

Lecture 5-3 : Test for Significance of Regression

노트 제목

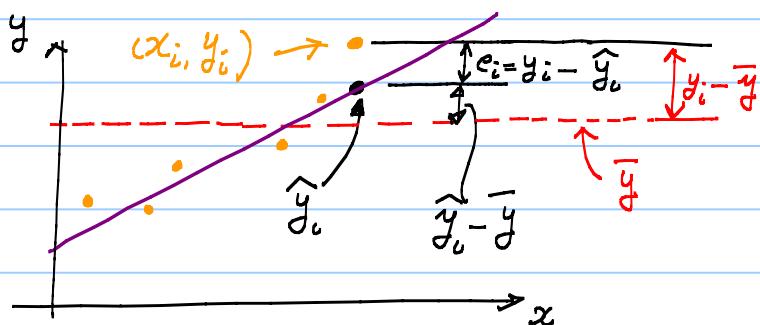
Assumption : $\varepsilon_i \sim N_{ID}(0, \sigma^2)$

$\rightarrow y_i$: normally and independently distributed with zero mean and σ^2 variance

$$(\text{Regress model: } \hat{y}_i = b_0 + \sum_{j=1}^k b_j x_{ij})$$

$$\Leftrightarrow \hat{y} = \mathbf{x} \cdot \mathbf{b}$$

Consider



- y_i : experimental data
- \hat{y}_i : estimated value by the regression model
- \bar{y} : mean value of y_i

$$\bar{y} = (\sum_{i=1}^n y_i) / n$$

① SS_T (Total Sum of Square)

$$SS_T = \sum_{i=1}^n (y_i - \bar{y})^2$$

Measure of the variability between y_i

where

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

(2)

Claim

$$\underline{SS_T = SS_R + SS_E}$$

Total variability
 ($y_i - \bar{y}$)
 ↗

Regression
 Sum of square
 $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$

↗ Sum of square
 of residual
 $\sum_{i=1}^n (y_i - \hat{y}_i)^2$

the part of
the variability
explained by regression
model

the part of the variability
explained by errors
(related to σ^2)

< proof of claim >

$$\begin{aligned}
 SS_T &= \sum_{i=1}^n (y_i - \bar{y})^2 \\
 &= \sum_{i=1}^n \{(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})\}^2 \\
 &= \sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\
 &\quad + 2 \sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) \\
 &\stackrel{\Delta}{=} SS_R + SS_E
 \end{aligned}$$

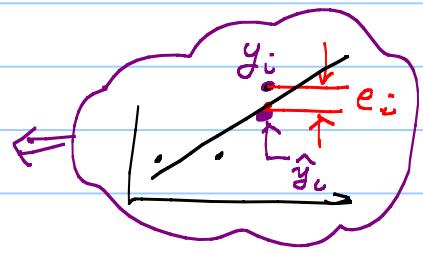
$\stackrel{\Delta}{=} SS_E$
 $\stackrel{\Delta}{=} SS_R$
 A = 0
 will be shown below

$$= SS_R + SS_E$$

(3)

Why $A = 0$?

$$\begin{aligned}
 A &= \sum (\hat{y}_i - \bar{y})(\hat{y}_i - \hat{\bar{y}}_i) \\
 &= \sum (\hat{y}_i - \bar{y}) e_i \\
 &= \sum_{i=1}^n \hat{y}_i e_i - \bar{y} \sum_{i=1}^n e_i
 \end{aligned}$$



$$\text{i)} \quad \sum_{i=1}^n e_i = \sum_{i=1}^n (y_i - \hat{y}_i)$$

$$\begin{aligned}
 &= \sum_{i=1}^n \left[y_i - (b_0 + \sum_{j=1}^k b_j x_{ij}) \right] \quad \text{due to Eq. (2b)} \\
 &= \sum_{i=1}^n y_i - n b_0 - \sum_{j=1}^k b_j \sum_{i=1}^n x_{ij} = 0 \quad \text{of Lecture 5-1}
 \end{aligned}$$

