

Solutions Manual

to accompany

PRINCIPLES OF STATISTICS FOR ENGINEERS AND SCIENTISTS 2nd ed.

**Prepared by
William Navidi**

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 2	21
Chapter 3	40
Chapter 4	69
Chapter 5	118
Chapter 6	141
Chapter 7	176
Chapter 8	207
Chapter 9	254
Chapter 10	284

Chapter 1

Section 1.1

1. (a) The population consists of all the times the process could be run. It is conceptual.
(b) The population consists of all the registered voters in the state. It is tangible.
(c) The population consists of all people with high cholesterol levels. It is tangible.
(d) The population consists of all concrete specimens that could be made from the new formulation. It is conceptual.
(e) The population consists of all bolts 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. (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. False

4. No. In the sample 1, 2, 4 the mean is $7/3$, which does not appear at all.

5. No. In the sample 1, 2, 4 the mean is $7/3$, which does not appear at all.
6. No. The median of the sample 1, 2, 4, 5 is 3.
7. The sample size can be any odd number.
8. Yes. For example, the list 1, 2, 12 has an average of 5 and a standard deviation of 6.08.
9. Yes. If all the numbers in the list are the same, the standard deviation will equal 0.
10. (a) Let X_1, \dots, X_{100} denote the 100 numbers of occupants.

$$\sum_{i=1}^{100} X_i = 70(1) + 15(2) + 10(3) + 3(4) + 2(5) = 152$$

$$\bar{X} = \frac{\sum_{i=1}^{100} X_i}{100} = \frac{152}{100} = 1.52$$

(b) The sample variance is

$$\begin{aligned} s^2 &= \frac{1}{99} \left(\sum_{i=1}^{100} X_i^2 - 100\bar{X}^2 \right) \\ &= \frac{1}{99} [(70)1^2 + (15)2^2 + (10)3^2 + (3)4^2 + (2)5^2 - 100(1.52)^2] \\ &= 0.87838 \end{aligned}$$

The standard deviation is $s = \sqrt{s^2} = 0.9372$.

Alternatively, the sample variance can be computed as

$$\begin{aligned} s^2 &= \frac{1}{99} \sum_{i=1}^{100} (X_i - \bar{X})^2 \\ &= \frac{1}{99} [70(1 - 1.52)^2 + 15(2 - 1.52)^2 + 10(3 - 1.52)^2 + 3(4 - 1.52)^2 + 2(5 - 1.52)^2] \\ &= 0.87838 \end{aligned}$$

- (c) The sample median is the average of the 50th and 51st value when arranged in order. Both these values are equal to 1, so the median is 1.
- (d) The first quartile is the average of the 25th and 26th value when arranged in order. Both these values are equal to 1, so the first quartile is 1. The third quartile is the average of the 75th and 76th value when arranged in order. Both these values are equal to 2, so the first quartile is 2.
- (e) Of the 100 cars, $15 + 10 + 3 + 2 = 30$ had more than the mean of 1.52 occupants, so the proportion is $30/100 = 0.3$.
- (f) The quantity that is one standard deviation greater than the mean is $1.52 + 0.9372 = 2.5472$. Of the 100 cars, $10 + 3 + 2 = 15$ had more than 2.8652 children, so the proportion is $15/100 = 0.15$.
- (g) The region within one standard deviation of the mean is $1.52 \pm 0.9372 = (0.5828, 2.4572)$. Of the 100 cars, $70 + 15 = 85$ are in this range, so the proportion is $85/100 = 0.85$.
11. The sum of the mens' heights is $20 \times 178 = 3560$. The sum of the womens' heights is $30 \times 164 = 4920$. The sum of all 50 heights is $3560 + 4920 = 8480$. Therefore the mean score for the two classes combined is $8480/50 = 169.6$.

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) $0.25(17) = 4.25$. 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: $Q_1 = 20.1$, $Q_3 = 21.3$

(d) The variance for A is

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

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 $s = \sqrt{0.55529} = 0.7542$.

(e) 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.

