Regression Models Course Project

Shadi

5/8/2020

Summary

The purpose of analysis is to explore which kind of transmission is better for MPG and to quantify the diffrence between the two kind of transmission.In order to answer these questions exploratory data analysis performed , and we used hypothesis testing and linear regression as methodologies to make inference. we then established both simple and multivariate linear regression analysis. However the result of the multivariable regression model is more promising as it includes the potential effect of other variables on MPG. Finally,we concluded that the manual car are better in terms of MPG as they have a higher MPG compared to autamtic car. see the Appenxix for full Model Summary and diagnostics.

Data

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). description for variable can be find in appendix

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1

looking at the data we see 11 variables with all numeric value.

Our variables of intrest here are 1. transmition (am) which form the information provided about the data we know that it take value 1 for manual and 0 for automatic. 2. MPG (mile per gallon).

Visualzing MPG by the vilon plot over two value of transmition would give us a good idea for our future hypothesis testing. (plot 1, appendix) It appears that automatic cars have a lower miles per gallon, and therefore a lower fuel efficiency, than manual cars do. But it is possible that this apparent pattern happened by random chance- that is, that we just happened to pick a group of automatic cars with low efficiency and a group of manual cars with higher efficiency. So to check whether that's the case, we have to use a statistical test.

In order to determine the relationship between the variables, and to findout which variables should be included in our model, we perform a correlation test and create a correlation heat map (plot 2, appendix).

Finally we take a look at the distribution of mpg to check if it look like normal. (plot 3, appendix)

Analyis

To addres the first question we run a simple t-test cheking if the automatic cars have a lower miles per gallon. our null hypothesis is: “there is no diffrenct between the two transmition”

t.statistic	df	p.value	lower.CL	upper.CL	automatic.mean	manual.mean
-3.767123	18.33225	0.0013736	-11.28019	-3.209684	17.14737	24.39231

Here the p-value is 0.001 meaning that we can reject the null.

quantifying the diffrence between the two transmission,we need to have a proper model, we conduct both simple linear and multivariable linear regression, and select the best model, to have the most accurate predicton

	Model 1	Model 2
(Intercept)	18.90 ***	17.15 ***
	[17.56, 20.23]	[15.30, 18.99]
wt	-3.83 ***	
	[-5.64, -2.03]	
qsec	2.19 **	
	[0.91, 3.47]	
am	2.94	7.24 **
	[-0.32, 6.19]	[3.17, 11.32]
N	32	32
R2	0.85	0.36

All continuous predictors are mean-centered and scaled by 1 standard deviation. Standard errors are heteroskedasticity robust. *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

In table below we can see that for linear regression, both intercept and transmission are significant.and the model shows that mpg for manual car are 7.25 scale higher than autmatic, but the value of the R2: 0.35 is low. For the multivariable linear regression, all the variable are significant and the R2 is near 1 which means that it is a better model compared to linear one.

Model selection

As we see erlier almost all of other variables have high correlation with MPG so in order to have a good predicter in our model using a stepwise selection method which consists of iteratively adding and removing predictors for find the subset of variables in the data set resulting in the best performing model would be best move. folowing code autamticly give us the best model

for selecting the best model one way is to look at the R2 value but to have a better result we conduct a likelihood ratio test and compare the two models.

Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
30	720.8966	NA	NA	NA	NA
28	169.2859	2	551.6107	45.61838	0

Finally, in order to intrepret the result we must make sure that the error dosen’t have hetroskesaticity and

colinearity this can be find at appandix(Plot 4).

Conclusion

Using model contruction, we have shown that, adjusted to other strong mpg predictors that we can find in the mtcars dataset, manual transmission is really the best transmission for mpg with mmpg=9.61+2.93 amManual-3.91 wt +1.22 qsec

