This lecture: R^2 and F-test

§1. R^2 and adjusted-R^2

- Multiple Regression with p predictors

$$Y_{n\times 1} = X_{n\times(p+1)}\,\beta_{(p+1)\times 1} + e_{n\times 1} \quad \text{(matrix form)}$$

OLS $\hat{\beta}_{(p+1)\times 1} = (X^T X)^{-1} X^T Y$

Fitted vaule $\hat{Y}_{n\times 1} = \begin{pmatrix} \hat{y}_1 = x_1^T \hat{\beta} \\ \vdots \\ \hat{y}_n = x_n^T \hat{\beta} \end{pmatrix} = X\hat{\beta}$

Residuals $\hat{e}_{n\times 1} = \begin{pmatrix} \hat{e}_1 \\ \vdots \\ \hat{e}_n \end{pmatrix}_{n\times 1} = Y_{n\times 1} - \hat{Y}_{n\times 1}$

- Residual Sum of Squares (RSS) $\quad RSS = \sum_{i=1}^{n} \hat{e}_i^2 = \hat{e}^T \underset{n\times 1}{\hat{e}}$

$$\left[\hat{\sigma}^2 = \frac{RSS}{n-(p+1)} \right]$$

- Total Sum squares (TSS or SYY)

$$TSS = \sum_{i=1}^{n} (Y_i - \bar{Y})^2 \quad \text{where } \bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$$

[Recall that the sample variance of $Y_1 \cdots Y_n$ is $\dfrac{TSS}{n-1} = \dfrac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}$]

- Regression Sum Squares (SS_{Reg})

$$SS_{Reg} = TSS - RSS.$$

☆ Definition of R^2:

$$R^2 = 1 - \frac{RSS}{TSS} = \frac{SS_{Reg}}{TSS}$$

R^2 measures the proportion of varibility of Y (Responses) that can be explained by X (predictors).
$\in [0,1]$

· $SS_{Reg} = TSS - RSS$ equals $\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$

$$TSS = SS_{Reg} + RSS$$

☆ $$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n} \underbrace{(Y_i - \hat{Y}_i)}_{\hat{e}_i}{}^2$$

why?

$$\sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y})^2$$

$$= \underbrace{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}_{RSS} + \underbrace{\sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2}_{SS_{Reg}} + \underbrace{2 \cdot \sum_{i=1}^{n} (Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y})}_{= 0 \ (*)}$$

For the (*) term above, we have

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) = \sum_{i=1}^{n} \hat{e}_i \hat{Y}_i - \sum_{i=1}^{n} \hat{e}_i \bar{Y}$$

$$= \hat{Y}_{n\times 1}^{T} \hat{e}_{n\times 1} - \bar{Y} \cdot \Big(\sum_{i=1}^{n} \hat{e}_i\Big) = 0$$

$\sum_{i=1}^{n} \hat{e}_i = 0$ shown in previous Lecture.

$$\hat{Y}^T \hat{e} = (X\hat{\beta})^T \hat{e} = \hat{\beta}^T X^T \hat{e} = 0$$

by the normal equations $X^T \hat{e} = 0$

Therefore, we have (☆) equation.

- A property of R^2 is that R^2 will always increase when adding new predictors into the model. $R^2 = 1 - \frac{RSS}{TSS}$

$\Leftrightarrow$ RSS will always decrease when adding new predictors.

