INSTRUCTOR'S SOLUTIONS MANUAL

STATISTICS

PRINCIPLES AND METHODS

Eighth Edition

Richard A. Johnson

Gouri K. Bhattacharyya

Richard A. Johnson

PREFACE

This instructor's manual is intended to provide the teacher with solutions to all of the exercises. Many are suitable for digital transfer and projection on a screen to facilate classroom discussion. We would greatly appreciate receiving your comments, corrections and suggestions for improvements.

K.T. Wu was of tremendous help in checking the solutions and in the preparation of earlier versions of this manuscript.

Richard A. Johnson

Contents

Chapter 1

INTRODUCTION

1.1 The *statistical population* consists of the entire set of wait/don't wait responses from all men in the United States while the *sample* consists of the responses of the 451 men contacted in the survey.

1.2 The sample is the collection of 150 number of passwords memorized by the visitors who agreed to participate. One specification of the statistical population is the collection of the number of passwords memorized by each visitor to the site.

1.3 The *statistical population* consists of the answers of all college students when asked their number of close friends. The *sample* is the collection of number of close friends reported by the twenty students who responded.

1.4 The *population* consists of the ratings of all tweets. The *sample* is te 4,220 rated tweets from the article.

1.5 The *statistical population* is the collection of yes/no answers to the question are you more stressed than last year. The *sample* is the responses of the 40 adults.

1.6 The *statistical population* is the rankings of 'favorite types of music' from all students attending a large U.S. university. The *sample* is the rankings from 56 students of a certain large U.S. university

1.7 Probably not a random sample of dog ownership among the cities' residents. People who belong to hiking clubs tend to participate in many outdoor activities that are conducive to dog ownership.

1.8 (a) Anecdotal. This is the result of an isolated action, sufficient data are not given to support this statement. item[(b)] Sample-based. The sample is the yes/no expresso-based purchase for the 47 drinks

recorded. The orders of a small group selected, hopefully at random, from a large group were used to make this observation.

(c) Sample-based. The sample is the 200 yes/no answers to thinking about food more than 17 times a day.

1.9 (a) This number was recorded from members of a class but one not randomly selected. It is likely recalled because the number is unusually large.

(b) Anecdotal. Just the opinion of one person.

(c) Sample-based. The sample is the 4837 yes/no answers to whether they used their cell phone while driving.

1.10 Answers will vary. From Table 1 , Appendix B, we start reading down starting in row 26 of columns 9 and 10. We obtain

57 19 32 29 21 91 8 75 72 62 58 85 41 3

17 7 60 86 3 15 89 41 93 84 23 87 78 1

Bicycles 1, 3, 7 and 8 are selected. It would be more efficient to use 1,21,41,61 and 81 to represent the first bicycle and so on. This would result in 12, 9, 1 and 11 being selected in the first five numbers.

1.11 Answers will vary. It is simpler to select 6 persons who will not go on the bus. Number the students from 1 to 50. In Table 1, we started in row 51 using columns 9 and 10. Reading downward, and ignoring 00 and numbers above 50, we selected students 23, 1, 44, 37, 19, and 26.

1.12 The term "on-time" is not well defined. This is not a yes/no question. Presumably, the bus reaches your stop multiple times per day. Moreover, a range of times should be specified for which the bus will characterized as being on-time . If it was one second late according to your watch, you would likely not characterize it as being late.

Purpose: Determine the percentage of time this semester the campus bus reaches your stop within one minute of the scheduled time

1.13 The notion of "comfortable " is not well defined. It is different for different people and it will depend on hand size and the position of controls. One improved statement of purpose is:

Purpose: Determine if over half the customers prefer a new style mouse to the one they currently use.

1.14 *Purpose* Determine the first choice of the campus population as their favorite campus pizza establishment.

1.15 *Purpose*: Determine the amount of time it takes those who use the Internet to make hotel reservations in San Francisco.

1.16 1.16 One-third of the calls took over 125 minutes to return (11 of 29). Seventeen of 29 took over 90 minutes to return.

1.17 At the lab, receptionist and x-ray.

1.18 (a) The variable of interest is whether or not a student can identify fake news

(b) The statistical population is the collection of yes/no responses answers to whether or not the student can identify fake news with one for each college student in the country (or from those colleges included in study).

(c) The sample is the collection of 2,300 yes/no answers from students included in the survey.

1.19 (a) A student at your college is the *unit*.

(b) The variable of interest is the total monthly entertainment expenses.

(c) The*statistical population* is the set of data consisting of the total monthly entertainment expenses for each student at your school.

1.20 (a) A person living in Chicago is the *unit*.

(b) The *variable of interest* is the characterization of a Chicago resident as eligible to vote or not.

(c) The *statistical population* consists of the collection of voter eligibility characterizations of each resident of the city of Chicago.

1.21 (a) The *statistical population* consists of the height measurements of male students on campus. The *sample* consists of the height measurements of the members of the basketball team.

(b) The sample is likely to be non-representative of the population, as basketball players tend to be much taller than the typical student

(c) There are many ways to choose a sample. One method is to use a table of random digits to select names from a student directory.

