

## Review 5

1. Least squares estimate and fitted equation
2. Testing hypothesis (F statistic and t statistic) and confidence interval
3. prediction
4.  $r^2, r_{XY}$ , and  $r_{Y\hat{Y}}$ .

- Least square estimate (point estimate):

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{s_{XY}}{s_{XX}}$$

and

$$b_0 = \bar{y} - b_1 \bar{x}.$$

- Interval estimate:

$$\begin{aligned} b_0 \pm (t_{n-2, \alpha/2} \cdot s_{B_0}) &= b_0 \pm \left[ t_{n-2, \alpha/2} \cdot \left( \frac{s^2 \sum_{i=1}^n x_i^2}{ns_{XX}} \right)^{1/2} \right] \\ b_1 \pm (t_{n-2, \alpha/2} \cdot s_{B_1}) &= b_1 \pm \left[ t_{n-2, \alpha/2} \cdot \left( \frac{s^2}{s_{XX}} \right)^{1/2} \right] \end{aligned}$$

- t-test:

$H_0: \beta_0 = c$ :

$$t = \frac{b_0 - c}{s_{B_0}} = \frac{b_0 - c}{\left( \frac{s^2 \sum_{i=1}^n x_i^2}{ns_{XX}} \right)^{1/2}} \Rightarrow \text{reject } H_0: |t| > t_{n-2, \alpha/2}$$

$H_0: \beta_1 = c$ :

$$t = \frac{b_1 - c}{s_{B_1}} = \frac{b_1 - c}{\left( \frac{s^2}{s_{XX}} \right)^{1/2}} \Rightarrow \text{reject } H_0: |t| > t_{n-2, \alpha/2}$$

**Note:**

$$s.e.(B_0) \equiv s_{B_0}, s.e.(B_1) \equiv s_{B_1}$$

● F-test:

$$H_0: \beta_1 = 0:$$

**ANOVA Table:**

Source	DF	SS	MS	F
Regression	1	$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$MSR = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{1} = MSR / MSE$	$f$
Residual (Error)	$n - 2$	$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$	$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2}$	
Total	$n - 1$	$SST = \sum_{i=1}^n (y_i - \bar{y})^2$		

Then

$$\text{reject } H_0: f > f_{1,n-2,\alpha}.$$

**Note:**

For ease of computation, the following equations can be used:

$$MSR = SSR = b_1 s_{XY} = b_1^2 s_{XX}.$$

$$H_0: \beta_0 = \beta_1 = 0:$$

**ANOVA Table:**

Source	DF	SS	MS	F
Regression	2	$\sum_{i=1}^n \hat{y}_i^2$	$\frac{\sum_{i=1}^n \hat{y}_i^2}{2}$	$f = \left( \frac{\sum_{i=1}^n \hat{y}_i^2}{2} \right) / s^2$
Residual (Error)	$n - 2$	$\sum_{i=1}^n (y_i - \hat{y}_i)^2$	$s^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2}$	
Total	$n$	$\sum_{i=1}^n y_i^2$		

Then,

$$\text{reject } H_0: f > f_{2,n-2,\alpha}.$$

**Note:**

$$\sum_{i=1}^n \hat{y}_i^2 = \sum_{i=1}^n y_i^2 - SSE = n\bar{y}^2 + b_1 s_{XY}.$$

- Prediction of  $E(y_p) = \beta_0 + \beta_1 x_p$

Point estimate:

$$\hat{y}_p = b_0 + b_1 x_p = \bar{y} + b_1(x_p - \bar{x})$$

Interval estimate:

$$\hat{y}_p \pm (t_{n-2, \alpha/2} s_{\hat{y}_p}) = [\hat{y}_p - t_{n-2, \alpha/2} s_{\hat{y}_p}, \hat{y}_p + t_{n-2, \alpha/2} s_{\hat{y}_p}],$$

$$s_{\hat{y}_p} = \sqrt{s^2 \left( \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)}$$

- $r^2, r_{XY}$ , and  $r_{Y\hat{Y}}$

$$r^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{b_1 s_{XY}}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{SSR}{SST}$$

$$r_{XY} = (\text{sign of } b_1) \sqrt{r^2}$$

$$r_{Y\hat{Y}} = \sqrt{r^2}$$

**Example 1:**

Suppose the model is  $y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \dots, 10, \epsilon_i \sim N(0, \sigma^2)$ ,

$$\sum_{i=1}^{10} x_i = 320, \sum_{i=1}^{10} y_i = 462, \sum_{i=1}^{10} x_i^2 = 11550,$$

$$\sum_{i=1}^{10} y_i^2 = 21710, \sum_{i=1}^{10} x_i y_i = 15390.$$

- Find the least square estimate and the fitted regression equation
- Provide an ANOVA table and use the F statistic to test  $H_0: \beta_1 = 0$  at  $\alpha = 0.05$ .
- Find the 95% confidence interval for  $\beta_1$  and use the confidence interval to test  $H_0: \beta_1 = 0$ .
- Find the ANOVA table for the hypothesis  $H_0: \beta_0 = \beta_1 = 0$  and use the F statistic to test  $H_0: \beta_0 = \beta_1 = 0$  at  $\alpha = 0.01$ .
- Determine  $r^2, r_{XY}$ , and  $r_{Y\hat{Y}}$ .
- Find the 95% confidence interval for  $\beta_0 + 35\beta_1$ .

