

#4.1.1

단순회귀에서 사용된 최소제곱법에 의한 추정방법으로 추정회귀식을 구해보면,

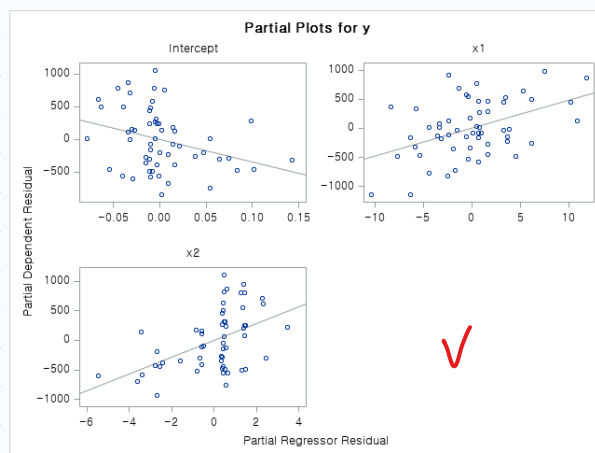
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

$$\hat{y} = -3444.835 + 48.444 x_1 + 141.124 x_2$$

-0.5

#4.1.2

· 추가변수검



```
data reg163;
infile 'C:\WREGDATA\reg163.dat';
input x1 x2 y;
run;

proc reg data=reg163;
model y=x1 x2 / partial;
run;
```

SAS 시스템

The REG Procedure  
Model: MODEL1  
Dependent Variable: y

Number of Observations Read	56
Number of Observations Used	56

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	6318298	3159149	16.04	<.0001
Error	53	10435686	196900		
Corrected Total	55	16753984			

Root MSE	443.73365	R-Square	0.3771
Dependent Mean	3413.03571	Adj R-Sq	0.3536
Coeff Var	13.00115		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-3444.83507	1405.75406	-2.45	0.0176
x1	1	48.44409	13.03427	3.72	0.0005
x2	1	141.12413	34.96089	4.04	0.0002

①  $H_0: \beta_1 = 0 \quad (Y = \beta_0 + \beta_2 X_2 + \epsilon)$   
 $H_1: \beta_1 \neq 0 \quad (Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon)$

$$t^* = \frac{\hat{\beta}_1}{s.e(\hat{\beta}_1)} = 3.72, |t^*| < t(0.025, 53) = 2.006$$

$$p\text{-value} = 2P(t(53) > 3.72) = 0.0005 < 0.05$$

→  $H_0$  기각,  $Y$ 에 대한  $X_1$ 의 추가적인 설명력은 유의하다. 즉,  $X_1$  추가



②  $H_0: \beta_2 = 0 \quad (Y = \beta_0 + \beta_1 X_1 + \epsilon)$   
 $H_1: \beta_2 \neq 0 \quad (Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon)$

$$t^* = \frac{\hat{\beta}_2}{s.e(\hat{\beta}_2)} = 4.04, |t^*| < t(0.025, 53) = 2.006$$

$$p\text{-value} = 2P(t(53) > 4.04) = 0.0002 < 0.05$$

→  $H_0$  기각,  $Y$ 에 대한  $X_2$ 의 추가적인 설명력은 유의하다. 즉,  $X_2$  추가



# 4.1.3

①  $X_1 \rightarrow X_2$  순서일때

②  $X_2 \rightarrow X_1$  순서일때

```
infile 'C:\WREGDATA\reg163.dat';
input x1 x2 y;
run;
```

```
proc reg data=reg163;
model y=x1 x2 / SSR;
run;
```

```
proc reg data=reg163;
model y=x2 x1 / SSR;
run;
```

SAS 시스템

The REG Procedure  
Model: MODEL1  
Dependent Variable: y

Number of Observations Read	56
Number of Observations Used	56

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	631.8298	315.9149	16.04	<.0001
Error	53	10435686	196900		
Corrected Total	55	16753984			

