Chapter V

Correlation and Regression (Solution)

5. 1.

Working Table

x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x}) \cdot (y - \bar{y})$	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$
125000	19	71100	3	213300	5055210000	9
100000	20	46100	4	184400	2125210000	16
40000	16	-13900	0	0	193210000	0
35000	16	-18900	0	0	357210000	0
41000	18	-12900	2	-25800	166410000	4
29000	12	-24900	-4	99600	620010000	16
35000	14	-18900	-2	37800	357210000	4
24000	12	-29900	-4	119600	894010000	16
50000	16	-3900	0	0	15210000	0
60000	17	6100	1	6100	37210000	1
539000	160	0	0	635000	9820900000	66

$\bar{x} = 53900, \qquad \bar{y} = 16$

$$r := \frac{635000}{\sqrt{9820900000 \cdot 66}} = 0.788725919 \approx 0.79.$$

Therefore, there is a high degree of correlation between the two factors.

5. 2.
We rank both sets of data, giving the largest value rank 1, the second largest value rank 2, etc.

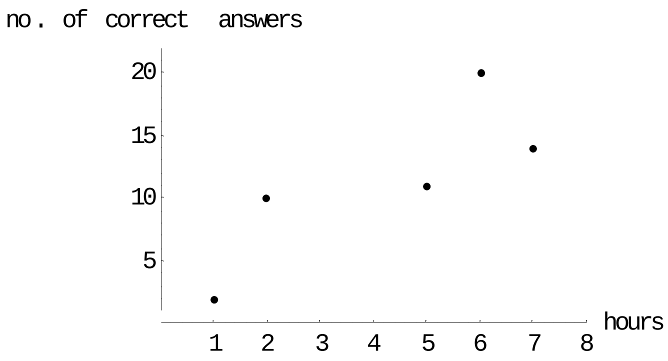
Business Mathematics [%]	68	54	19	72	50	44	92	37
Statistics [%]	51	76	32	85	62	25	74	59
R_i	3	4	8	2	5	6	1	7
$R_i^{'}$	6	2	7	1	4	8	3	5
$\left(R_i - R_i^{'}\right)^2$	9	4	1	1	1	4	4	4

$$\rho := 1 - \frac{6 \cdot 28}{7 \cdot 8 \cdot 9} \approx 0.67$$

Therefore, there is a moderate degree of correlation between the marks in the two subjects.

5. 3.

a)



b)

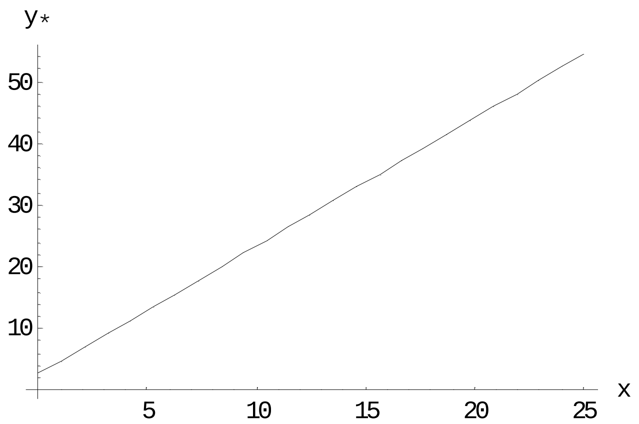
x_i	y_i	x_i^2	$x_i \cdot y_i$
1	2	1	2
2	10	4	20
6	20	36	120
7	14	49	98
5	11	25	55
21	57	115	295

$$y^* = a_0 + a_1 \cdot x$$

$$\begin{aligned} 5a_0 \quad + 21a_1 &= \quad 57 \\ 21a_0 + 115a_1 &= 295 \end{aligned} \quad \Rightarrow \quad a_0 = 2.71, \quad a_1 = 2.07$$

$$y^* = 2.71 + 2.07x$$

c)



d)

$$y^*(3) = 2.71 + 2.07 \cdot 3 = 8.92.$$

5. 4.

