Supplementary Exercises 2.S.1–2.S.24

2.S.1 A sample of four students had the following heights (in cm): 180, 182, 179, 176. Suppose a fifth student were added to the group. How tall would that student have to be to make the mean height of the group equal 181?

2.S.2 A botanist grew 15 pepper plants on the same greenhouse bench. After 21 days, she measured the total stem length (cm) of each plant, and obtained the following values:[53]

12.4	12.2	13.4
10.9	12.2	12.1
11.8	13.5	12.0
14.1	12.7	13.2
12.6	11.9	13.1

(a) Construct a dotplot for these data, and mark the positions of the quartiles.

(b) Calculate the interquartile range.

(c) Are there any outliers? Briefly justify your answer.

2.S.3 In a behavioral study of the fruitfly *Drosophila melanogaster*, a biologist measured, for individual flies, the total time spent preening during a 6-minute observation period. The following are the preening times (sec) for 20 flies:[54]

34	24	10	16	52
76	33	31	46	24
18	26	57	32	25
48	22	48	29	19

(a) Determine the median and the quartiles.

(b) Determine the interquartile range.

(c) Construct a boxplot of the data.

2.S.4 To calibrate a standard curve for assaying protein concentrations, a plant pathologist used a spectrophotometer to measure the absorbance of light (wavelength 500 nm) by a protein solution. The results of 27 replicate assays of a standard solution containing 60 μg protein per ml water were as follows:[55]

0.111	0.115	0.115	0.110	0.099
0.121	0.107	0.107	0.100	0.110
0.106	0.116	0.098	0.116	0.108
0.098	0.120	0.123	0.124	0.122
0.116	0.130	0.114	0.100	0.123
0.119	0.107			

Construct a frequency distribution and display it as a table and as a histogram.

2.S.5 Refer to the absorbance data of Exercise 2.S.4.

(a) Determine the median, the quartiles, and the interquartile range.

(b) How large must an observation be to be an outlier?

2.S.6 The midrange is defined as the average of the minimum and maximum of a distribution. Is the midrange a robust statistic? Why or why not?

2.S.7 Twenty patients with severe epilepsy were observed for 8 weeks. The following are the numbers of major seizures suffered by each patient during the observation period:[56]

5	0	9	6	0	0	5	0	6	1
5	0	0	0	0	7	0	0	4	7

(a) Determine the median number of seizures.

(b) Determine the mean number of seizures.

(c) Construct a histogram of the data. Mark the positions of the mean and the median on the histogram.

(d) What feature of the frequency distribution suggests that neither the mean nor the median is a meaningful summary of the experience of these patients?

2.S.8 Consider the histogram from Exercise 2.3.13. By "reading" the histogram, estimate the percentage of observations that are less than 45. Is this percentage closest to 10%, 30%, 50%, 70%, 90%? (*Note*: The frequency scale is not given for this histogram, because there is no need to calculate the number of observations in each class. Rather, the percentage of observations that are less than 45 can be estimated by looking at area.)

2.S.9 Consider the histogram from Exercise 2.3.15. By "reading" the histogram, estimate the percentage of observations that are greater than 25. Is this percentage closest to 10%, 30%, 50%, 70%, 90%? (*Note*: The frequency scale is not given for this histogram, because there is no need to calculate the number of observations in each class. Rather, the percentage of observations that are greater than 25 can be estimated by looking at area.)

2.S.10 Calculate the SD of each of the following fictitious samples:

(a) 11, 8, 4, 10, 7

(b) 23, 29, 24, 21, 23

(c) 6, 0, −3, 2, 5

2.S.11 To study the spatial distribution of Japanese beetle larvae in the soil, researchers divided a 12- × 12-foot section of a cornfield into 144 one-foot squares. They counted the number of larvae Y in each square, with the results shown in the following table.[57]

Number of larvae	Frequency (Number of squares)
0	13
1	34
2	50
3	18
4	16
5	10
6	2
7	1
Total	144

(a) The mean and SD of Y are $\bar{y} = 2.23$ and $s = 1.47$. What percentage of the observations are within
  (i) 1 SD of the mean?
  (ii) 2 SDs of the mean?

(b) Determine the total number of larvae in all 144 squares. How is this number related to $\bar{y}$?

(c) Determine the median value of the distribution.

2.S.12 One measure of physical fitness is maximal oxygen uptake, which is the maximum rate at which a person can consume oxygen. A treadmill test was used to determine the maximal oxygen uptake of nine college women before and after participation in a 10-week program of vigorous exercise. The accompanying table shows the before and after measurements and the change (after–before); all values are in ml O_2 per mm per kg body weight.[58]

	Maximal oxygen uptake		
Participant	Before	After	Change
1	48.6	38.8	−9.8
2	38.0	40.7	2.7
3	31.2	32.0	0.8
4	45.5	45.4	−0.1
5	41.7	43.2	1.5
6	41.8	45.3	3.5
7	37.9	38.9	1.0
8	39.2	43.5	4.3
9	47.2	45.0	−2.2

The following computations are to be done on the *change* in maximal oxygen uptake (the right-hand column).