1.23 No, because a self-selection bias is likely to exist since only people who are interested in this particular exam are likely to answer, and such people perceive such a problem with the values.

1.24 The *statistical population* consists of the entire collection of yes/no responses to program purchase from every reader while the *sample* consists of the yes/no responses from those readers who actually sent in the completed form. This is apt to be a very biased sample because persons who have not purchased the program are not as likely to take the time to fill out the form and send it in as readers who have purchased the program

1.25 The newspaper is suggesting that the *statistical* is the collection of preferences for each adult in the city while the *sample* is the collection of preferences of the particular persons who sent in their votes. This sample is apt to be nonrepresentative because those persons in the sample are self-selected. Only the few who feel very strongly positive will likely send in a vote.

1.26 (a) Sample-based. The sample is the characteristic of lying regularly or not lying regularly reported by the 200 students.

(b) Anecdotal. Data are not given to support this statement. item[(c)] Sample-based. The sample is the purchase/not purchase data for the 50 persons interviewed.

1.27 (a) Anecdotal. No data given. class was not randomly selected.

(b) Sample-based. The yes/no answer regarding multiple credit cards, for each of the 22 students, is the sample on which the statement is based.

(c) Sample-based. The yes/no answer regarding destination outside the continental United States, for each of the 55 people at the airport, is the sample on which the statement is based.

1.28 The term "too long" is not well defined. By asking a number of people, we may determine that 5 minutes is too long. Further, the time will not be the same for all people. One improved statement of the purpose is

Purpose: Determine if over half the customers take over 5 minutes to get cash during the lunch hour.

1.29 Answers vary. First number the boats 0 to 9. In Table 9, we started in row 5 and column 12. Reading downward, we obtain 8 8 9 2 so boats 2, 8, and 9 are selected.

1.30 We select 3 of 9 sites using Table 1, Appendix B. Reading across in row 10 starting at column 17, we read one digit at a time and obtain 7 0 6 9. Sites 7, 6 and 9 are the three to visit.

1.31 Answers will vary. It is simpler to select 6 persons who will not go on the bus. Number the students from 1 to 50. In Table 1, we started in row 51 using columns 9 and 10. Reading downward, and ignoring 00 and numbers above 50, we selected students 23, 1, 44, 37, 19, and 26.

1.32 Answers will vary. We started in row 10 and read down column 9 and then down column 6 from the top ignoring 8 and 9. The 20 pairs of random digits are

4,0	1, 2	4, 2	5, 2	2,1
5,1	3, 2	2,0	7,6	5, 4
0,1	2,5	2,7	3,6	5, 4
2, 4	5,7	3,5	6,7	7, 2

(a) $4/20 = .20$

(b) $9/20 = .45$

(c) $7/20 = .35$

1.33 (a) The statement must refer to an average amount per person. Clearly, some persons create much more garbage and others less.

(b) Most likely from a sample of garbage, the average was 4.44 pounds per person per day. Certainly, the average for the whole population of the United States is unknown.

(c) You would prefer a nation wide sample. If restricted to households, you could conceivably use random numbers to selected from census listings.

1.34 (a) The miniture poodles could never be observed even if the greatly outnumber the Great Danes. Only the big dogs can volunteer to show they were inside the fence.

(b) Persons who call-in their opinions are self selected because they have strong opinions. This is analogous to the big dogs who are the only volunteers to show they were inside the fence.

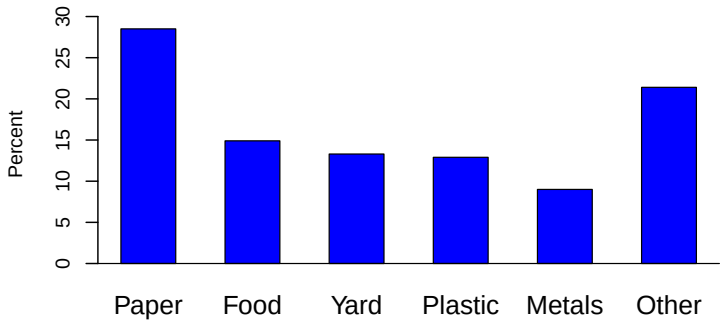
Chapter 2

ORGANIZATION AND DESCRIPTION OF DATA

2.1 (a) The percentage of other is

$$100 - (28.5 + 14.9 + 13.3 + 12.9 + 9.0) = 21.4$$

(b The pareto chart is



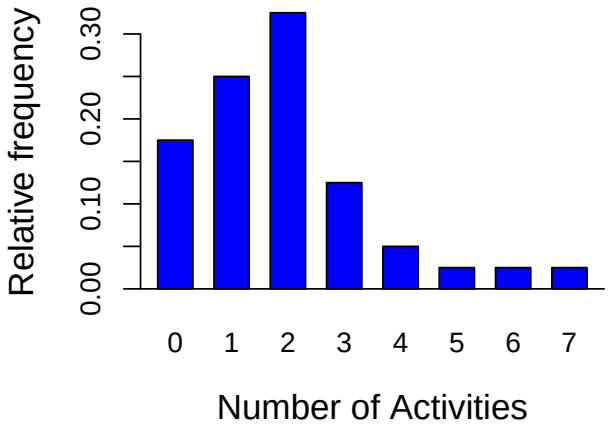
(c) Paper board is 28.5 %, the top two are 28.5 + 14.9 = 43.4 and the top five 78.6 percent.