1.

$$x = a_0 \cdot r^{a_1}$$

$$\lg x = \lg a_0 + a_1 \cdot \lg r$$

Let

$$X := \lg x, \qquad A_0 := \lg a_0, \qquad A_1 := a_1, \qquad R := \lg r.$$

We now have to solve the system of equations:

$$n \cdot A_0 + A_1 \cdot \sum_i \lg r_i = \sum_i \lg x_i$$
$$A_0 \cdot \sum_i \lg r_i + A_1 \cdot \sum_i (\lg r_i)^2 = \sum_i \lg r_i \cdot \lg x_i$$

Working Table

r_i	x_i	$\lg r_i$	$\lg x_i$	$(\lg x_i)^2$	$\lg x_i \cdot \lg r_i$
5.0	1.54	0.69897000	0.18752072	0.488559067	0.13107136
5.5	1.64	0.74036269	0.21484385	0.548136912	0.15906237
6.0	1.70	0.77815125	0.23044892	0.605519368	0.17932412
6.5	1.80	0.81291336	0.25527251	0.660828125	0.20751443
7.0	1.86	0.84509804	0.26951294	0.714190697	0.22776486
7.5	1.90	0.87506126	0.2787536	0.765732215	0.24392648
8.0	1.98	0.90308999	0.29666519	0.815571525	0.26791536
		5.65364659	1.73301773	4.598537909	1.41657898

$$7A_0 \quad + 5.65A_1 = 1.73$$
$$5.65A_0 + 4.60A_1 = 1.42$$

$$A_0 = -0.234 \quad \Rightarrow \quad a_0 = 0.583$$
$$A_1 = 0.596 \quad \Rightarrow \quad a_1 = 0.596$$

$$x = 0.583 \cdot r^{0.596}$$

2.

$$x(8.3) = 0.583 \cdot 8.3^{0.596} \approx 2.06$$

5. 5.

1.

$$C(x) = a_0 \cdot e^{a_1 \cdot x}$$

$$\ln C(x) = \ln a_0 + a_1 \cdot x$$

Let

$$Y := \ln C(x), \quad A_0 := \ln a_0, \quad A_1 := a_1, \quad X := x.$$

We now have to solve the system of equations:

$$n \cdot A_0 + A_1 \cdot \sum_i X_i = \sum_i Y_i$$
$$A_0 \cdot \sum_i X_i + A_1 \cdot \sum_i (X_i)^2 = \sum_i X_i \cdot Y_i$$

Working Table

X_i	$C(x_i)$	$X_i^{\,2}$	Y_i	$X_i \cdot Y_i$
10	102.06	100	4.625560876	46.25560876
15	103.06	225	4.635311343	69.52967014
20	104.10	400	4.645351976	92.90703951
25	105.14	625	4.655292795	116.38231988
30	106.20	900	4.665324109	139.95972326
35	107.21	1225	4.674789528	163.61763348
135		3475	27.901630626	628.65199504

$$6A_0 + 135A_1 = 27.90$$
$$135A_0 + 3475A_1 = 628.65$$

$$A_0 \approx 4.65 \quad \Rightarrow \quad a_0 \approx 104.58$$
$$A_1 \approx 0.0021 \quad \Rightarrow \quad a_1 \approx 0.0.0021$$

$$C(x) = 104.58 \cdot e^{0.0021x}$$

2.

$$C(37) = 104.58 \cdot e^{37 \cdot 0.0021} \approx 113.03$$

5. 6.
Yes, by applying regression analysis:

x_i	y_i	x_i^2	$x_i \cdot y_i$
20	89	400	1780
16	72	256	1152
20	93	400	1860
18	84	324	1512
17	81	289	1377
16	75	256	1200
15	70	225	1050
17	82	289	1394
15	69	225	1035
16	83	256	1328
15	80	225	1200
17	83	289	1411
16	81	256	1296
17	84	289	1428
14	76	196	1064
249	1202	4175	20087

$$y^* = a_0 + a_1 \cdot x$$

$$\begin{aligned} 15a_0 + 249a_1 &= 1202 \\ 249a_0 + 4175a_1 &= 20087 \end{aligned} \quad \Rightarrow \quad a_0 \approx 26.74, \quad a_1 \approx 3.22,$$

$$y^* = 26.74 + 3.22x$$

5. 7.
1.

$$x = a_0 \cdot r^{a_1}$$

$$\lg x = \lg a_0 + a_1 \cdot \lg r$$

Let

$$X := \lg x, \qquad A_0 := \lg a_0, \qquad A_1 := a_1, \qquad R := \lg r.$$

$$\begin{aligned} n \cdot A_0 + A_1 \cdot \sum_i \lg r_i &= \sum_i \lg x_i \\ A_0 \cdot \sum_i \lg r_i + A_1 \cdot \sum_i (\lg r_i)^2 &= \sum_i \lg r_i \cdot \lg x_i \end{aligned}$$