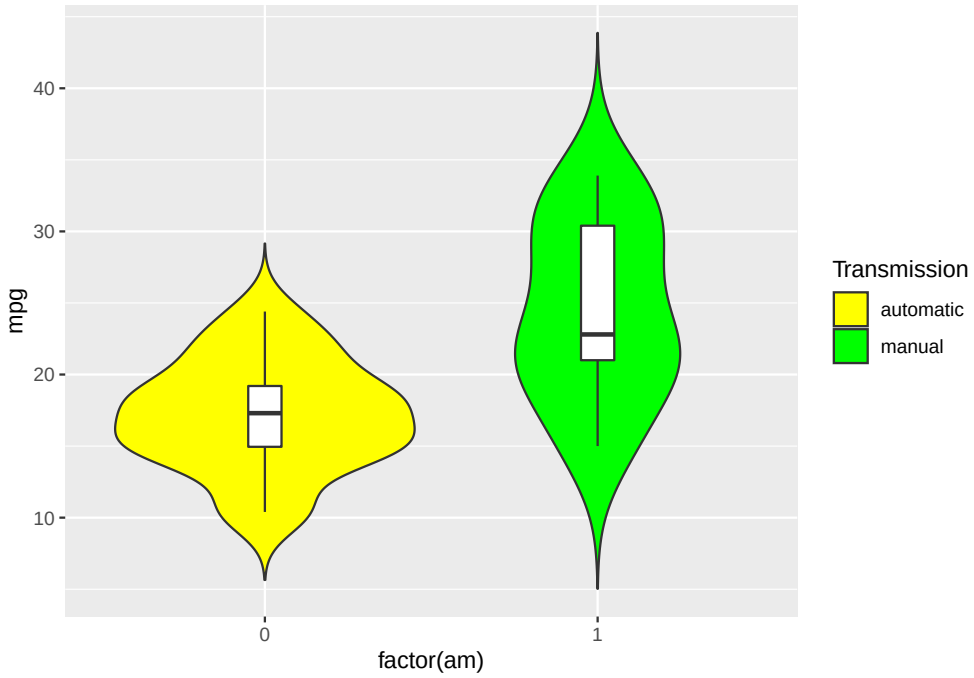
Appendix

residual plot and some diagnostics

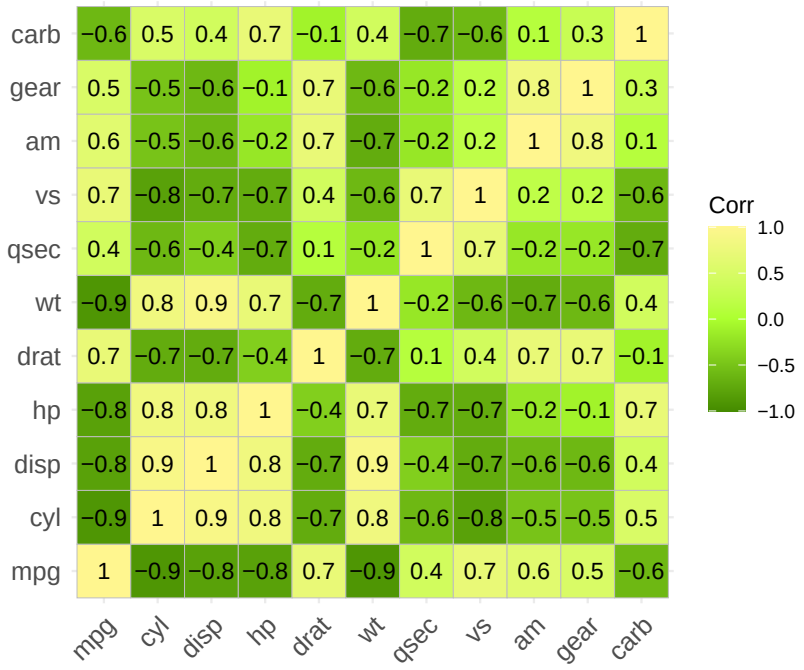
data description

1-**mpg**: Miles/(US) gallo 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

