

Transmission Type and Fuel economy

Wang Yihong

Saturday, June 21, 2014

Forgive me to use the packages, All I want are prettier tables.

Executive Summary

This report provides an analysis of relation between the transmission types(automatic vs manual) and MPG(miles per gallon).Methods include t test for two groups, linear regression and data visualization.Results show that, in this sample, compared to manual transmission cars, automatic cars consume less fuel on average.

Descriptive Statistics

```
library(stargazer);  
stargazer(mtcars,title = "Summary Statistics")
```

% Table created by stargazer v.5.1 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
% Date and time: Sun, Jun 22, 2014 - 12:08:39 AM

Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
mpg	32	20.090	6.027	10.400	33.900
cyl	32	6.188	1.786	4	8
disp	32	230.700	123.900	71.100	472.000
hp	32	146.700	68.560	52	335
drat	32	3.597	0.535	2.760	4.930
wt	32	3.217	0.978	1.513	5.424
qsec	32	17.850	1.787	14.500	22.900
vs	32	0.438	0.504	0	1
am	32	0.406	0.499	0	1
gear	32	3.688	0.738	3	5
carb	32	2.812	1.615	1	8

T test of mpg for automatic and manual cars:

```
dif = t.test(mtcars$mpg[mtcars$am==0],mtcars$mpg[mtcars$am==1])
```

In average, automatic cars have lower mpg than manual transmission cars. 17.1474 vs 24.3923.The t statistic -3.7671 indicates that the difference between the two group is significant at 0.0014 level.

Model Selection

I study the correlation table to determine the variables that are correlated to the dependent variable mpg. (I know I'm not supposed to use the packages, but I just want a prettier table. Actually, I can use cor and cor.test to check the correlation.)

```
#source("https://gist.github.com/nofacetou/42ed588f76b32a8e96ec/raw/CorTab.R")
source("CorTab.R")
require(xtable)
ct = xtable(corstars1(mtcars),digits=3,caption = "Correlation Table")
#print(ct, type="html")
print(ct)
```

% latex table generated in R 3.1.0 by xtable 1.7-3 package % Sun Jun 22 00:08:40 2014

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear
mpg										
cyl	-0.85***									
disp	-0.85***	0.90***								
hp	-0.78***	0.83***	0.79***							
drat	0.68***	-0.70***	-0.71***	-0.45**						
wt	-0.87***	0.78***	0.89***	0.66***	-0.71***					
qsec	0.42*	-0.59***	-0.43*	-0.71***	0.09	-0.17				
vs	0.66***	-0.81***	-0.71***	-0.72***	0.44*	-0.55***	0.74***			
am	0.60***	-0.52**	-0.59***	-0.24	0.71***	-0.69***	-0.23	0.17		
gear	0.48**	-0.49**	-0.56***	-0.13	0.70***	-0.58***	-0.21	0.21	0.79***	
carb	-0.55**	0.53**	0.39*	0.75***	-0.09	0.43*	-0.66***	-0.57***	0.06	0.27

Table 2: Correlation Table

Then I use the backward selection to determine the final model(results are untabulated):

```
summary(step(lm(data=mtcars,
                mpg~cyl+wt+hp+disp+drat+qsec+factor(vs)+factor(am)+gear+carb),
          direction="backward",test="F"))
```

The best model(has lowest AIC) I've got from the above results is:

```
fit <- lm(mpg ~ wt + qsec + factor(am), data = mtcars)
stargazer(fit,title="Regression Results", omit.stat=c("LL","ser","f"), no.space=TRUE)
```

% Table created by stargazer v.5.1 by Marek Hlavac, Harvard University. E-mail: hlavac at fas.harvard.edu
 % Date and time: Sun, Jun 22, 2014 - 12:08:40 AM

Then the average difference in mpg between the automatic transmission cars and manual transmission cars is 2.9358,at the same level of weight and 1/4 mile time.

```
sumCoef <- summary(fit)$coefficients
sumCoef[4,1] + c(-1, 1) * qt(.975, df = fit$df) * sumCoef[4, 2]
```

```
## [1] 0.04573 5.82594
```