Working Table

r_i	x_i	$\lg r_i$	$\lg x_i$	$(\lg x_i)^2$	$\lg x_i \cdot \lg r_i$
5.0	1.54	0.69897000	0.18752072	0.488559067	0.13107136
5.5	1.64	0.74036269	0.21484385	0.548136912	0.15906237
6.0	1.70	0.77815125	0.23044892	0.605519368	0.17932412
6.5	1.80	0.81291336	0.25527251	0.660828125	0.20751443
7.0	1.86	0.84509804	0.26951294	0.714190697	0.22776486
7.5	1.90	0.87506126	0.2787536	0.765732215	0.24392648
8.0	1.98	0.90308999	0.29666519	0.815571525	0.26791536
		5.65364659	1.73301773	4.598537909	1.41657898

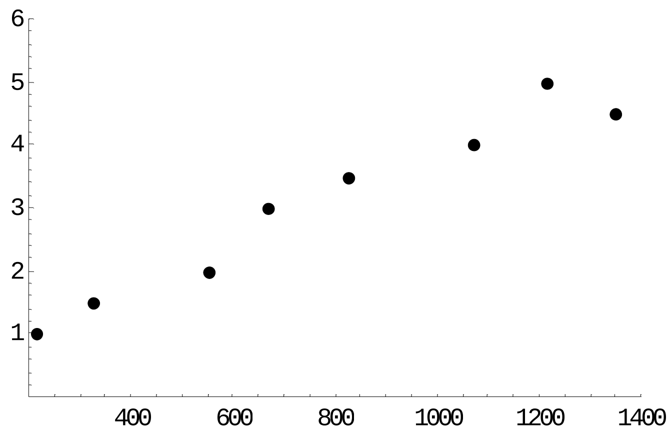
$$7A_0 \quad + 5.65A_1 = 1.73$$
$$5.65A_0 + 4.60A_1 = 1.42$$

$$A_0 = -0.234 \quad \Rightarrow \quad a_0 = 0.583$$
$$A_1 = 0.596 \quad \Rightarrow \quad a_1 = 0.596$$

$$x = 0.583 \cdot r^{0.596}$$

5. 8.

1.



2.

Working Table

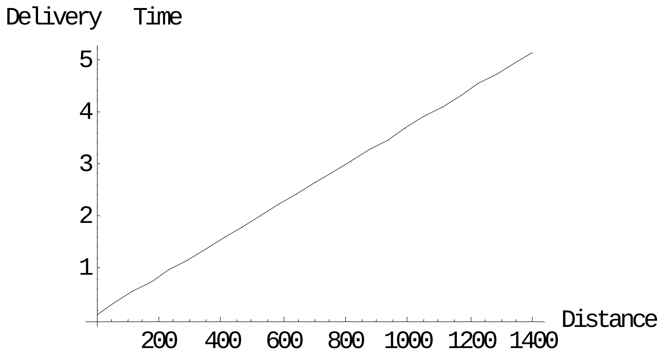
x_i	y_i	x_i^2	$x_i \cdot y_i$	y_i^2
825	3.5	680625	2887,5	12.25
215	1.0	46225	215,0	1.00
1070	4.0	1144900	4280,0	16.00
550	2.0	302500	1100,0	4.00
480	1.0	230400	480,0	1.00
920	3.0	846400	2760,0	9.00
1350	4.5	1822500	6075,0	20.25
325	1.5	105625	487,5	2.25
670	3.0	448900	2010,0	9.00
1215	5.0	1476225	6075,0	25.00
7620	28.5	7104300	26370,0	99.75

$$\begin{cases} 10a_0 + 7620a_1 = 28.5 \\ 7620a_0 + 7104300a_1 = 26370 \end{cases}$$

$$a_0 = 0.118129074 \approx 0.1181, \qquad a_1 = 0.003585132449 \approx 0.0036$$

$$y^* = 0.1181 + 0.0036x.$$

3.



4.

