Sec.10-1 (p.553)

For Exercises 11 through 27, perform the following steps

a. Draw the scatter plot for the variables.

b. Compute the value of the correlation coefficient.

c. State the hypotheses.

d. Test the significance of the correlation coefficient at α = 0.05. Using Table I or the P-value method.

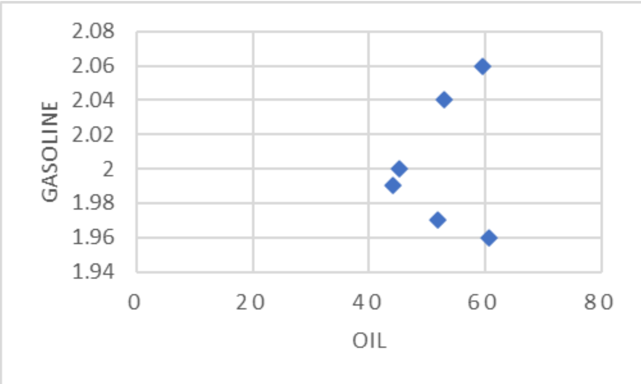
e. Give a brief explanation of the type of relationship.

Assume all assumptions have been met.

12. The average gasoline price per gallon and the cost of a barrel of oil are shown for a random selection of weeks in 2015. Is there a linear relationship between the variables?

Oil ($)	51.91	60.65	59.56	52.86	45.12	44.21
Gasoline ($)	1.97	1.96	2.06	2.04	2.00	1.99

a. scatter plot



b. correlation coefficient

x	y	xy	x^2	y^2
51.91	1.97	102.263	2694.648	3.881
60.65	1.96	118.874	3678.423	3.842
59.56	2.06	122.694	3547.394	4.244
52.86	2.04	107.834	2794.180	4.162
45.12	2.00	90.240	2035.814	4.000
44.21	1.99	87.978	1954.524	3.960
$\sum x = 314.31$	$\sum y = 12.02$	$\sum xy = 629.883$	$\sum x^2 = 16704.983$	$\sum y^2 = 24.089$

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$
$$= \frac{6 \cdot 629.883 - 314.31 \cdot 12.02}{\sqrt{(6 \cdot 16704.983 - 314.31^2)(6 \cdot 24.089 - 12.02^2)}}$$
$$= 0.147$$

c. hypothesis

$$H_0: \rho = 0 \ \ v.s. H_1: \rho \neq 0$$

d. test with $\alpha = 0.05$

$$d.f. = n - 2 = 4 \Rightarrow R = \{|r| \geq 0.811\}$$

$0.147 \ngeq 0.811 \Rightarrow$ 不拒絕H_0

There is not enough evidence to conclude that there is a significant linear relationship between the average gasoline price per gallon and the cost of a barrel of oil at $\alpha = 0.05$.
我們沒有足夠的證據表明在 $\alpha = 0.05$ 時，每加侖汽油價格與每桶石油的成本之間存在顯著的線性關係。

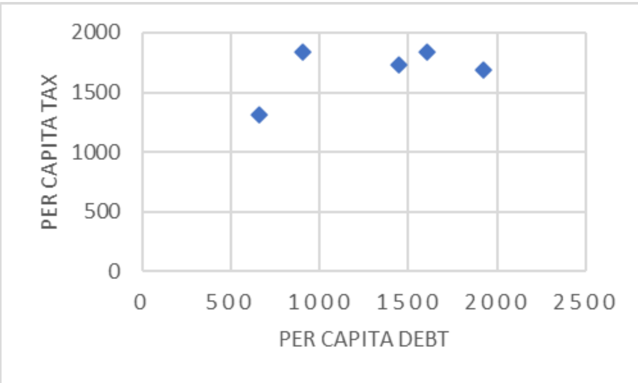
e. explaination of the type of relationship
The linear correlation coefficient suggests a weak positive linear relationship between the average gasoline price per gallon and the cost of a barrel of oil.
線性相關係數表明每加侖平均汽油價格與每桶石油成本之間呈現弱的正相關。

16. An economics student wishes to see if there is a relationship between the amount of state debt per capita and the amount of tax per capita at the state level. Based on the following data, can she or he conclude that per capita state debt and per capita state taxes are related? Both amounts are in dollars and represent five randomly selected states.