Root MSE	443.73385	R-Square	0.3771
Dependent Mean	3413.03571	Adj R-Sq	0.3536
Coeff Var	13.00115		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-3444.83507	1405.75406	-2.45	0.0176
x1	1	48.44409	13.03427	3.72	0.0005
x2	1	141.12413	34.96089	4.04	0.0002

SAS 시스템

The REG Procedure  
Model: MODEL1  
Dependent Variable: y

Number of Observations Read	56
Number of Observations Used	56

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	631.8298	315.9149	16.04	<.0001
Error	53	10435686	196900		
Corrected Total	55	16753984			

Root MSE	443.73385	R-Square	0.3771
Dependent Mean	3413.03571	Adj R-Sq	0.3536
Coeff Var	13.00115		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	-3444.83507	1405.75406	-2.45	0.0176
x2	1	141.12413	34.96089	4.04	0.0002
x1	1	48.44409	13.03427	3.72	0.0005

② 단변회귀SSR에  
설명변수  $X_1$ 이 추가될때  
더해지는 값

③ 설명변수  $X_1$ 이 추가된 모형에  
 $X_2$ 가 추가될때 더해지는 값

① 단변회귀SSR

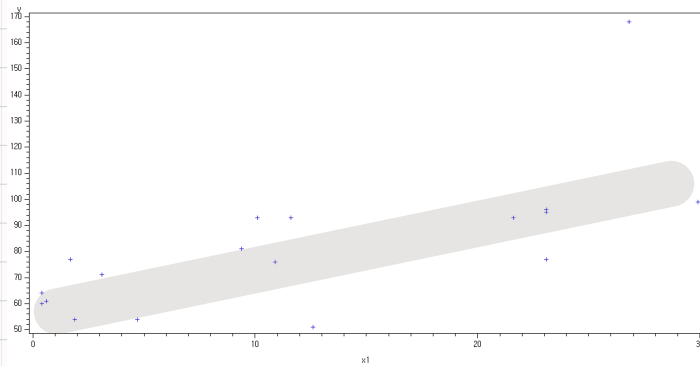
③ 설명변수  $X_2$ 가 추가된  
모형에  $X_1$ 이 추가될때  
더해지는 값

② 단변회귀SSR에  
설명변수  $X_2$ 가 추가  
될때 더해지는 값

- 각각 더해지는 값은 각 설명변수가  $Y$ 에 대해 추가적으로 설명하는 양이다.
- ① + ② + ③의 값은 설명변수가 추가되는 순서에 관계없이 같다.
- $X_1$ 이 추가된 후  $X_2$ 가 설명하는 양보다  $X_2$ 가 추가된 후  $X_1$ 이 설명하는 양이 비교적 작음을 볼 수 있다.

## # 4.3.1

### ① Y와 X1의 산점도



Y와 X1의 관계: X1이 증가할 때 Y도 증가하는 선형관계를 보인다.

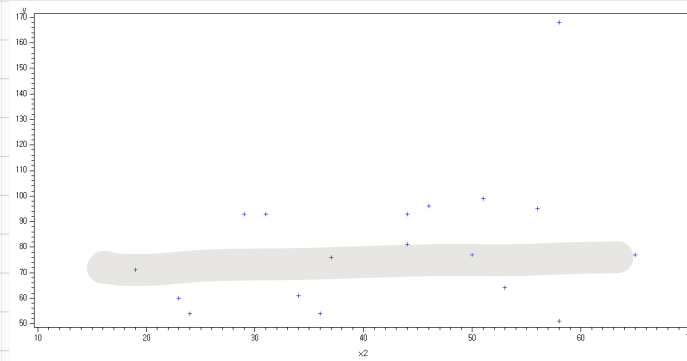
```
data reg165;
  infile 'C:\WREGDATA\reg165.dat';
  input x1 x2 y;
  run;
```

```
proc reg data=reg165;
  model y=x1;
  run;
```

```
proc gplot data=reg165;
  plot y*x1;
  run;
```

