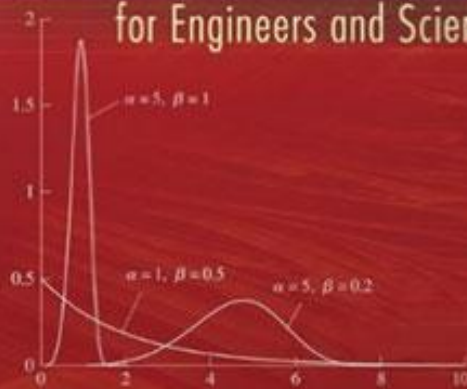


SOLUTIONS MANUAL

Third Edition

Statistics

for Engineers and Scientists



William Navidi

Solutions Manual

to accompany

STATISTICS FOR ENGINEERS AND SCIENTISTS, 3rd ed.

Prepared by
William Navidi

PROPRIETARY AND CONFIDENTIAL

This Manual is the proprietary property of The McGraw-Hill Companies, Inc. (“McGraw-Hill”) and protected by copyright and other state and federal laws. By opening and using this Manual the user agrees to the following restrictions, and if the recipient does not agree to these restrictions, the Manual should be promptly returned unopened to McGraw-Hill: **This Manual is being provided only to authorized professors and instructors for use in preparing for the classes using the affiliated textbook. No other use or distribution of this Manual is permitted. This Manual may not be sold and may not be distributed to or used by any student or other third party. No part of this Manual may be reproduced, displayed or distributed in any form or by any means, electronic or otherwise, without the prior written permission of McGraw-Hill.**

PROPRIETARY MATERIAL. © The McGraw-Hill Companies, Inc. All rights reserved. No part of this Manual may be displayed, reproduced or distributed in any form or by any means, without the prior written permission of the publisher, or used beyond the limited distribution to teachers and educators permitted by McGraw-Hill for their individual course preparation. If you are a student using this Manual, you are using it without permission.

Table of Contents

Chapter 1	1
Chapter 3	108
Chapter 4	146
Chapter 5	232
Chapter 6	269
Chapter 7	340
Chapter 8	386
Chapter 9	422
Chapter 10	464

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.

2. (iii). It is very unlikely that students whose names happen to fall at the top of a page in the phone book will differ systematically in height from the population of students as a whole. It is somewhat more likely that engineering majors will differ, and very likely that students involved with basketball intramurals will differ.

3. (a) False
(b) True

4. (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 0.12 or more, even though the sample proportion was only 0.11.
(c) Finding 2 defective circuits in the sample.

6. (a) False
- (b) True
- (c) True
7. A good knowledge of the process that generated the data.
8. (a) An observational study
- (b) It is not well-justified. Because the study is observational, there could be differences between the groups other than the level of exercise. These other differences (confounders) could cause the difference in blood pressure.
9. (a) A controlled experiment
- (b) It is well-justified, because it is based on a controlled experiment rather than an observational study.

Section 1.2

1. False
2. No. In the sample 1, 2, 4 the mean is $7/3$, which does not appear at all.
3. No. In the sample 1, 2, 4 the mean is $7/3$, which does not appear at all.
4. No. The median of the sample 1, 2, 4, 5 is 3.
5. The sample size can be any odd number.

6. Yes. For example, the list 1, 2, 12 has an average of 5 and a standard deviation of 6.08.
7. Yes. If all the numbers in the list are the same, the standard deviation will equal 0.
8. The mean increases by \$50; the standard deviation is unchanged.
9. The mean and standard deviation both increase by 5%.

10. (a) Let X_1, \dots, X_{100} denote the 100 numbers of children.

$$\sum_{i=1}^{100} X_i = 27(0) + 22(1) + 30(2) + 12(3) + 7(4) + 2(5) = 156$$

$$\bar{X} = \frac{\sum_{i=1}^{100} X_i}{100} = \frac{156}{100} = 1.56$$

(b) The sample variance is

$$\begin{aligned} s^2 &= \frac{1}{99} \left(\sum_{i=1}^{100} X_i^2 - 100\bar{X}^2 \right) \\ &= \frac{1}{99} [(27)0^2 + (22)1^2 + (30)2^2 + (12)3^2 + (7)4^2 + (2)5^2 - 100(1.56^2)] \\ &= 1.7034 \end{aligned}$$

The standard deviation is $s = \sqrt{s^2} = 1.3052$.