Per capita debt x	1924	907	1445	1608	661
Per capita tax y	1685	1838	1734	1842	1317

amount of state debt per capita 人均州債務數額;
amount of tax per capita at the state level 州級人均稅收數額;

a. scatter plot



b. correlation coefficient

x	y	xy	x^2	y^2
1924	1685	3241940	3701776	2839225
907	1838	1667066	822649	3378244
1445	1734	2505630	2088025	3006756
1608	1842	2961936	2585664	3392964
661	1317	870537	436921	1734489
$\sum x = 6545$	$\sum y = 8416$	$\sum xy = 11247109$	$\sum x^2 = 9635035$	$\sum y^2 = 14351678$

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} = 0.518$$

c. hypothesis

$$H_0: \rho = 0 \ \ v.s. H_1: \rho \neq 0$$

d. test with $\alpha = 0.05$

$$d.f. = n - 2 = 3 \Rightarrow R = \{|r| \geq 0.878\}$$

$0.518 \ngeq 0.878 \Rightarrow$ 不拒絕H_0

There is not enough evidence to conclude that there is a significant linear relationship between the amount of state debt per capita and the amount of tax per capita at the state level at α = 0.05.

我們沒有足夠的證據表明在 α = 0.05 時，人均州債務數額與州級人均稅收數額之間存在顯著的線性關係。

e. explaination of the type of relationship

The linear correlation coefficient suggests a positive linear relationship between the amount of state debt per capita and the amount of tax per capita at the state level.

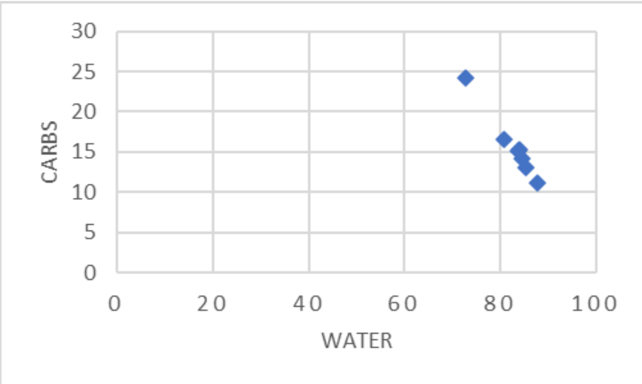
線性相關係數表明人均州債務數額與州級人均稅收額之間存在正相關。

20. Here are the number of grams of water and the number of grams of carbohydrates for a random selection of raw foods. Is there a linear relationship between the variables?

Water	83.93	80.76	87.66	85.20	72.85	84.61	83.81
Carbs	15.25	16.55	11.10	13.01	24.27	14.13	15.11

carbohydrates 碳水化合物; raw foods 生食;

a. scatter plot



b. correlation coefficient

x	y	xy	x^2	y^2
83.93	15.25	1279.933	7044.245	232.563
80.76	16.55	1336.578	6522.178	273.903
87.66	11.1	973.026	7684.276	123.210
85.2	13.01	1108.452	7259.040	169.260
72.85	24.27	1768.070	5307.123	589.033
84.61	14.13	1195.539	7158.852	199.657
83.81	15.11	1266.369	7024.116	228.312
$\sum x = 578.82$	$\sum y = 109.42$	$\sum xy = 8927.967$	$\sum x^2 = 47999.83$	$\sum y^2 = 1815.938$

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} = -0.993$$

c. hypothesis

$$H_0: \rho = 0 \;\; v.s. H_1: \rho \neq 0$$

d. test with $\alpha = 0.05$

$$d.f. = n - 2 = 5 \Rightarrow R = \{|r| \geq 0.754\}$$

$|-0.993| > 0.754 \Rightarrow$ 拒絕H_0

There is enough evidence to conclude that there is a significant linear relationship between the number of grams of water and the number of grams of carbohydrates at α = 0.05.

我們有足夠的證據表明在 α = 0.05 時，水的克數與碳水化合物的克數之間存在顯著的線性關係。

e. explaination of the type of relationship

The linear correlation coefficient suggests a strong negative linear relationship between the number of grams of water and the number of grams of carbohydrates.