$$\text{ii)} \quad \sum_{i=1}^n \hat{y}_i e_i = \hat{y}^T e$$

$$\begin{aligned}
 &= (\underline{x}^T \underline{b})^T (\underline{y} - \hat{\underline{y}}) \\
 &= \underline{b}^T \underline{x}^T \underline{y} - \underline{b}^T \underline{x}^T \hat{\underline{y}} \\
 &= \underline{b}^T \underline{x}^T \underline{y} - \underline{b}^T \underline{x}^T (\underline{x} \underline{b}) \\
 &= \underline{b}^T (\underline{x}^T \underline{y} - \underline{x}^T \underline{x} \underline{b})
 \end{aligned}$$

(B) { will show that
this term $\equiv 0$

Notation:

- $\underline{y} = \{y_1, \dots, y_n\}^T$
- $\hat{\underline{y}} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n\}^T$
- $\underline{x} = \underline{x} \underline{b}$
- $\underline{x} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ \vdots & & & \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \quad \underline{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_k \end{bmatrix}$
- $m \times (k+1)$
- $= n \times p$
- $p \times 1$

$$\begin{aligned}
 \underline{e} &= \{e_1, \dots, e_n\}^T \\
 &= \{y_1 - \hat{y}_1, \dots, y_n - \hat{y}_n\}^T
 \end{aligned}$$

(4)

To Show ③, Recall that \underline{b} was selected to satisfy

$$\frac{\partial L}{\partial \beta} \Big|_{\beta=\underline{b}} = 0 \Leftrightarrow \underbrace{\underline{x}^T \underline{y} - \underline{x}^T \underline{x} \underline{b}}_0 = 0$$

where

$$L = \underline{\varepsilon}^T \underline{\varepsilon}$$

$$= (\underline{y} - \underline{x}\underline{\beta})^T (\underline{y} - \underline{x}\underline{\beta})$$

$$L = \underline{\varepsilon}^T \underline{\varepsilon}$$

$$= (\underline{y} - \underline{x}\underline{\beta})^T (\underline{y} - \underline{x}\underline{\beta})$$

$$\left[\begin{array}{l} \underline{y} = \underline{x}\underline{\beta} + \underline{\varepsilon} \\ \text{↑ measured response} \end{array} \right]$$

Thus, ③ holds and

$$\boxed{\sum \hat{y}_i e_i = \hat{y}^T e \equiv 0} \quad \Leftarrow$$

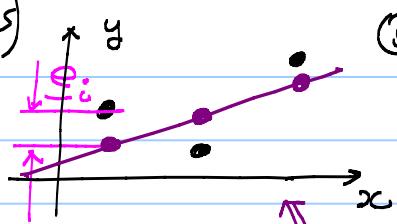
From ① and ③,

$$A \equiv 0.$$



(sum of residual squares) ⑤

EXAMINE $SS_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum e_i^2$

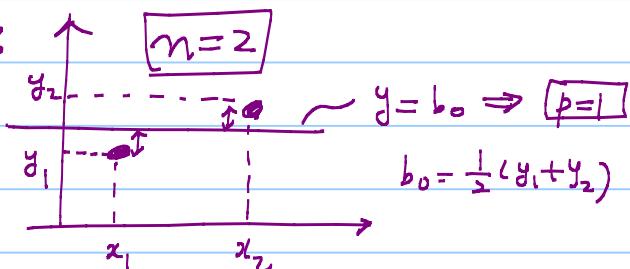


$\sigma^2 \triangleq \frac{E(SS_E)}{n-p}$ Degree of freedom of SS_E

In the above example,
 $p=2$ because the regression model contains 2 parameters
 (b_0, b_1) as in $y = b_0 + b_1 x$

More on DOF

CONSIDER: $n=2$



$$\begin{aligned} SS_E &= (b_0 - y_1)^2 + (b_0 - y_2)^2 = \left[\frac{1}{2}(y_1 + y_2) - y_1 \right]^2 + \left[\frac{1}{2}(y_1 + y_2) - y_2 \right]^2 \\ &= \frac{(y_2 - y_1)^2}{4} + \frac{1}{4}(y_1 - y_2)^2 = \frac{1}{2}(y_1 - y_2)^2 \end{aligned}$$