2.2 The frequency table for blood type is

Blood type	Frequency	Relative Frequency
O	16	0.40 = 16/40
A	18	0.45 = 18/40
B	4	0.10 = 4/40
AB	2	0.05 = 2/40
Total	40	1.00

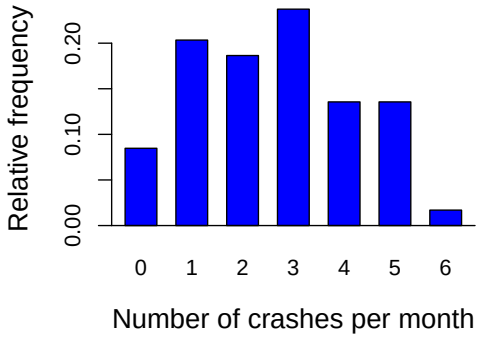
2.3 The frequency table and bar chart for number of activities are

Number of Activities	Frequency	Relative Frequency
0	7	7/40 = 0.175
1	10	10/40 = 0.250
2	13	13/40 = 0.325
3	5	5/40 = 0.125
4	2	2/40 = 0.050
5	1	1/40 = 0.025
6	1	1/40 = 0.025
7	1	1/40 = 0.025
Total	40	1.000



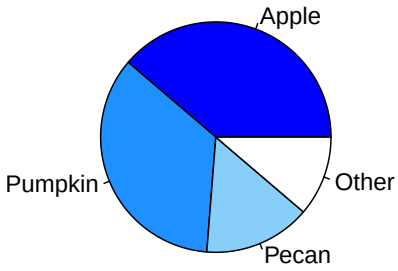
2.4 The frequency table and bar chart are

Number of Crashes	Frequency	Relative Frequency
0	5	5/ 59
1	12	12/59
2	13	11/59
3	5	14/59
4	2	8/ 59
5	1	8 /59
6	1	1/59
Total	59	1.000



2.5 The relative frequencies and pie chart are

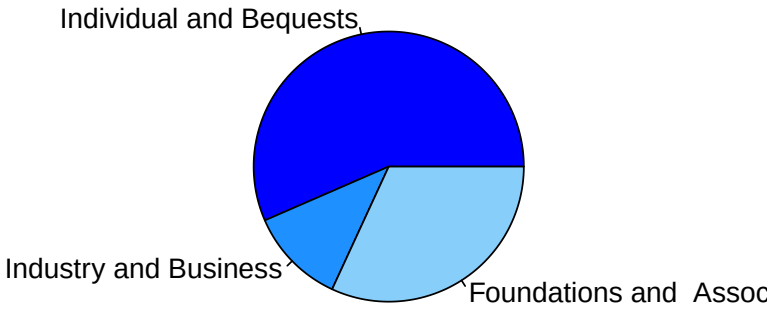
$$31/80 = .39 \quad 28/80 = .35 \quad 12/80 = .15 \quad 9/80 = .11$$



2.6 The table of relative frequencies for the money raised (in million dollars) is

Source	Frequency	Relative frequency
Individuals and bequests	117	0.565 = 117/207
Industry and business	24	0.116 = 24/207
Foundations and associations	66	0.319 = 66/207
Total	207	1.000

The pie chart for the university fund drive is



2.7 There are overlapping classes in the grouping. A report of 3 burglaries will fall in two classes.

2.8 If your team scores 6 goals, that is not included in any class.

2.9 There is a gap. The response 5 close friends does not fall in any class. The last class should be 5 or more.

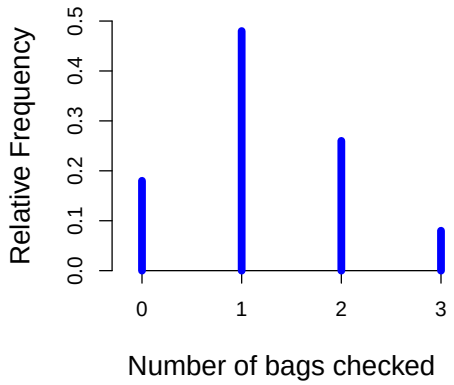
2.10 A light weight kicker cannot be assigned to a class.

2.11 (a) Yes. (b) Yes. (c) Yes. (d) No. (e) No.

2.12 The frequency table of the survey response is

Response	Frequency	Relative Frequency
1	14	0.28 = 14/50
2	13	0.26 = 13/50
3	7	0.14 = 7/50
4	16	0.32 = 16/50
Total	50	1.00

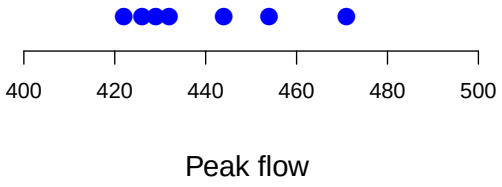
2.13 (a) The relative frequencies are 9/50 = .18, .48, .26, and .8 for 0, 1, 2, and 3 bags , respectively.