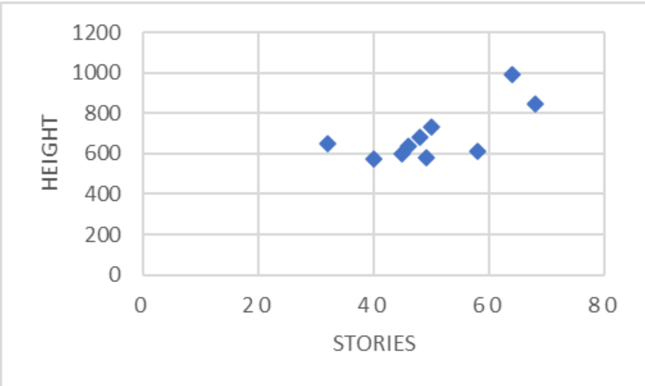
線性相關係數表明水的克數和碳水化合物的克數之間存在很強的負相關。

26. An architect wants to determine the relationship between the heights (in feet) of buildings and the number of stories in the buildings. The data for a sample of 10 buildings in Chicago are shown. Explain the relationship (if any).

Stories x	64	68	50	48	32	46	58	45	49	40
Height y	995	844	732	679	648	635	610	600	583	573

architect 建築師; feet 呎; number of stories in the buildings 建築物樓層數;

a. scatter plot



b. correlation coefficient

x	y	xy	x^2	y^2
64	995	63680	4096	990025
68	844	57392	4624	712336
50	732	36600	2500	535824
48	679	32592	2304	461041
32	648	20736	1024	419904
46	635	29210	2116	403225
58	610	35380	3364	372100
45	600	27000	2025	360000
49	583	28567	2401	339889
40	573	22920	1600	328329
$\sum x = 500$	$\sum y = 6899$	$\sum xy = 354077$	$\sum x^2 = 26054$	$\sum y^2 = 4922673$

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} = 0.696$$

c. hypothesis

$$H_0: \rho = 0 \;\; v.s. H_1: \rho \neq 0$$

d. test with $\alpha = 0.05$

$$d.f. = n - 2 = 8 \Rightarrow R = \{|r| \geq 0.632\}$$

$0.696 > 0.632 \Rightarrow$ 拒絕H_0

There is enough evidence to conclude that there is a significant linear relationship between the heights of buildings and the number of stories in the buildings at α = 0.05.

我們有足夠的證據表明在 α = 0.05 時，建築物的高度與建築物的樓層數之間存在顯著的線性關係。

e. explaination of the type of relationship

The linear correlation coefficient suggests a positive linear relationship between the heights of buildings and the number of stories in the buildings.

線性相關係數表明建築物高度與樓層數之間呈現正相關。

Sec. 10-2 (p.562)

For Exercises 11 through 27, use the same data as for the corresponding exercises in Section 10-1.

For each exercise, find the equation of the regression line and find the y' value for the specified x value.

Remember that no regression should be done when r is not significant.

14. Number of fires and number of acres burned are as follows:

Fires x	72	69	58	47	84	62	57	45
Acres y	62	42	19	26	51	15	30	15

Find y' when $x = 60$ fires.

number of acres burned 燒毀面積;

a. correlation coefficient

x	y	xy	x^2	y^2
72	62	4464	5184	3844
69	42	2898	4761	1764
58	19	1102	3364	361
47	26	1222	2209	676
84	51	4284	7056	2601
62	15	930	3844	225
57	30	1710	3249	900
45	15	675	2025	225
$\sum x = 494$	$\sum y = 260$	$\sum xy = 17285$	$\sum x^2 = 31692$	$\sum y^2 = 10596$

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} = 0.771$$

b. hypothesis

$$H_0: \rho = 0 \;\; v.s. H_1: \rho \neq 0$$

c. test with $\alpha = 0.05$

$$d.f. = n - 2 = 6 \Rightarrow R = \{|r| \geq 0.707\}$$

$0.771 > 0.707 \Rightarrow$ 拒絕H_0

There is enough evidence to conclude that there is a significant linear relationship between the number of fires and the number of acres burned at $\alpha = 0.05$.

