.123
## - as.factor(cyl)   2    95.351 278.32 79.614
## - wt               1    81.527 264.50 81.450
##
```

```
## Step:  AIC=69.67
## mpg ~ wt + as.factor(cyl)
##
##                   Df Sum of Sq    RSS    AIC
## <none>                         183.06 69.673
## - as.factor(cyl)  2    95.263 278.32 76.149
## - wt              1   118.204 301.26 82.149
```

```
##
## Call:
## lm(formula = mpg ~ wt + as.factor(cyl), data = mtcars)
##
## Coefficients:
##     (Intercept)               wt  as.factor(cyl)6  as.factor(cyl)8
##          33.991           -3.206           -4.256           -6.071
```

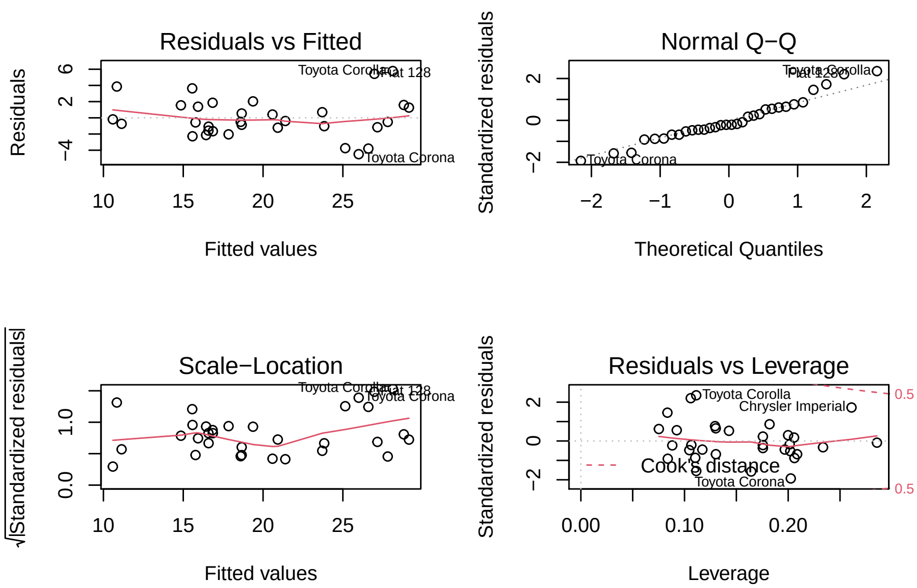

Figure 4: Residual plots for model diagnostics

```
## Potentially influential observations of
##   lm(formula = mpg ~ wt + as.factor(cyl) + as.factor(am), data = mtcars) :
##
##                     dfb.1_ dfb.wt dfb.a.()6 dfb.a.()8 dfb.a.()1 dffit cov.r
## Lincoln Continental  0.04  -0.05   0.02      0.02     -0.02     -0.05  1.69_*
## Toyota Corolla       0.35  -0.23  -0.30     -0.11      0.14      0.92  0.43_*
##                     cook.d hat
## Lincoln Continental  0.00   0.29
## Toyota Corolla       0.14   0.11
```

Table 1

	Model 1	Model 2
(Intercept)	33.75 ***	33.99 ***
	[27.98, 39.53]	[30.12, 37.86]
wt	-3.15 **	-3.21 ***
	[-5.01, -1.29]	[-4.75, -1.66]
as.factor(cyl)6	-4.26 **	-4.26 **
	[-7.15, -1.36]	[-7.09, -1.42]
as.factor(cyl)8	-6.08 **	-6.07 ***
	[-9.53, -2.62]	[-9.46, -2.69]
as.factor(am)1	0.15	
	[-2.52, 2.82]	
N	32	32
R2	0.84	0.84

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.