

Solutions Manual

to accompany

STATISTICS FOR ENGINEERS AND SCIENTISTS

by

William Navidi

Table of Contents

Chapter 1	1
Chapter 2	13
Chapter 3	53
Chapter 4	72
Chapter 5	115
Chapter 6	133
Chapter 7	168
Chapter 8	190
Chapter 9	209
Chapter 10	229

Chapter 1

Section 1.1

1. (a) The population consists of all the bolts in the shipment. It is tangible.
(b) The population consists of all measurements that could be made on that resistor with that ohmmeter. It is conceptual.
(c) The population consists of all residents of the town. It is tangible.
(d) The population consists of all welds that could be made by that process. It is conceptual.
(e) The population consists of all parts manufactured that day. It is tangible.
3. (a) False
(b) True
5. (a) No. What is important is the population proportion of defectives; the sample proportion is only an approximation. The population proportion for the new process may in fact be greater or less than that of the old process.
(b) No. The population proportion for the new process may be 0.10 or more, even though the sample proportion was only 0.09.
(c) Finding 2 defective bottles in the sample.
7. A good knowledge of the process that generated the data.

Section 1.2

1. False
3. No. In the sample 1, 2, 4 the mean is $7/3$, which does not appear at all.
5. The sample size can be any odd number.
7. Yes. If all the numbers in the list are the same, the standard deviation will equal 0.
9. The mean and standard deviation both increase by 5%.
11. The total number of points scored in the class of 30 students is $30 \times 72 = 2160$. The total number of points scored in the class of 40 students is $40 \times 79 = 3160$. The total number of points scored in both classes combined is $2160 + 3160 = 5320$. There are $30 + 40 = 70$ students in both classes combined. Therefore the mean score for the two classes combined is $5320/70 = 76$.
13. (a) All would be multiplied by 2.54.

(b) Not exactly the same, because the measurements would be a little different the second time.
15. (a) The sample size is $n = 16$. The tertiles have cutpoints $(1/3)(17) = 5.67$ and $(2/3)(17) = 11.33$. The first tertile is therefore the average of the sample values in positions 5 and 6, which is $(44 + 46)/2 = 45$. The second tertile is the average of the sample values in positions 11 and 12, which is $(76 + 79)/2 = 77.5$.

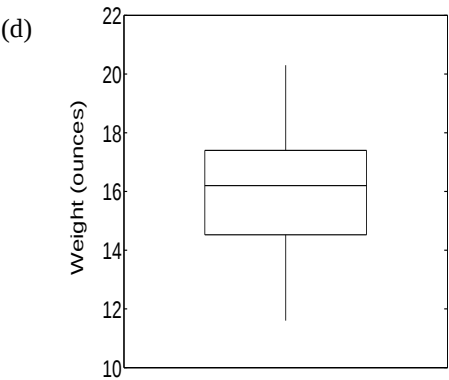
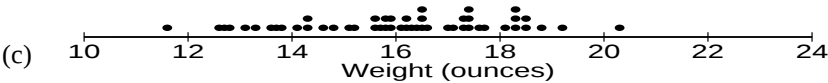
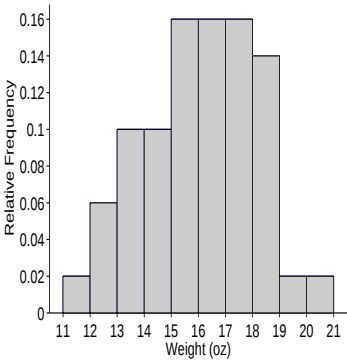
(b) The sample size is $n = 16$. The quintiles have cutpoints $(i/5)(17)$ for $i = 1, 2, 3, 4$. The quintiles are therefore the averages of the sample values in positions 3 and 4, in positions 6 and 7, in positions 10 and 11, and in positions 13 and 14. The quintiles are therefore $(23 + 41)/2 = 32$, $(46 + 49)/2 = 47.5$, $(74 + 76)/2 = 75$, and $(82 + 89)/2 = 85.5$.

Section 1.3

1. (a)

Stem	Leaf
11	6
12	678
13	13678
14	13368
15	126678899
16	122345556
17	013344467
18	1333558
19	2
20	3

(b) Here is one histogram. Other choices for the endpoints are possible.



