MEASURES OF ASSOCIATION BETWEEN TWO VARIABLES

Prepared by: Jasmine A. Tulin

MEASURES OF ASSOCIATION BETWEEN TWO VARIABLES

Prepared by: Jasmine A. Tulin

SCATTER CHARTS; COVARIANCE

- **Scatter Charts**: Useful graph for analyzing the relationship between two variables.
- **Covariance**: Descriptive measure of the linear association between two variables.
  - Sample covariance for a sample of size *n* with the observations (x_1, y_1), (x_2, y_2), and so on:

$$s_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}$$

  - Population covariance, $\sigma_{xy} = \frac{\sum(x_i - \mu_x)(y_i - \mu_y)}{N}$

CORRELATION COEFFICIENT

- **Correlation coefficient**: Measures the relationship between two variables.
  - Not affected by the units of measurement for *x* and *y*.
  - Sample correlation coefficient denoted by r_{xy}.
    - $r_{xy} = \frac{s_{xy}}{s_x s_y}$
      - s_{xy} = sample covariance = $\frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n-1}$
      - s_x = sample standard deviation of *x* = $\sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$
      - s_y = sample standard deviation of *y* = $\sqrt{\frac{\sum(y_i - \bar{y})^2}{n-1}}$

Interpretation of Correlation Coefficient

- $-1 \leq r \leq +1$

r value	Relationship between the *x* and *y* variables
< 0	Negative linear
Near 0	No linear relationship
> 0	Positive linear

Table 2.14 - Data for Bottled Water Sales at Queensland Amusement Park for a Sample of 14 Summer Days

High Temperature (°F)	Bottled Water Sales (cases)
78	23
79	22
80	24
80	22
82	24
83	26
85	27
86	25
87	28
87	26
88	29
88	30
90	31
92	31

Figure 2.23 - Chart Showing the Positive Linear Relation Between Sales and High Temperatures

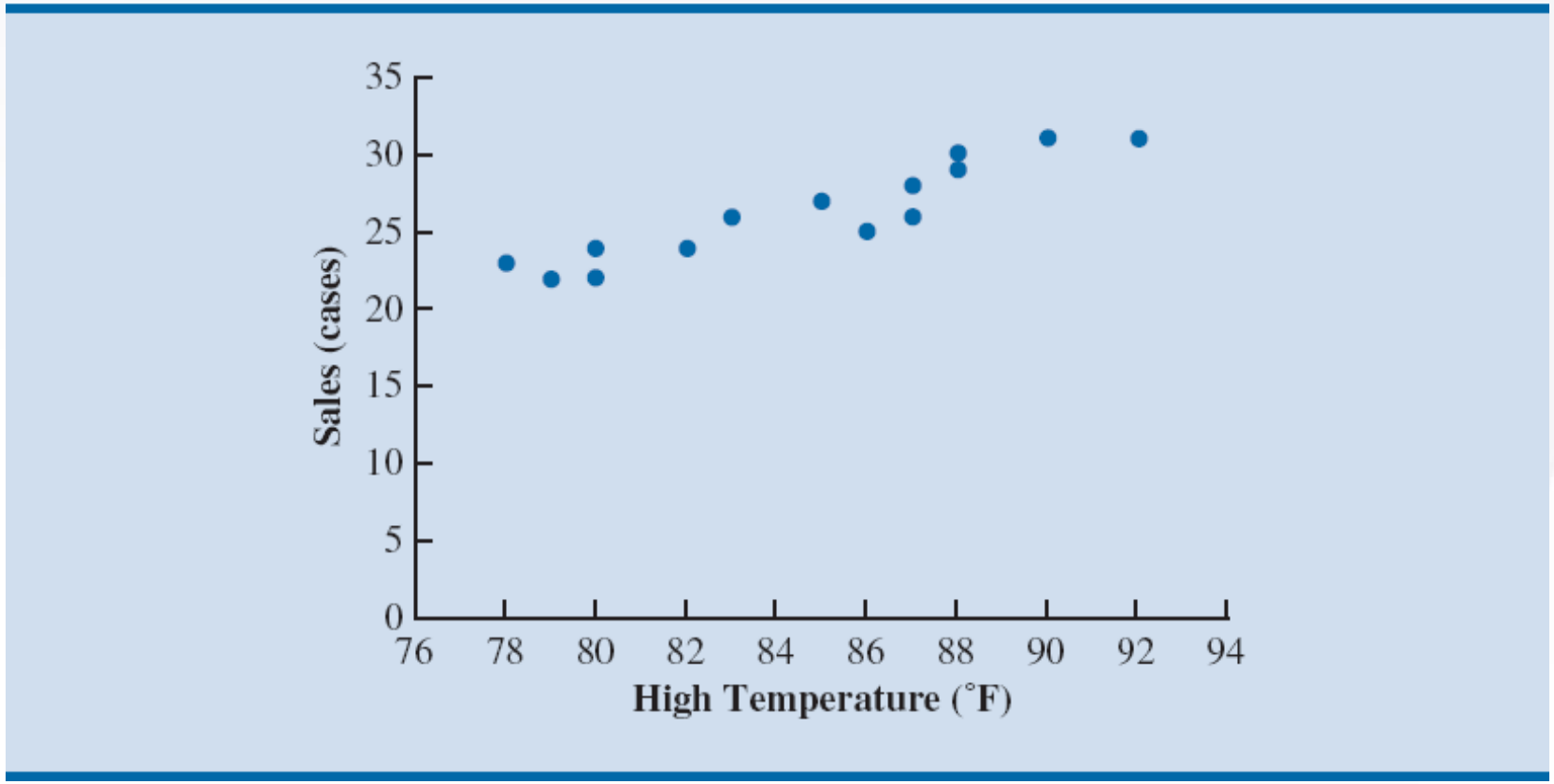


Table 2.15 - Sample Covariance Calculations for Daily High Temperature and Bottled Water Sales at Queensland Amusement Park