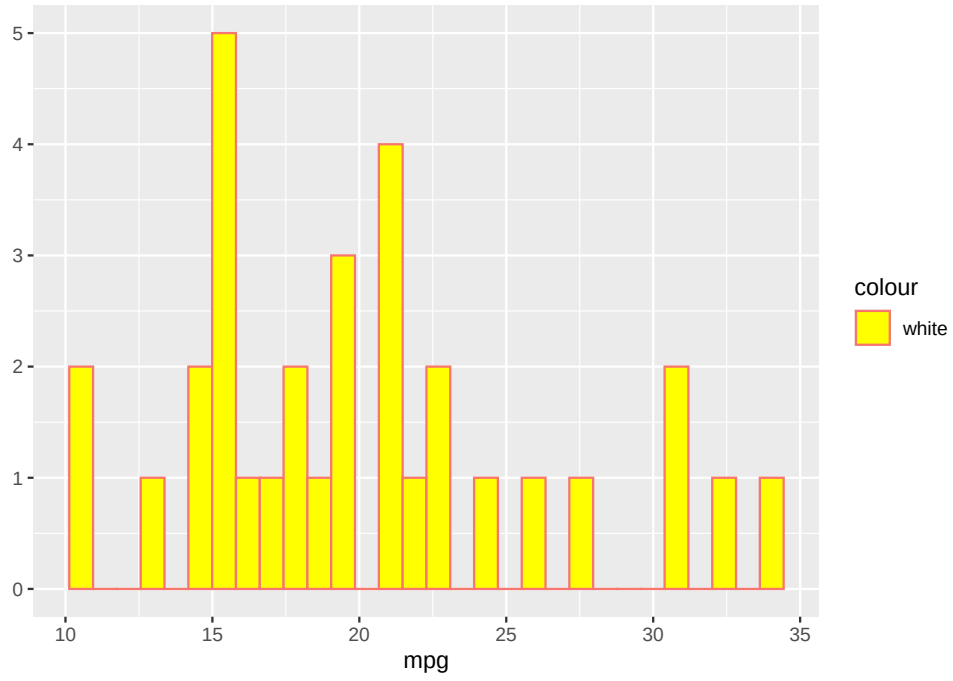
plot 1



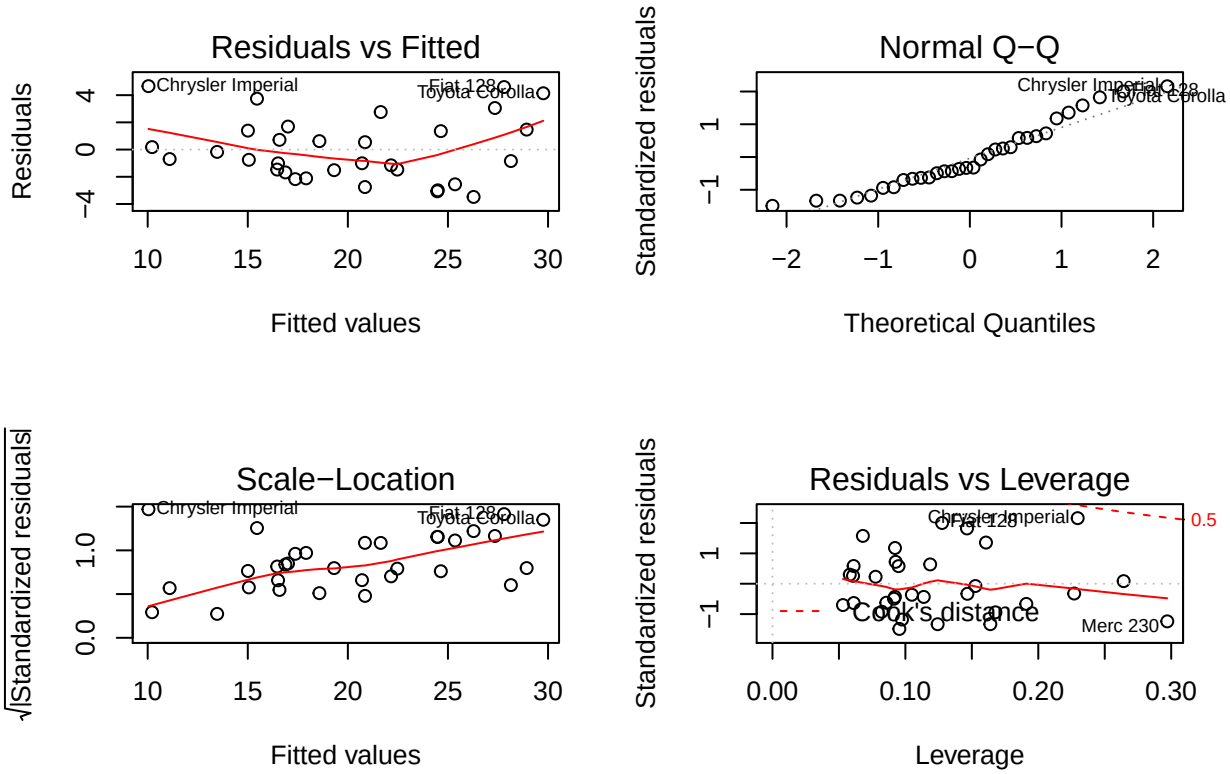
plot 2


carb −0.6 0.5 0.4 0.7 −0.1 0.4 −0.7 −0.6 0.1 0.3 1
gear 0.5 −0.5 −0.6 −0.1 0.7 −0.6 −0.2 0.2 0.8 1 0.3
am 0.6 −0.5 −0.6 −0.2 0.7 −0.7 −0.2 0.2 1 0.8 0.1
vs 0.7 −0.8 −0.7 −0.7 0.4 −0.6 0.7 1 0.2 0.2 −0.6
qsec 0.4 −0.6 −0.4 −0.7 0.1 −0.2 1 0.7 −0.2 −0.2 −0.7
wt −0.9 0.8 0.9 0.7 −0.7 1 −0.2 −0.6 −0.7 −0.6 0.4
drat 0.7 −0.7 −0.7 −0.4 1 −0.7 0.1 0.4 0.7 0.7 −0.1
hp −0.8 0.8 0.8 1 −0.4 0.7 −0.7 −0.7 −0.2 −0.1 0.7
disp −0.8 0.9 1 0.8 −0.7 0.9 −0.4 −0.7 −0.6 −0.6 0.4
cyl −0.9 1 0.9 0.8 −0.7 0.8 −0.6 −0.8 −0.5 −0.5 0.5
mpg 1 −0.9 −0.8 −0.8 0.7 −0.9 0.4 0.7 0.6 0.5 −0.6
mpg cyl disp hp drat wt qsec vs am gear carb
Corr
1.0
0.5
0.0
−0.5
−1.0

plot 3


5
4
3
2
1
0
10
15
20
25
30
35
mpg
colour
white

plot 4


Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Fiat 128
Toyota Corolla
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Residuals vs Leverage
Cook's distance
Leverage
Merc 230
0.5