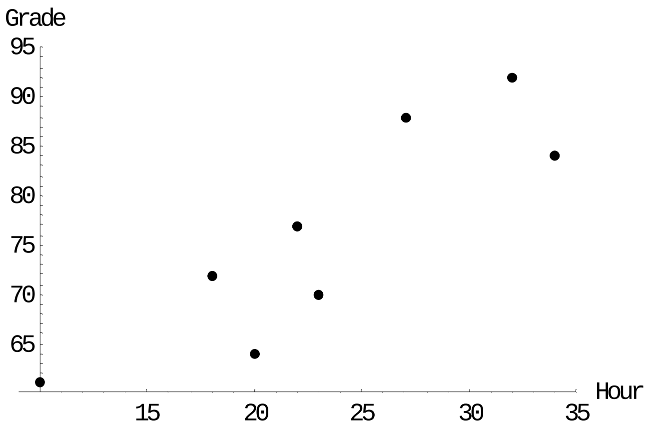
$$r = \frac{10 \cdot 26370 - 7620 \cdot 28.50}{\sqrt{\left(10 \cdot 7104300 - 7620^2\right) \cdot \left(10 \cdot 99.75 - 28.5^2\right)}} = 0.948942769$$

The two factors distance and delivery time move in the same direction,

$$r^2 = 0.90049238 \approx 0.90.$$

Distance accounts for up to 90% of the development of delivery time,

5. 9.
1.



2.

Working Table

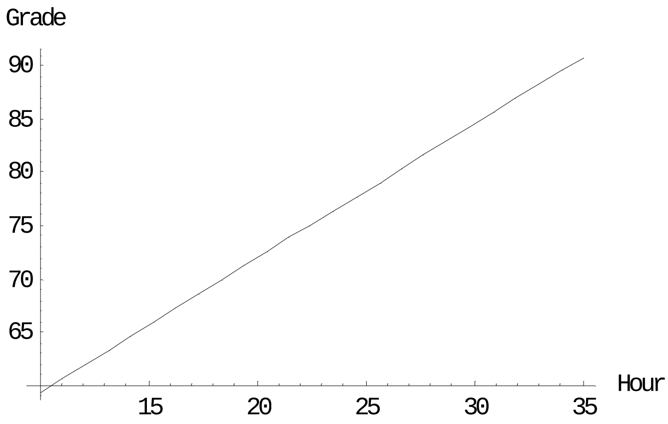
x_i	y_i	x_i^2	$x_i \cdot y_i$	y_i^2
20	64	400	1280	4096
10	61	100	610	3721
34	84	1156	2856	7056
23	70	529	1610	4900
27	88	729	2376	7744
32	92	1024	2944	8464
18	72	324	1296	5184
22	77	484	1694	5929
186	608	4746	14666	47094

$$\begin{cases} 8a_0 + 186a_1 = 608 \\ 186a_0 + 4746a_1 = 14666 \end{cases}$$

$a_0 \approx 46.7655, \qquad a_1 \approx 1.2574$

$$y^* = 46.7655 + 1.2574x.$$

3.



4.

$$r = \frac{8 \cdot 14666 - 186 \cdot 608}{\sqrt{\left(8 \cdot 4746 - 186^2\right) \cdot \left(8 \cdot 47094 - 608^2\right)}} \approx 0.8673.$$

The two factors study hours and examination grade move in the same direction,

$$r^2 \approx 0.7522.$$

Study hours account for up to 75% of the achieved examination grade.

5. 10.

1.

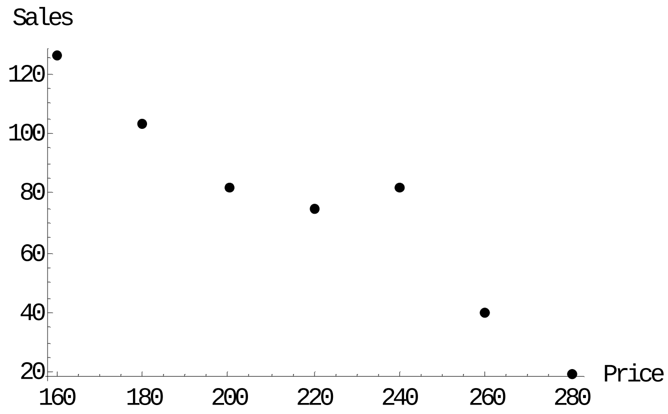
Working Table 1

Price	Increase [%]
160	-
180	12.5000
200	11.1111
220	10.0000
240	9.0909
260	8.3333
280	7.6923

$$\bar{x}_g = \sqrt[6]{(1+0.1250)(1+0.1111)(1+0.1000)(1+0.0909)(1+0.0833)(1+0.0769)} \approx 1.0978 .$$

The average rate of increase of the prices was, therefore, approximately equal to 9.78%

2.