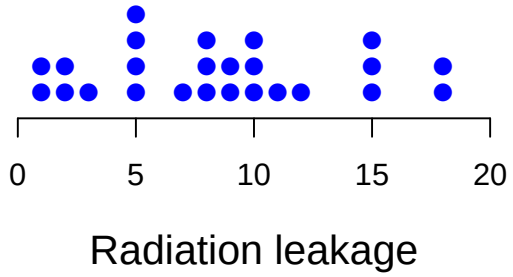
(b) Almost half of the passengers check exactly one bag. The longest tail is to the right.

(c) The proportion of passengers who don't check a bag is 9/50 = .18.

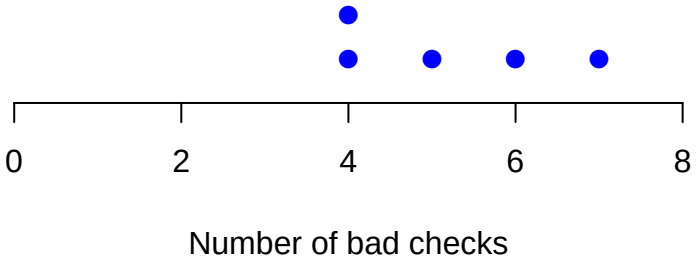
2.14 The dot diagram of peak flow measurements is



2.15 The dot diagram for radiation leakage is



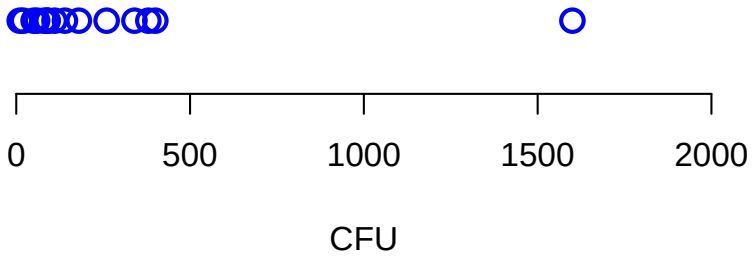
2.16 The dot diagram for number of bad checks received is



2.17 (a) The dot diagram of the number of CFU's is given on the next page.

(b) There is a long tail to the right with one extremely large value 1600 CFU units.

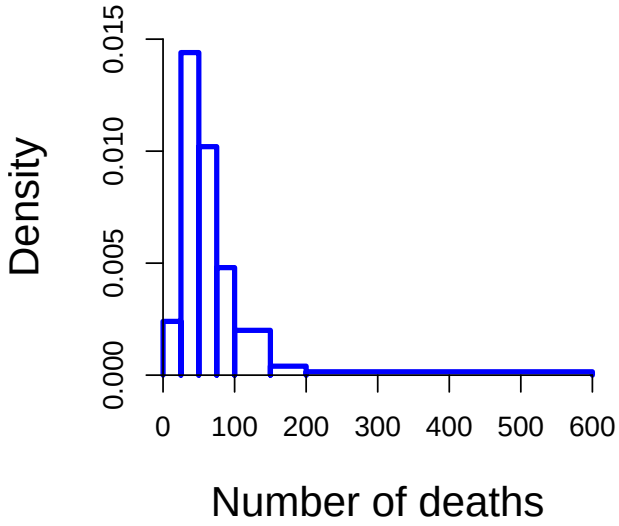
(c) $1 / 15 = .067$



2.18 (a) and (b). The frequency distribution of toronado fatalities and the heights are given in the table below.

Class Interval	Relative Frequency	Relative Frequency	Rectangle Height
[0, 25)	3/50 = .06	.06/25	= .0024
[25, 50)	18/50 = .36	.36/25	= .0144
[50, 75)	14/50 = .28	.28/25	= .0102
[75, 100)	6/50 = .12	.12/25	= .0048
[100, 150)	5/50 = .10	.10/50	= .0020
[150, 200)	1/50 = .02	.02/50	= .0004
[200, 600)	3/50 = .06	.06/400	= .00015

