Solution Section 2.2 – Least-Squares Regression

Exercise

A physician measured the weights and cholesterol levels of a random sample of men. The regression equation is $\hat{y} = -116 + 2.44x$, where x represents weight (in pounds). What does the symbol $\hat{y}$ represent? What does the predictor variable represent? What does the response variable represent?

Solution

The symbol $\hat{y}$ represents the predicted cholesterol level. The predictor variable x represents weight. The response variable represents cholesterol level.

Exercise

In what sense is the regression line the straight line that "best" fits the points in a scatterplot?

Solution

The regression line is the best fit for the points of a scatterplot in the sense that it minimizes the sum of the squared differences between the observed y values and the y values predicted by the regression line.

Exercise

In a study, the total weight (in pounds) of garbage discarded in one week and the household size were recorded for 62 households. The linear correlation coefficient is $r = 0.759$ and the regression equation $\hat{y} = 0.445 + 0.119x$, where x represents the total weight of discarded garbage. The mean of the 62 garbage weights is 27.4 lb. and the 62 households have a mean size of 3.71 people. What is the best predicted number of people in a household that discards 50 lb. of garbage?

Solution

For n = 62, the critical value = ± 0.254.

Since r = 0.759 > 0.254, use the regression line for prediction.

$$\hat{y}\Big|_{x=50} = 0.445 + 0.119(50)\backslash$$

$$\underline{= 6.4 \text{ } people|}$$

Exercise

A sample of 8 mother/daughter pairs of subjects was obtained, and their heights (in inches) were measured. The linear correlation coefficient is 0.693 and the regression equation $\hat{y} = 69 - 0.0849x$, where x represents the height of the mother. The mean height of the mothers is 63.1 in. and the mean height of the daughters is 63.3 in. Find the best predicted height of a daughter given that the mother has a height of 60 in.

Solution

For $n = 8$, the critical value $= \pm 0.707$.

Since $r = 0.693 < 0.707$, use the regression line for prediction. $\hat{y} = \bar{y}$

$$\hat{y}\Big|_{x=60} = \bar{y} = \underline{= 63.3\ in}$$

Exercise

A sample of 40 women is obtained, and their heights (in inches) and pulse rates (in beats per minute) are measured. The linear correlation coefficient is 0.202 and the equation of the regression line is $\hat{y} = 18.2 + 0.920x$, where x represents height. The mean of the 40 heights is 63.2 in. and the mean of the 40 pulse rates is 76.3 beats per minute. Find the best predicted pulse rate of a woman who is 70 in. tall.

Solution

For $n = 40$, the critical value $= \pm 0.312$.

Since $r = 0.202 < 0.312$, use the regression line for prediction. $\hat{y} = \bar{y}$

$$\hat{y}\Big|_{x=70} = \bar{y}\ \underline{= 76.3\ \ beats / \min}$$

Exercise

Heights (in inches) and weights (in pounds) are obtained from a random sample of 9 supermodels. The linear correlation coefficient is 0.360 and the equation of the regression line is $\hat{y} = 31.8 + 1.23x$, where x represents height. The mean of the 9 heights is 69.3 in. and the mean of the 9 weights is 117 lb. Find the best predicted weight of a supermodel with a height of 72 in.?

Solution

For $n = 9$, the critical value $= \pm 0.666$.

Since $r = 0.360 < 0.666$, use the regression line for prediction. $\hat{y} = \bar{y}$

$$\hat{y}\Big|_{x=72}\ \underline{= 117\ \ lbs}$$

Exercise

Find the equation of the regression line for the given data below

x	10	8	13	9	11	14	6	4	12	7	5
y	9.14	8.14	8.74	8.77	9.26	8.10	6.13	3.10	9.13	7.26	4.74

Examine the scatterplot and identify a characteristic of the data that is ignored by the regression line