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	59.25896	7.41999	7.99	<.0001
x1	1	1.84344	0.47890	3.85	0.0014

### ② Y와 X2의 산점도



Y와 X2의 관계: 선형관계는 보이나 X2는 Y에게 영향을 주지 않는다.

```
data reg165;
  infile 'C:\WREGDATA\reg165.dat';
  input x1 x2 y;
  run;
```

```
proc reg data=reg165;
  model y=x2;
  run;
```

```
proc gplot data=reg165;
  plot y*x2;
  run;
```

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	51.70125	20.44692	2.53	0.0224
x2	1	0.70234	0.46319	1.52	0.1489

## # 4.3.2

### ① X1 → X2 순서일 때

SAS 시스템

The REG Procedure

Model: MODEL1

Dependent Variable: y

Number of Observations Read	18
Number of Observations Used	18

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	5975.66853	2987.83427	6.99	0.0072
Error	15	6413.94258	427.59617		
Corrected Total	17	12390			

Root MSE	20.67840	R-Square	0.4823
Dependent Mean	81.27778	Adj R-Sq	0.4133
Coeff Var	25.44164		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Type I SS
Intercept	1	56.25102	16.31074	3.45	0.0036	118909
x1	1	1.78977	0.55674	3.21	0.0058	5957.02249
x2	1	0.08665	0.41494	0.21	0.8374	18.64604

```
data reg165;
  infile 'C:\WREGDATA\reg165.dat';
  input x1 x2 y;
  run;
```

```
proc reg data=reg165;
  model y=x1 x2 / SS1;
  run;
```

### ② X2 → X1 순서일 때

SAS 시스템

The REG Procedure

Model: MODEL1

Dependent Variable: y

Number of Observations Read	18
Number of Observations Used	18

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	5975.66853	2987.83427	6.99	0.0072
Error	15	6413.94258	427.59617		
Corrected Total	17	12390			

Root MSE	20.67840	R-Square	0.4823
Dependent Mean	81.27778	Adj R-Sq	0.4133
Coeff Var	25.44164		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Type I SS
Intercept	1	56.25102	16.31074	3.45	0.0036	118909
x2	1	0.08665	0.41494	0.21	0.8374	1556.70846
x1	1	1.78977	0.55674	3.21	0.0058	4418.96007

```
data reg165;
  infile 'C:\WREGDATA\reg165.dat';
  input x1 x2 y;
  run;
```

```
proc reg data=reg165;
  model y=x2 x1 / SS1;
  run;
```

## # 4.3.1과 비교해 보면

SAS 시스템  
The REG Procedure  
Model: MODEL1  
Dependent Variable: y

Number of Observations Read 18  
Number of Observations Used 18

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	5957.02249	5957.02249	14.82	0.0014
Error	16	6432.58862	402.03679		
Corrected Total	17	12390			

Root MSE	20.05096	R-Square	0.4808
Dependent Mean	81.27778	Adj R-Sq	0.4484
Coef Var	24.6954		

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	59.25896	7.41999	7.99	<.0001
x1	1	1.84344	0.47890	3.85	0.0014

$H_0: \beta_0 = 0$  기각

SAS 시스템  
The REG Procedure  
Model: MODEL1  
Dependent Variable: y

Number of Observations Read 18  
Number of Observations Used 18

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1556.70846	1556.70846	2.30	0.1489
Error	16	10833	677.05642		
Corrected Total	17	12390			

Root MSE	26.02031	R-Square	0.1256
Dependent Mean	81.27778	Adj R-Sq	0.0710
Coef Var	32.01405		

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	1	51.70125	20.44692	2.53	0.0224
x2	1	0.70234	0.46319	1.52	0.1489

