Solution *Section* 2.1 – Scatter Diagrams and Correlation

Exercise

For each of several randomly selected years, the total number of points scored in the Super Bowl football game and the total number of new cars sold in The U.S. are recorded. For this sample of paired data

a) What does r represent?

b) What does ρ represent?

c) With our doing any research or calculations, estimate the value of r.

Solution

a) r = the correlation in the sample. In this context, r is the linear correlation coefficient computed using the chosen paired (points in Super Bowl, number of new cars sold) values for the randomly selected years in the sample.

b) ρ = the correlation in the population. In this context, ρ is the linear correlation coefficient computed using the paired (points in Super Bowl, number of new cars sold) values for every year there has been a Super Bowl.

c) Since there is no relationship between the number of points scored in a Super Bowl and the number of new cars sold that year, the estimated value of r is 0.

Exercise

The heights (in inches) of a sample of eight mother/daughter pairs of subjects measured. Using Excel with the paired mother/daughter heights, the linear correlation coefficient is found to be 0.693. Is there sufficient evidence to support the claim that there is a linear correlation between the heights of mothers and the heights of their daughters? Explain.

Solution

From the table for $n = 8, \quad C.V. = \pm 0.707$. Therefore $r = 0.693$ indicates a significant (positive) linear correlation. Yes; there is sufficient evidence to support the claim that there is a linear correlation between the heights of mothers and the heights of their daughters,

Exercise

The heights and weights of a sample of 9 supermodels were measured. Using a TI calculator, the linear correlation coefficient is found to be 0.360. Is there sufficient evidence to support the claim that there is a linear correlation between the heights and weights of supermodels? Explain.

Solution

From the table for $n = 9, \quad C.V. = \pm 0.666$. Therefore $r = 0.360$ does not indicate a significant linear correlation. No; there is not sufficient evidence to support the claim that there is a linear correlation between the heights and weights of supermodels.

Exercise

Given the table 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

a) Construct a scatterplot

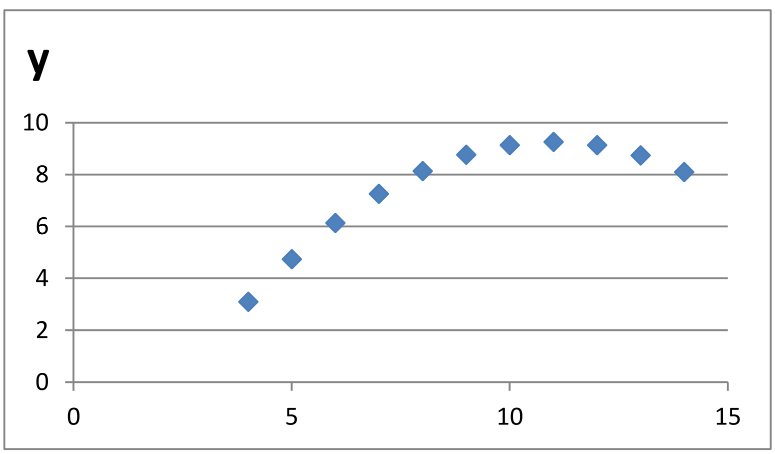
b) Find the value of linear correlation coefficient *r* and then determine whether there is sufficient evidence to support the claim of a linear correlation between the 2 variables.

c) Identify the feature of the data that would be missed if part (*b*) was completed without constructing the scatterplot.

Solution

a) Excel produces the following

x	***y***	***xy***	x^2	y^2
10	9.14	91.4	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	8.1	113.4	196	65.6100
6	6.13	36.78	36	37.5769
4	3.1	12.4	16	9.6100
12	9.13	109.56	144	83.3569
7	7.26	50.82	49	52.7076
5	4.74	23.7	25	22.4676
99	82.51	797.59	1001	660.176



b) $r = \dfrac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2} \cdot \sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}}$

$$= \frac{(11)(797.59)-(99)(82.51)}{\sqrt{(11)(1001)-(99)^2}\cdot\sqrt{(11)(660.1763)-(82.51)^2}}$$

$$= 0.816$$

From table A-5; $n = 11$, $C.V. = \pm 0.602$

Therefore r = 0.816 indicates a significant (positive) linear correlation. Yes; there is sufficient evidence to support the claim that there is a linear correlation between the 2 variables.

