**Exploring the Relationship Between Fuel Consumption, Engine Displacement, and Performance Accessories in Automobiles**

410978004 郭依璇(Yi-Hsuan Kuo)

Table of Contents

1. Background, Motivation, and Research Objectives
2. Data Description
3. Multivariate Analysis Methods
4. Analysis Results
5. Comprehensive Conclusions and Recommendations
6. References

I. Background, Motivation, and Research Objectives

In an era where nearly everyone drives and environmental awareness is on the rise, young people are concerned about which cars consume less fuel and can achieve environmental protection by using less fuel. Additionally, they care about engine displacement, as less exhaust emissions contribute to a better environment. Therefore, this study aims to explore which cars have lower fuel consumption and lower engine displacement.

II. Data Description

1. mpg: Miles/(US) gallon
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

The dataset comprises a total of 32 observations (obs=32).

The data for mtcars originates from the 1974 Motor Trend US magazine and is also built into R, encompassing 32 vehicles and 11 associated parameters.

Y : Car Performance

= mpg

= disp

X : Car measures

= qsec

= hp

= drat

= wt

B : background

= vs

= am

= gear

= carb

= cyl

Data Source: <https://gist.github.com/seankross/a412dfbd88b3db70b74b>

III. Multivariate Analysis Methods

Step 1: Canonical Correlation Analysis

* Use canonical correlation analysis to examine if there is a relationship between the Y construct and the X construct.

Step 2: Clustering

* Perform clustering on、、 to identify features of the X construct associated with low mpg and low disp.

Step 3: MANOVA on Clusters

* Treat clusters as categories (with cluster=1 and cluster=2), and perform MANOVA (Multivariate Analysis of Variance) with background variables against the clusters. If we reject H0, it indicates differences between the background variables and the different clusters, suggesting that this background is related to mpg and disp, warranting further analysis.

Step 4: Constructing Multiple Regression Models

* Build a multiple regression model with y\_1 and y\_2 as dependent variables and x\_1 to x\_4 along with the significant background variables from the previous step as predictors.

Step 5: Backward Selection

* Use backward selection to refine the model, retaining variables significantly related to mpg and disp.

Step 6: Analysis of the Final Model

* Examine various statistical measures of the final model and conduct an analysis.

Step 7: Summary

* Consolidate findings from the above steps for a comprehensive conclusion.

IV. Analysis Results

canonical

= mpg

= disp

= qsec

= hp

= drat

= wt

: 0.4381 – 0.6013

: 0.1860 – 0.2003 + 0.1764 - 0.6522

= 0.97

: -1.8326 – 1.7857

: 1.0527 – 1.3515 + 1.0901 – 0.1707

= 0.04

Test for the significance of the canonical relations with

vs

F = 20.29 ; p-value < 0.0001 < 0.05 → reject at , at least the first pair of canonical variables are important.

vs

F = 0.02 ; p-value = 0.9969 > 0.05 → do not reject at , only the first pair of canonical variables are important.

Thus, the first canonical variates are good summary measures.

: Given the superior performance of mpg and disp, it is named as the performance index of the vehicle.

: Given the superior performance of wt, it is named as the weight index.

RM(Y|X) = (0.8454+0.8902)/2 = 0.8687

RM(X|Y) = (0.1974+0.6653+0.5267+0.8355)/4 = 0.5562

The Y construct can be explained by the X construct by approximately 87%.