3.

x_i	y_i	x_i^2	$x_i \cdot y_i$	y_i^2
160	126	25600	20160	15876
180	103	32400	18540	10609
200	82	40000	16400	6724
220	75	48400	16500	5625
240	82	57600	19680	6724
260	40	67600	10400	1600
280	20	78400	5600	400
1540	528	350000	107280	47558

$$\begin{cases} 7a_0 + 1540a_1 = 528 \\ 1540a_0 + 350000a_1 = 107280 \end{cases} \Rightarrow a_0 \approx 249.8571, \quad a_1 \approx -0.7929$$

$$y^* = 249.8571 - 0.7929x$$

4,

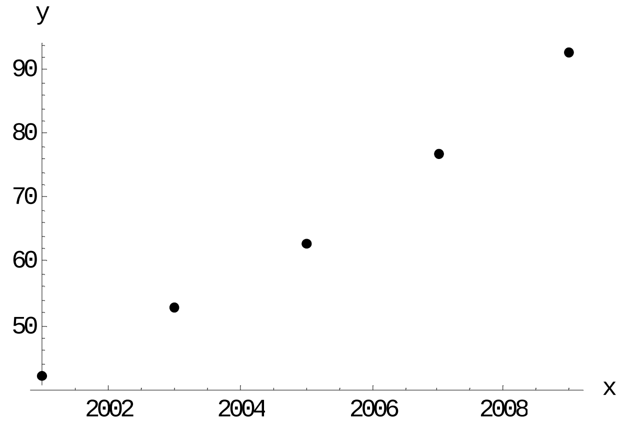
$$r = \frac{7 \cdot 107280 - 1540 \cdot 528}{\sqrt{(7 \cdot 350000 - 1540^2)(7 \cdot 47558 - 528^2)}} \approx -0.9543$$

$$r^2 \approx 0.9106$$

The sales are to about 91% determined by the prices.

5. 11.

1.



2.

2.

$$y^* = a_0 \cdot a_1^x$$

$$\lg y^* = \lg a_0 \cdot a_1^x$$

$$\lg y^* = \lg a_0 + \lg a_1^x$$

$$\lg y^* = \lg a_0 + x \cdot \lg a_1$$

$$Y^* = A_0 + A_1 x \quad \text{with } A_0 := \lg a_0,\ \ A_1 := \lg a_1,\ \ Y_i := \lg y_i$$

$$\begin{cases} nA_0 \qquad + A_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} Y_i \\ A_0 \sum_{i=1}^{n} x_i + A_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i Y_i \end{cases}$$

Working Table

	x_i	y_i	x_i^2	$\lg y_i$	$x_i \cdot \lg y_i$
2001	-2	42	4	1.62324929	-3.24649858
2003	-1	51	1	1.70757018	-1.70757018
2005	0	63	0	1.79934055	0.0000000
2007	1	77	1	1.88649073	1.88649073
2009	2	93	4	1.96848295	3.9369659
	0		10	8.98513369	0.86938787

$$\begin{cases} 5A_0 \qquad = 8.98513369 \\ \qquad 10A_1 = 0.86938787 \end{cases} \Rightarrow$$

$A_0 = 1.797026738 \approx 1.7970, \quad A_1 = 0.08693878700 \approx 0.0869$

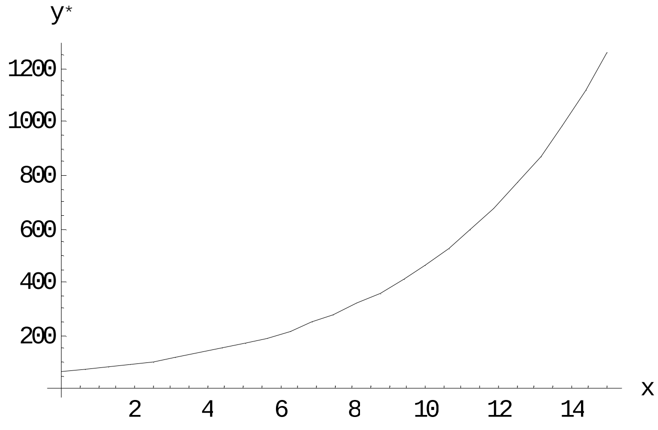
$a_0 = 62.66524443 \approx 62.6652, \quad a_1 = 1.22162746200 \approx 1.2216$

$y^* = 62.6654 \cdot 1.2216^x$

3.

$y^*_{2011} = y^*(3) \approx 114$

4.



(Last revised: 17.08.2010)