Alternatively, the sample variance can be computed as

$$\begin{aligned} s^2 &= \frac{1}{99} \sum_{i=1}^{100} (X_i - \bar{X})^2 \\ &= \frac{1}{99} [27(0 - 1.56)^2 + 22(1 - 1.56)^2 + 30(2 - 1.56)^2 + 12(3 - 1.56)^2 + 7(4 - 1.56)^2 + 2(5 - 1.56)^2] \\ &= 1.7034 \end{aligned}$$

(c) The sample median is the average of the 50th and 51st value when arranged in order. Both these values are equal to 2, so the median is 2.

- (d) The first quartile is the average of the 25th and 26th value when arranged in order. Both these values are equal to 0, so the first quartile is 0.
- (e) Of the 100 women, $30 + 12 + 7 + 2 = 51$ had more than the mean of 1.56 children, so the proportion is $51/100 = 0.51$.
- (f) The quantity that is one standard deviation greater than the mean is $1.56 + 1.3052 = 2.8652$. Of the 100 women, $12 + 7 + 2 = 21$ had more than 2.8652 children, so the proportion is $21/100 = 0.21$.
- (g) The region within one standard deviation of the mean is $1.56 \pm 1.3052 = (0.2548, 2.8652)$. Of the 100 women, $22 + 30 = 52$ are in this range, so the proportion is $52/100 = 0.52$.
11. The total height of the 20 men is $20 \times 178 = 3560$. The total height of the 30 women is $30 \times 164 = 4920$. The total height of all 50 people is $3560 + 4920 = 8480$. There are $20 + 30 = 50$ people in total. Therefore the mean height for both groups put together is $8480/50 = 169.6$ cm.

12. (a) The mean for A is

$$(18.0 + 18.0 + 18.0 + 20.0 + 22.0 + 22.0 + 22.5 + 23.0 + 24.0 + 24.0 + 25.0 + 25.0 + 25.0 + 25.0 + 26.0 + 26.4)/16 = 22.744$$

The mean for B is

$$(18.8 + 18.9 + 18.9 + 19.6 + 20.1 + 20.4 + 20.4 + 20.4 + 20.4 + 20.5 + 21.2 + 22.0 + 22.0 + 22.0 + 22.0 + 23.6)/16 = 20.700$$

The mean for C is

$$(20.2 + 20.5 + 20.5 + 20.7 + 20.8 + 20.9 + 21.0 + 21.0 + 21.0 + 21.0 + 21.0 + 21.5 + 21.5 + 21.5 + 21.5 + 21.6)/16 = 20.013$$

The mean for D is

$$(20.0 + 20.0 + 20.0 + 20.0 + 20.2 + 20.5 + 20.5 + 20.7 + 20.7 + 20.7 + 21.0 + 21.1 + 21.5 + 21.6 + 22.1 + 22.3)/16 = 20.806$$

- (b) The median for A is $(23.0 + 24.0)/2 = 23.5$. The median for B is $(20.4 + 20.4)/2 = 20.4$. The median for C is $(21.0 + 21.0)/2 = 21.0$. The median for D is $(20.7 + 20.7)/2 = 20.7$.

(c) $0.20(16) = 3.2 \approx 3$. Trim the 3 highest and 3 lowest observations.

The 20% trimmed mean for A is

$$(20.0 + 22.0 + 22.0 + 22.5 + 23.0 + 24.0 + 24.0 + 25.0 + 25.0 + 25.0)/10 = 23.25$$

The 20% trimmed mean for B is

$$(19.6 + 20.1 + 20.4 + 20.4 + 20.4 + 20.5 + 21.2 + 22.0 + 22.0)/10 = 20.70$$

The 20% trimmed mean for C is

$$(20.7 + 20.8 + 20.9 + 21.0 + 21.0 + 21.0 + 21.0 + 21.0 + 21.5 + 21.5)/10 = 21.04$$

The 20% trimmed mean for D is

$$(20.0 + 20.2 + 20.5 + 20.5 + 20.7 + 20.7 + 20.7 + 21.0 + 21.1 + 21.5)/10 = 20.69$$

(d) $0.25(17) = 4.25$. Therefore the first quartile is the average of the numbers in positions 4 and 5. $0.75(17) = 12.75$. Therefore the third quartile is the average of the numbers in positions 12 and 13.

