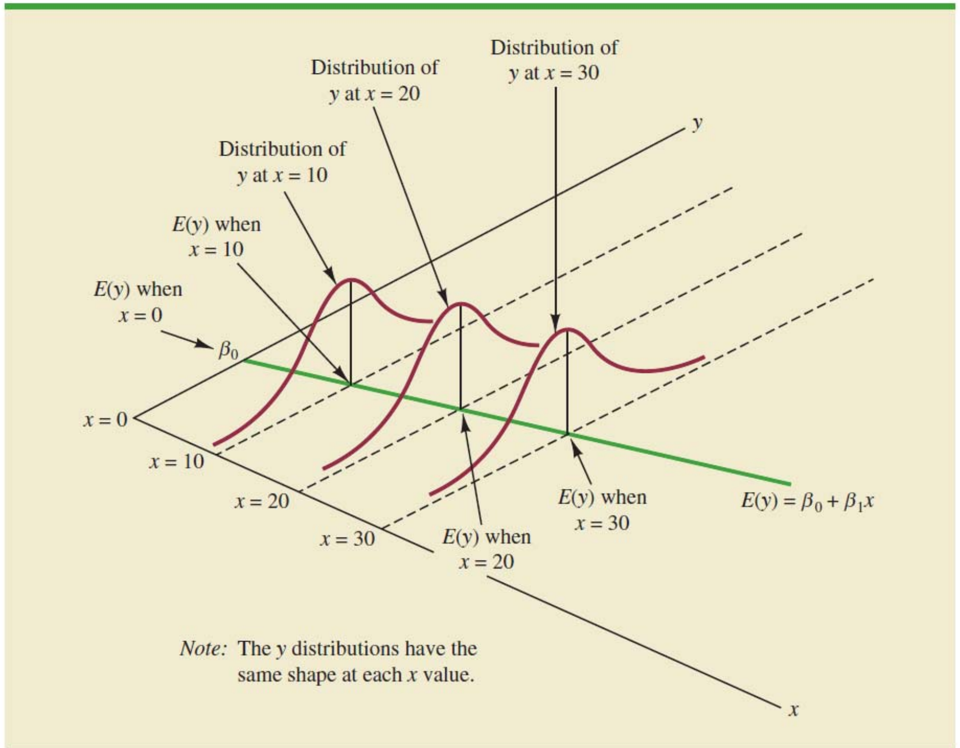
14.4 Model Assumptions

FIGURE 14.6 ASSUMPTIONS FOR THE REGRESSION MODEL



- Even with a large value of r^2, the estimated egression should not be useful until further analysis of the appropriateness of the assumed model has been conducted. →14.5

14.5 Testing for Significance (p.676)

- Estimate of σ^2 ➔ the variance of ε (殘差), also represents the variance of the y values about the regression line

$$SSE = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - b_0 - b_1 x_i)^2$$

- MSE provides the estimate of σ^2

$$s^2 = \mathrm{MSE} = \frac{SSE}{n-2}$$

- Standard error of the estimate

$$s = \sqrt{\mathrm{MSE}} = \sqrt{\frac{SSE}{n-2}}$$

To test for a significant regression relationship, we must conduct a hypothesis test to determine whether the value of β_1 is zero.

- Two tests commonly used are the t test and F test.

- Both the t test and F test require an estimate of σ^2, the variance of ε in the regression model.

- t test

Hypotheses:

$$H_0\colon \beta_1 = 0$$
$$H_a\colon \beta_1 \neq 0$$

Test Statistic:

$$t = \frac{b_1}{s_{b_1}} \quad \text{where} \quad s_{b_1} = \frac{s}{\sqrt{\sum(x_i - \bar{x})^2}}$$

Rejection Rule:

Reject H_0 if p –value ≤ α

Reject H_0 if $t \leq -t_{\alpha/2}$ or if $t \geq t_{\alpha/2}$

where:

$t_{\alpha/2}$ is based on a t distribution with n – 2 degrees of freedom

For Armand's Pizza Parlors

$$t = \frac{b_1}{s_{b_1}} \text{ where } s_{b_1} = \frac{s}{\sqrt{\sum(x_i - \bar{x})^2}}$$

$$s = \sqrt{\text{MSE}} = \sqrt{\frac{SSE}{n-2}}$$

TABLE 14.3 CALCULATION OF SSE FOR ARMAND'S PIZZA PARLORS

Restaurant i	x_i = Student Population (1000s)	y_i = Quarterly Sales ($1000s)	Predicted Sales $\hat{y}_i = 60 + 5x_i$	Error $y_i - \hat{y}_i$	Squared Error $(y_i - \hat{y}_i)^2$
1	2	58	70	−12	144
2	6	105	90	15	225
3	8	88	100	−12	144
4	8	118	100	18	324
5	12	117	120	−3	9
6	16	137	140	−3	9
7	20	157	160	−3	9
8	20	169	160	9	81
9	22	149	170	−21	441
10	26	202	190	12	144
					SSE = 1530

TABLE 14.2 CALCULATIONS FOR THE LEAST SQUARES ESTIMATED REGRESSION EQUATION FOR ARMAND'S PIZZA PARLORS

Restaurant i	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$	$(x_i - \bar{x})^2$
1	2	58	−12	−72	864	144
2	6	105	−8	−25	200	64
3	8	88	−6	−42	252	36
4	8	118	−6	−12	72	36
5	12	117	−2	−13	26	4
6	16	137	2	7	14	4
7	20	157	6	27	162	36
8	20	169	6	39	234	36
9	22	149	8	19	152	64
10	26	202	12	72	864	144
Totals	140	1300			2840	568
	Σx_i	Σy_i			$\Sigma(x_i - \bar{x})(y_i - \bar{y})$	$\Sigma(x_i - \bar{x})^2$

$$s_{b_1} = \frac{13.829}{\sqrt{568}} = .5803$$

At the α = .01 level of significance, the test statistic is

$$t = \frac{b_1}{s_{b_1}} = \frac{5}{.5803} = 8.62$$

With 8 degrees of freedom, p-value = .000 < α = .01. We reject H_0 and conclude that β_1 is not equal to zero.