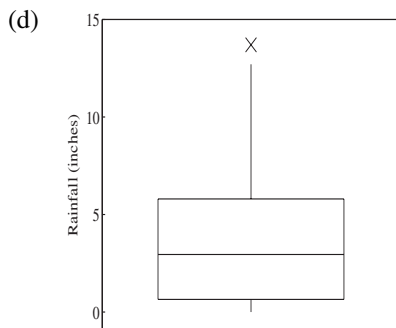
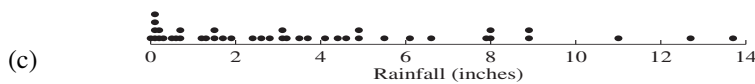
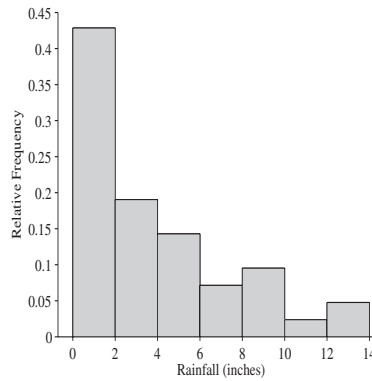
(f) Other things being equal, a smaller standard deviation is better. With any measurement method, the result is somewhat different each time a measurement is made. When the standard deviation is small, a single measurement is more valuable, since we know that subsequent measurements would probably not 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) Let S_0 be the sum of the original 10 numbers and let S_1 be the sum after the change. Then $S_0/10 = 70,000$, so $S_0 = 700,000$.
Now $S_1 = S_0 - 100,000 + 1,000,000 = 1,600,000$, so the new mean is $S_1/10 = 160,000$.
- (b) The median is unchanged at 55,000.
- (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 $(1/9)[T_0 - 10(70,000^2)] = 60,000^2 = 3.6 \times 10^9$, so $T_0 = 8.14 \times 10^{10}$. Let T_1 be the sum of the squares after the change.
Then $T_1 = T_0 - 100,000^2 + 1,000,000^2 = 1.0714 \times 10^{12}$.
The new standard deviation is $\sqrt{(1/9)[T_1 - 10(160,000^2)]} = 310,000$.
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.

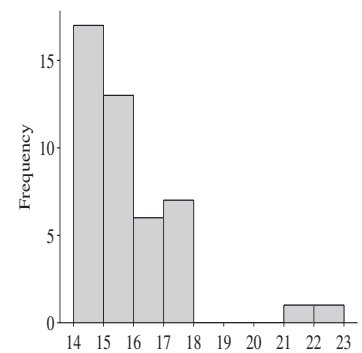


The boxplot shows one outlier.

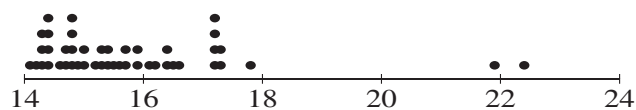
2. (a)

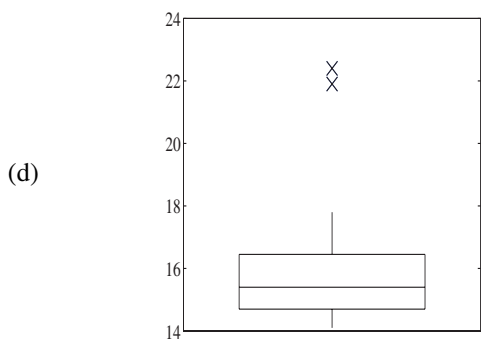
Stem	Leaf
14	123334444467788889
15	0023344567799
16	124456
17	2222338
18	
19	
20	
21	9
22	4

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



(c)





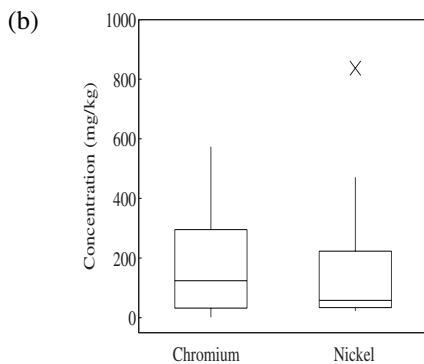
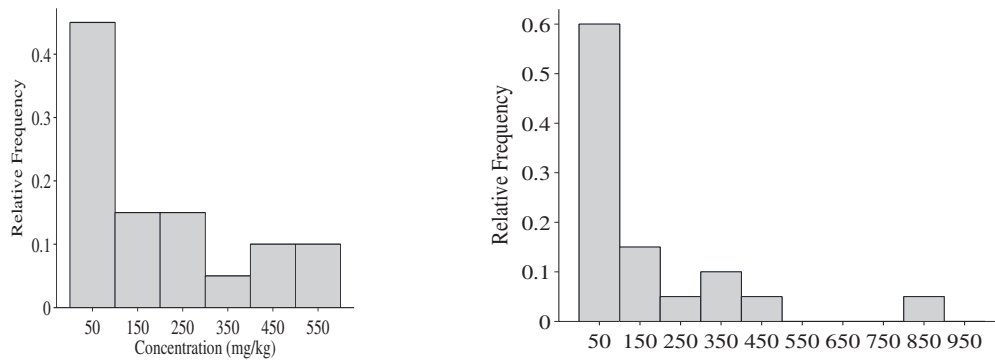
The boxplot shows two outliers.