A: $Q_1 = 21.0$, $Q_3 = 25.0$; B: $Q_1 = 19.85$, $Q_3 = 22.0$; C: $Q_1 = 20.75$, $Q_3 = 21.5$; D: $Q_1 = 20.1$, $Q_3 = 21.3$

(e) The variance for A is

$$s^2 = \frac{1}{15} [18.0^2 + 18.0^2 + 18.0^2 + 20.0^2 + 22.0^2 + 22.0^2 + 22.5^2 + 23.0^2 + 24.0^2 + 24.0^2 + 25.0^2 + 25.0^2 + 25.0^2 + 26.0^2 + 26.4^2 - 16(22.744^2)] = 8.2506$$

The standard deviation for A is $s = \sqrt{8.2506} = 2.8724$.

The variance for B is

$$s^2 = \frac{1}{15} [18.8^2 + 18.9^2 + 18.9^2 + 19.6^2 + 20.1^2 + 20.4^2 + 20.4^2 + 20.4^2 + 20.4^2 + 20.5^2 + 21.2^2 + 22.0^2 + 22.0^2 + 22.0^2 + 22.0^2 + 23.6^2 - 16(20.700^2)] = 1.8320$$

The standard deviation for B is $s = \sqrt{1.8320} = 1.3535$.

The variance for C is

$$s^2 = \frac{1}{15} [20.2^2 + 20.5^2 + 20.5^2 + 20.7^2 + 20.8^2 + 20.9^2 + 21.0^2 + 21.0^2 + 21.0^2 + 21.0^2 + 21.0^2 + 21.5^2 + 21.5^2 + 21.5^2 + 21.5^2 + 21.6^2 - 16(20.013^2)] = 0.17583$$

The standard deviation for C is $s = \sqrt{0.17583} = 0.4193$.

The variance for D is

$$s^2 = \frac{1}{15} [20.0^2 + 20.0^2 + 20.0^2 + 20.0^2 + 20.2^2 + 20.5^2 + 20.5^2 + 20.7^2 + 20.7^2 + 20.7^2 + 21.0^2 + 21.1^2 + 21.5^2 + 21.6^2 + 22.1^2 + 22.3^2 - 16(20.806^2)] = 0.55529$$

The standard deviation for D is $s = \sqrt{0.55529} = 0.7542$.

- (f) Method A has the largest standard deviation. This could be expected, because of the four methods, this is the crudest. Therefore we could expect to see more variation in the way in which this method is carried out, resulting in more spread in the results.
- (g) Other things being equal, a smaller standard deviation is better. With any measurement method, the result is somewhat different each time a measurement is made. When the standard deviation is small, a single measurement is more valuable, since we know that subsequent measurements would probably not be much different.
13. (a) All would be divided by 2.54.
- (b) Not exactly the same, because the measurements would be a little different the second time.
14. (a) We will work in units of \$1000. Let S_0 be the sum of the original 10 numbers and let S_1 be the sum after the change. Then $S_0/10 = 70$, so $S_0 = 700$. Now $S_1 = S_0 - 100 + 1000 = 1600$, so the new mean is $S_1/10 = 160$.
- (b) The median is unchanged at 55.
- (c) Let X_1, \dots, X_{10} be the original 10 numbers. Let $T_0 = \sum_{i=1}^{10} X_i^2$. Then the variance is $(1/9)[T_0 - 10(70^2)] = 20^2 = 400$, so $T_0 = 52,600$. Let T_1 be the sum of the squares after the change. Then $T_1 = T_0 - 100^2 + 1000^2 = 1,042,600$. The new standard deviation is $\sqrt{(1/9)[T_1 - 10(160^2)]} = 295.63$.
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 $(82 + 89)/2 = 85.5$.

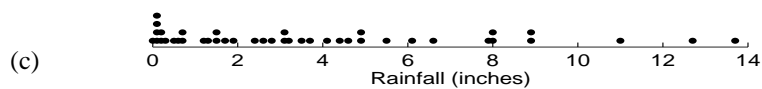
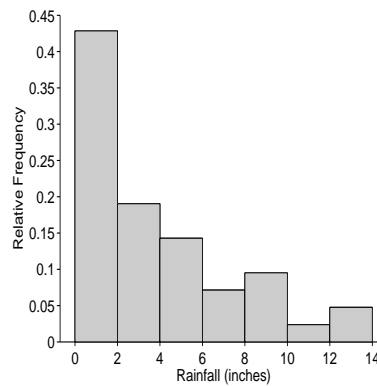
16. (a) Seems certain to be an error.