我們有足夠的證據表明在 $\alpha = 0.05$ 時，火災次數與燒燬面積之間存在顯著的線性關係。

d. r 顯著時計算迴歸方程式並代入

$$a = \frac{\sum y \sum x^2 - \sum x \sum xy}{n\sum x^2 - (\sum x)^2} = -31.46$$

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = 1.036$$

Hence, the equation of the regression line $y' = a + bx$ is

$$y' = -31.46 + 1.036 \cdot x$$

When $x = 60$, $y' = -31.46 + 1.036 \times 60 = 30.7$

20. Here are the number of grams of water and the number of grams of carbohydrates for a random selection of raw foods (100 g each).

Water	83.93	80.76	87.66	85.20	72.85	84.61	83.81
Carbs	15.25	16.55	11.10	13.01	24.27	14.13	15.11

Find y' for $x = 75$.

a. correlation coefficient

x	y	xy	x^2	y^2
83.93	15.25	1279.933	7044.245	232.563
80.76	16.55	1336.578	6522.178	273.903
87.66	11.1	973.026	7684.276	123.210
85.2	13.01	1108.452	7259.040	169.260
72.85	24.27	1768.070	5307.123	589.033
84.61	14.13	1195.539	7158.852	199.657
83.81	15.11	1266.369	7024.116	228.312
$\sum x = 878.820$	$\sum y = 109.420$	$\sum xy = 8927.967$	$\sum x^2 = 47999.830$	$\sum y^2 = 1815.938$

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} = -0.993$$

b. hypothesis

$$H_0: \rho = 0 \;\; v.s. H_1: \rho \neq 0$$

c. test with $\alpha = 0.05$

$$d.f. = n - 2 = 5 \Rightarrow R = \{|r| \geq 0.754\}$$

$|-0.993| > 0.754 \Rightarrow$ 拒絕H_0

There is enough evidence to conclude that there is a significant linear relationship between the number of grams of water and the number of grams of carbohydrates at $\alpha = 0.05$.

我們有足夠的證據表明在 $\alpha = 0.05$ 時，水的克數與碳水化合物的克數之間存在顯著的線性關係。

d. r 顯著時計算迴歸方程式並代入

$$a = \frac{\sum y \sum x^2 - \sum x \sum xy}{n\sum x^2 - (\sum x)^2} = 87.408$$

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = -0.868$$

Hence, the equation of the regression line $y' = a + bx$ is

$$y' = 87.408 - 0.868 \cdot x$$

When $x = 75$, $y' = 87.408 - 0.868 \times 75 = 22.308$

For Exercises 28 through 33, do a complete regression analysis by performing these steps.

a. Draw a scatter plot.

b. Compute the correlation coefficient.

c. State the hypotheses.

d. Test the hypotheses at a = 0.05. Use Table I, or use the P-value method.

e. Determine the regression line equation if r is significant.

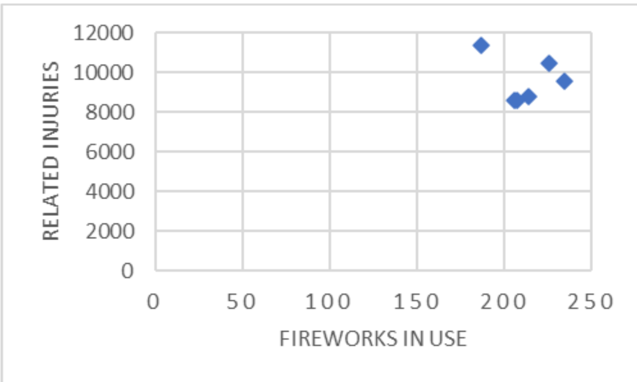
f. Plot the regression line on the scatter plot, if appropriate.

g. Summarize the results.

28. These data were obtained for the years 2009 through 2014 and indicate the number of fireworks (in millions) used and the related injuries. Predict the number of injuries if 200 million fireworks are used during a given year.

