**Exploring relationships among variables**

Dataset used for demo :

**mtcars**

Variables in the dataset :

**mpg** Miles/(US) gallon

**cyl** Number of cylinders

**disp** Displacement (cu.in.)

**hp** Gross horsepower

**drat** Rear axle ratio

**wt** Weight (1000 lbs)

**qsec** 1/4 mile time

**vs** Engine (0 = V-shaped, 1 = straight)

**am** Transmission (0 = automatic, 1 = manual)

**gear** Number of forward gears

**carb** Number of carburetors

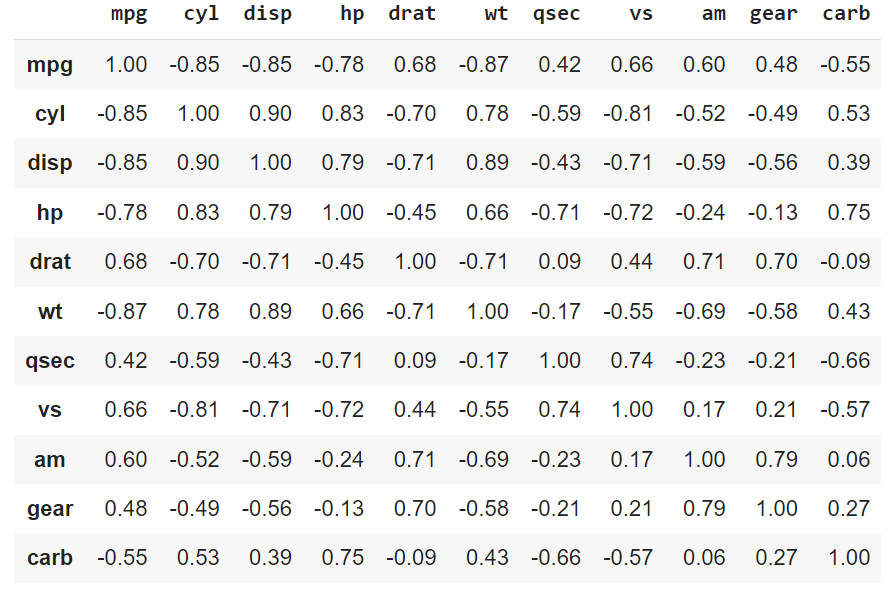
* Visual exploration

|  |  |  |
| --- | --- | --- |
| **data(mtcars)**  **plot(mtcars$hp, mtcars$mpg)**  **plot(mtcars$wt, mtcars$mpg)**  **plot(mtcars$cyl, mtcars$mpg)**  **plot(mtcars$drat, mtcars$mpg)** | | |
|  | | |  | |
|  | | |  | |
| **pairs(mtcars[,c('mpg','hp','wt','cyl','drat')])** | | |



* Correlation analysis

|  |
| --- |
| **round(cor(mtcars), digits = 2)** |

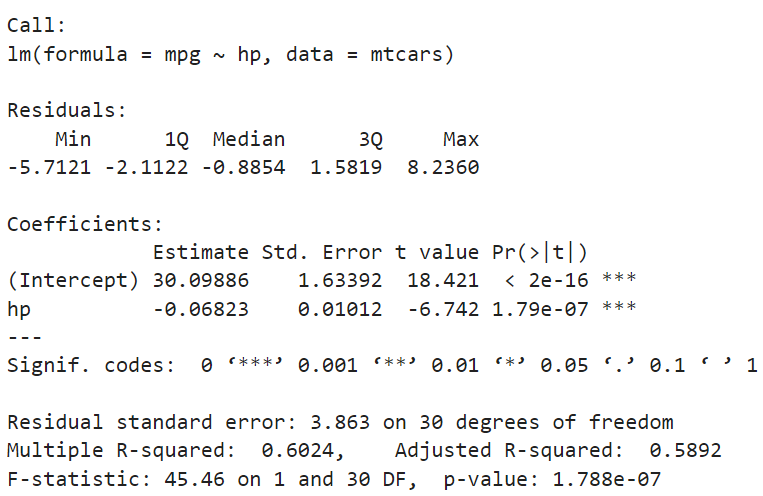


|  |
| --- |
| **library(corrplot)**  **corrplot(cor(mtcars), method = "circle")** |



* Exploring relationships with simple regression models

|  |
| --- |
| **model1 <- lm(mpg ~ hp, data = mtcars)**  **summary(model1)** |



**How to interpret the results?**

1. Intercept Term:

* Represents the predicted value of mpg when hp is 0 (not meaningful in this context).
* In this model, the intercept is 30.099, suggesting a car with 0 horsepower would have a predicted fuel efficiency of 30.099 mpg (hypothetical).