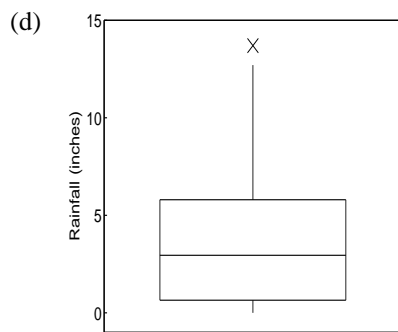
(b) Could be correct.

Section 1.3

	Stem	Leaf
	0	011112235677
1. (a)	1	235579
	2	468
	3	11257
	4	14699
	5	5
	6	16
	7	9
	8	0099
	9	
	10	
	11	0
	12	7
	13	7

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



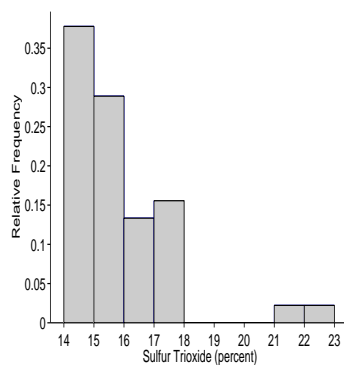


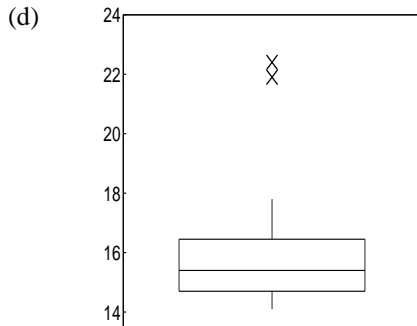
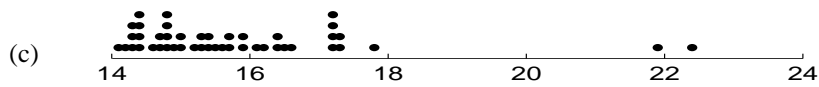
The boxplot shows one outlier.

2. (a)

Stem	Leaf
14	123334444467788889
15	0023344567799
16	124456
17	2222338
18	
19	
20	
21	9
22	4

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





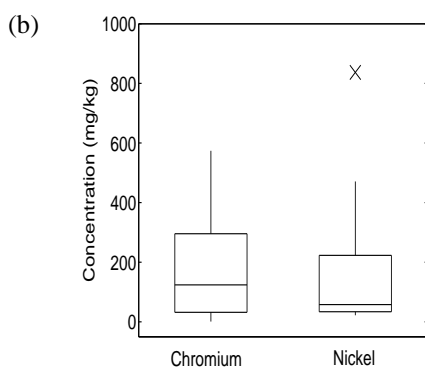
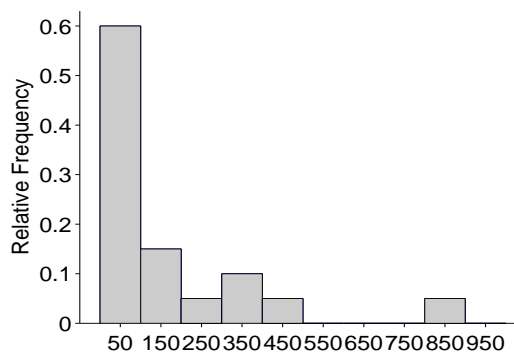
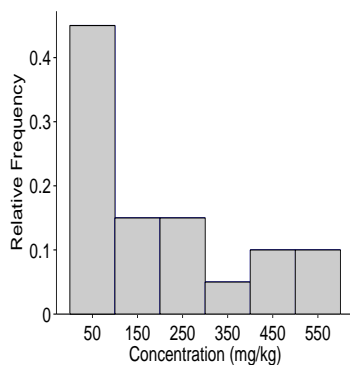
The boxplot shows 2 outliers.