* This indicates that there is a correlation between X and Y constructs, warranting further analysis.

cluster

Cluster = 1

| **Variable** | **N** | **Mean** | **Standard Deviation** | | **Minimum** | | **Maximum** | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | **mpg** | | **disp** | | **hp** | | **drat** | | **wt** | | **qsec** | | |  | | --- | | 16 | | 16 | | 16 | | 16 | | 16 | | 16 | | |  | | --- | | 15.6812500 | | 339.1500000 | | 196.5000000 | | 3.1906250 | | 3.9165000 | | 17.1543750 | | |  | | --- | | 2.9267089 | | 73.9587092 | | 58.8217647 | | 0.3670689 | | 0.7435644 | | 1.5332622 | | | |  | | --- | | 10.4000000 | | 225.0000000 | | 105.0000000 | | 2.7600000 | | 3.1700000 | | 14.5000000 | | | |  | | --- | | 21.4000000 | | 472.0000000 | | 335.0000000 | | 4.2200000 | | 5.4240000 | | 20.2200000 | | |
|  | |  | |  | |  | |

Cluster = 2

| **Variable** | **N** | **Mean** | **Standard Deviation** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- |
| |  | | --- | | **mpg** | | **disp** | | **hp** | | **drat** | | **wt** | | **qsec** | | |  | | --- | | 16 | | 16 | | 16 | | 16 | | 16 | | 16 | | |  | | --- | | 24.5000000 | | 122.2937500 | | 96.8750000 | | 4.0025000 | | 2.5180000 | | 18.5431250 | | |  | | --- | | 5.0026660 | | 34.5671997 | | 31.0115570 | | 0.3233265 | | 0.6185472 | | 1.7943568 | | |  | | --- | | 17.8000000 | | 71.1000000 | | 52.0000000 | | 3.6200000 | | 1.5130000 | | 15.5000000 | | |  | | --- | | 33.9000000 | | 167.6000000 | | 175.0000000 | | 4.9300000 | | 3.4400000 | | 22.9000000 | |

Comparison of Cluster 1 and Cluster 2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | mpg | disp | hp | drat | wt | qsec |
| Cluster1 | Low | High | High | Low | High | Low |
| Cluster2 | High | Low | Low | High | Low | High |

→ From this, it can be observed that the group with low mpg has higher hp and wt, and lower drat and qsec; whereas the group with low disp has higher drat and qsec, and lower hp and wt. These results can be compared with the final results of the multiple regression.

MANOVA

= vs

= am

= gear

= carb

= cyl

= cluster1

= cluster2

Set = 0.1

**X** = **μ** + + i=1,2

Objective: The goal is to test whether fuel efficiency (mpg) and displacement (disp) have a causal relationship with the variables vs, am, gear, and carb, cyl.

(1) FOR vs:

(1) H0: ; Ha: at least one

F = 0 ; p-value = 1 > 0.1 => not reject H0 at = 0.1

There is insufficient evidence to demonstrate a significant relationship between vs and the cluster.

(2) FOR am:

(2) H0’: ; Ha’: at least one

F = 7.42 ; p-value = 0.0106 < 0.1 => reject H0’ at = 0.1

There is sufficient evidence to indicate a significant relationship between am and the cluster.

(3) FOR gear:

(3) H0’’: ; Ha’’: at least one

F = 4.03 ; p-value = 0.0537 < 0.1 => reject H0’’ at = 0.1

There is sufficient evidence to indicate a significant relationship between gear and the cluster.

(4) FOR carb:

(3) H0’’’: ; Ha’’’: at least one

F = 0.05 ; p-value = 0.8308 > 0.1 => do not reject H0’’’ at = 0.1

There is insufficient evidence to demonstrate a significant relationship between carb and the cluster.

(5) FOR cyl:

(3) H0’’’’: ; Ha’’’’: at least one

F = 0.98 ; p-value = 0.3303 > 0.1 => do not reject H0’’’’ at = 0.1

There is insufficient evidence to demonstrate a significant relationship between cyl and the cluster.

Conclusion: Therefore, only am and gear have a causal relationship with mpg and disp

Multivariate Multiple regression

Objective: The aim is to test whether fuel efficiency (mpg) and displacement (disp) are related to qsec, hp, drat, wt, am, and gear.

= mpg

= disp

= qsec

= hp

= drat

= wt

= am

= gear

Set

(3)

(i) Model1

，0.8329

， 0.9119

Under the condition of , perform backward selection. The remaining variable in the end is model2.