[Solution:]

(a)

$$s_{XX} = \sum_{i=1}^{10} x_i^2 - 10\bar{x}^2 = 11550 - 10 \cdot \left(\frac{320}{10}\right)^2 = 1310$$

and

$$s_{XY} = \sum_{i=1}^{10} x_i y_i - 10\bar{x}\bar{y} = 15390 - 10 \cdot \left(\frac{320}{10}\right) \cdot \left(\frac{462}{10}\right) = 606.$$

Then, the least squares estimate is

$$b_1 = \frac{s_{XY}}{s_{XX}} = \frac{606}{1310} = 0.4626,$$

$$b_0 = \bar{y} - b_1 \bar{x} = \left(\frac{462}{10}\right) - 0.4626 \cdot \left(\frac{320}{10}\right) = 31.3968.$$

The fitted regression equation is

$$\hat{y} = 31.3968 + 0.4626x.$$

(b)

$$SSR = b_1 s_{XY} = 0.4626 \cdot 606 = 280.3356.$$

$$SST = \sum_{i=1}^{10} y_i^2 - 10\bar{y}^2 = 21710 - 10 \cdot \left(\frac{462}{10}\right)^2 = 365.6$$

and

$$SSE = SST - SSR = 365.6 - 280.3356 = 85.2644.$$

The ANOVA table is

Source	DF	SS	MS	F
Regression	1	$SSR=280.3356$	$MSR = \frac{SSR}{1}$ $= 280.3356$	$f = \frac{MSR}{MSE}$ $= \frac{280.3356}{10.6581}$ $= 26.3026$
Residual (Error)	$n - 2$ $= 8$	$SSE=85.2644$	$MSE = \frac{SSE}{8}$ $= 10.6581$	
Total	9	365.6		

Since

$$f = 26.3026 > 5.32 = f_{1,8,0.05} = f_{1,n-2,\alpha},$$

we reject  $H_0$ .

(c) The 95% confidence interval for  $\beta_1$  is

$$b_1 \pm \left[ t_{n-2, \alpha/2} \cdot \left( \frac{s^2}{s_{xx}} \right)^{1/2} \right] = 0.4626 \pm \left[ t_{8, 0.025} \cdot \left( \frac{10.6581}{1310} \right)^{1/2} \right] = [0.25, 0.67].$$

Since

$$0 \notin [0.25, 0.67],$$

we reject  $H_0$ .

(d) Since

$$\sum_{i=1}^{10} \hat{y}_i^2 = \sum_{i=1}^{10} y_i^2 - SSE = 21710 - 85.2644 = 21624.74,$$

the ANOVA table for  $H_0: \beta_0 = \beta_1 = 0$  is

Source	DF	SS	MS	F
Regression	2	$\sum_{i=1}^{10} \hat{y}_i^2$ $= 21624.74$	$\frac{21624.74}{2}$ $= 10812.37$	$f$ $= \frac{10812.37}{10.6581}$ $= 1014.474$
Residual (Error)	$n - 2$ $= 8$	$SSE = 85.2644$	$MSE = \frac{SSE}{8}$ $= 10.6581$	
Total	10	21710		

Since

$$f = 1014.474 > 8.65 = f_{2, 8, 0.05} = f_{2, n-2, \alpha},$$

we reject  $H_0$ .

(e)

$$r^2 = \frac{SSR}{SST} = \frac{280.3356}{365.6} = 0.77.$$

$$r_{XY} = (\text{sign of } b_1) \sqrt{r^2} = 0.88$$

and

$$r_{YY} = \sqrt{r^2} = 0.88.$$

(f)

Since  $x_p = 35$ , the 95% confidence interval for  $\beta_0 + 35\beta_1$  is

$$\begin{aligned}\hat{y}_p &\pm \left( t_{n-2, \alpha/2} \sqrt{s^2 \left( \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)} \right) \\ &= (31.3968 + 0.4626 \cdot 35) \pm \left( t_{8, 0.025} \cdot \sqrt{10.6581 \cdot \left( \frac{1}{10} + \frac{(35 - 32)^2}{1310} \right)} \right) \\ &= [45.1267, 50.0489].\end{aligned}$$

**Example 2:**

Suppose we have the following data for the model,

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \dots, 7, \epsilon_i \sim N(0, \sigma^2),$$

$y_i$	5	10	10	25	10	30	22
$x_i$	-1	1	1	0	-1	-1	1

- (a) Find the fitted regression equation and the residual of the second observation.
- (b) Provide an ANOVA table for  $H_0: \beta_1 = 0$  and use F statistic to test the hypothesis at  $\alpha = 0.05$ .
- (c) Provide an ANOVA table for  $H_0: \beta_0 = \beta_1 = 0$  and use F statistic to test the hypothesis at  $\alpha = 0.05$ .
- (d) Use the confidence interval method to test  $H_0: \beta_1 = 1$  at  $\alpha = 0.05$ .
- (e) Use the t-statistic to test  $H_0: \beta_0 \geq 15$  at  $\alpha = 0.1$ .
- (f) Determine  $r^2, r_{XY}$ , and  $r_{Y\hat{Y}}$ .
- (g) Find the 90% confidence interval for  $E(y_p)$  at  $x_p = 1.5$ .