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

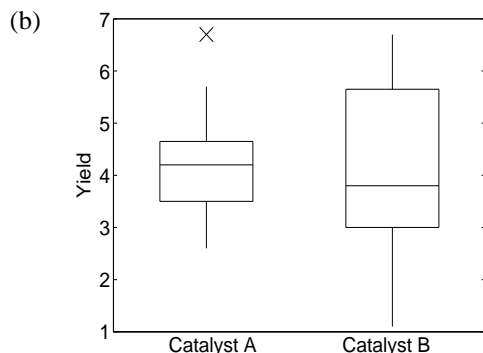
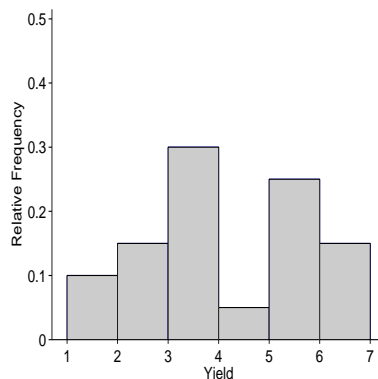
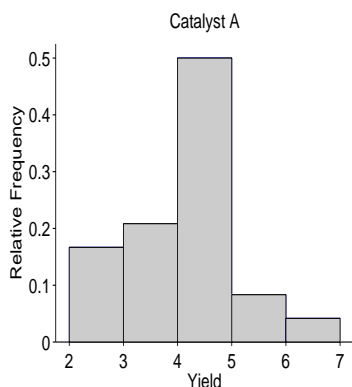
There are 23 stems in this plot. An advantage of this plot over the one in Figure 1.6 is that the values are given to the tenths digit instead of to the ones digit. A disadvantage is that there are too many stems, and many of them are empty.

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



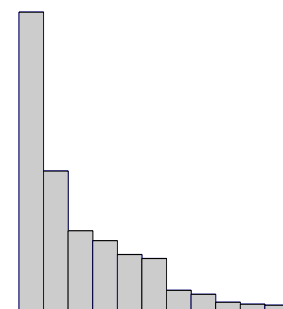
- (c) The concentrations of nickel are on the whole lower than the concentrations of chromium. The nickel concentrations are highly skewed to the right, which can be seen from the median being much closer to the first quartile than the third. The chromium concentrations are somewhat less skewed. Finally, the nickel concentrations include an outlier.

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

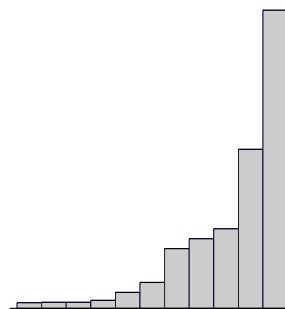


(c) The yields for catalyst B are considerably more spread out than those for catalyst A. The median yield for catalyst A is greater than the median for catalyst B. The median yield for B is closer to the first quartile than the third, but the lower whisker is longer than the upper one, so the median is approximately equidistant from the extremes of the data. Thus the yields for catalyst B are approximately symmetric. The largest yield for catalyst A is an outlier; the remaining yields for catalyst A are approximately symmetric.

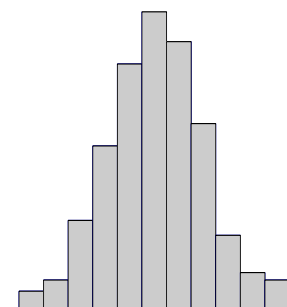
6. (a) The histogram should be skewed to the right. Here is an example.



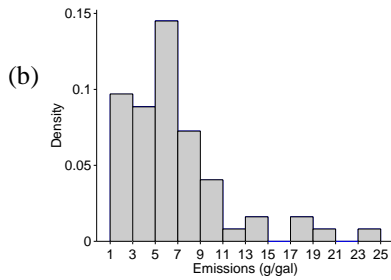
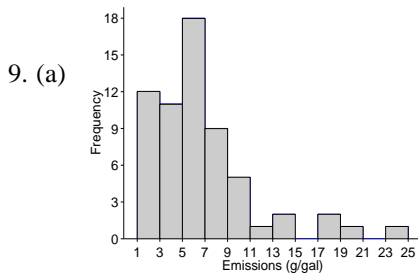
(b) The histogram should be skewed to the left. Here is an example.



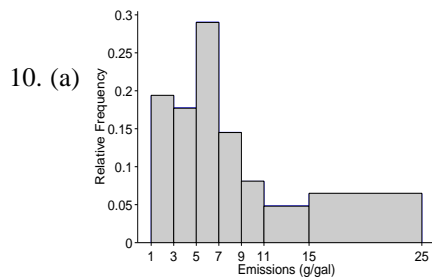
(c) The histogram should be approximately symmetric. Here is an example.