Model2

，0.8497

， 0.9111

For mpg Model2

Test the whole model

vs

= 0.15 ; p-value < 0.001 → reject at , the model is significant.

Test for individual independent variable

vs

= 0.61 ; p-value = 0.0002 < 0.1 → reject at , is significant.

vs

= 0.48 ; p-value < 0.0001 < 0.1 → reject at , is significant.

vs

= 0.87; p-value = 0.0467 < 0.1 → reject at , is significant.

**We remove nothing out of model**

For disp Model2

Test the whole model

vs

= 0.09 ; p-value < 0.001 → reject at , the model is significant.

Test for individual independent variable

vs

= 0.85 ; p-value = 0.0384 < 0.1 → reject at , is significant.

vs

= 0.87 ; p-value =0.0556 < 0.1 → reject at , is significant.

vs

= 0.55; p-value < 0.0001 < 0.1→ reject at , is significant.

vs

= 0.71; p-value = 0.0026 < 0.1→ reject at , is significant.

**We remove nothing out of model**

Additionally, although the R-squared value of model 2 is close to that of model 1, the model has been significantly simplified.

∴the model we suggest is

Conclusion:

For mpg:

Fuel efficiency (mpg) is related to qsec, wt, and am.

For every one-second increase in the fastest time to travel a quarter mile from a standstill (qsec), the average fuel efficiency increases by 1.23 miles per gallon.

For every 1 (1000 lbs) increase in the weight of the car (wt), the average fuel efficiency decreases by 3.92 miles per gallon.

If the car has a manual transmission (am=1), the average fuel efficiency increases by 2.94 miles per gallon.

For disp:

Displacement (disp) is related to qsec, hp, wt, and gear.

* For every one-second increase in the fastest time to travel a quarter mile from a standstill (qsec), the average displacement decreases by 14.77 cubic inches.
* For every one-unit increase in total horsepower (hp), the average displacement increases by 0.47 cubic inches.
* For every 1 (1000 lbs) increase in the weight of the car (wt), the average displacement increases by 66.86 cubic inches.

For each additional forward gear (gear), the average displacement decreases by 43.74 cubic inches.

V. Comprehensive Conclusion and Recommendations

Based on the cluster results, the group with low mpg has higher hp and wt, and lower drat and qsec; whereas the group with low disp has higher drat and qsec, but lower hp and wt.

According to the results of the multiple regression, it is known that the higher the wt, the lower the mpg (the coefficient of x\_4 is negative); the lower the qsec, the lower the mpg (the coefficient of x\_1 is positive), which is consistent with the cluster results. When qsec is higher, disp is lower (the coefficient of x\_1 is negative); and lower hp and wt lead to lower disp (the coefficients of x\_2 and x\_4 are positive), which also matches the cluster results. If one wishes to know the characteristics of the continuous variables (x\_3) that are not significant in the multiple regression, the cluster results can provide preliminary inferences.

Unfortunately, this dataset is not suitable for dimensionality reduction through PCA or Factor Analysis. After attempting to apply PCA to variables x\_1 to x\_4, it was found that there was only one significant PCA, and only the PCA1 coefficient for x\_2 was 0.99, with the other variables being less than 0.01, which could not adequately represent the data. Fortunately, by abandoning PCA and FA, such favorable results were obtained.

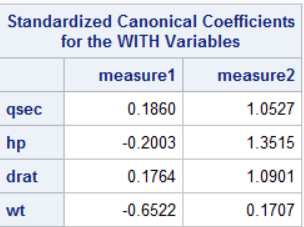
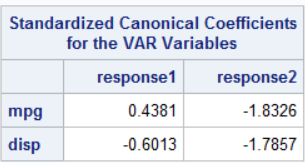
VI. References