Solution

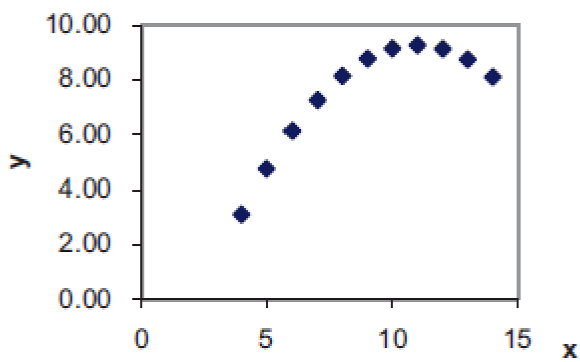
x	**y**	**xy**	x^2	y^2
10	9.14	91.40	100	83.5396
8	8.14	65.12	64	66.2596
13	8.74	113.62	169	76.3876
9	8.77	78.93	81	76.9129
11	9.26	101.86	121	85.7476
14	8.10	113.40	196	65.61
6	6.13	36.78	36	37.5769
4	3.10	12.40	16	9.61
12	9.13	109.56	144	83.3569
7	7.26	50.82	49	52.7076
5	4.74	23.70	25	22.4676
99	82.51	797.59	1001	660.1763

$$\bar{x} = \frac{\sum x}{n} = \frac{99}{11} = 9.0 \qquad \bar{y} = \frac{\sum y}{n} = \frac{82.52}{11} = 7.5$$

$$b_1 = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2}$$

$$= \frac{11(797.59) - (99)(82.51)}{11(1001) - (99)^2}$$

$$\underline{= 0.50}$$



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

$$= 7.50 - 0.5(9)$$

$$= 3$$

$$\hat{y} = b_0 + b_1 x$$

$$\underline{= 3.0 + 0.5x}$$

The scatterplot indicates that the relationship between the variables is quadratic, not linear.

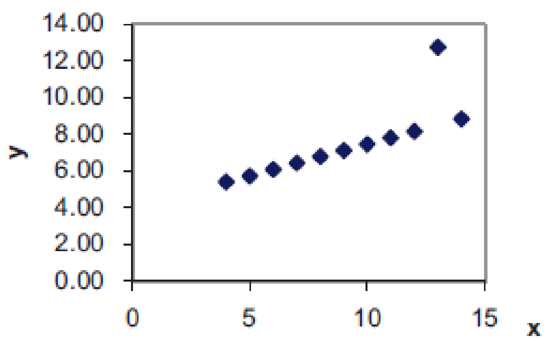
Exercise

Find the equation of the regression line for the given data below

x	10	8	13	9	11	14	6	4	12	7	5
y	7.46	6.77	12.74	7.11	7.81	8.84	6.08	5.39	8.15	6.42	5.73

Examine the scatterplot and identify a characteristic of the data that is ignored by the regression line

Solution



x	y	xy	x^2	y^2
10	7.46	74.60	100	55.6516
8	6.77	54.16	64	45.8329
13	112.74	165.62	169	162.3076
9	7.11	63.99	81	50.5521
11	7.81	85.91	121	60.9961
14	8.84	123.76	196	78.1456
6	6.08	36.48	36	36.9664
4	5.39	21.56	16	29.0521
12	8.15	97.80	144	66.4225
7	6.42	44.94	49	41.2164
5	5.73	28.65	25	32.8329
99	82.50	797.47	1001	659.9762

$$\bar{x} = \frac{\sum x}{n} = \frac{99}{11} = 9.0 \qquad \bar{y} = \frac{\sum y}{n} = \frac{82.52}{11} = 7.5$$

$$b_1 = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2}$$

$$= \frac{11(797.47) - (99)(82.50)}{11(1001) - (99)^2}$$

$$\underline{= 0.50}$$

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

$$= 7.50 - 0.5(9)$$

$$= 3$$

$$\hat{y} = b_0 + b_1 x$$

$$\underline{= 3.0 + 0.5x}$$

The scatterplot indicates that the relationship between the variables is essentially a perfect straight line except for one point, which is likely an error or an outlier.