- F test (p.679)

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0$$

Test statistics: $F = \frac{\text{MSR}}{\text{MSE}}$

思考的邏輯：

1. 如果虛無假設無法被推翻，表示斜率是 0 ，估計迴歸方程式算出來的 predicted yi 應該就在截距值附近，與 yi 及 y 平均值也都相當接近，因此 MSR/MSE 的值應該接近於 1。
2. 反之，若推翻虛無假設，對立假設成立，斜率不是 0 ，估計迴歸方程式算出來的 predicted yi 應該可以造成較大的 SSR 與較小的 SSE，因此 MSR/MSE 的值就會大於 1

p.680

- Hypotheses: $H_0\!: \beta_1 = 0$
  $H_a\!: \beta_1 \neq 0$

- Test Statistic: $F = \dfrac{MSR}{MSE}$

- Rejection Rule:
  Reject H_0 if the *p*-value ≤ 0.05 or if $F \geq F_\alpha$
  where F_α is based on an F distribution with 1 degree of freedom in the numerator and *n* – 2 degrees of freedom in the denominator.

For Armand's Pizza Parlors

$$F = \frac{MSR}{MSE} = \frac{SSR/1}{\frac{SSE}{n-2}} = \frac{14200}{1530/8} = 74.25$$

檢定統計量 F 值 74.25 之 p-value 為

小於 alpha 0.01，所以拒絕虛無假設，對立假設成立，母體斜率(β1)不等於 0，我們所估計出來的迴歸方程式是一條有效的迴歸方程式

或者利用 critical value 檢定，我們算出的檢定統計量 74.25 大於臨界值 $F_{0.01(1,8)} = 11.26$ (查表得知)，所以拒絕虛無假設，對立假設成立。

課堂練習：Exercise 24 (延續 Exercise 2、16)

24. The data from exercise 2 follow.

x_i	3	12	6	20	14
y_i	55	40	55	10	15

a. Compute the mean square error using equation (14.15).
b. Compute the standard error of the estimate using equation (14.16).
c. Compute the estimated standard deviation of b_1 using equation (14.18).
d. Use the t test to test the following hypotheses ($\alpha = .05$):

$$H_0: \beta_1 = 0$$
$$H_a: \beta_1 \neq 0$$

e. Use the F test to test the hypotheses in part (d) at a .05 level of significance. Present the results in the analysis of variance table format.

	Xi	Yi	Predicted Yi	Deviation (Yi - Predicted Yi)	Sqared (Yi-Predicted Yi)	Yi-Ymean	Squared (Yi-Y mean)	Predicted Yi-Ymean	Squared (Predicted Yi-Ymean)	Squared (Xi-X mean)
	3	55	59	-4	16	20	400	24	576	64
	12	40	32	8	64	5	25	-3	9	1
	6	55	50	5	25	20	400	15	225	25
	20	10	8	2	4	-25	625	-27	729	81
	14	15	26	-11	121	-20	400	-9	81	9
Mean	11	35			230		1850		1620	180
					SSE		SST		SSR	

a. s^2 = MSE = SSE/(n – 2) = 230/3 = 76.6667

b. $s = \sqrt{\text{MSE}} = \sqrt{76.6667} = 8.7560$

c. $\Sigma(x_i - \bar{x})^2 = 180$

$$s_{b_1} = \frac{s}{\sqrt{\Sigma(x_i - \bar{x})^2}} = \frac{8.7560}{\sqrt{180}} = 0.6526$$

d. $t = \frac{b_1}{s_{b_1}} = \frac{-3}{.653} = -4.59$

Using t table (3 degrees of freedom), area in tail is less than .01; p-value is less than .02.

Using Excel, the p-value corresponding to t = –4.59 is .0193.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$.

e. MSR = SSR/1 = 1,620

F = MSR/MSE = 1,620/76.6667 = 21.13

Using F table (1 degree of freedom numerator and 3 denominator), p-value is less than .025.

Using Excel, the p-value corresponding to F = 21.13 is .0193.

Because p-value $\leq \alpha$, we reject H_0: $\beta_1 = 0$.

Chapter 14 Homework (14.1~14.5)

題組一：E3、E17、E25

題組二：E18、E26

請同學先行練習，另外提供檢討影片

6/8	6/15
1. Chapter 13 小考(9:00-10:30) 2. 觀看 ch14.1~14.2、Ch14.3 兩段影片 3. 回家作業: 課本練習題 3、17	1. 8:50-9:10 登入 E-course，點名 2. 觀看課堂練習講解影片 3、17 3. 觀看 Ch14.4~14.5 影片 4. 課堂練習: 課本練習題 25、18、26 + 觀看講解影片 5. 教師公告 6/22 期末考(webassign 平台、選擇題題組)