Pair	Exposed, X	Control, Y	Difference, D = X − Y	Pair	Exposed, X	Control, Y	Difference, D = X − Y
1	38	16	22	18	10	13	−3
2	23	18	5	19	45	9	36
3	41	18	23	20	39	14	25
4	18	24	−6	21	22	21	1
5	37	19	18	22	35	19	16
6	36	11	25	23	49	7	42
7	23	10	13	24	48	18	30
8	62	15	47	25	44	19	25
9	31	16	15	26	35	12	23
10	34	18	16	27	43	11	32
11	24	18	6	28	39	22	17
12	14	13	1	29	34	25	9
13	21	19	2	30	13	16	−3
14	17	10	7	31	73	13	60
15	16	16	0	32	25	11	14
16	20	16	4	33	27	13	14
17	15	24	−9				

Source: D. Morton et al., "Lead Absorption in Children of Employees in a Lead-Related Industry," *American Journal of Epidemiology*, vol. 115, 1982, pp. 549–555. Used with permission.

For each of the three variables ("exposed," "control," and "difference"), draw the frequency histogram and describe the shape of the distribution.

1.13 Cadmium (Cd) is a bluish-white metal used as a protective coating for iron, steel, and copper. Overexposure to cadmium dust can damage the lungs, kidney, and liver. The federal standard for cadmium dust in the workplace is 200 $\mu g/m^3$. The following (simulated) data were obtained by measuring the levels of cadmium dust at 5-min intervals over a 4-h period.

181	190	192	197	201	204
199	189	200	198	192	189
192	204	199	196	195	199
198	202	193	194	197	189
197	195	209	201	195	202
190	190	193	196	200	196
206	192	197	198	195	198
200	193	195	193	198	195

(a) Make an ordered stem-and-leaf plot of the data using an appropriate stem and leaf.

(b) Draw the frequency histogram. Do the data appear to be symmetrically distributed?

(c) Comparing these data with the federal standard, would you work in this factory? Justify your answer.

original one by computing $y_i = g(x_i)(i = 1, \ldots, n)$, where $g(x)$ is a function. The most frequently used functions are the *linear functions* defined by

$$y = g(x) = ax + b$$

We now give some examples.

1. To transform x, in pounds, to y, in kilograms, we set $a = 1/2.2$ and $b = 0$.
2. To transform x, in inches, to y, in centimeters, we set $a = 2.54$ and $b = 0$.
3. To transform x, in degrees Celsius, to y, in degrees Fahrenheit, we set $a = 9/5$ and $b = 32$.

We compute the mean, variance, and standard deviation for data transformed by the linear function $g(x) = ax + b$ by applying the following proposition. The proof is given in Sec. 1.7.

PROPOSITION 1.1

The sample mean, sample variance, and sample standard deviation of the transformed data $y_i = ax_i + b(i = 1, 2, \ldots, n)$ are given by

$$\bar{y} = a\bar{x} + b$$
$$s_y^2 = a^2 s_x^2$$
$$s_y = |a| s_x$$

■

---

example 1.26 Compute the sample mean, sample variance, and sample standard deviation for the girls' weights (in kilograms) in the StatLab data set (Table 1.5).

Solution. The weight y in kilograms is given by $y = x/2.2$, so $a = 1/2.2$ and $b = 0$. The sample mean, sample variance, and sample standard deviation of the original data are given by (see Example 1.25)

$$\bar{x} = 68.92, \qquad s_x^2 = 168.54, \qquad s_x = 12.92$$

Therefore, the sample mean, sample variance, and sample standard deviation of the transformed data (in kilograms) are given by

$$\bar{y} = \frac{68.92}{2.2} = 31.33, \qquad s_y^2 = \frac{168.54}{(2.2)^2} = 34.82, \qquad s_y = \frac{12.92}{2.2} = 5.90$$

■

PROBLEMS

1.32 The Analytical Methods Committee of the Royal Society of Chemistry reported the following results on the determination of tin in foodstuffs. The samples were boiled with hydrochloric acid under reflux for different times.

Refluxing time (min)	Tin found (mg/kg)					
30	55	57	59	56	56	59
75	57	55	58	59	59	59

Source: "The Determination of Tin in Organic Matter by Atomic Absorption Spectrometry," *The Analyst*, vol. 108, 1983, pp. 109–115. Used with permission.

Compute the sample means, sample variances, and sample standard deviations for the 30- and 75-min refluxing times.

two different methods. We are interested in determining whether the two methods yield similar results.

(a) Compute the sample means and sample standard deviations for each variable (x, y, d) in the data set.

(b) Compute the box plots for each variable (x, y, d) in the data set.

(c) Using the results from parts (*a*) and (*b*), answer, as best you can, the following question: Do the data indicate a difference between the two methods of measuring the nitrate concentration? Justify your answer.

1.37 The frequency distribution for the fathers' heights listed in Table 1.5 is given in the following table, where height $= x$ and freq. $= f(x)$.

Height	64.9	65	65.5	66	67	67.8	68	68.1	68.5	69	70
Freq.	1	2	1	1	1	1	1	1	1	3	3
Height	70.1	70.5	70.8	71	71.8	72	72.8	73	73.5	74	75
Freq.	1	1	1	6	1	2	1	1	2	2	2

Compute the sample mean and sample standard deviation of the heights of the fathers. [*Hint:* Use Eqs. (1.14) and (1.15).]

1.38 The frequency distribution for the girls' scores on the Raven progressive matrices test (Table 1.5) is given in the following table, where score $= x$ and freq. $= f(x)$.

Score	14	15	19	20	21	24	25	27	28	29	31	32
Freq.	1	1	1	1	1	1	3	1	1	3	1	2
Score	33	34	35	36	37	38	39	40	42	43	45	
Freq.	1	3	3	1	2	3	1	2	1	1	1	

Compute the sample mean and sample standard deviation of the girls' scores. [*Hint:* Use Eqs. (1.14) and (1.15).]

1.39 Find the sample mean, variance, and standard deviation of the bearing lifetimes from Prob. 1.18.

1.40 With reference to the failure times of the air-conditioning systems from Prob. 1.27, find the sample mean, variance, and standard deviation for:

(a) Airplane B7912

(b) Airplane B7913

(c) Airplane B7914

(d) Airplane B8045

(e) How do you explain the fact that in each case the sample mean is greater than the sample median?

1.41 Refer to the battery data of Prob. 1.5. Compute the sample mean and sample standard deviation.

1.42 Compute the sample mean, sample variance, and sample standard deviation for the capacitances listed in Prob. 1.14.