Exercise

Find the equation of the regression line for the given data below

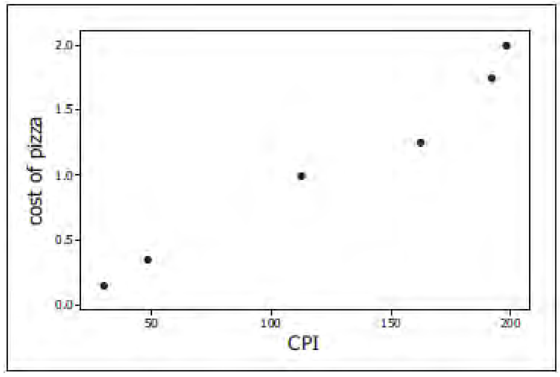
CPI	30.2	48.3	112.3	162.2	191.9	197.8
Cost of Pizza	0.15	0.35	1.00	1.25	1.75	2.00

Let the first variable be the predictor (x) variable. Find the best indicated predicted cost of a slice of pizza when the Consumer Price Index (CPI) is 182.5 (in the year 2000).

Solution

Excel produces the following

x	y	xy	x^2	y^2
30.2	0.15	4.53	912.04	0.0225
48.3	0.35	16.905	2332.89	0.1225
112.3	1.00	112.3	12611.29	1.00
162.2	1.25	202.75	26308.84	1.5625
191.9	1.75	335.825	36825.61	3.0625
197.8	2.00	395.60	39124.84	4.00
742.7	6.50	1067.91	118115.5	9.77



$$\bar{x} = \frac{\sum x}{n} = \frac{742.7}{6} = 123.78 \qquad \bar{y} = \frac{\sum y}{n} = \frac{6.50}{6} = 1.08$$

$$b_1 = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2}$$

$$= \frac{6(1067.91) - (742.7)(6.5)}{6(118115.5) - (742.7)^2}$$

$$= 0.01005$$

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

$$= 1.08 - 0.0101(123.78)$$

$$= -0.1616$$

$$\hat{y} = b_0 + b_1 x$$

$$= -0.162 + 0.0101x$$

$$\hat{y}_{182.5} = -0.162 + 0.0101(182.5)$$

$$= \$1.67$$

Exercise

Find the equation of the regression line for the given data below

CPI	30.2	48.3	112.3	162.2	191.9	197.8
Subway fare	0.15	0.35	1.00	1.35	1.5	2.00

Let the first variable be the predictor (x) variable. Find the best indicated predicted cost of a slice of pizza when the Consumer Price Index (CPI) is 182.5 (in the year 2000).

Solution

x	y	xy	x^2	y^2
30.2	0.15	4.53	912.04	0.0225
48.3	0.35	16.905	2332.89	0.1225
112.3	1.00	112.3	12611.29	1.00
162.2	1.35	218.97	26308.84	1.8225
191.9	1.50	287.85	36825.61	2.25
197.8	2.00	395.60	39124.84	4.00
742.7	6.35	1036.155	118115.51	9.2175

$$\bar{x} = \frac{\sum x}{n} = \frac{742.7}{6} = 123.78 \qquad \bar{y} = \frac{\sum y}{n} = \frac{6.35}{6} = 1.06$$

$$\begin{aligned} b_1 &= \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2} \\ &= \frac{6(1036.155) - (742.7)(6.35)}{6(118115.51) - (742.7)^2} \\ &= 0.00955 \end{aligned}$$

$$\begin{aligned} b_0 &= \bar{y} - b_1\bar{x} \\ &= 1.06 - 0.00955(123.78) \\ &= -0.124 \end{aligned}$$

$$\begin{aligned} \hat{y} &= b_0 + b_1 x \\ &= -0.124 + 0.00955x \end{aligned}$$

$$\begin{aligned} \hat{y}_{182.5} &= -0.124 + 0.00955(182.5) \\ &= \$1.62 \end{aligned}$$

Exercise

Listed below are systolic blood pressure measurements (in mm HG) obtained from the same woman.

Right Arm	102	101	94	79	79
Left Arm	175	169	182	146	144

Find the best predicted systolic blood pressure in the left arm given that the systolic blood pressure in the right arm is 100 mm Hg.