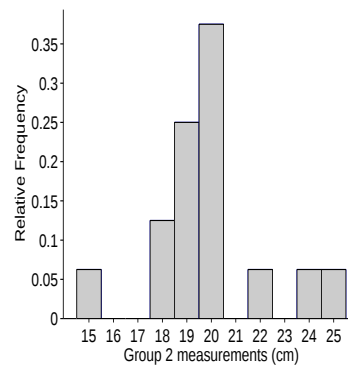
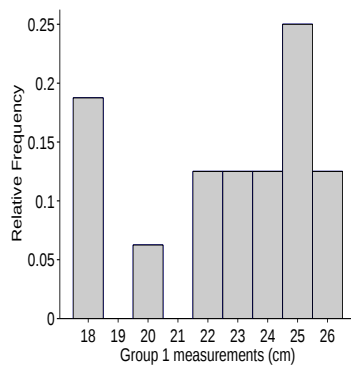
The boxplot shows no outliers.

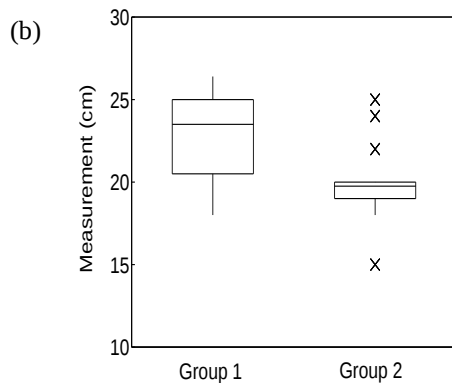
3.

Stem	Leaf
1	1588
2	00003468
3	0234588
4	0346
5	2235666689
6	00233459
7	113558
8	568
9	1225
10	1
11	
12	2
13	06
14	
15	
16	
17	1
18	6
19	9
20	
21	
22	
23	3

There are 23 stems in this plot. An advantage of this plot over the one in Figure 1.6 is that the values are given to the tenths digit instead of to the ones digit. A disadvantage is that there are too many stems, and many of them are empty.

5. (a) Here are histograms for each group. Other choices for the endpoints are possible.



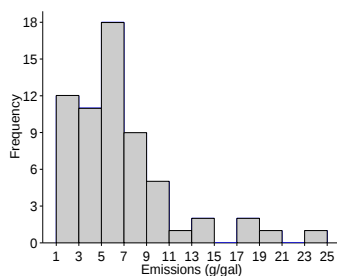


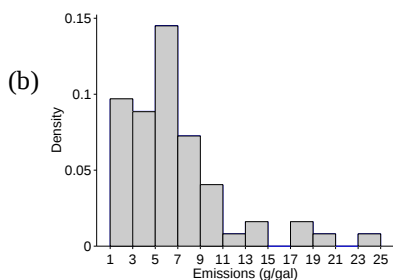
- (c) The measurements in Group 1 are generally larger than those in Group 2. The measurements in Group 1 are not far from symmetric, although the boxplot suggests a slight skew to the left since the median is closer to the third quartile than the first. There are no outliers. Most of the measurements for Group 2 are highly concentrated in a narrow range, and skewed to the left within that range. The remaining four measurements are outliers.

7. (a) The proportion is the sum of the relative frequencies (heights) of the rectangles above 130. This sum is approximately $0.12 + 0.045 + 0.045 + 0.02 + 0.005 + 0.005 = 0.24$. This is closest to 25%.

- (b) The height of the rectangle over the interval 130–135 is greater than the sum of the heights of the rectangles over the interval 140–150. Therefore there are more women in the interval 130–135 mm.

9. (a)





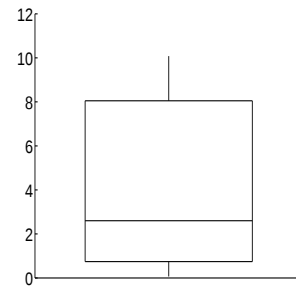
(c) Yes, the shapes of the histograms are the same.

11. Any point more than 1.5 IQR (interquartile range) below the first quartile or above the third quartile is labeled an outlier. To find the IQR, arrange the values in order: 4, 10, 20, 25, 31, 36, 37, 41, 44, 68, 82. There are $n = 11$ values. The first quartile is the value in position $0.25(n + 1) = 3$, which is 20. The third quartile is the value in position $0.75(n + 1) = 9$, which is 44. The interquartile range is $44 - 20 = 24$. So 1.5 IQR is equal to $(1.5)(24) = 36$. There are no points less than $20 - 36 = -16$, so there are no outliers on the low side. There is one point, 82, that is greater than $44 + 36 = 80$. Therefore 82 is the only outlier.
13. The figure on the left is a sketch of separate histograms for each group. The histogram on the right is a sketch of a histogram for the two groups combined. There is more spread in the combined histogram than in either of the separate ones. Therefore the standard deviation of all 200 heights is greater than 2.5 in. The answer is (ii).