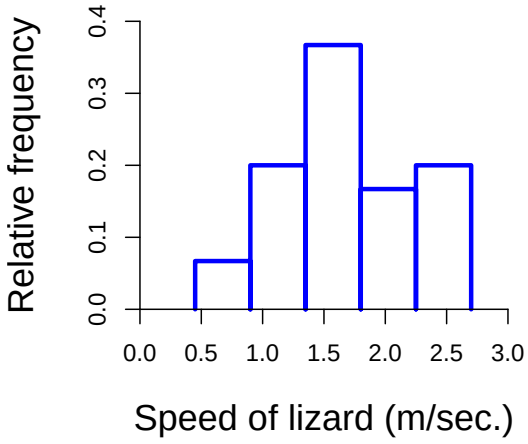
(c) The density histogram is



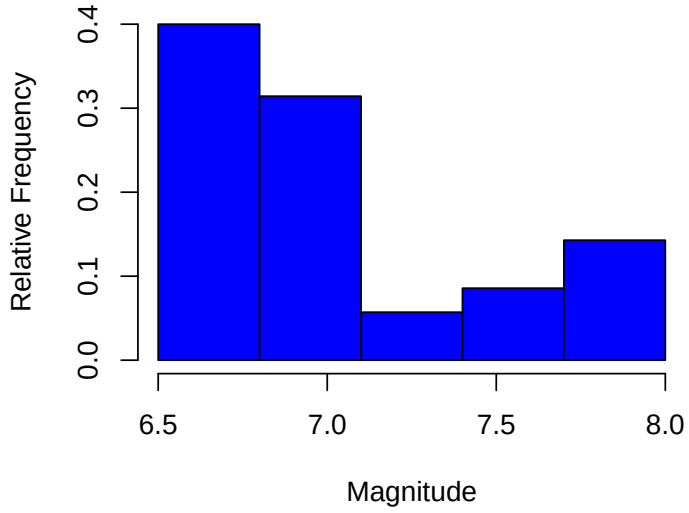
2.19 (a) In the following frequency distribution of lizard speed (in meters per second), the left endpoint is included in the class interval but not the right endpoint.

Class Interval			Frequency	Relative Frequency
.45	to	.90	2	0.067
.90	to	1.35	6	0.200
1.35	to	1.80	11	0.367
1.80	to	2.25	5	0.167
2.25	to	2.70	6	0.200
		Total	30	1.001

(b) All of the class intervals are of length .45 so we can graph rectangles whose heights are the relative frequency.



2.20 The frequencies are 14, 11, 2, 3, 5 and the equal interval relative frequency histogram is

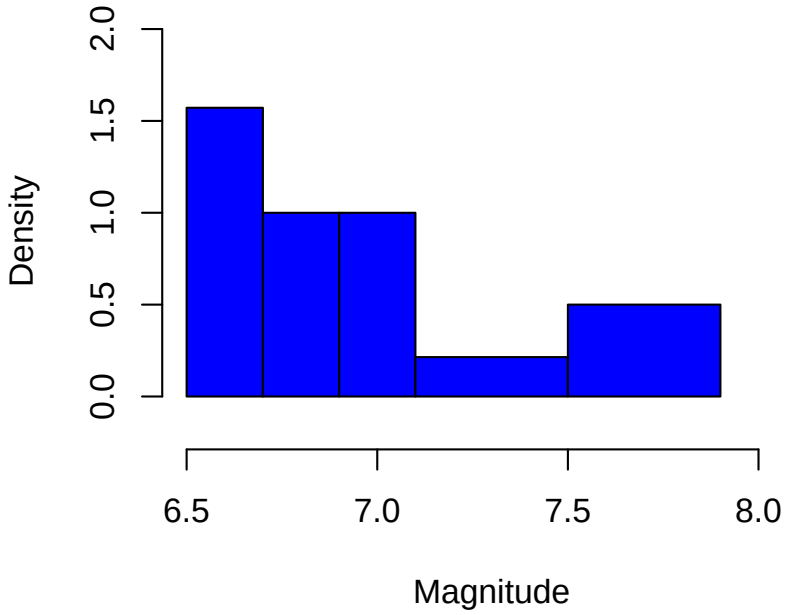


2.21 The values for density, relative frequency/ width, are

$$\frac{11}{35}/.2 = 1.57 \quad \frac{7}{35}/.2 = 1.00 \quad \frac{7}{35}/.2 = 1.00$$

$$\frac{3}{35}/.4 = .21 \quad \frac{7}{35}/.4 = .50$$

and the density histogram is



2.22 The stem-and-leaf display of the scores is

9	58
10	6
11	559
12	6
13	135678
14	344557
15	2478
16	01222567
17	14688
18	24
19	04

2.23 The stem-and-leaf display of the amount of iron present in the oil is

0	6
1	2234455567777889
2	000000222445567799
3	022444566
4	1167
5	12

2.24 The corresponding measurements are

246 268 293 319 344 371 382 397 405 426 443 490 504 568 613

2.25 The double-stem display of the amount of iron present in the oil is

0	6
1	22344
1	55567777889
2	00000022244
2	5567799
3	022444
3	566
4	11
4	67
5	12

2.26 The corresponding measurements are

18 20 20 21 22 22 23 23 24 24 24 25 25 25 26 26 27 29 30

2.27 Since $35 \times .5 = 17.5$, the median is in the 18-th position among the ordered data

6.6	6.6	6.6	6.6	6.6	6.6	6.7	6.7	6.7	6.7	6.7	6.8
6.8	6.8	6.9	6.9	6.9	6.9	7.0	7.0	7.0	7.0	7.1	7.1
7.1	7.2	7.2	7.5	7.6	7.7	7.8	7.8	7.8	7.9	7.9	

so the median = 6.9. Next, $35 \times .25 = 8.75$ so the first quartile is in the 9-th position. $Q_1 = 6.7$ Counting down from the top nine places, $Q_3 = 7.2$. The minium is 6.6 and the maximum is 7.9. The five number summary is

min	Q_1	median	Q_3	max
6.6	6.7	6.9	7.2	7.9

2.28 (a)

$$\overline{x} = \frac{\sum x_i}{n} = \frac{2 + 10 + 3 + 6 + 4}{5} = 5$$

Since $5 \times .5 = 2.5$, the median is in the third position among the ordered observations 2 3 4 6 10. The median = 4.