Solution

x	y	xy	x^2	y^2
102	175	17850	10404	30625
101	169	17069	10201	28561
94	182	17108	8836	33124
79	146	11534	6241	21316
79	144	11376	6241	20736
455	816	74937	41923	134362

$$\overline{x} = \frac{\sum x}{n} = \frac{455}{5} = 91.0 \qquad \overline{y} = \frac{\sum y}{n} = \frac{816}{5} = 163.2$$

$$b_1 = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2}$$

$$= \frac{5(74937) - (455)(816)}{5(41923) - (455)^2}$$

$$= 1.315$$

$$b_0 = \overline{y} - b_1\overline{x}$$

$$= 163.2 - 1.315(91)$$

$$= 43.56$$

$$\hat{y} = b_0 + b_1 x$$

$$= 43.6 + 1.31x$$

$$\hat{y}_{182.5} = \overline{y} = 163.2 \ \ mmHg$$

No significant correlation

Exercise

Find the best predicted height of runner-up Goldwater, given that the height of the winning presidential candidate is 75 in. Is the predicted height of Goldwater close to his actual height of 72 in.?

Winner	69.5	73	73	74	74.5	74.5	71	71
Runner-Up	72	69.5	70	68	74	74	73	76

Solution

x	***y***	***xy***	x^2	y^2
69.5	72	5004	4830.25	5184
73	69.5	5073.5	5329	4830.25
73	70	5110	5329	4900
74	68	5032	5476	4624
74.5	74	5513	5550.25	5476
74.5	74	5513	5550.25	5476
71	76	5183	5041	5329
71	76	5396	5041	5776
580.5	576.5	41824.5	42146.75	41595.25

$$\bar{x} = \frac{\sum x}{n} = \frac{580.5}{8} = 72.56 \qquad \bar{y} = \frac{\sum y}{n} = \frac{576.5}{8} = 72.06$$

$$b_1 = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2}$$

$$= \frac{8(41824.5) - (580.5)(576.5)}{8(42146.75) - (580.5)^2}$$

$$\underline{= -0.321}$$

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

$$= 72.06 - (-0.321)(72.56)$$

$$\underline{= 95.38}$$

$$\hat{y} = b_0 + b_1 x$$

$$\underline{= 95.4 - 0.321x}$$

$$\hat{y}_{182.5} = \bar{y} \ \underline{= 72.1 \ \ in.}$$

No significant correlation

Exercise

Find the best predicted amount of revenue (in millions of dollars), given that the amount has a size 87 thousand ft^2. How does the result compare to the actual revenue of \$65.1 million?

Size	160	227	140	144	161	147	141
Revenue	189	157	140	127	123	106	101

Solution

x	y	xy	x^2	y^2
160	189	30240	25600	35721
227	157	35639	51529	24649
140	140	19600	19600	19600
144	127	18288	20736	16129
161	123	19803	25921	15129
147	106	15582	21609	11236
141	101	14241	19881	10201
1120	943	153393	184876	132665

$$\bar{x} = \frac{\sum x}{n} = \frac{1120}{7} = 160.0 \qquad \bar{y} = \frac{\sum y}{n} = \frac{943}{7} = 134.71$$

$$b_1 = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2}$$

$$= \frac{7(153393) - (1120)(943)}{7(184876) - (1120)^2}$$

$$\underline{= 0.443}$$

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

$$= 134.71 - (0.443)(160)$$

$$\underline{= 63.87}$$

$$\hat{y} = b_0 + b_1 x$$

$$\underline{= 63.9 + 0.443x}$$

$$\hat{y}_{182.5} = \bar{y} \ \underline{= 134.7 \ \ million \ \$}$$ ***No significant correlation***

The predicted value is far from the actual value. Since there is no significant correlation, the mean is used for all predictions – but the $x = 87$ thousand ft^2 is well outside the range of x values used to construct the predictive regression equation.

Exercise

Find the best predicted new mileage rating of a jeep given that old rating is 19 mi/gal. Is the predicted value close to the actual value of 17 mi/gal?