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

Exercise

Given the table 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

a) Construct a scatterplot

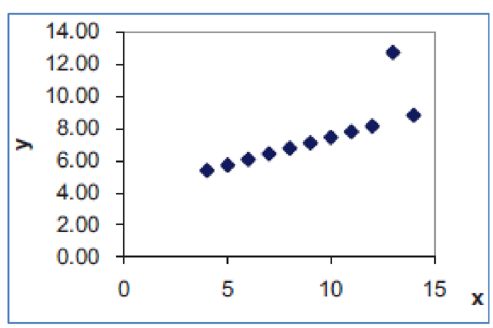
b) Find the value of linear correlation coefficient r and then determine whether there is sufficient evidence to support the claim of a linear correlation between the 2 variables.

c) Identify the feature of the data that would be missed if part (b) was completed without constructing the scatterplot.

Solution

a) Excel produces the following

x	**y**	**xy**	x^2	y^2
10	7.46	74.60	100	55.6516
8	6.77	54.16	64	45.8329
13	12.74	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



b) $$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right)-\left(\sum x\right)^2}\cdot\sqrt{n\left(\sum y^2\right)-\left(\sum y\right)^2}}$$

$$= \frac{(11)(797.59)-(99)(82.5)}{\sqrt{(11)(1001)-(99)^2}\cdot\sqrt{(11)(659.9762)-(82.5)^2}}$$

$= 0.816$

From table A-5; $n = 11$, $C.V. = \pm 0.602$

Therefore $r = 0.816$ indicates a significant (positive) linear correlation. Yes; there is sufficient evidence to support the claim that there is a linear correlation between the 2 variables.

c) 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 oulier.

Exercise

The paired values of the Consumer Price Index (CPI) and the cost of a slice of pizza are shown 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

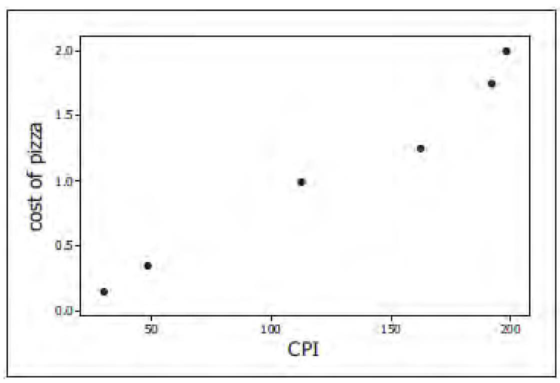
a) Construct a scatterplot

b) Find the value of linear correlation coefficient r.

Solution

a) 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



b) $$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right)-\left(\sum x\right)^2}\cdot\sqrt{n\left(\sum y^2\right)-\left(\sum y\right)^2}}$$

$$= \frac{(6)(1067.91)-(742.7)(6.5)}{\sqrt{(6)(118115.5)-(742.7)^2}\cdot\sqrt{(6)(9.77)-(6.5)^2}}$$

$= 0.985$

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

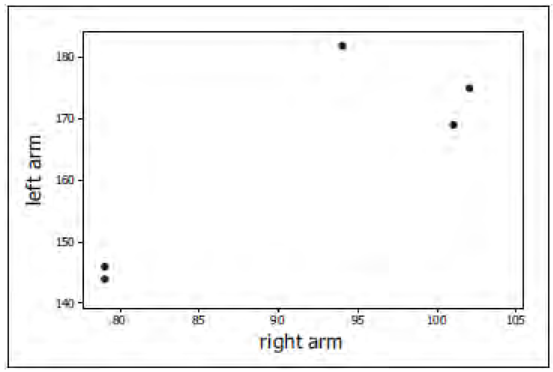
a) Construct a scatterplot

b) Find the value of linear correlation coefficient r.