[solution:]

(a)

$$\sum_{i=1}^7 x_i = 0, \sum_{i=1}^7 y_i = 112, \sum_{i=1}^7 x_i^2 = 6, \sum_{i=1}^7 y_i^2 = 2334, \sum_{i=1}^7 x_i y_i = -3.$$

$$s_{XX} = \sum_{i=1}^7 x_i^2 - 7\bar{x}^2 = 6 - 0 = 6$$

and

$$s_{XY} = \sum_{i=1}^7 x_i y_i - 7\bar{x}\bar{y} = -3 - 0 = -3.$$

Then, the least squares estimate is

$$b_1 = \frac{s_{XY}}{s_{XX}} = \frac{-3}{6} = -0.5,$$

$$b_0 = \bar{y} - b_1 \bar{x} = 16 - (-0.5) \cdot 0 = 16.$$

The fitted regression equation is

$$\hat{y} = 16 - 0.5x$$

and

$$e_2 = 10 - 16 + 0.5 \cdot 1 = -5.5.$$

(b)

$$SSR = b_1 s_{XY} = (-0.5) \cdot (-3) = 1.5.$$

$$SST = \sum_{i=1}^7 y_i^2 - 7\bar{y}^2 = 2334 - 7 \cdot 16^2 = 542$$

and

$$SSE = SST - SSR = 542 - 1.5 = 540.5.$$

The ANOVA table is

Source	DF	SS	MS	F
Regression	1	$SSR=1.5$	$MSR = \frac{SSR}{1} = 1.5$	$f = \frac{MSR}{MSE}$ $= \frac{1.5}{108.1} = 0.014$
Residual (Error)	$n - 2$ $= 5$	$SSE=540.5$	$MSE = \frac{SSE}{5}$ $= 108.1$	
Total	6	542		

Since

$$f = 0.014 < 6.61 = f_{1,5,0.05} = f_{1,n-2,\alpha},$$

we do **not** reject  $H_0$ .

(c)

Since

$$\sum_{i=1}^{10} \hat{y}_i^2 = n\bar{y}^2 + b_1 s_{XY} = 7 \cdot 16^2 + (-0.5) \cdot (-3) = 1793.5,$$

the ANOVA table for  $H_0: \beta_0 = \beta_1 = 0$  is

Source	DF	SS	MS	F
Regression	2	$\sum_{i=1}^7 \hat{y}_i^2 = 1793.5$	$\frac{1793.5}{2} = 896.75$	$f = \frac{896.75}{108.1} = 8.3$
Residual (Error)	$n - 2 = 5$	$SSE = 540.5$	$MSE = \frac{SSE}{5} = 108.1$	
Total	7	2334		

Since

$$f = 8.3 > 5.79 = f_{2,5,0.05} = f_{2,n-2,\alpha},$$

we reject  $H_0$ .

(d) The 95% confidence interval for  $\beta_1$  is

$$b_1 \pm \left[ t_{n-2,\alpha/2} \cdot \left( \frac{s^2}{s_{XX}} \right)^{1/2} \right] = -0.5 \pm \left[ t_{5,0.025} \cdot \left( \frac{108.1}{6} \right)^{1/2} \right] = [-11.41, 10.41].$$

Since

$$1 \in [-11.41, 10.41],$$

we do not reject  $H_0$ .

(e) The  $t$  statistic is

$$t = \frac{b_0 - c}{\left( \frac{s^2 \sum_{i=1}^n x_i^2}{ns_{XX}} \right)^{1/2}} = \frac{16 - 15}{\left( \frac{108.1 \cdot 6}{7 \cdot 6} \right)^{1/2}} = 0.25.$$

Since

$$t = 0.25 > -1.476 = -t_{5,0.1} = -t_{n-2,\alpha}$$

we do not reject  $H_0$ .

(f)

$$r^2 = \frac{SSR}{SST} = \frac{1.5}{542} = 0.0028.$$

$$r_{XY} = (\text{sign of } b_1) \sqrt{r^2} = -0.053$$

and

$$r_{Y\hat{Y}} = \sqrt{r^2} = 0.053.$$

(g) Since  $x_p = 1.5$ , the 90% confidence interval for  $E(y_p)$  is

$$\begin{aligned}
& \hat{y}_p \pm \left( t_{n-2,\alpha/2} \sqrt{s^2 \left( \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)} \right) \\
& = (b_0 + b_1 x_p) \pm \left( t_{5,0.05} \cdot \sqrt{108.1 \cdot \left( \frac{1}{7} + \frac{(1.5 - 0)^2}{6} \right)} \right) \\
& = (16 - 0.5 \cdot 1.5) \pm (2.015 \cdot 7.48) \\
& = [0.17, 30.33].
\end{aligned}$$