Multiple Linear Regression

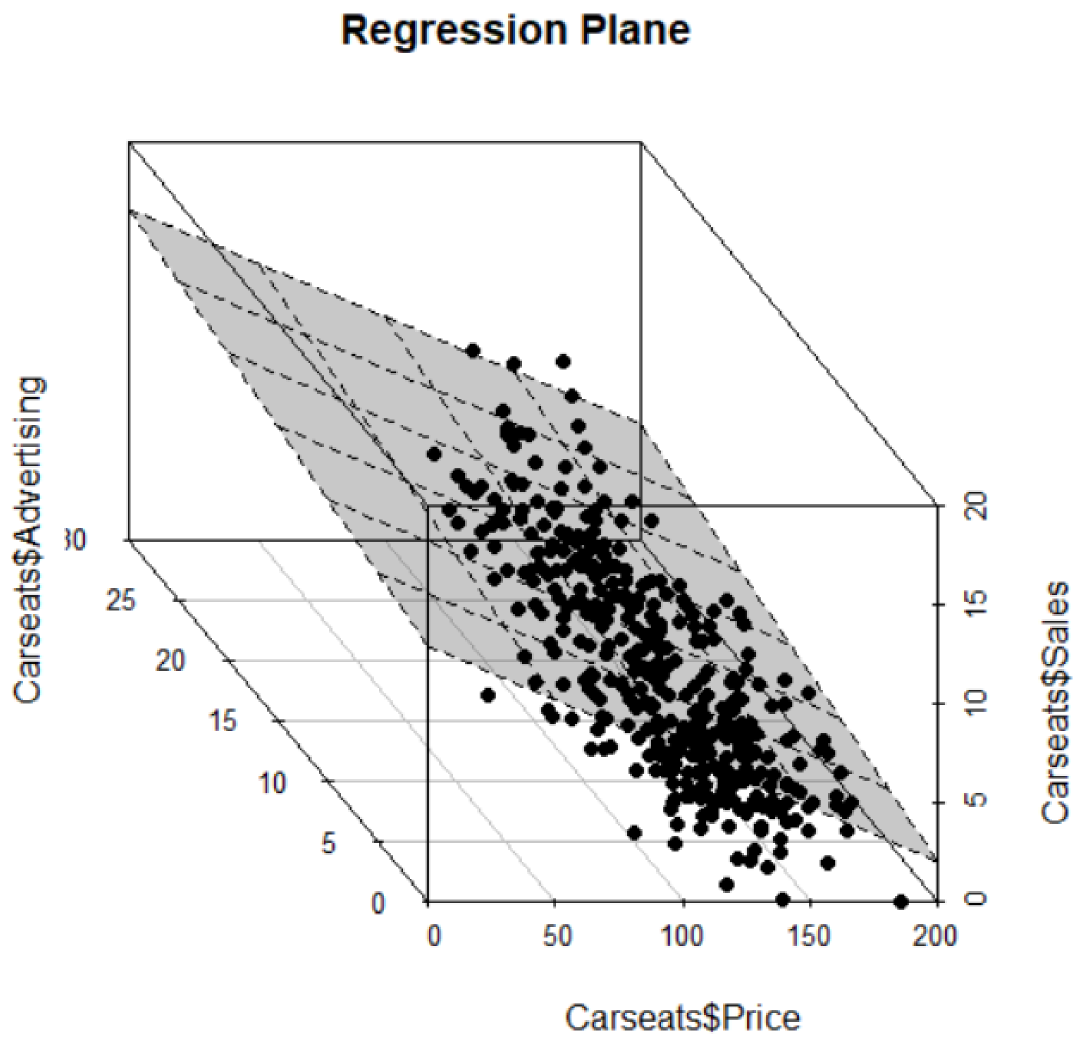
- Simple Linear Regression predict Response for a single predictor X, for example TV
- Multiple Linear Regression deals with multiple predictors, even fitting separate Simple regression to each predictor X this will make us miss some key correlations and associations between predictors and Response

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon$$

For example :

$$\text{sales} = \beta_0 + \beta_1 TV + \beta_2 Radio + \beta_3 Newspaper$$

Geometric Interpretation (Regression Plane)



a

- Unlike Simple Linear Regression Multiple Linear regression have multiple predictors which is draw as a hyperplane

Estimating The Regression Coefficients :

- The Coefficients in the multiple regression are unknown $\beta_0, \beta_1 \ldots, \beta_p$
- So we estimate them as as in the Simple Linear Regression

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p$$

Multiple Linear Regression Matrix Form :

- The multiple Regression formula can be written in a matrix form making it better to work with and derive the Coefficients
- $$Y = X\beta + \varepsilon$$

  With :
  - $Y \rightarrow$ Vector of dependent variables Response
  - $X \rightarrow$ Matrix of $n * p$ dimensions + **intercept** β_0
  - $\beta \rightarrow$ Vector of Coefficients (to be estimated)
  - $\varepsilon \rightarrow$ Vector of Error terms

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{p1} \\ 1 & X_{12} & X_{22} & \dots & X_{p2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \dots & X_{pn} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

- $E(\varepsilon) = 0$
- $var(\varepsilon) = \sigma^2 I_{n*n}$
- $\beta's$ are called partial regression coefficients cause β_1 is the expected change in Y per Unite change in X_1, While holding other $X's$ constant