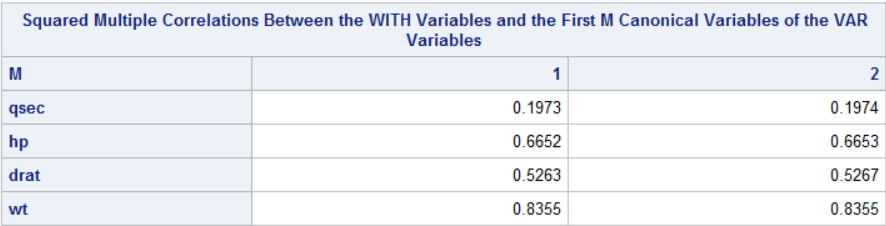
The built-in data frame in R: mtcars

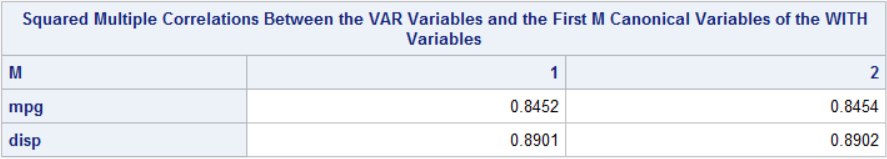
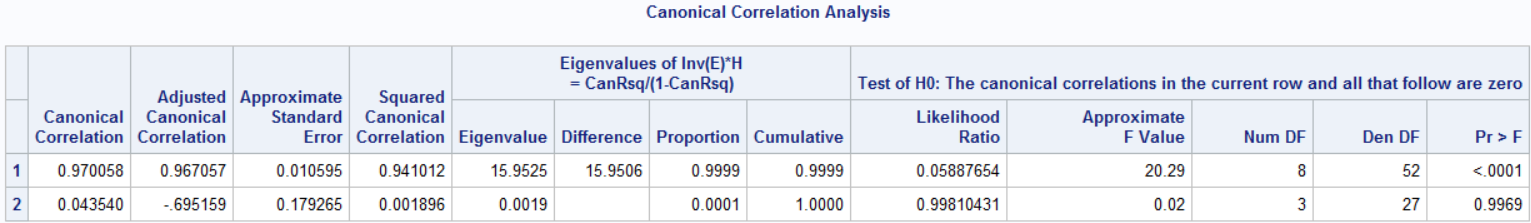
<https://2formosa.blogspot.com/2017/10/R-builtin-dataframe-mtcars.html>

**SAS output:**

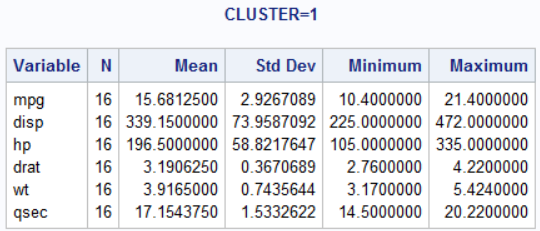
canonical

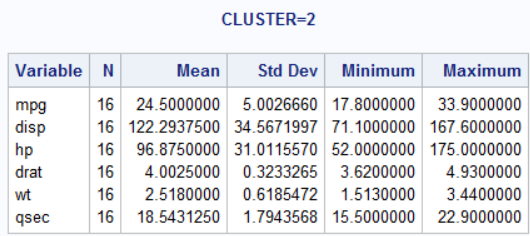




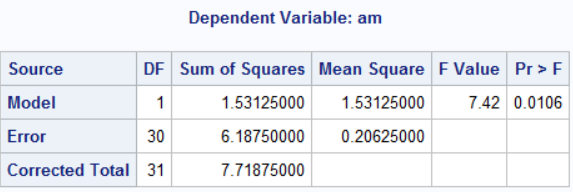
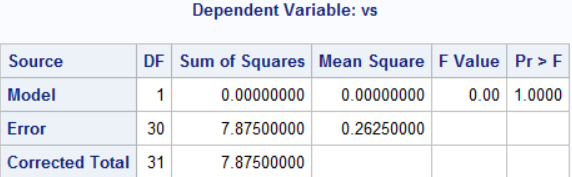