(b)

$$\overline{x} = \frac{\sum x_i}{n} = \frac{3 + 2 + 7 + 4}{4} = 4$$

Since $4 \times .5 = 2$, the median is the average of the values in the second and third positions among the ordered observations 2 3 4 7. The median = $(3 + 4)/2 = 3.5$.

2.29 (a)

$$\overline{x} = \frac{\sum x_i}{n} = \frac{3 + 6 + 2 + 5 + 4}{5} = 4$$

Since $5 \times .5 = 2.5$, the median is in the third position among the ordered observations 2 3 4 5 6. The median = 4.

(b)

$$\overline{x} = \frac{\sum x_i}{n} = \frac{4 + 3 + 8 + 5}{4} = 4$$

Since $4 \times .5 = 2$, the median is the average of the values in the second and third positions among the ordered observations 3 4 5 8. The median = $(4 + 5)/2 = 4.5$.

(c)

$$\overline{x} = \frac{\sum x_i}{n} = \frac{-4 + 0 - 3 - 2 + 2 - 1 + 5)}{7} = 1$$

Since $7 \times .5 = 3.5$, the median is in the fourth position among the ordered observations -4 -3 -2 -1 -1 0 2. The median = -1.

2.30

$$\overline{x} = \frac{\sum x_i}{n} = \frac{6.3 + 6.9 + 5.7 + 5.4 + 5.6 + 5.5 + 6.6 + 6.5}{8}$$
$$= 6.0625$$

Since $8 \times .5 = 4$, the median is the average of values in the fourth and in fifth position among the ordered observations

5.4 5.5 5.6 5.7 6.3 6.5 6.6 6.9. The median = 6.0.

2.31 (a) $\bar{x} = 3810/15 = 254$.

(b) The ordered observations are

10 20 50 60 80 90 90 110
140 180 260 340 380 400 1600

so the median is 110 CFU units. The one very large observation makes the sample mean much larger.

2.32 (a) The ordered monthly incomes are: 2275, 2350, 2425, 2450, 2475, 2650, 4700.

$$\bar{x} = \frac{19325}{7} = 2760.71 \quad , \quad \text{median } = 2450.$$

(b) For a typical salary, the median is best. Only one person makes more then the mean.

2.33 The mean is $956/12 = 79.67$. The claim ignores variability and is not true. It is certainly unpleasant with daily maximum temperature $105^0 F$ in July.

2.34 The sample mean is

$$\bar{x} = \frac{100 + 45 + 6 + 130 + 30}{5} = \frac{305}{5} = 61 \text{ minutes}$$

2.35 (a) $\bar{x} = 212/25 = 8.48$

(b) The sample median is 8. Since the sample mean and median are about the same, either of them can be used as an indication of radiation leakage.

2.36 (a)

$$\overline{x} = \frac{\sum x_i}{n} = \frac{4789}{8} = 598.6 \quad \text{feet}$$

(b)

$$\overline{x} = \frac{\sum x_i}{8} = \frac{3961}{8} = 565.9 \quad \text{feet}$$

The mean has been reduce by 32.4 feet. The one very tall building exerts too much influence on the mean.

2.37 The sample mean weight is 118.05 oz and the sample median is 117.00 oz. The average weight is 118.05 oz while about half weigh less than 117 oz and about half more.

2.38 (a) $\bar{x}$

$$= \frac{0\,(7) + 1\,(10) + 2\,(13) + 3\,(5) + 4\,(2) + 5\,(1) + 6\,(1) + 7\,(1,)}{40}$$

$$= 1.92 \text{ activities .}$$

(b) Sample median is 2 activites

(c) The larger activities 5, 6, and 7 did not overly influence the value of the mean.

2.39 (a) $\bar{x} = \dfrac{1(7) + 2(9) + 3(5) + 4(5) + 5(3)}{30} = 2.6$ returns.

(b) Sample median is 2 returns.

2.40 (a) Sample median $= (240 + 248)/2 = 244$(seconds).

(b) $\bar{x} = 1239/6 = 206.5$(seconds).

2.41 (a) $\bar{x} = 271/40 = 6.775$.

(b) Sample median $= (6 + 7)/2 = 6.5$. Both the sample mean and the sample median give a good indication of the amount of mineral lost.

2.42 (a) Sample median for males $= (45.8 + 48.3)/2 = 47.05$.