(a) Calculate the mean and the SD.

(b) Determine the median.

(c) Eliminate participant 1 from the data and repeat parts (a) and (b). Which of the descriptive measures display robustness and which do not?

2.S.13 A veterinary anatomist investigated the spatial arrangement of the nerve cells in the intestine of a pony. He removed a block of tissue from the intestinal wall, cut the block into many equal sections, and counted the number of nerve cells in each of 23 randomly selected sections. The counts were as follows.[59]

35 19 33 34 17 26 16 40
28 30 23 12 27 33 22 31
28 28 35 23 23 19 29

(a) Determine the median, the quartiles, and the interquartile range.

(b) Construct a boxplot of the data.

2.S.14 Exercise 2.S.13 asks for a boxplot of the nerve-cell data. Does this graphic support the claim that the data came from a reasonably symmetric distribution?

2.S.15 A geneticist counted the number of bristles on a certain region of the abdomen of the fruitfly *Drosophila melanogaster*. The results for 119 individuals were as shown in the table.[60]

Number of bristles	Number of flies	Number of bristles	Number of flies
29	1	38	18
30	0	39	13
31	1	40	10
32	2	41	15
33	2	42	10
34	6	43	2
35	9	44	2
36	11	45	3
37	12	46	2

(a) Find the median number of bristles.

(b) Find the first and third quartiles of the sample.

(c) Make a boxplot of the data.

(d) The sample mean is 38.45 and the SD is 3.20. What percentage of the observations fall within 1 SD of the mean?

(e) What percentage of the observations fall between the quartiles identified in part (b)?

2.S.16 The carbon monoxide in cigarettes is thought to be hazardous to the fetus of a pregnant woman who smokes. In a study of this hypothesis, blood was drawn from pregnant women before and after smoking a cigarette. Measurements were made of the percent of blood hemoglobin bound to carbon monoxide as carboxyhemoglobin (COHb). The results for 10 women are shown in the table.[61]

Subject	Blood COHb (%)		
	Before	After	Increase
1	1.2	7.6	6.4
2	1.4	4.0	2.6
3	1.5	5.0	3.5
4	2.4	6.3	3.9
5	3.6	5.8	2.2
6	0.5	6.0	5.5
7	2.0	6.4	4.4
8	1.5	5.0	3.5
9	1.0	4.2	3.2
10	1.7	5.2	3.5

(a) Calculate the mean and SD of the *increase* in COHb.

(b) Calculate the mean COHb before and the mean after. Is the mean increase equal to the increase in means?

(c) Determine the median increase in COHb.

(d) Repeat part (c) for the before measurements and for the after measurements. Is the median increase equal to the increase in medians?

2.S.17 (Computer problem) A medical researcher in India obtained blood specimens from 31 young children, all of whom were infected with malaria. The following data, listed in increasing order, are the numbers of malarial parasites found in 1 ml of blood from each child.[62]

100	140	140	271	400	435	455	770
826	1,400	1,540	1,640	1,920	2,280	2,340	3,672
4,914	6,160	6,560	6,741	7,609	8,547	9,560	10,516
14,960	16,855	18,600	22,995	29,800	83,200	134,232	

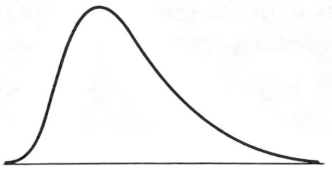
(a) Construct a frequency distribution of the data, using a class width of 10,000; display the distribution as a histogram.

(b) Transform the data by taking the logarithm (base 10) of each observation. Construct a frequency distribution of the transformed data and display it as a histogram. How does the log transformation affect the shape of the frequency distribution?

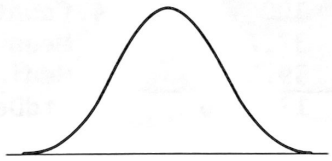
(c) Determine the mean of the original data and the mean of the log-transformed data. Is the mean of the logs equal to the log of the mean?

(d) Determine the median of the original data and the median of the log-transformed data. Is the median of the logs equal to the log of the median?

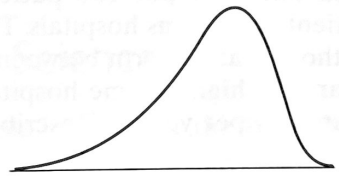
2.S.18 Rainfall, measured in inches, for the month of June in Cleveland, Ohio, was recorded for each of 41 years.[63] The values had a minimum of 1.2, an average of 3.6, and an SD of 1.6. Which of the following is a rough histogram for the data? How do you know?



I

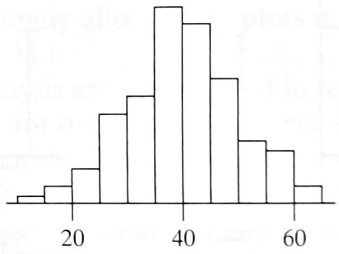


II

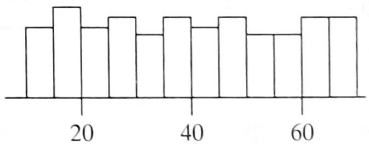


III

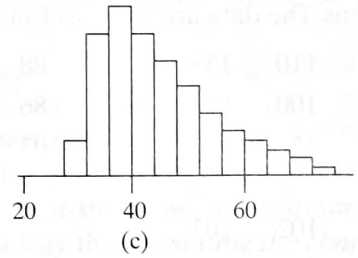
2.S.19 The following histograms (a), (b), and (c) show three distributions.