SS_E : depends only on one Single measure

$$\therefore DOF = 1$$

⑥

Estimator of $\sigma^2 \doteq \hat{\sigma}^2$

$$\hat{\sigma}^2 = \frac{SS_E}{n-p}$$

The notion of
Degree of freedom
is useful to properly
estimate σ^2

Analysis of variance for significance of Regression

Source of Variation	Sum of Square	Degree of freedom	Mean Square	F ₀
Total	SS _T	n-1		
regression	SS _E	n-K-1	$MS_E = \frac{SS_E}{n-K-1}$	$\frac{MS_R}{MS_E}$
Error or residual	SS _R	K	$MS_R = \frac{SS_R}{K}$	

Remark 1: $SS_T = \sum_{i=1}^n (y_i - \bar{y})^2$
 $\bar{y} = \frac{1}{n} \sum y_i$ (1 equation used to get \bar{y})

\therefore DOF for $SS_T = n-1$

(7)

Remark 2: $SS_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ used i.e.g. ($\bar{y} = \sum \hat{y}_i$)

$\hat{y}_i = b_0 + \sum_{j=1}^k b_j x_{ij}$; has $(k+1)$ coeffs = $(k+1)$ DoF in the model

$$\therefore \text{DoF of } SS_R = (k+1) - 1 = k$$

Remark 3: $SS_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2$

has "n" data points to be determined by selected $\Rightarrow \text{DoF} = n$

$$\frac{\partial L}{\partial \beta} \Big|_{\beta=0} = 0 \Leftrightarrow (k+1) \text{ eqs.}$$

$$\therefore \text{DoF of } SS_E = n - (k+1)$$

Remark 4: To compare the variation from the regression model and the variation from the error (or residual), better to normalize them wrt DoF

$$\Rightarrow \text{compare } MS_R = SS_R / k \text{ and } MS_E = SS_E / (n-k-1)$$

(8)

$$F_0 \triangleq \frac{MS_R}{MS_E} \quad \begin{matrix} \leftarrow \text{normalized variation from regression model} \\ \leftarrow \text{normalized variation from error} \end{matrix}$$

$$= \frac{SS_R / k}{SS_E / (n-k-1)} \quad \begin{matrix} \leftarrow F_{\text{test}} \end{matrix}$$

④ We say that the regression model is significant by $(1-\alpha) \times 100\%$ if

$$F_0 > F_{\alpha, k, n-k-1}$$

↑ ↓
 DOF of SS_E
 DOF of SS_R

example: $F_{\alpha, n-j}$ $\alpha = 0.05$, $k = 2$, $n = 14$
 $y = a + bx$ 14 data pts
 linear model

$$F_{\alpha, k, n-k-1} = F_{0.05, 2, 11} = 3.98$$

↑ from Table.

(9)

Alternative Method to estimate the result by regression model

→ The coefficient of multiple determination

$$R^2 = \frac{SS_R}{SS_T} \quad (\text{if it is close to 1, the result is satisfactory})$$

$$= \frac{SS_T - SS_E}{SS_T} = 1 - \frac{SS_E}{SS_T}$$

(Measure of variability reduction by the use of regressor variables x_1, \dots, x_k)

If $R^2 = 0.98$, it means that 98% of the variation of y is explained by the use of x_1, \dots, x_k

Better to use adjusted R^2_{adj}

$$R_{adj}^2 = 1 - \frac{SS_E / (n-p)}{SS_T / (n-1)}$$

$b = k+1$

(usually $0.9 \leq R_{adj}^2 \leq 1 \rightarrow \text{acceptable}$)

(6)