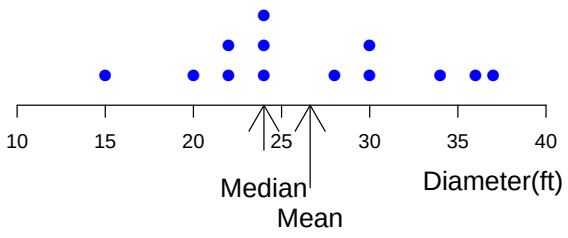
(b) Sample median for females $= 30.3$.

(c) Sample median for the combined set of males and females $= 38.6$.

2.43 Sample median $= (176 + 187)/2 = 181.5$(minutes).

2.44 In the previous problem, the sample mean $= 1862/10 = 186.2$(minutes). The total time for 10 games is $10\overline{x} = 1862$ minutes and this is meaningful. However, $10\times$ median ignores the actual times of the long games and is therefore meaningless.

2.45 (a) The dot diagram for the diameters (in feet) of the Indian mounds in southern Wisconsin is



(b) $\bar{x} = 346/13 = 26.62$. Sample median $= 24$.

(c) $13/4 = 3.25$, so we count in 4 observations. $Q_1 = 22$ and $Q_3 = 30$.

2.46 $16/4 = 4$, an integer, so we average the 4-th and 5-th observations. Q_1 = (9 + 12)/2 = 10.5 and Q_3 = (28 + 30)/2 = 29. The median, or $Q_2 = (15 + 20)/2 = 17.5$ days.

2.47 (a) Median = (152 + 154)/2 = 153.

(b) $40/4 = 10$, so we need to count in 10 observations. The 11-th smallest observation also satisfies the definition.

$$Q_1 = \frac{135 + 136}{2} = 135.5, \quad Q_3 = \frac{166 + 167}{2} = 166.5$$

2.48 $\bar{x} = 2283/25 = 91.32$ calls per shift.

2.49 The ordered data are

50	57	68	69	72	73	73	80	82	91
92	93	94	96	96	100	102	104	105	106
108	109	118	118	127					

Since the number of observations is 25, the median or second quartile is the 13th ordered observation in the list. The first quartile is the 7th observation.

$$Q_1 = 73 \quad Q_2 = 94 \quad Q_3 = 105$$

2.50 (a) The ordered data are

0.50	0.76	1.02	1.04	1.20	1.24	1.28	1.29	1.36	1.49
1.55	1.56	1.57	1.57	1.63	1.70	1.72	1.78	1.78	1.92
1.94	2.10	2.11	2.17	2.47	2.52	2.54	2.57	2.66	2.67

Since the number of observations is 30, the median or second quartile is the average of the 15th and 16th in the list. Sample median = (1.63+1.70)/2 = 1.665 meters per second. Because $30/4 = 7.5$, the first quartile is the 8th ordered observation.

$$Q_1 = 1.29 \quad Q_2 = 1.665 \quad Q_3 = 2.11$$

(b) Since .9(30) = 27, the 90th percentile is the average of the 27th and 28th observation in the ordered list. Sample 90th percentile = (2.54 + 2.57)/2 = 2.555.

2.51 (a) The ordered observations are

10	20	50	60	80	90	90	110
140	180	260	340	380	400	1600	

Since the sample size is 15, the median is the 8th observation 110. To obtain Q_1, we find $15/4 = 3.75$ so the first quantile if the 4th observation in the ordered list.

$$Q_1 = 60 \quad Q_3 = 340$$

(b) The 90-th percentile requires us to count in at least $.9(15) = 13.5$ or 14 observations. The 90-th sample percentile= 400.

2.52 (a) The mean of the original data set is

$$\bar{x} = \frac{4+8+8+7+9+6}{6} = \frac{42}{6} = 7$$

Adding $c = 4$ to the original data set we get: 8, 12, 12, 11, 13, 10. The mean of the new data set is

$$\overline{x+4} = \frac{8+12+12+11+13+10}{6} = \frac{66}{6} = 11$$

which equals $\bar{x} + c = 7 + 4$. Multiplying the original data set by $d = 2$ we get: 8, 16, 16, 14, 18, 12. The mean of the new data set is

$$\overline{2x} = \frac{8+16+16+14+18+12}{6} = \frac{84}{6} = 14$$

which equals $d \times \bar{x} = 2(7)$.

(b) The median of the original data set is

$$\text{median} = \frac{7+8}{2} = 7.5$$

When $c = 4$ is added to the original data set, the median of the new data set is

$$\text{median of } (x+4) = \frac{11+12}{2} = 11.5$$

which equals (median + c) = 7.5 + 4. When the original data set is multiplied by $d = 2$, the median of the new data set is

$$\text{median of } 2x = \frac{14+16}{2} = 15$$

which equals ($d \times$ median) = 2(7.5).

2.53 (a) The ordered data are 73, 74, 76, 76, 80. The median is 76^0F and the mean is $\bar{x} = 379/5 = 75.8^0F$.