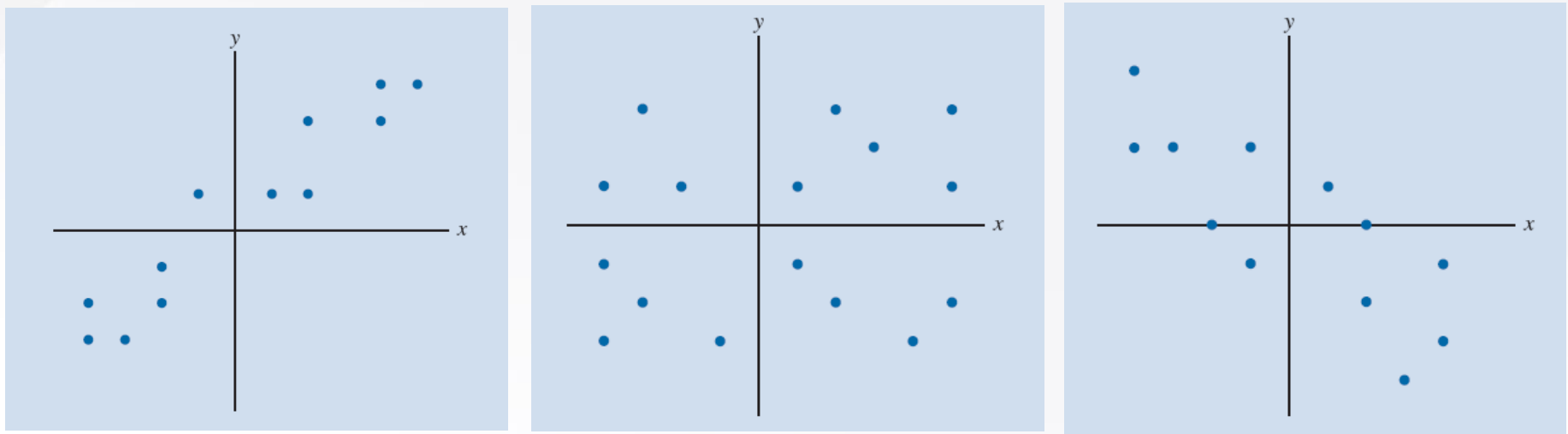
	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
	78	23	−6.6	−3.3	21.78
	79	22	−5.6	−4.3	24.08
	80	24	−4.6	−2.3	10.58
	80	22	−4.6	−4.3	19.78
	82	24	−2.6	−2.3	5.98
	83	26	−1.6	−0.3	0.48
	85	27	0.4	0.7	0.28
	86	25	1.4	−1.3	−1.82
	87	28	2.4	1.7	4.08
	87	26	2.4	−0.3	−0.72
	88	29	3.4	2.7	9.18
	88	30	3.4	3.7	12.58
	90	31	5.4	4.7	25.38
	92	31	7.4	4.7	34.78
Totals	1185	368	0.6	−0.2	166.42

$$\bar{x} = 84.6$$

$$\bar{y} = 26.3$$

$$s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n - 1} = \frac{166.42}{14 - 1} = 12.8$$

Figure 2.25 - Scatter Diagrams and Associated Covariance Values for Different Variable Relationships



(a)
s_{xy} **Positive:**
(*x* and *y* are positively linearly related)

(b)
s_{xy} **Approximately 0:**
(*x* and *y* are not linearly related)

(c)
s_{xy} **Negative:**
(*x* and *y* are negatively linearly related)

Computation of Correlation Coefficient

- Illustration - To determine the sample correlation coefficient for bottled water sales at Queensland Amusement Park:

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{12.8}{(4.36)(3.15)} = 0.93$$

- There is a very strong linear relationship between high temperature and sales.

Figure 2.24 - Calculating Covariance and Correlation Coefficient for Bottled Water Sales Using Excel

	A	B
1	High Temperature (degrees F)	Bottled Water Sales (cases)
2	78	23
3	79	22
4	80	24
5	80	22
6	82	24
7	83	26
8	85	27
9	86	25
10	87	28
11	87	26
12	88	29
13	88	30
14	90	31
15	92	31
16		
17	Covariance:	=COVARIANCE.S(A2:A15,B2:B15)
18	Correlation Coefficient:	=CORREL(A2:A15,B2:B15)

	A	B
1	High Temperature (degrees F)	Bottled Water Sales (cases)
2	78	23
3	79	22
4	80	24
5	80	22
6	82	24
7	83	26
8	85	27
9	86	25
10	87	28
11	87	26
12	88	29
13	88	30
14	90	31
15	92	31
16		
17	Covariance:	12.80
18	Correlation Coefficient:	0.93