TEST of Individual Regression Coefficients
and group of coefficients



use t-test

$$\text{Recall } \text{Cov}(\hat{b}) = \sigma^2 (\hat{X}^\top \hat{X})^{-1}$$

$$\rightarrow \left\{ \begin{array}{l} \text{Cov}(b_{jj}) \Rightarrow \text{Define it as } [se(b_j)]^2 \\ se(b_j) = \sqrt{\sigma^2 C_{jj}} \end{array} \right.$$

Standard error of
the regression coefficient

with $C_{jj} = [\hat{X}^\top \hat{X}]_{jj}^{-1}$

Thus introduce

$$t_0 = \frac{b_j}{se(b_j)}$$

for each b_j

(the larger t_0 is, the better the corresponding b_j is)

$$\text{If } |t_0| > t_{\frac{\alpha}{2}, \underbrace{n-k-1}_{\text{DoF}}}, \quad \left(\left(t_{\frac{\alpha}{2}, k} \right)^2 = F_{\alpha, b, k} \right)$$

the use of variable x_j is $(1-\alpha)\%$ significant.

(11)

example : if $\alpha = 0.05$ (95% reliability of each coeff.)
and $n=14$, $k=2$ ($\Leftrightarrow p=3$; 3 coeffs to be
determined)

$$t_{\frac{\alpha}{2}, n-k-1} = t_{0.025, 11} = 2.220$$

\therefore for 95% significance, $|t_0| > 2.220$

(12)

Example 2.1

The transistor gain in an integrated circuit device between emitter and collector (hFE) is related to two variables that can be controlled at the deposition process, emitter drive-in time (ξ_1 , in minutes), and emitter dose (ξ_2 , units of 10^{14} ions). Fourteen samples were observed following deposition, and the resulting data are shown in Table 2.2. We will fit a linear regression model using gain as the response and emitter drive-in time and emitter dose as the regressor variables.

Table 2.2 Data on Transistor Gain (y) for Example 2.1

Observation	ξ_1 (drive-in time, minutes)	ξ_2 (dose, 10^{14} ions)	x_1	x_2	y (gain or hFE)
1	195	4.00	-1	-1	1004
2	255	4.00	1	-1	1636
3	195	4.60	-1	0.6667	852
4	255	4.60	1	0.6667	1506
5	225	4.20	0	-0.4444	1272
6	225	4.10	0	-0.7222	1270
7	225	4.60	0	0.6667	1269
8	195	4.30	-1	-0.1667	903
9	255	4.30	1	-0.1667	1555
10	225	4.00	0	-1	1260
11	225	4.70	0	0.9444	1146
12	225	4.30	0	-0.1667	1276
13	225	4.72	0	1	1225
14	230	4.30	0.1667	-0.1667	1321

Solution:

The least squares fit with the regression coefficients reported to one decimal place is

$$\hat{y} = 1242.3 + 323.4x_1 - 54.8x_2$$

$$\hat{y} = 1242.3 + 323.4\left(\frac{\xi_1 - 225}{30}\right) - 54.8\left(\frac{\xi_2 - 4.36}{0.36}\right)$$

$$\hat{y} = -520.1 + 10.781\xi_1 - 152.15\xi_2$$

Table 2.3 Observations, Fitted Values, Residuals, and Other Summary Information for Example 2.1

Observation	y_i	\hat{y}_i	e_i	h_{ii}	r_i	t_i	D_i
1	1004.0	973.7	30.3	0.367	1.092	1.103	0.231
2	1636.0	1620.5	15.5	0.358	0.553	0.535	0.057
3	852.0	882.4	-30.4	0.317	-1.052	-1.057	0.171
4	1506.0	1529.2	-23.2	0.310	-0.801	-0.787	0.096
5	1272.0	1266.7	5.3	0.092	0.160	0.153	0.001
6	1270.0	1281.9	-11.9	0.133	-0.365	-0.350	0.007
7	1269.0	1205.8	63.2	0.148	1.960	2.316	0.222
8	903.0	928.0	-25.0	0.243	-0.823	-0.810	0.072
9	1555.0	1574.9	-19.9	0.235	-0.651	-0.633	0.043
10	1260.0	1297.1	-37.1	0.197	-1.185	-1.209	0.115
11	1146.0	1190.6	-44.6	0.217	-1.442	-1.527	0.192
12	1276.0	1251.4	24.6	0.073	0.730	0.714	0.014
13	1225.0	1187.5	37.5	0.233	1.225	1.256	0.152
14	1321.0	1305.3	15.7	0.077	0.466	0.449	0.006