1. Slope Term (hp):

* Represents the change in mpg for a one-unit increase in hp.
* The slope is -0.068, indicating that for each additional horsepower, fuel efficiency decreases by 0.068 mpg.

1. R-squared:

* A measure of how well the model fits the data (0 to 1).
* In this model, R-squared is 0.602, meaning 60.2% of the variation in mpg can be explained by the linear relationship with hp.
* Higher the value better the model.

1. P-values [Pr(>|t|)]:

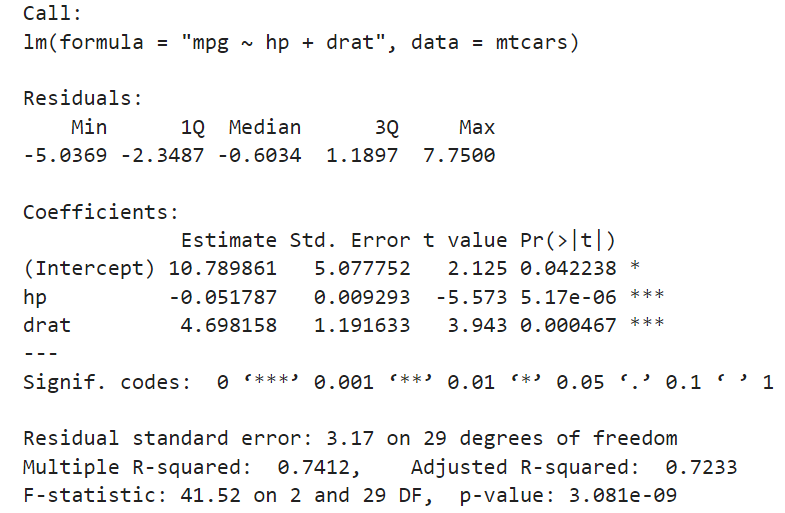
* The p-value represents the probability that the coefficient is zero (has no effect).
* In social science, in general, p-value < 0.05 indicates statistically significant relationship between variables.
* Exploring relationships with regression models

|  |
| --- |
| **plot(mtcars$hp, mtcars$mpg)**  **abline(model1, col='red')** |



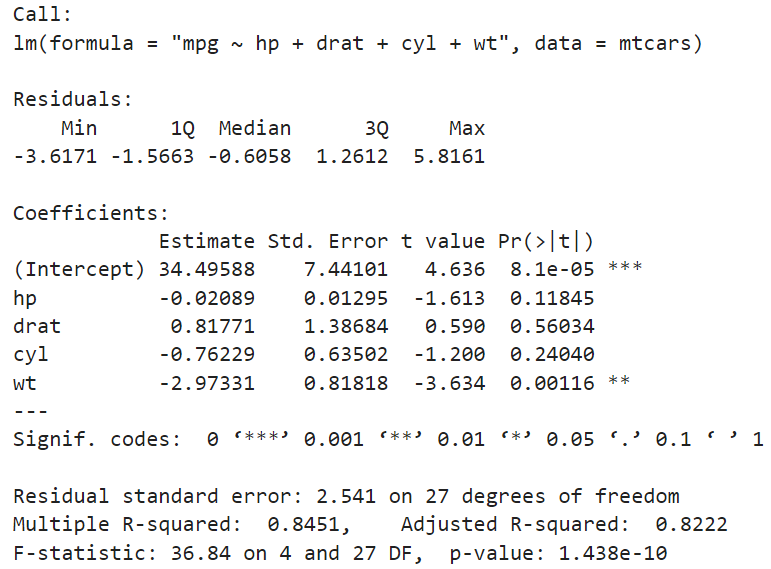
* Exploring relationships with multiple regression models.

|  |
| --- |
| **model2 <- lm(mpg ~ hp + drat, data = mtcars)**  **summary(model2)** |



Let’s interpret the results of model2.

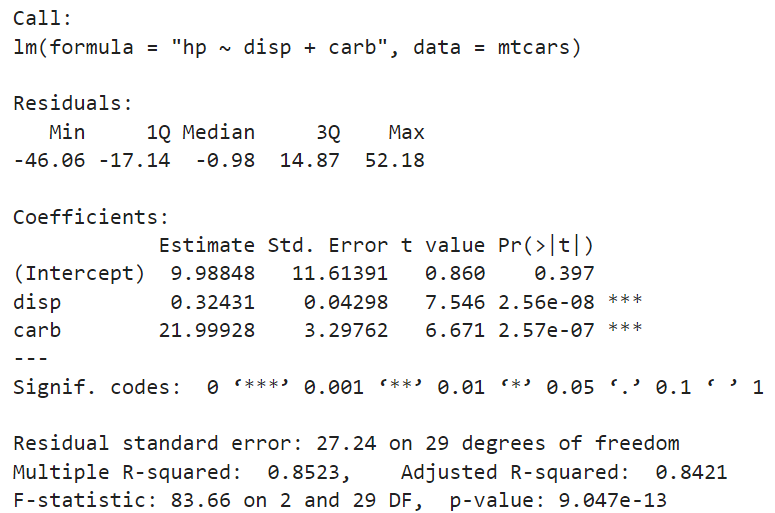
|  |
| --- |
| **model3 <- lm(mpg ~ hp + drat + cyl + wt, data = mtcars)**  **summary(model3)** |