(a)



(b)



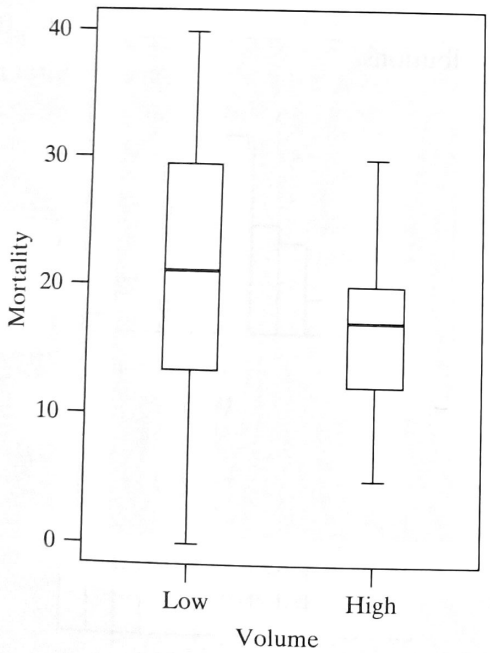
(c)

The accompanying computer output shows the mean, median, and SD of the three distributions, plus the mean, median, and SD for a fourth distribution. Match

the histograms with the statistics. Explain your reasoning. (One set of statistics will not be used.)

1. Count	100	2. Count	100
Mean	41.3522	Mean	39.6761
Median	39.5585	Median	39.5377
StdDev	13.0136	StdDev	10.0476
3. Count	100	4. Count	100
Mean	37.7522	Mean	39.6493
Median	39.5585	Median	39.5448
StdDev	13.0136	StdDev	17.5126

2.S.20 The following boxplots show mortality rates (deaths within one year per 100 patients) for heart transplant patients at various hospitals. The low-volume hospitals are those that perform between 5 and 9 transplants per year. The high-volume hospitals perform 10 or more transplants per year.[64] Describe the distributions, paying special attention to how they compare to one another. Be sure to note the shape, center, and spread of each distribution.

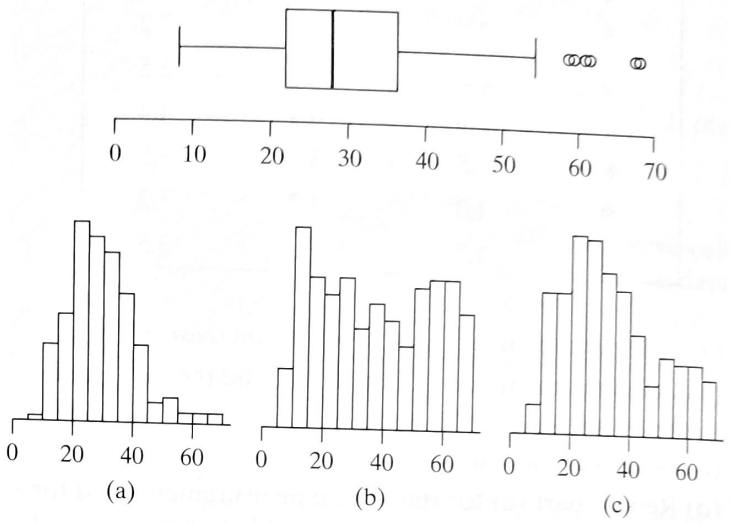


2.S.21 (Computer problem) Physicians measured the concentration of calcium (nM) in blood samples from 38 healthy persons. The data are listed as follows:[65]

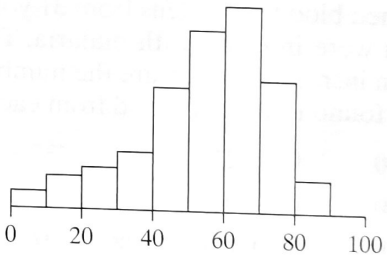
95	110	135	120	88	125
112	100	130	107	86	130
122	122	127	107	107	107
88	126	125	112	78	115
78	102	103	93	88	110
104	122	112	80	121	126
90	96				

Calculate appropriate measures of the center and spread of the distribution. Describe the shape of the distribution and any unusual features in the data.

2.S.22 The following boxplot shows the same data that are shown in one of the three histograms. Which histogram goes with the boxplot? Explain your answer.



2.S.23 Here is a histogram.



Explain why the mean is less than the median of the distribution.

2.S.24 Consider the histogram from Exercise 2.S.23. By "reading" the histogram, estimate the percentage of observations that are less than 40. Is this percentage closest to 10%, 20%, 50%, 80%, or 90%? *Note:* The frequency scale is not given for this histogram, because there is no need to calculate the number of observations in each class. Rather, the percentage of observations that are less than 40 can be estimated by looking at area.