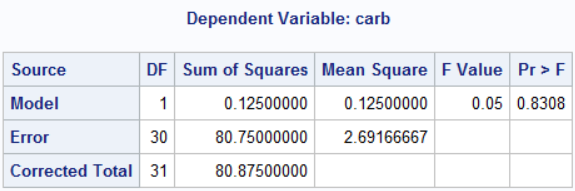
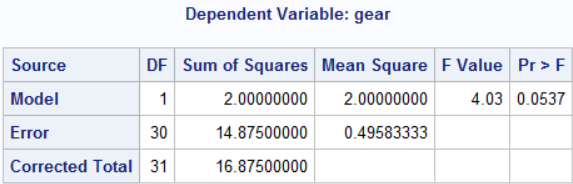
Cluster

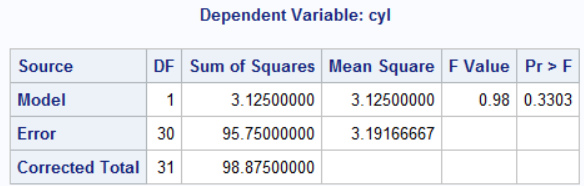




MANOVA



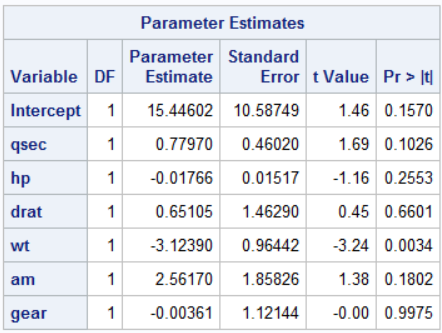




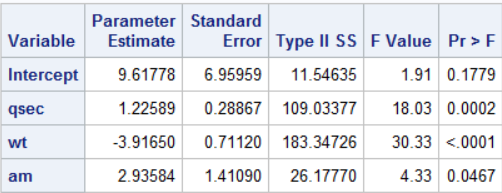
Multiple regression

For mpg

The original model

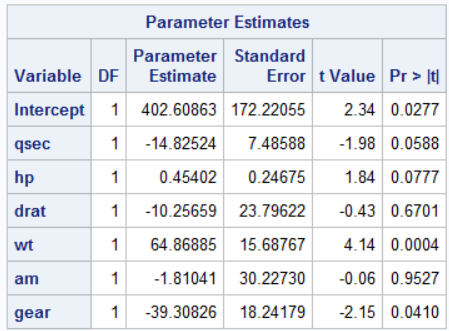


After performing backward selection

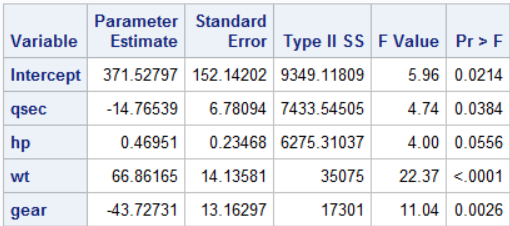


For disp

The original model

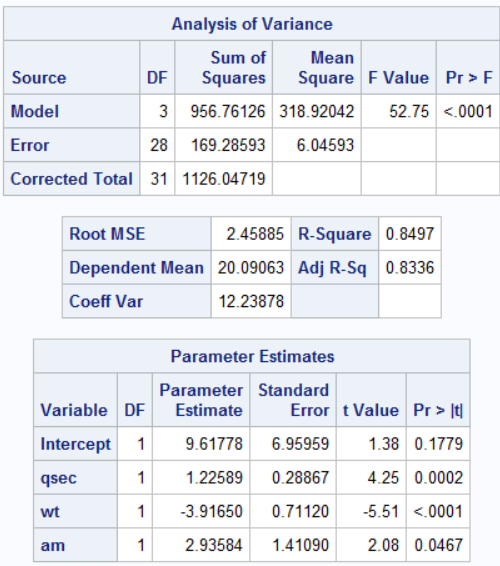


After performing backward selection



The final regression model's various statistical measures

For mpg



for disp