Let’s interpret the results of model3.

* Why **hp, drat**, and **cyl** have insignificant coefficient estimates? : Because all the independent variables are highly correlated to each other. This leads to Multicollinearity problem in the estimation.

|  |
| --- |
| **model4 <- lm(hp ~ disp + carb, data = mtcars)**  **summary(model4)** |

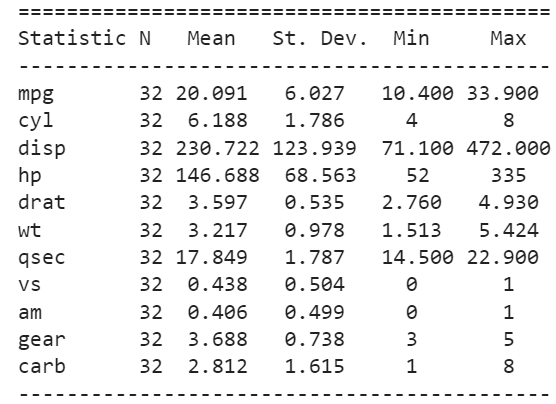


Let’s interpret the results of model4.

**Presenting the model results using "stargazer"**

* Descriptive statistics using "stargazer"

**stargazer(mtcars, type = 'text')**



* Presenting the models results (model1, model2, model3, model4)