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

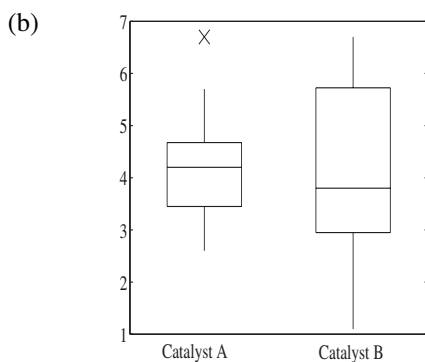
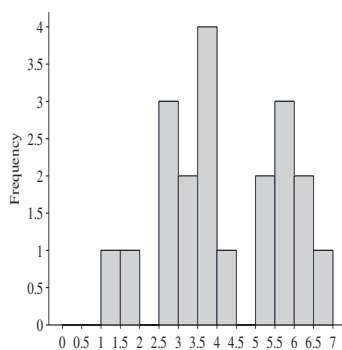
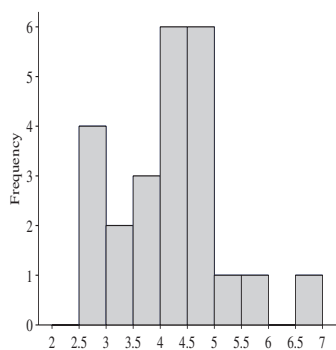
There are 23 stems in this plot. An advantage of this plot over the one in Figure 1.6 is that the values are given to the tenths digit instead of to the ones digit. A disadvantage is that there are too many stems, and many of them are empty.

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



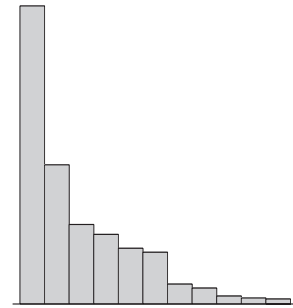
- (c) The concentrations of nickel are on the whole lower than the concentrations of chromium. The nickel concentrations are highly skewed to the right, which can be seen from the median being much closer to the first quartile than the third. The chromium concentrations are somewhat less skewed. Finally,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

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

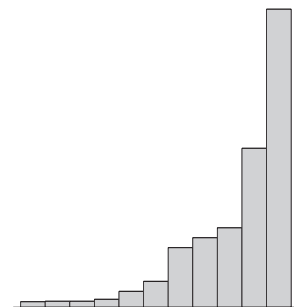


- (c) The results for Catalyst B are noticeably more spread out than those for Catalyst A. The median yield for catalyst A is greater than the median for catalyst B. The median yield for B is closer to the first quartile than the third, but the lower whisker is longer than the upper one, so the median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

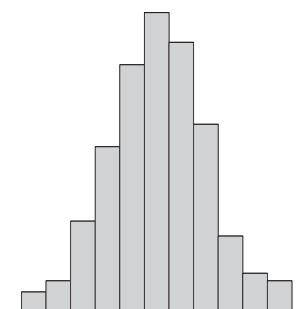
6. (a) The histogram should be skewed to the right. Here is an example.



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

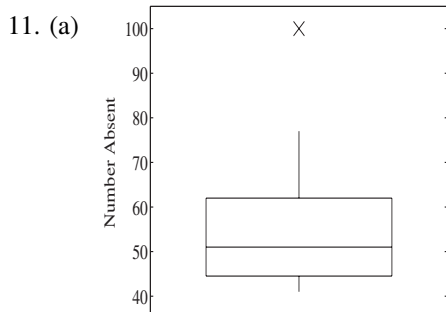


(c) The histogram should be approximately symmetric. Here is an example.



7. (a) The proportion is the sum of the relative frequencies (heights) of the rectangles above 240. This sum is approximately $0.14 + 0.10 + 0.05 + 0.01 + 0.02$. This is closest to 30%.