Solution

a) Excel produces the following

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



b) $$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2} \cdot \sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}}$$

$$= \frac{(5)(74937) - (455)(816)}{\sqrt{(5)(41923) - (255)^2} \cdot \sqrt{(5)(134362) - (816)^2}}$$

$$= 0.867$$

Exercise

Listed below are costs (in dollars) of air fares for different airlines from NY to San Francisco. The costs are based on tickets purchased 30 days in advance and one day in advance.

30 Days	244	260	264	264	278	318	280
One Day	456	614	567	943	628	1088	536

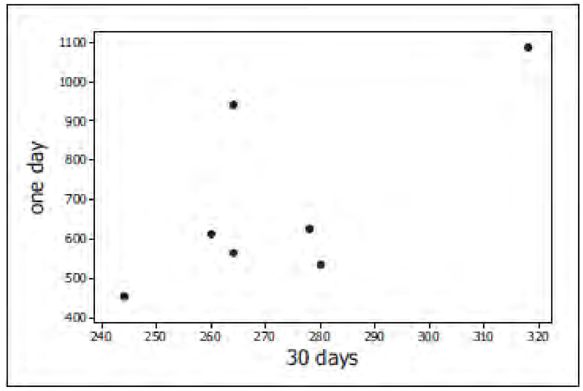
a) Construct a scatterplot

b) Find the value of linear correlation coefficient r.

Solution

a) Excel produces the following

x	***y***	***xy***	x^2	y^2
244	456	111264	59536	207936
260	614	159640	67600	376996
264	567	149688	69696	889249
264	943	248952	69696	889249
278	628	174584	77284	394384
318	1088	345984	101124	1183744
280	536	150080	78400	287296
1908	4832	1340192	523336	3661094



b) $r = \dfrac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2} \cdot \sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}}$

$= \dfrac{(7)(1340192) - (1908)(4832)}{\sqrt{(7)(523336) - (1908)^2} \cdot \sqrt{(7)(3661094) - (4832)^2}}$

$\underline{= 0.709}$

Exercise

Listed below are repair costs (in dollars) for cars crashed at 6 mi/h in full-front crash tests and the same cars crashed at 6 mi/f in full-rear crash tests.

Front	936	978	2252	1032	3911	4312	3469
Rear	1480	1202	802	3191	1122	739	2767

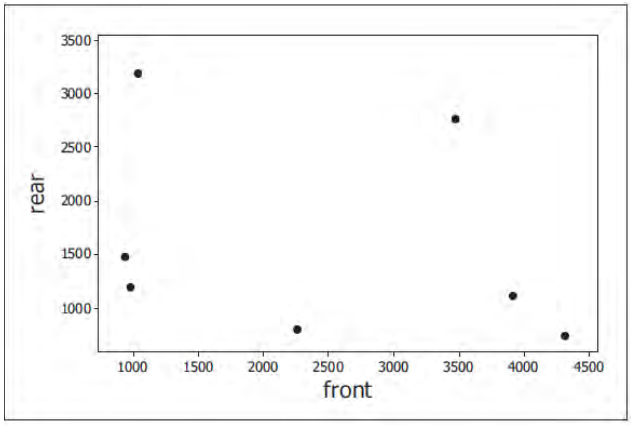
a) Construct a scatterplot

b) Find the value of linear correlation coefficient *r*.

Solution

a) Excel produces the following

x	***y***	***xy***	x^2	y^2
936	1480	1385280	876096	2190400
978	1202	1175556	956484	1444804
2252	802	1806104	5071504	643204
1032	3191	3293112	1065024	10182481
3911	1122	4388142	15295921	1258884
4312	739	3186568	18593344	546121
3469	2767	9598723	12033961	7656289
16890	11303	24833485	53892334	23922183



b) $$r = \frac{n\sum xy - \left(\sum x\right)\left(\sum y\right)}{\sqrt{n\left(\sum x^2\right) - \left(\sum x\right)^2} \cdot \sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}}$$

$$= \frac{(7)(24833485) - (16890)(11303)}{\sqrt{(7)(53892344) - (16890)^2} \cdot \sqrt{(7)(23922183) - (11303)^2}}$$

$$\underline{= -0.283}$$

Exercise

For the following data:

x	2	4	7	7	9
y	1.6	2.1	2.4	2.6	3.2

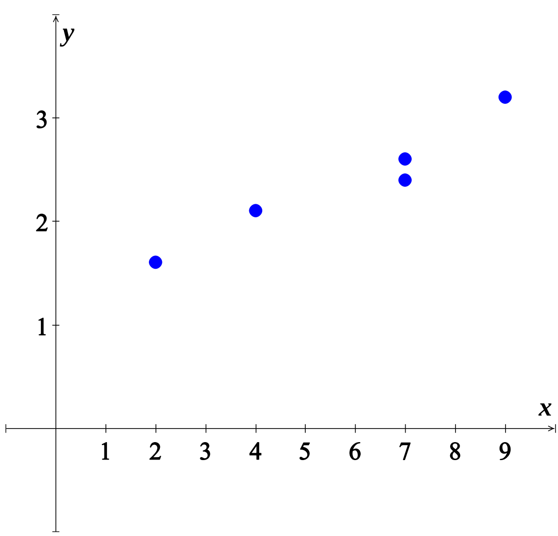
a) Draw a scatter diagram

b) Compute the correlation coefficient

c) Comment on the type of the relation that appears to exist between x and y.

Solution

a)



b) $r \approx 0.968$

c) The linear correlation is close to **1**, so a positive linear relation exists between *x* and *y*.

Exercise

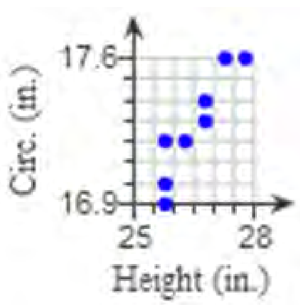
A pediatrician wants to determine the relation that may exist between a child's height and head circumference. She randomly selects 8 children, measures their height and head cieecumference, and obtains the data shown in the table.

Height (in.)	27.25	25.75	26.25	25.75	27.75	26.75	25.75	26.75
Head Circumference (in.)	17.6	17	17.2	16.9	17.6	17.3	17.2	17.4

a) Draw a scatter diagram
b) Compute the correlation coefficient
c) If the pediatrician wants to use height to predict head circumference, determine which variable is the explanatory variable and which is the response variable.
d) Does a linear relation exist between height and head circumference?

Solution

a)



b) $r \approx 0.968$

c) The explanatory variable is height and the response is head circumference.

d) Yes, there appears to be a positive linear association because r is positive and is greater than the critical value.

Exercise

An engineer wanted to determine how the weight of a car affects gas mileage. The accompanying data represent the weight of various domestic cars and their mileages in the city for the 2008 model year. Suppose that we add Car 12 to the original data. Car 12 weights 3,305 *lbs*. and gets 19 miles per gallon.

Car	***Weight (lbs)***	***Miles / Gal***
1	3,775	21
2	3,964	17
3	3,470	21
4	3,175	22
5	2,580	27
6	3,730	18
7	2,605	26
8	3,772	17
9	3,310	20
10	2,991	25
11	2,752	26

a) Draw a scatter diagram

b) Compute the correlation coefficient.

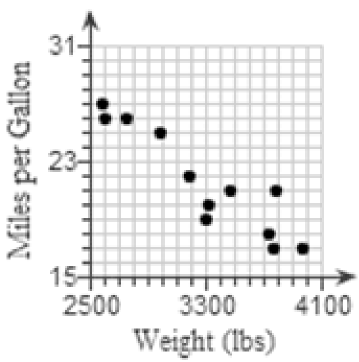
c) Compute the correlation coefficient with Car 12 included

d) Compare the correlation coefficient in part (*b*) & (*c*), and why are the results reasonable.

e) Suppose that we add Car 13 (a hybrid car) to the original data (remove the Car 12). Car 13 weighs 2,890 *lbs*. and gets 60 miles per gallon. Compute the linear coefficient with Car 13 included,

Solution

a)



b) $r \approx -0.946$

c) $r \approx -0.925$

d) The results are reasonable because the Car 12 follows the overall pattern of the data.

e) $r \approx -0.502$