**library(stargazer)**

**stargazer(model1, model2, model3, model4, type = "text",**

**digits = 2,**

**covariate.labels = c('Gross horsepower (hp)',**

**'Rear axle ratio (dart)',**

**'Number of cylinders (cyl)',**

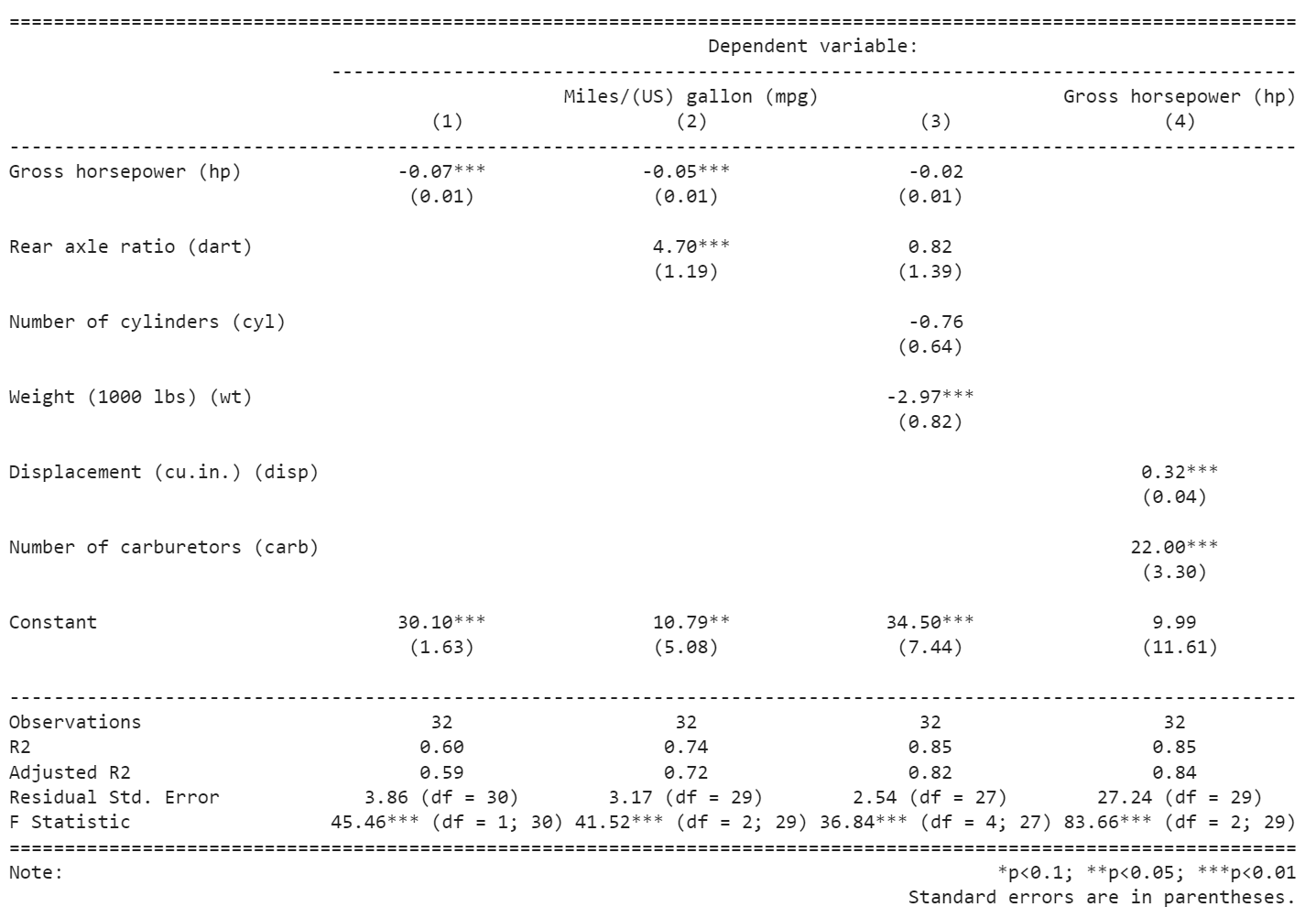
**'Weight (1000 lbs) (wt)',**

**'Displacement (cu.in.) (disp)',**

**'Number of carburetors (carb)'),**

**dep.var.labels = c("Miles/(US) gallon (mpg)", "Gross horsepower (hp)"),**

**notes = "Standard errors are in parentheses.")**



Saving results to a html file named ‘model\_results.html’.

**stargazer(model1, model2, model3, model4, type = "html", out = 'model\_results.html',**

**digits = 2,**

**covariate.labels = c('Gross horsepower (hp)',**

**'Rear axle ratio (dart)',**

**'Number of cylinders (cyl)',**

**'Weight (1000 lbs) (wt)',**

**'Displacement (cu.in.) (disp)',**

**'Number of carburetors (carb)'),**

**dep.var.labels = c("Miles/(US) gallon (mpg)", "Gross horsepower (hp)"),**

**notes = "Standard errors are in parentheses.")**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | | | |
|  | *Dependent variable:* | | | |
|  |  | | | |
|  | Miles/(US) gallon (mpg) | | | Gross horsepower (hp) |
|  | (1) | (2) | (3) | (4) |
|  | | | | |
| Gross horsepower (hp) | -0.07\*\*\* | -0.05\*\*\* | -0.02 |  |
|  | (0.01) | (0.01) | (0.01) |  |
|  |  |  |  |  |
| Rear axle ratio (dart) |  | 4.70\*\*\* | 0.82 |  |
|  |  | (1.19) | (1.39) |  |
|  |  |  |  |  |
| Number of cylinders (cyl) |  |  | -0.76 |  |
|  |  |  | (0.64) |  |
|  |  |  |  |  |
| Weight (1000 lbs) (wt) |  |  | -2.97\*\*\* |  |
|  |  |  | (0.82) |  |
|  |  |  |  |  |
| Displacement (cu.in.) (disp) |  |  |  | 0.32\*\*\* |
|  |  |  |  | (0.04) |
|  |  |  |  |  |
| Number of carburetors (carb) |  |  |  | 22.00\*\*\* |
|  |  |  |  | (3.30) |
|  |  |  |  |  |
| Constant | 30.10\*\*\* | 10.79\*\* | 34.50\*\*\* | 9.99 |
|  | (1.63) | (5.08) | (7.44) | (11.61) |
|  |  |  |  |  |
|  | | | | |
| Observations | 32 | 32 | 32 | 32 |
| R2 | 0.60 | 0.74 | 0.85 | 0.85 |
| Adjusted R2 | 0.59 | 0.72 | 0.82 | 0.84 |
| Residual Std. Error | 3.86 (df = 30) | 3.17 (df = 29) | 2.54 (df = 27) | 27.24 (df = 29) |
| F Statistic | 45.46\*\*\* (df = 1; 30) | 41.52\*\*\* (df = 2; 29) | 36.84\*\*\* (df = 4; 27) | 83.66\*\*\* (df = 2; 29) |
|  | | | | |
| *Note:* | \*p<0.1; \*\*p<0.05; \*\*\*p<0.01 | | | |
|  | Standard errors are in parentheses. | | | |