E.g.

$$\begin{array}{llll} Y \sim X_1 & \longrightarrow R_1^2 & . & RSS_1 \\ & \quad \wedge & & \vee \\ Y \sim X_1 + X_2 & \rightarrow R_2^2 & . & RSS_2 \end{array}$$

☆ Adjusted-R^2: adjust for the number of predictors in the model

$$= 1 - \underbrace{(1-R^2)}_{\frac{RSS}{TSS}} \frac{n-1}{n-(p+1)}$$

$$= 1 - \frac{RSS/[n-(p+1)]}{TSS/(n-1)}$$

$\hat{\sigma}^2$ estimated from OLS

Sample variance of responses

Adjusted-R^2 is useful for selecting "true" models

feature selection

E.g.:

Case ①

(true) $Y \sim X_1$ $\quad R_1^2 = 1 - \frac{RSS_1}{TSS}$; Adjusted-$R_1^2$ $= 1 - \frac{RSS_1/(n-2)}{TSS/(n-1)}$

$\wedge$ $\quad \vee$ (very likely)

$Y \sim X_1 + X_2$ $\quad R_2^2 = 1 - \frac{RSS_2}{TSS}$; Ad-$R_2^2$ $= 1 - \frac{RSS_2/(n-3)}{TSS/n-1}$

in this case, $RSS_1 - RSS_2$ will be very "small".

(X_2 not important)

Case ②

$Y \sim X_1$ $\quad R_1^2$ $\quad$ Adj-R_1^2

$\wedge$ $\quad \wedge$ (very likely)

(true) $Y \sim X_1 + X_2$ $\quad R_2^2$ $\quad$ Adj-R_2^2

in this case, $RSS_1 - RSS_2$ will be "large"

(X_2 is important)

(Model selection will be discussed in later lectures

Feature selection)

§2. F - test

compare $\overbrace{\text{Model 1}}^{H_0}$ v.s $\overbrace{\text{Model 2}}^{H_A}$

where model 1 is a submodel of Model 2

E.g. $M_1: Y \sim X_1$, $M_2: Y \sim X_1 + X_2$ ✓ ; $M_1: Y \sim X_1 + X_3$, $M_2: Y \sim X_1 + X_2$ ✗

① Overall F-test (the F-test in the last row R summary output)

For model : $Y_i = \beta_0 + \beta_1 x_{i,1} + \cdots + \beta_p x_{i,p} + e_i$, $i = 1, \ldots, n$

$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$ v.s H_A: at least one β_j, $j = 1, \cdots, p$ is not 0.

$\Updownarrow$

Mean functions $H_0: E[Y_i | X_i] = \beta_0$

$H_A: E[Y_i | X_i] = \beta_0 + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p}$

Procedure (overall F-test)

Step 1 : Fit the model under H_0 ($Y \sim 1$)

get $RSS_{H_0} = TSS = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$

(practice problem)

and $df_{H_0} = n-1$

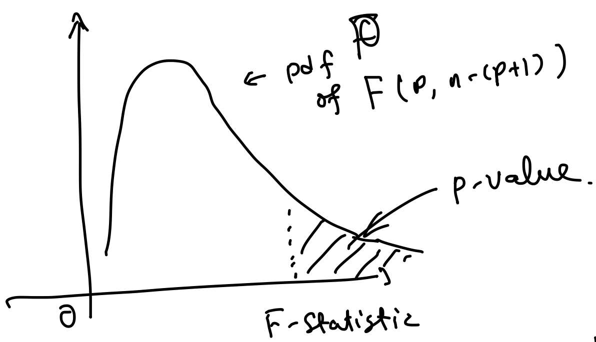
Step 2 : Fit the model under H_A ($Y \sim X_1 + \cdots + X_p$)

get RSS_{H_A} and $df_{H_A} = n - (p+1)$

Step 3 : Compute or F-test statistic as

$$\text{F-statistic} = \frac{(RSS_{H_0} - RSS_{H_A})/(df_{H_0} - df_{H_A})}{RSS_{H_A}/df_{H_A}}$$

when H_0 is true and statistical errors are normally distributed → F-distribution with d.f.s $(df_{H_0} - df_{H_A},\ df_{H_A})$



For overall F-test,

$$df_{H_0} - df_{H_A} = (n-1) - (n-(p+1)) = p$$

and $df_{H_A} = n-(p+1)$.

Rej H_0 if p-value $< \alpha$ (e.g. 5%)

Accept H_0 otherwise.