Fireworks in use x	213.9	205.9	234.1	207.5	186.4	225.3
Related injuries y	8800	8600	9600	8600	11400	10500

a. scatter plot



b. correlation coefficient

	x	y	xy	x^2	y^2
	213.9	8800	1882320	45753.21	77440000
	205.9	8600	1770740	42394.81	73960000
	234.1	9600	2247360	54802.81	92160000
	207.5	8600	1784500	43056.25	73960000
	186.4	11400	2124960	34744.96	129960000
	225.3	10500	2365650	50760.09	110250000
$\sum =$	1273.1	57500	12175530	271512.13	557730000

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} = -0.260$$

c. hypothesis

$$H_0: \rho = 0 \ \ v.s. H_1: \rho \neq 0$$

d. test with $\alpha = 0.05$

$$d.f. = n - 2 = 4 \Rightarrow R = \{|r| \geq 0.811\}$$

$|-0.260| \ngeq 0.811 \Rightarrow$ 不拒絕H_0

There is not enough evidence to conclude that there is a significant linear relationship between the number of fireworks used and the related injuries at α = 0.05.

我們沒有足夠的證據表明在 α = 0.05 時，使用的煙火數量與相關傷害之間存在顯著的線性關係。

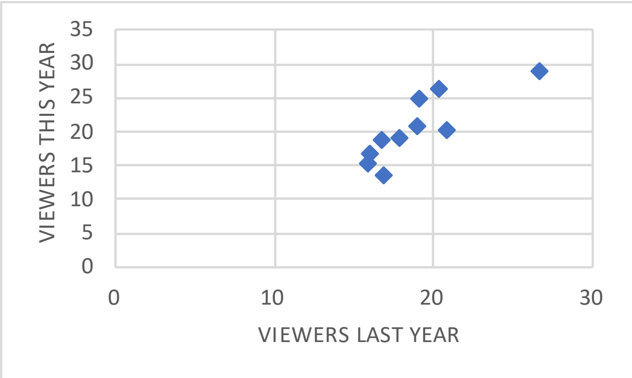
Hence, no regression should be done.

32. A television executive selects 10 television shows and compares the average number of viewers the show had last year with the average number of viewers this year. The data (in millions) are shown. Describe the relationship.

Viewers last year x	26.6	17.85	20.3	16.8	20.8	16.7	19.1	18.9	16.0	15.8
Viewers this year y	28.9	19.2	26.4	13.7	20.2	18.8	25.0	21.0	16.8	15.3

television executive 電視主管;

a. scatter plot



b. correlation coefficient

	x	y	xy	x^2	y^2
	26.6	28.9	768.74	707.56	835.21
	17.85	19.2	342.72	318.623	368.64
	20.3	26.4	535.92	412.09	696.96
	16.8	13.7	230.16	282.24	187.69
	20.8	20.2	420.16	432.64	408.04
	16.7	18.8	313.96	278.89	353.44
	19.1	25	477.5	364.81	625
	18.9	21	396.9	357.21	441
	16	16.8	268.8	256	282.24
	15.8	15.3	241.74	249.64	234.09
$\sum =$	188.85	205.3	3996.6	3659.703	4432.31

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} = 0.839$$

c. hypothesis

$$H_0: \rho = 0 \;\; v.s. H_1: \rho \neq 0$$

d. test with $\alpha = 0.05$

$$d.f. = n - 2 = 8 \Rightarrow R = \{|r| \geq 0.632\}$$

$0.839 > 0.632 \Rightarrow$ 拒絕H_0

There is enough evidence to conclude that there is a significant linear relationship between the average number of viewers the show had last year and the average number of viewers this year at α = 0.05.

我們有足夠的證據表明在 α = 0.05 時，去年的平均觀眾人數與今年的平均觀眾人數之間存在顯著的線性關係。

e. r 顯著時計算迴歸方程式

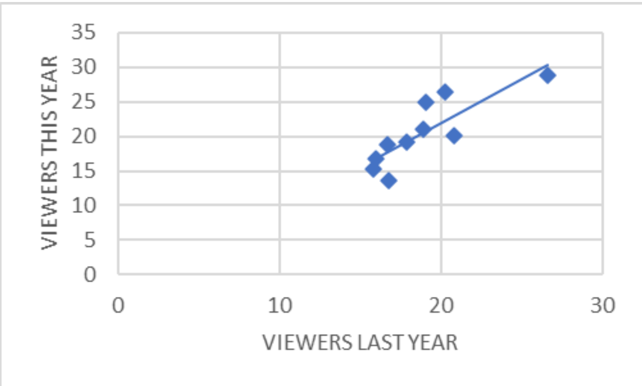
$$a = \frac{\sum y \sum x^2 - \sum x \sum xy}{n\sum x^2 - (\sum x)^2} = -3.668$$