Old	16	27	17	33	28	24	18	22	20	29	21
New	15	24	15	29	25	22	16	20	18	26	19

Solution

x	y	xy	x^2	y^2
16	15	240	256	225
27	24	648	729	576
17	16	272	289	256
33	29	957	1089	841
28	25	700	784	625
24	22	528	576	484
18	16	288	324	256
22	20	440	484	400
20	18	360	400	324
29	26	754	841	676
21	19	399	441	361
255	230	5586	6213	5024

$$\overline{x} = \frac{\sum x}{n} = \frac{255}{11} = 23.18 \qquad \overline{y} = \frac{\sum y}{n} = \frac{230}{11} = 20.82$$

$$\begin{aligned} b_1 &= \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2} \\ &= \frac{11(5586) - (255)(230)}{11(6213) - (255)^2} \\ &= 0.863 \end{aligned}$$

$$\begin{aligned} b_0 &= \overline{y} - b_1\overline{x} \\ &= 20.82 - (0.863)(23.18) \\ &= 0.808 \end{aligned}$$

$$\begin{aligned} \hat{y} &= b_0 + b_1 x \\ &= 0.808 + 0.863x \end{aligned}$$

$$\hat{y}_{182.5} = 0.808 + 0.863(19) = 17.2 \ \ mpg$$

Yes; the predicted value is close to the actual value of 17 mpg.

Exercise

Find the best predicted temperature for a recent year in which the concentration (in parts per million) of CO_2 is 370.9. Is the predicted temperature close to the actual temperature of 14.5° C??

CO_2	314	317	320	326	331	339	346	354	361	369
Temperature	13.9	14.0	13.9	14.1	14.0	14.3	14.1	14.5	14.5	14.4

Solution

x	***y***	***xy***	x^2	y^2
314	13.9	4364.6	985696	193.21
317	14	4438	100489	196
320	13.9	4448	102400	193.21
326	14.1	4596.6	106276	198.81
331	14	4634	109561	196
339	14.3	4847.7	114921	204.49
346	14.1	4878.6	119716	198.81
354	14.5	5133	125316	210.25
361	14.5	5234.5	130321	210.25
369	14.4	5313.6	136161	207.36
3377	141.7	47888.6	1143757	2008.39

$$\bar{x} = \frac{\sum x}{n} = \frac{3377}{10} = 337.7 \qquad \bar{y} = \frac{\sum y}{n} = \frac{141.7}{10} = 14.17$$

$$\begin{aligned} b_1 &= \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2} \\ &= \frac{10(47888.6) - (3377)(141.7)}{10(1143757) - (3377)^2} \\ &= 0.0109 \end{aligned}$$

$$\begin{aligned} b_0 &= \bar{y} - b_1\bar{x} \\ &= 14.17 - (0.0109)(337.7) \\ &= 10.48 \end{aligned}$$

$$\begin{aligned} \hat{y} &= b_0 + b_1 x \\ &= 0.10.5 + 0.0109x \end{aligned}$$

$$\hat{y}_{182.5} = 10.5 + 0.0109(370.9) = 14.5\ \ ^{\circ}C$$

Yes; the predicted temperature is equal to the actual temperature of 14.5 °C..

Exercise

Find the best predicted IQ score of someone with a brain size of 1275 cm^3

Brain Size	965	1029	1030	1285	1049	1077	1037	1068	1176	1105
IQ	90	85	86	102	103	97	124	125	102	114