- (b) The height of the rectangle over the interval 240–260 is greater than the sum of the heights of the rectangles over the interval 280–340. Therefore there are more men in the interval 240–260 mg/dL.
8. The relative frequencies of the rectangles shown are 0.05, 0.1, 0.15, 0.25, 0.2, 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.
9. 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.
10. 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.



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

12. (a) False

(b) True

(c) False

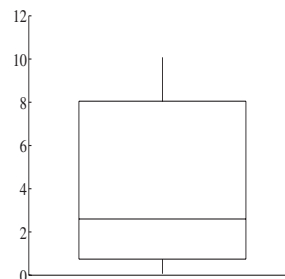
(d) False

(e) True

(f) False

13. (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.



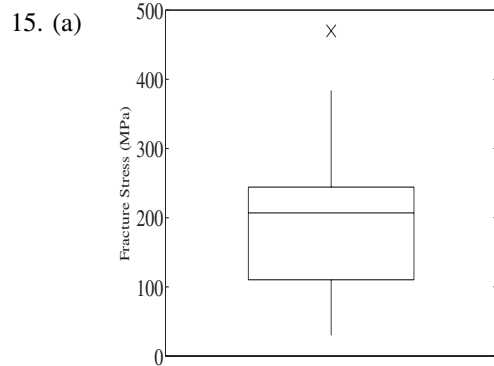
(c) No. The minimum value of -2.235 is an “outlier,” since it is more than 1.5 times the interquartile range below the first quartile. The lower whisker should extend to the smallest point that is not an outlier, but the value of this point is not given.

14. (a) (4)

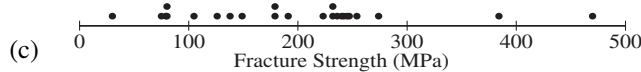
(b) (2)

(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.

Supplementary Exercises for Chapter 1

1. The mean and standard deviation both increase by 5%.

2. The mean is increased by \$2000. The standard deviation is 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 $11.61/15 = 0.774$.
- (b) No, it is not possible to determine the value of the mean after the change, since the original mean is unknown.
- (c) The median is the eighth number when the list is arranged in order, and this is unchanged.
- (d) It is not possible to tell by how much the standard deviation changes, 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.

8. (a) The sum of the numbers 284.34, 173.01, 229.55, 312.95, 215.34, 188.72, 144.39, 172.79, 139.38, 197.81, 303.28, 256.02, 658.38, 105.14, 295.24, 170.41 is 3846.75. The mean is therefore $3846.75/16 = 240.4219$.

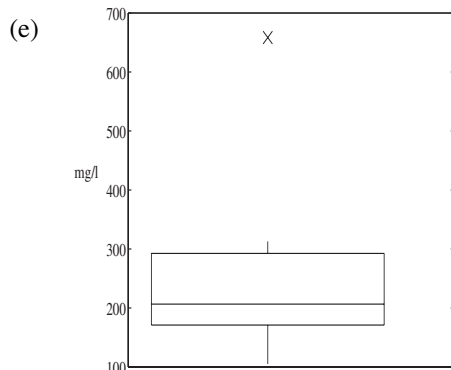
(b) The 16 values arranged in increasing order are:

105.14, 139.38, 144.39, 170.41, 172.79, 173.01, 188.72, 197.81,
215.34, 229.55, 256.02, 284.34, 295.24, 303.28, 312.95, 658.38

The median is the average of the 8th and 9th numbers, which is $(197.81 + 215.34)/2 = 206.575$.

(c) $0.25(17) = 4.25$, so the first quartile is the average of the 4th and 5th numbers, which is $(170.41 + 172.79)/2 = 171.60$.

(d) $0.75(17) = 12.75$, so the third quartile is the average of the 12th and 13th numbers, which is $(284.34 + 295.24)/2 = 289.79$.



The median is closer to the first quartile than to the third quartile, which indicates that the sample is skewed a bit to the right. In addition, the sample contains an outlier.

9. Statement (i) is true. The sample is skewed to the right.
10. (a) False. The length of the whiskers is at most 1.5 IQR.
- (b) False. The length of the whiskers is at most 1.5 IQR.
- (c) True. A whisker extends to the most extreme data point that is within 1.5 IQR of the first or third quartile.
- (d) True. A whisker extends to the most extreme data point that is within 1.5 IQR of the first or third quartile.
11. (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.
12. (i) It would be skewed to the right. The mean is greater than the median. Also note that half the values are between 0 and 0.10, so the left-hand tail is very short.