$$b = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = 1.281$$

Hence, the equation of the regression line $y' = a + bx$ is

$$y' = -3.668 + 1.281 \cdot x$$

f. 在散佈圖上畫出迴歸線



g. summarize

The linear correlation coefficient suggests a positive linear relationship between the average number of viewers the show had last year and the average number of viewers this year.

線性相關係數表明該節目去年的平均觀眾人數與今年的平均觀眾人數之間存在正相關。

Sec. 10-3 (p.581)

For Exercises 8 through 13, find the coefficients of determination and nondetermination and explain the meaning of each.

8. $r = 0.62$

coefficients of determination

$$r^2 = 0.62^2 = 0.3844$$

38.44% of the variation of y is due to the variation of x.

38.44% 的 y 變異是由 x 的變異所引起的

coefficients of nondetermination

$$1 - r^2 = 0.6156$$

61.56% is due to chance.

61.56%的 y 變異是隨機的

12. $r = 0.12$

coefficients of determination

$$r^2 = 0.12^2 = 0.0144$$

1.44% of the variation of y is due to the variation of x.

1.44% 的 y 變異是由 x 的變異所引起的

coefficients of nondetermination

$$1 - r^2 = 0.9856$$

98.56% is due to chance.

98.56%的 y 變異是隨機的

16. Compute the standard error of the estimate for Exercise 14 in Section 10-1. The regression line equation was found in Exercise 14 in Section 10-2.

Fires x	72	69	58	47	84	62	57	45
Acres y	62	42	19	26	51	15	30	15

20. For the data in Exercises 14 in Sections 10-1 and 10-2 and 16 in Section 10-3, find the 95% prediction interval when x = 60 years.

x	y	xy	x^2	y^2
72	62	4464	5184	3844
69	42	2898	4761	1764
58	19	1102	3364	361
47	26	1222	2209	676
84	51	4284	7056	2601
62	15	930	3844	225
57	30	1710	3249	900
45	15	675	2025	225
$\sum x = 494$	$\sum y = 260$	$\sum xy = 17285$	$\sum x^2 = 31692$	$\sum y^2 = 10596$

$$y' = -31.46 + 1.036 \cdot x$$

$$s_{est} = \sqrt{\frac{\sum(y - y')^2}{n-2}} = 12.055$$

$$s_{est} = \sqrt{\frac{\sum y^2 - a\sum y - b\sum xy}{n-2}} = 12.03$$

$$95\% CI \Rightarrow t_{\left(\frac{0.05}{2}, 8-2\right)} = t_{(0.025,6)} = 2.447$$

$$x = 60 \Rightarrow y' = 30.7$$

$$y' \pm t_{\frac{\alpha}{2}} \cdot s_{est} \cdot \sqrt{1 + \frac{1}{n} + \frac{n(x - \bar{X})^2}{n\sum x^2 - (\sum x)^2}}$$

$$\Rightarrow 30.7 \pm 2.447 \cdot 12.03 \cdot 1.062$$

$$\Rightarrow -0.563 < y < 61.963$$

Hence the 95% prediction interval when x = 60 is (−0.559, 31.259).

18. Compute the standard error of the estimate for Exercise 16 in Section 10-1. The regression line equation was found in Exercise 16 in Section 10-2.

Per capita debt x	1924	907	1445	1608	661
Per capita tax y	1685	1838	1734	1842	1317

Since r is not significant, the standard error should not be calculated. (10-1 #16)

22. For the data in Exercises 16 in Sections 10-1 and 10-2 and 18 in Section 10-3, find the 98% prediction interval when x = 47 years.

Since r is not significant, the prediction interval should not be calculated. (10-1 #16)