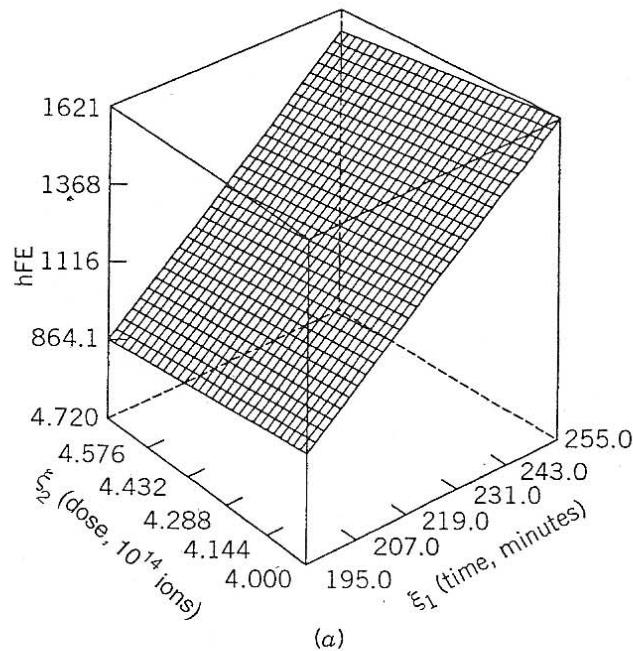
↑ ↑ ↑

$n=14$

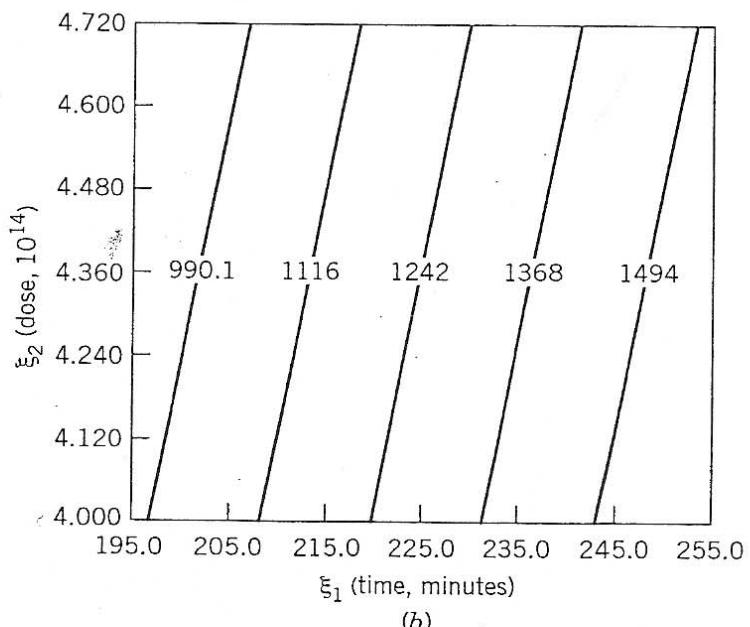
$n=2$

($p=3$)

13



(a)



(b)

Figure 2.1 (a) Response surface for gain, Example 2.1. (b) The gain contour plot.