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.
8. The relative frequencies of the rectangles shown are 0.2, 0.3, 0.15, 0.1, and 0.1. The sum of these relative frequencies is 0.85. Since the sum of all the relative frequencies must be 1, the missing rectangle has a height of 0.15.



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

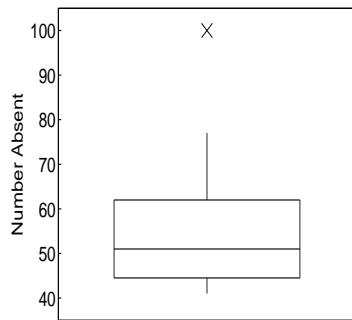


(b) No

(c) The class interval widths are unequal.

(d) The classes 11–<15 and 15–<25

11. (a)



(b) Yes. The value 100 is an outlier.

12. The mean, the median, and the first and third quartiles are indicated directly on a boxplot, and the interquartile range can be computed as the difference between the first and third quartiles.

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).



14. (a) False

(b) True

(c) False

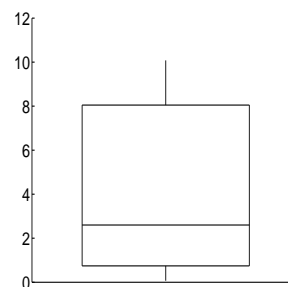
(d) False

(e) True

(f) False

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.



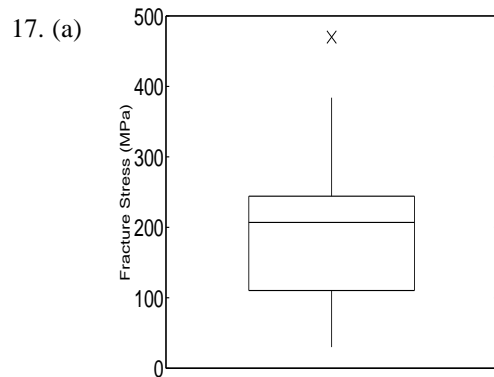
(c) No. The minimum value of -2.235 is an “outlier,” since it is more than 1.5 times the interquartile range below the first quartile. The lower whisker should extend to the smallest point that is not an outlier, but the value of this point is not given.

16. (a) (4)

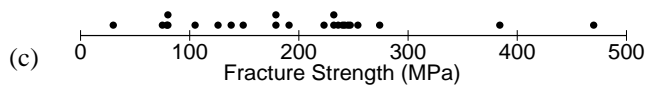
(b) (2)

(c) (1)

(d) (3)



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



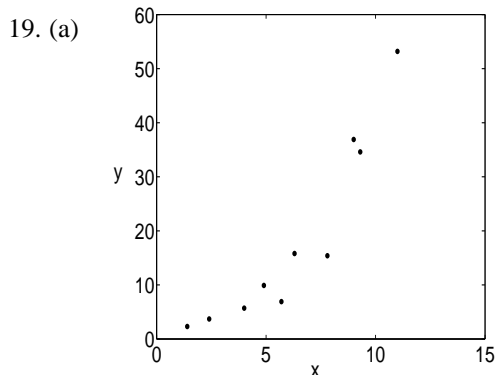
(d) The dotplot indicates that the value 384 is detached from the bulk of the data, and thus could be considered to be an outlier.

18. (a) iii

(b) i

(c) iv

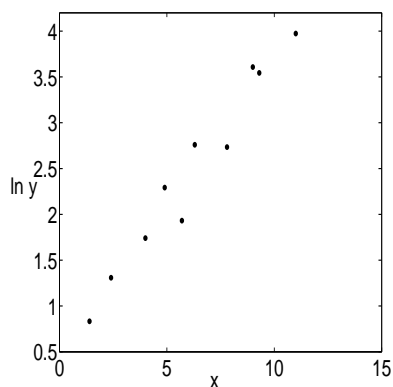
(d) ii



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.

2. (a) The mean will increase by 50 g.
- (b) The standard deviation will be unchanged.
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.
4. (a) No. This could well be sampling variation.
- (b) Yes. It is virtually impossible for sampling variation to be this large.
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.
6. (a) The sum of the numbers decreases by $12.9 - 1.29 = 11.61$, so the mean decreases by $11.61/15 = 0.774$.