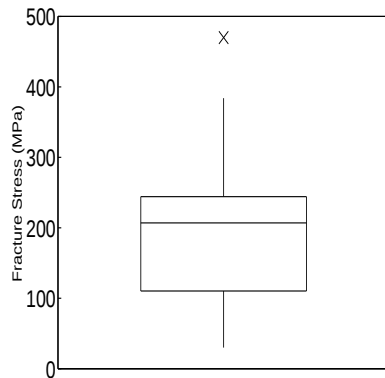
15. (a) $\text{IQR} = 3\text{rd quartile} - 1\text{st quartile}$. A: $\text{IQR} = 6.02 - 1.42 = 4.60$, B: $\text{IQR} = 9.13 - 5.27 = 3.86$

- (b) Yes, since the minimum is within 1.5 IQR of the first quartile and the maximum is within 1.5 IQR of the third quartile, there are no outliers, and the given numbers specify the boundaries of the box and the ends of the whiskers.

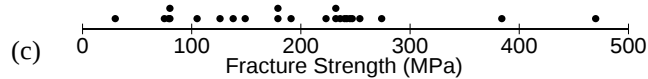


- (c) No. The minimum value of -2.235 is an “outlier,” since it is more than 1.5 times the interquartile range below the first quartile. The lower whisker should extend to the smallest point that is not an outlier, but the value of this point is not given.

17. (a)

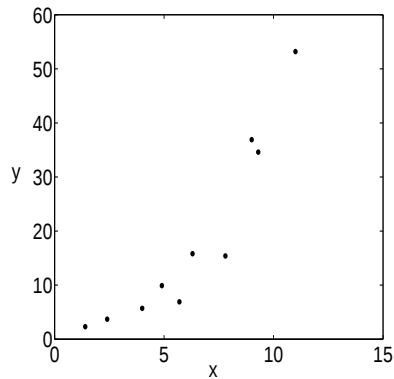


- (b) The boxplot indicates that the value 470 is an outlier.



- (d) The dotplot indicates that the value 384 is detached from the bulk of the data, and thus could be considered to be an outlier.

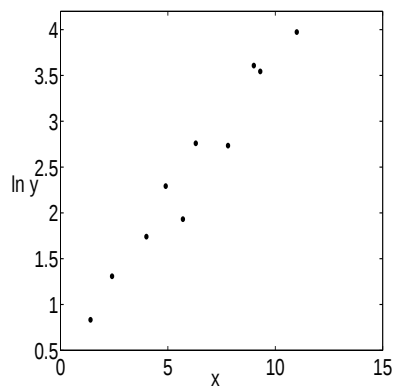
19. (a)



The relationship is non-linear.

(b)

x	1.4	2.4	4.0	4.9	5.7	6.3	7.8	9.0	9.3	11.0
$\ln y$	0.83	1.31	1.74	2.29	1.93	2.76	2.73	3.61	3.54	3.97



The relationship is approximately linear.

(c) It would be easier to work with x and $\ln y$, because the relationship is approximately linear.

Supplementary Exercises for Chapter 1

1. (a) The mean will be divided by 2.2.

(b) The standard deviation will be divided by 2.2.

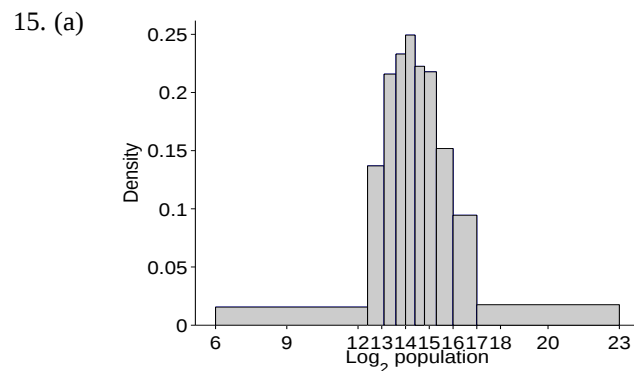
3. (a) False. The true percentage could be greater than 5%, with the observation of 4 out of 100 due to sampling variation.

-
- (b) True
- (c) False. If the result differs greatly from 5%, it is unlikely to be due to sampling variation.
- (d) True. If the result differs greatly from 5%, it is unlikely to be due to sampling variation.
5. (a) It is not possible to tell by how much the mean changes, because the sample size is not known.
- (b) If there are more than two numbers on the list, the median is unchanged. If there are only two numbers on the list, the median is changed, but we cannot tell by how much.
- (c) It is not possible to tell by how much the standard deviation changes, both because the sample size is unknown and because the original standard deviation is unknown.
7. (a) The mean decreases by 0.774.
- (b) The value of the mean after the change is $25 - 0.774 = 24.226$.
- (c) The median is unchanged.
- (d) It is not possible to tell by how much the standard deviation changes, because the original standard deviation is unknown.
9. Statement (i) is true. The sample is skewed to the right.
11. (a) Incorrect, the total area is greater than 1.
- (b) Correct. The total area is equal to 1.
- (c) Incorrect. The total area is less than 1.