(14)

Table 2.5 Test for Significance of Regression, Example 2.3

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
SS_R : Regression	651,996.1	2	325,983.0	267.2	4.74×10^{-10}
SS_E : Error	13,421.1	11	1220.1		
SS_T : Total	665,387.2	13			

$$R^2 = \frac{SS_R}{SS_T} = \frac{651,966.1}{665,387.2} = 0.9798$$

$$\begin{aligned} R_{\text{adj}}^2 &= 1 - \frac{n-1}{n-p}(1-R^2) \\ &= 1 - \frac{13}{11}(1-0.9798) \\ &= 0.9762 \end{aligned}$$

X

$$\hat{\sigma}^2 = \frac{SS_E}{n-p} = \frac{13,421.1}{14-3} = 1220.1$$

diagonal elements of $(X'X)^{-1}$ corresponding to β_1 and β_2 are $C_{11} = 0.165954$ and $C_{22} = 0.142876$, respectively, so the two t -statistics are computed as follows:

$$\begin{aligned} \text{For } H_0: \beta_1 = 0: \quad t_0 &= \frac{b_1}{\sqrt{\hat{\sigma}^2 C_{11}}} \\ &= \frac{323.4}{\sqrt{(1220.1)(0.165954)}} = \frac{323.4}{14.2} = 22.73 \end{aligned}$$

$$\begin{aligned} \text{For } H_0: \beta_2 = 0: \quad t_0 &= \frac{b_2}{\sqrt{\hat{\sigma}^2 C_{22}}} = \frac{-54.8}{\sqrt{(1220.1)(0.142876)}} \\ &= \frac{-54.8}{13.2} = -4.15 \end{aligned}$$

The absolute values of these t -statistics are compared with $t_{0.025, 11} = 2.201$ (assuming that we select $\alpha = 0.05$). Both t -statistics are larger than this criterion. Consequently we conclude that $\beta_1 \neq 0$, which implies that x_1 contributes significantly to the model given that x_2 is included, and that $\beta_2 \neq 0$, which implies that x_2 contributes significantly to the model given that x_1 is included.

(75)

A Few More Words

Q:

Variance of $\hat{y}(\underline{x})$ for an arbitrary point \underline{x}_0 ?
 (not the points used for experiment)

T

Recall : $\hat{y} = b_0 + b_1 x_1 + \dots + b_n x_n = \underline{x}^T \underline{b}$

where $\underline{x} = \{1, x_1, x_2, \dots, x_n\}^T$

Define : $\underline{x}_0 = \{1, x_{01}, x_{02}, \dots, x_{0n}\}^T$

At $\underline{x} = \underline{x}_0$

$\cdot \hat{y}(\underline{x}_0) = \underline{x}_0^T \underline{b}$

$\cdot \text{Var}[\hat{y}(\underline{x}_0)] = E[(\hat{y}(\underline{x}_0) - E(\hat{y}(\underline{x}_0)))^2]$ (a)

$= E[\underline{x}_0^T \underline{b}] = \underline{x}_0^T E(\underline{b}) = \underline{x}_0^T \underline{\beta}$

↑
not r.v.

$= E[\underbrace{\underline{x}_0^T (\underline{b} - \underline{\beta})}_{\text{scalar}} \underbrace{(\underline{b} - \underline{\beta})^T \underline{x}_0}_{\text{scalar} = (\text{scalar})^T}]$

(use $\underline{b} - \underline{\beta} = \underline{X}^{-1} \underline{\varepsilon}$ ← from lecture 5-1)

76

$$= E \left[\underline{x}_0^T \underline{X}^{-1} \underbrace{\underline{\varepsilon} \underline{\varepsilon}^T}_{\text{random variables}} \underline{X}^{-T} \underline{x}_0 \right]$$

$$= \underline{x}_0^T \underline{X}^{-1} E(\underline{\varepsilon} \underline{\varepsilon}^T) \underline{X}^{-T} \underline{x}_0$$

$$= \underline{x}_0^T \underline{X}^{-1} \sigma^2 \mathbb{I} \underline{X}^{-T} \underline{x}_0$$

$$= \sigma^2 \underline{x}_0^T (\underline{X}^{-1} \underline{X}^{-T}) \underline{x}_0$$

$$= \sigma^2 \underline{x}_0^T (\underline{X}^T \underline{X})^{-1} \underline{x}_0$$