$H_0: \beta_0 = 0$  기각 불가

x1과 x2를 각각 따로 단항회귀의 검정을 했을 때는,

x1은 y에 영향을 주지만 x2는 영향을 주지 않는다는

결과가 나왔다. 두 변수를 다항회귀의 검정을 하면

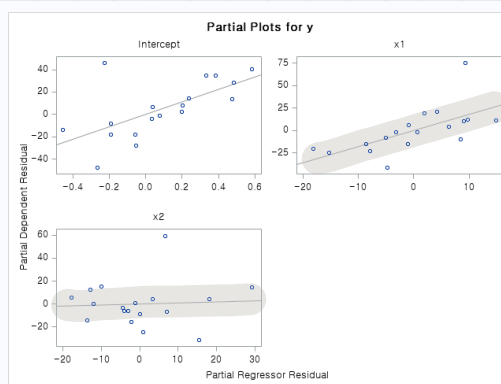
x1과 x2 중 적어도 하나는 y를 설명하는데 영향을 준다는

결과가 나왔다.

// ✓

## # 4.3.3

## 추가변수검정



```
data reg165;
infile 'C:\WREGDATA\reg165.dat';
input x1 x2 y;
run;
```

```
proc reg data=reg165;
model y=x1 x2 / partial;
run;
```

```
proc reg data=reg165;
model y=x2 x1 / partial;
run;
```

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t	Type I SS
Intercept	1	1.12165	0.21086	5.32	<.0001	45.81365
x1	1	0.04170	0.02098	1.99	0.0632	0.35273
x2	1	0.00468	0.01828	0.26	0.8010	0.00179

$$\textcircled{1} \begin{cases} H_0: \beta_1 = 0 & (Y = \beta_0 + \beta_2 X_2 + \epsilon) \\ H_1: \beta_1 \neq 0 & (Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon) \end{cases}$$

$$t^* = \frac{\hat{\beta}_1}{s.e(\hat{\beta}_1)} = 1.99, |t^*| < t(0.025, 15) = 2.132$$

$$p\text{-value} = 2P(t(15) > 1.99) = 0.0632 > 0.05$$

→  $H_0$  기각 불가, y에 대한 x1의 추가적인 설명력은 유의하지 않다. ✓



$$\textcircled{2} \begin{cases} H_0: \beta_2 = 0 & (Y = \beta_0 + \beta_1 X_1 + \epsilon) \\ H_1: \beta_2 \neq 0 & (Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon) \end{cases}$$

$$t^* = \frac{\hat{\beta}_2}{s.e(\hat{\beta}_2)} = 0.26, |t^*| < t(0.025, 15) = 2.132$$

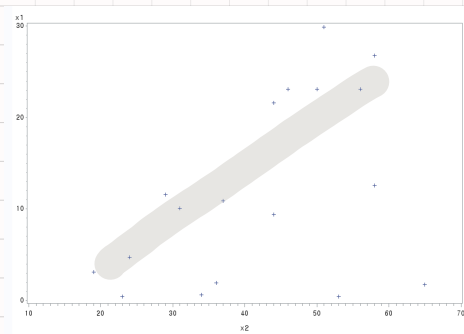
$$p\text{-value} = 2P(t(15) > 0.26) = 0.8010 > 0.05$$

→  $H_0$  기각 불가, y에 대한 x2의 추가적인 설명력은 유의하지 않다. ✓



# 4.3.4

•  $x_1$ 과  $x_2$ 의 산점도



```
data reg165;  
infile 'C:\REGDATA\reg165.dat';  
input x1 x2 y;  
run;  
  
proc gplot data=reg165;  
plot x1=x2;  
run;
```

→ 선형관계를 가짐을 확인할 수 있다. ✓

•  $h_{11}, h_{22}, \dots, h_{nn}$ 은  $H = X(X^T X)^{-1} X^T$ 의 대각원소이며 또  $h_{ii} = p'$ 이다. ✓

각 자료에 대한  $h_{ii}$ 값을 구해 확인해보면

$h_{ii}$  구하고, 비교하기 - 0.5