(d) Correct. The total area is equal to 1.

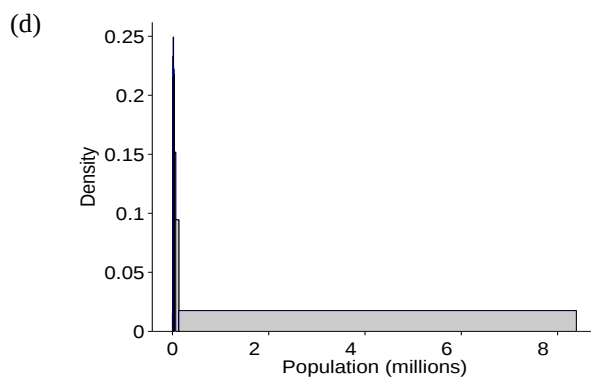
13. (a) Skewed to the left. The 85th percentile is much closer to the median (50th percentile) than the 15th percentile is. Therefore the histogram is likely to have a longer left-hand tail than right-hand tail.

(b) Skewed to the right. The 15th percentile is much closer to the median (50th percentile) than the 85th percentile is. Therefore the histogram is likely to have a longer right-hand tail than left-hand tail.



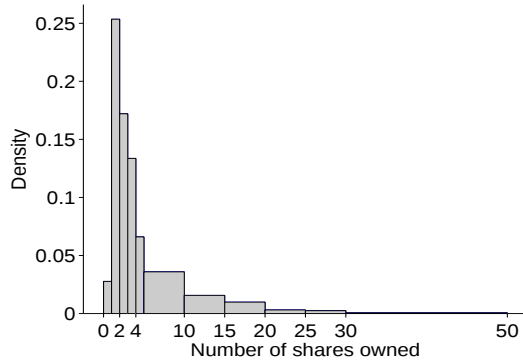
(b) 0.14

(c) Approximately symmetric



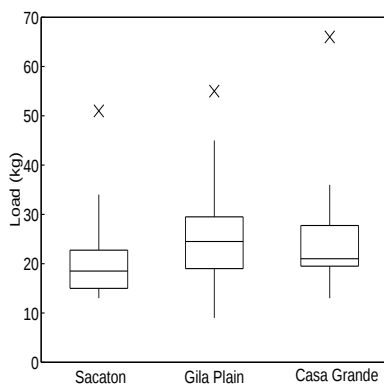
The data on the raw scale are skewed so much to the right that it is impossible to see the features of the histogram.

17. (a)



- (b) The sample size is 651, so the median is approximated by the point at which the area to the left is $0.5 = 325.5/651$. The area to the left of 3 is $295/651$, and the area to the left of 4 is $382/651$. The point at which the area to the left is $325.5/651$ is $3 + (325.5 - 295)/(382 - 295) = 3.35$.
- (c) The sample size is 651, so the first quartile is approximated by the point at which the area to the left is $0.25 = 162.75/651$. The area to the left of 1 is $18/651$, and the area to the left of 2 is $183/651$. The point at which the area to the left is $162.75/651$ is $1 + (162.75 - 18)/(183 - 18) = 1.88$.
- (d) The sample size is 651, so the third quartile is approximated by the point at which the area to the left is $0.75 = 488.25/651$. The area to the left of 5 is $425/651$, and the area to the left of 10 is $542/651$. The point at which the area to the left is $488.25/651$ is $5 + (488.25 - 425)/(542 - 425) = 7.70$.

19. (a)



(b) Each sample contains one outlier.

- (c) In the Sacaton boxplot, the median is about midway between the first and third quartiles, suggesting that the data between these quartiles are fairly symmetric. The upper whisker of the box is much longer than the lower whisker, and there is an outlier on the upper side. This indicates that the data as a whole are skewed to the right.

In the Gila Plain boxplot data, the median is about midway between the first and third quartiles, suggesting that the data between these quartiles are fairly symmetric. The upper whisker is slightly longer than the lower whisker, and there is an outlier on the upper side. This suggests that the data as a whole are somewhat skewed to the right. In the Casa Grande boxplot, the median is very close to the first quartile. This suggests that there are several values very close to each other about one-fourth of the way through the data. The two whiskers are of about equal length, which suggests that the tails are about equal, except for the outlier on the upper side.