Solution

x	y	xy	x^2	y^2
965	90	86850	931225	8100
1029	85	87465	1058841	7225
1030	86	88580	1060900	7396
1285	102	131070	1651225	10404
1049	103	108047	1100401	10609
1077	97	104469	1159929	9409
1037	124	128588	1075369	15376
1068	125	133500	1140624	15625
1176	102	119952	1382976	10404
1105	114	125970	1221025	12996
10821	1028	1114491	11782515	107544

$$\bar{x} = \frac{\sum x}{n} = \frac{10821}{10} = 1082.1 \qquad \bar{y} = \frac{\sum y}{n} = \frac{1028}{10} = 102.8$$

$$b_1 = \frac{n\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{n\left(\sum x^2\right) - \left(\sum x\right)^2}$$

$$= \frac{10(1114491) - (10821)(1028)}{10(1178251) - (10821)^2}$$

$$\underline{= 0.0286}$$

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

$$= 102.80 - (0.0286)(1082.1)$$

$$\underline{= 71.83}$$

$$\hat{y} = b_0 + b_1 x$$

$$\underline{= 71.8 - 0.0286x}$$

$$\hat{y}_{182.5} = \bar{y} \ \underline{= 102.8}$$

No significant correlation

Exercise

Listed below are the word counts for men and women.

Male

27531	15684	5638	27997	25433	8077	21319	17572	26429	21966	11680	10818
12650	21683	19153	1411	20242	10117	20206	16874	16135	20734	7771	6792
26194	10671	13462	12474	13560	18876	13825	9274	20547	17190	10578	14821
15477	10483	19377	11767	13793	5908	18821	14069	16072	16414	19017	37649
17427	46978	25835	10302	15686	10072	6885	20848				

Female

20737	24625	5198	18712	12002	15702	11661	19624	13397	18776	15863	12549
17014	23511	6017	18338	23020	18602	16518	13770	29940	8419	17791	5596
11467	18372	13657	21420	21261	12964	33789	8709	10508	11909	29730	20981
16937	19049	20224	15872	18717	12685	17646	16255	28838	38154	25510	34869
24480	31553	18667	7059	25168	16143	14730	28117				

Find the best predicted word count of a woman given that her male partner speaks 6,000 words in a day.

Solution

Using Excel spread sheet - ***Regression***

	Coefficients
Intercept	13438.884
X Variable 1	0.302

$\hat{y} = 13439 + 0.302x$

$\hat{y}\big|_{6000} = 13439 + 0.302(6000)$

$= 15,248 \ \ \textit{words per week}$

Exercise

According the least-squares property, the regression line minimizes the sum of the squares of the residuals. Listed below are the paired data consisting of the first 6 pulse and the first systolic blood pressures of males.

Pulse (***x***)	68	64	88	72	64	110
Systolic (***y***)	125	107	126	110	72	107

a) Find the equation of the regression line.

b) Identify the residuals, and find the sum of squares of the residuals.

c) Show that the equation $\hat{y} = 70 + 0.5x$ results in a larger sum of squares of residuals.

Solution

x = pulse rate

y = systolic blood pressures

a) Using Excel spread sheet – ***Data Analysis*** - ***Regression***

The equation of the regression line: $\hat{y} = 71.678 + 0.5956x$

	Coefficients
Intercept	71.678
X Variable 1	0.5956

b) $y - \hat{y} =$ residuals for the regression line

x	y	$\hat{y}$	$y-\hat{y}$	$(y-\hat{y})^2$
68	125	112.208	12.792	163.635
64	107	109.824	-2.824	7.975
88	126	124.128	1.872	3.504
72	110	114.592	-4.592	21.086
64	110	109.824	0.176	0.031
72	107	114.592	-7.592	57.638
428	685	684.997	0.003	253.866

The table indicates that the sum of the squares of the residuals is 253.866

c) $y-v=$ residuals for the regression line where $v=70+0.5x$

x	y	v	$y-v$	$(y-v)^2$
68	125	104.000	21.000	441.000
64	107	102.000	5.000	25.000
88	126	114.000	12.000	144.000
72	110	106.000	4.000	16.000
64	110	102.000	8.000	64.000
72	107	106.000	1.000	1.000
428	685	634.0	51.0	691.0

The table indicates that the sum of the squares of the residuals is 691, which is greater the the 253.866 of the least squares regression equation.

Exercise

The scatter diagram for the data set below

$\boldsymbol{x}$	0	2	3	5	5	5
$\boldsymbol{y}$	7.3	5.1	6	4	5.3	3.6

Given that $\bar{x}=3.333$, $s_x=2.0655911$, $\bar{y}=5.217$, $s_y=1.3467244$, and $r=-0.8363944$, determine the least squares regression line.