Least Squares Estimator :

Multiple Regression often reveal how much a predictor X_i effect the prediction Response Y, that the Simple regression don't address

- Due to the slop in the Simple Linear Regression represent the average increase in Y without association with the other predictors
- In Multiple Regression the average increase in Y associated with increasing X_1 while holding the others X fixed
- Multiple Regression can suggest a no relationship between Y and a Predictor X

Deriving The coefficients estimates Using OLS method Ordinary Least Squares

Assessing the Accuracy of the Coefficients $\hat{\beta}_p$

Same as in the Simple Linear Regression we use :

1. Standard Error / Variance
2. Confidence intervals
3. Hypothesis testing (F-test)

Standard Error of $\hat{\beta}_p$

The Standard Error is the square root of its variance of $\hat{\beta}_j$ is how much $\hat{\beta}_j$ will vary from the mean or the expected value of $\hat{\beta}_j$ we found that its unbiased in Standard Error Derivation

$$E[\hat{\beta}] = \beta$$

We also Derived the the Standard Error of $\hat{\beta}$ in Standard Error Derivation and got:

$$\text{Var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$$

$$SE(\hat{\beta}) = \sigma \sqrt{(X^T X)^{-1}}$$

- σ^2 is almost unknown in all practical situations
- We use the Sample standard deviation S^2
  - $S^2 = \frac{\sum e_i^2}{n-p} = \frac{e^T e}{n-p} = MSE$

Confidence Interval

- Constructing a how Confident we are on the estimated coefficients $\hat{\beta}_j$
  From Ordinary Least Squares and Standard Error Derivation we know

  $$\hat{\beta} = (X^T X)^{-1} X^T Y$$

  $$\text{Var}(\hat{\beta}) = \sigma^2 (X^T X)^{-1} \ \rightarrow \text{SE}(\hat{\beta}) = \sigma \sqrt{(X^T X)^{-1}}$$

  Since σ^2 is most of the time :

  $$\text{SE}(\hat{\beta}) = S \sqrt{(X^T X)^{-1}}$$

  We construct the following **confidence interval**:

  $$t_{n-p} = \frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta})}$$

- $n-p$ Degrees of Freedom
- $\hat{\beta}_j - \beta_j$ How far our estimate to the real coefficient

$$P(\hat{\beta}_j - t \,.\, \text{SE}(\hat{\beta}_j) < \beta_j < \hat{\beta}_j + t \,.\, \text{SE}(\hat{\beta}_j)) = 1 - \alpha$$

and get :

$$\hat{\beta}_j \pm t_{\frac{\alpha}{2},n-p} \,.\, \mathrm{SE}(\hat{\beta}_j)$$

Hypothesis Testing (F-Statistics)

- The question asked is : Is there a relationship between the Response and the Predictors X
- We check this using the hypothesis Testing

$$H_0 : \beta_1 = \beta_2 = \dots \beta_p = 0 \;\; \text{There is no realtionship between the predictors and the reponse}$$

$$H_a : \text{alteast one } \beta_j \neq 0$$

- To test this Hypothesis we use $F - statistics$ test

$$F = \frac{\frac{TSS-RSS}{p}}{\frac{RSS}{(n-p-1)}}$$

- $TSS = \sum(y_i - \bar{y})^2 \rightarrow$ total sum squared
- $RSS = \sum(y_i - \hat{y}_i)^2 \rightarrow$ Residual sum squared
- We divide by RSS to have a proportion of difference
- P number of predictors to explain Y
- -1 is the intercept β_0
- $TSS - RSS$ is the explained variance by the regression

If the linear model assumptions are correct:

$$E\left\{\frac{RSS}{(n-p-1)}\right\} = \sigma^2$$

- That the Expected value of the **unexplained variance** is due to irreducible error ε

if H_0 is true

$$E\left\{\frac{TSS-RSS}{p}\right\} = \sigma^2$$

- Which means that the predictors X didn't effect the outcome response and have no relationship between each other
- Cause if there was a relationship between the predictors and response its gonna be :

$$E\left\{\frac{TSS-RSS}{p}\right\} > \sigma^2$$

The $F - test$ give us evidence to either reject or accept the **null hypothesis**, How big the $F - statistic$ should be to reject H_0

- Depends on the n and p
- if n number of Observation is large little larger than 1 is enough
- if n number is small $\rightarrow$ need larger F-Statistic

Sometimes we want to test a particular **subset** of predictors coefficients are zero

$$H_0 : \beta_{p-q+1} = \beta_{p-q+2} = \dots \beta_p = 0$$

$$F = \frac{(RSS_0 - RSS)}{q} \times \frac{n-p-1}{RSS}$$

- $RSS_0 \rightarrow$ Residual sum of squares for the new model that only conclude q coefficients we want to test
  WHY F-TEST:
- when number of the variables is large $p = 100$ in the H_0 there is a 5% chance of p-value being below 0.05 by chance
- That's why individual t-test each predictor X can lead to wrong assumptions
- F-Test avoids that by deciding and adjusting for the number of predictors $\frac{1}{p}$
- Nothing I learned till now will help if number of variables $p > n$