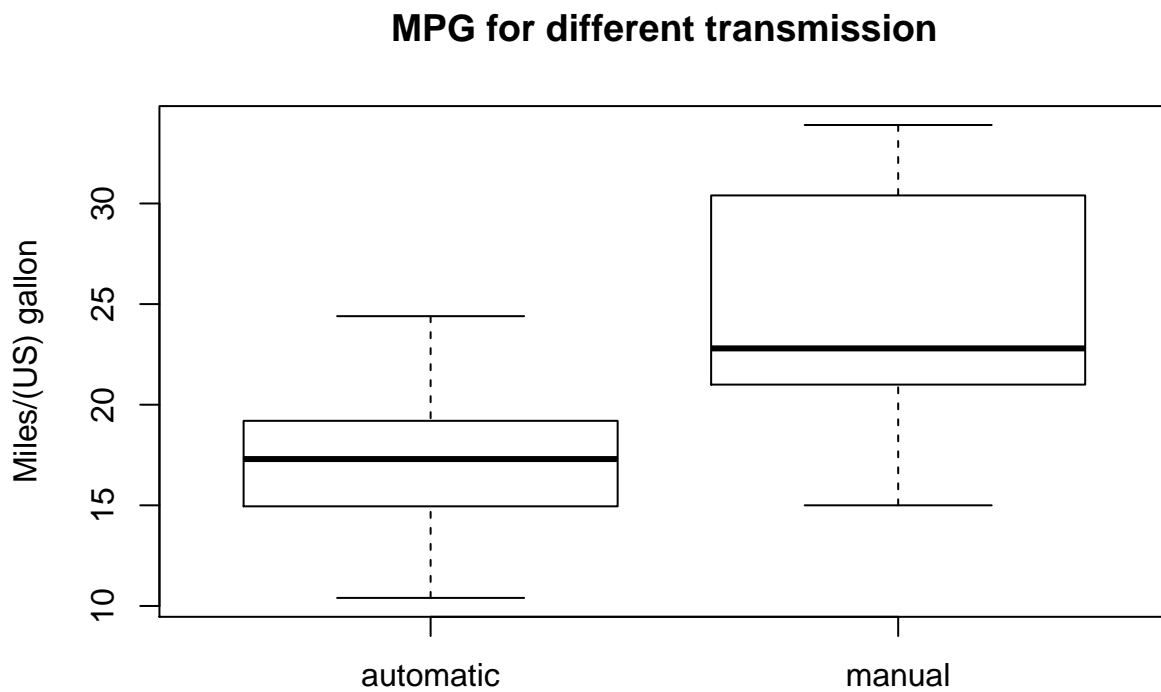
With 95% confidence, I estimate that a manual transmission car consumes more fuel, from 0.04573 to 5.82594 mpg, compared to an automatic transmission car, given the same weight and 1/4 mile time.

Table 3: Regression Results

<i>Dependent variable:</i>	
	mpg
wt	−3.917*** (0.711)
qsec	1.226*** (0.289)
factor(am)1	2.936** (1.411)
Constant	9.618 (6.960)
Observations	32
R ²	0.850
Adjusted R ²	0.834
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

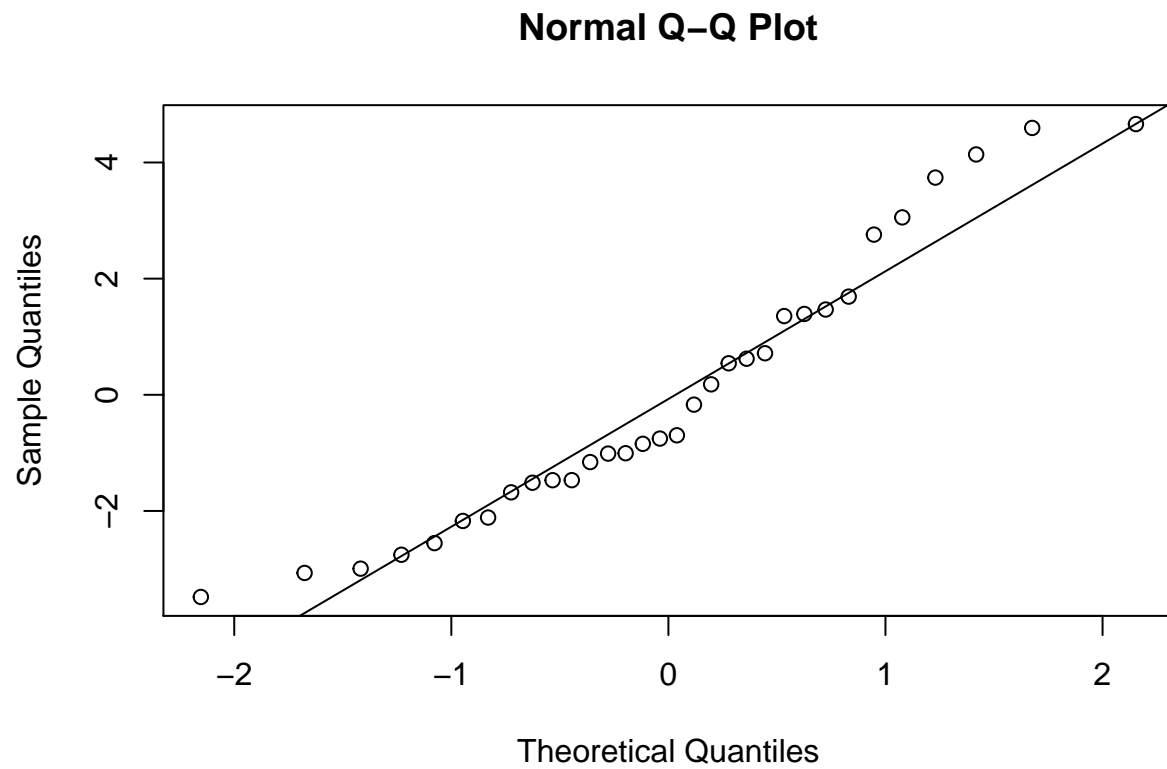
Appendix

```
boxplot(mpg~am,data=mtcars, names=c("automatic","manual"),
        ylab = "Miles/(US) gallon",main="MPG for different transmission")
```



Model Diagnostic

```
qqnorm(residuals(fit))  
qqline(residuals(fit))
```



The residuals for the model is approximately normally distributed.