Solution

$$b_1=r\cdot\frac{s_y}{s_x}=-0.8363944\frac{1.3467244}{2.0655911}\ \underline{\approx -.54531}$$

$$b_0=\bar{y}-b_1\bar{x}=5.217-(-.54531)(3.333)\ \underline{=7.0345}$$

$$\hat{y}=b_0+b_1x$$

$$\underline{\hat{y}=-0.5453x+7.0345}$$

Exercise

The scatter diagram for the data set below

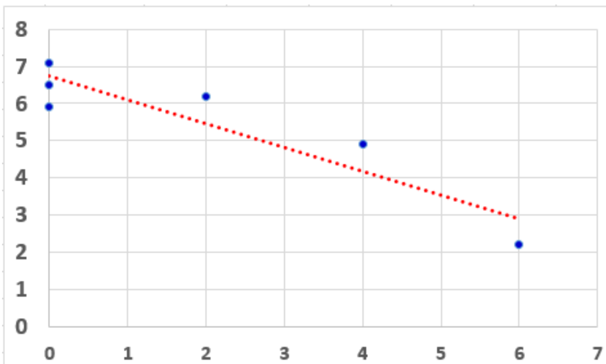
x	0	0	0	2	4	6
y	7.1	5.9	6.5	6.2	4.9	2.2

a) Determine the least squares regression line.
b) Graph the least-squares regression line on the scatter diagram

Solution

a) $\underline{\hat{y} = -0.6375x + 6.7417}$ (***using excel***)

b)



Exercise

The scatter diagram for the data set below

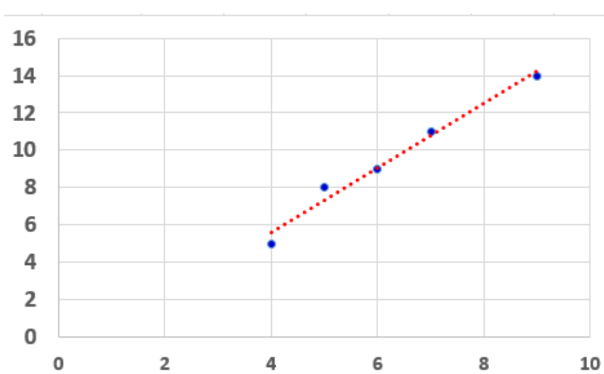
x	4	5	6	7	9
y	5	8	9	11	14

a) Determine the least squares regression line.
b) Graph the least-squares regression line on the scatter diagram.
c) Compute the sum of the squared residuals for the least-squares regression line found in part (*a*).

Solution

a) $\underline{\hat{y} = 1.73x - 1.324}$

b)



c) Using excel regression

ANOVA		
	df	*SS*
Regressio	1	44.28108
Residual	3	0.918919

The sum of the squared residuals for the least-squares regression line is **0.919**.

Exercise

A student at a junior college conducted a survey of 20 randomly selected full-time students to determine the relation between the number of hours of video game playing each week, *x*, and grade-point average, *y*. She found that a linear relation exists between the two variables. The least-squares regression line that describes this relation is $\hat{y} = -0.0531x + 2.9213$.

a) Predict the grade-point average of a student who plays video games 8 hours per week.

b) Interpret the slope

c) Interpret the appropriate *y*-intercept.

d) A student who plays video games 7 hours per week has a grade-point average of 2.67. Is the student grade-point average above or below average among all students who play video games 7 hours per week.

Solution

a) $\hat{y} = -0.0531(8) + 2.9213 \approx 2.50$

b) For each additional hour that a student spends playing video games in a week, the grade-point average will decrease by 0.0531 points, on average.

c) The grade-point average of a student who does not play video games in 2.9213

d) $\hat{y} = -0.0531(7) + 2.9213 \approx 2.55$

The student's grade-point average is above average for those who play video games 7 hours per week.