(b) The mean of $({}^0F - 32)$ is $\bar{x} - 32$ by property (i) of Exercise 2.50 with $c = -32$. By property (ii)

$$\text{mean of } \frac{5}{9}({}^0F - 32) = \frac{5}{9}(\text{ mean of } ({}^0F - 32) \text{ })$$

$$= \frac{5}{9}(\bar{x} - 32) = \frac{5}{9}(75.8 - 32) = 24.33{}^0C$$

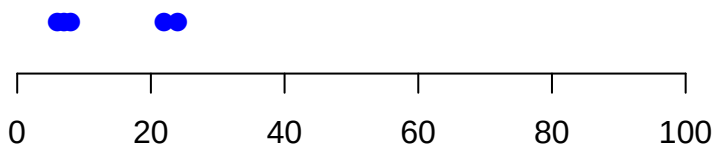
By similar properties for the median

$$\text{median of } \frac{5}{9}({}^0F - 32) = \frac{5}{9}(\text{ median of } ({}^0F) - 32 \text{ }) = \frac{5}{9}(76 - 32) = 24.44{}^0C$$

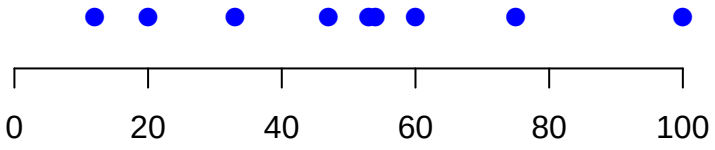
2.54 (a) A person with superior ability should expect to make above average salary, Company A is their best choice

(b) A person with medium abilities should expect to paid near the middle or median salary. Company B is their best choice.

2.55 (a)



L. Apopka (Testosterone)



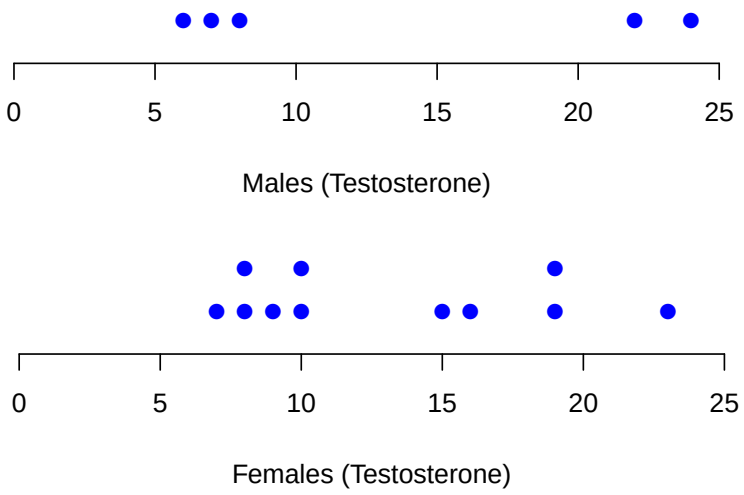
L. Woodruff (Testosterone)

(b) Lake Apopka:$\bar{x} = 67/5 = 13.40$
Lake Woodruff: $\bar{x} = 454/9 = 50.44$

(c) From the dot diagrams, the males in Lake Apopka have lower levels of testosterone and their sample mean is only about one-third of that for males in (un-contaminated) Lake Woodruff.

This finding is consistent with the environmentalists' concern that the contamination has affected the testosterone levels and reproductive abilities.

2.56 (a)



(b) Males $\bar{x} = 67/5 = 13.40$ Females $\bar{x} = 144/11 = 13.09$

(c) The dot diagrams of the amount of testosterone seem to be quite similar for males and females although there is a gap in the male diagram. The two means are nearly the same which suggests that the insecticide contamination has pushed hormone concentrations far out of balance because, ordinarily, males should have higher testosterone concentrations.

2.57 (a) $\bar{x} = \frac{\sum x_i}{n} = \frac{6+2+4}{3} = 4$ so the deviations $x_i - \bar{x}$ are 2 -2 0. They clearly add to 0.

(b)

$$s^2 = \frac{\sum (x_i - \bar{x})}{n-1} = \frac{2^2 + (-2)^2 + 0^2}{2} = 4$$

so $s = \sqrt{(4)} = 2$.

(c)
```
x=c(6,2,4)    mean(x)    var(x)      sd(x)
                 4          4           2
```

2.58 (a) $\bar{x} = \frac{\sum x_i}{n} = \frac{5+9-2}{3} = 4$ so the deviations $x_i - \bar{x}$ are 1 -5 -60. They clearly add to 0.

(b)

$$s^2 = \frac{\sum (x_i - \bar{x})}{n-1} = \frac{1^2 + (5)^2 + (-6)^2}{2} = 31$$

so $s = \sqrt{31} = 5.57$.

(c)

`x=c(5,9,-2)`	`mean(x)`	`var(x)`	`sd(x)`
	`4`	`31`	`5.568`

2.59 (a) We carry out all the necessary calculations in the following table.

	x	$x - \overline{x}$	$(x - \overline{x})^2$
	8	0	0
	6	-2	4
	14	6	36
	4	-4	16
Total	32	62	56

(b) The variance and standard deviation are

$$s^2 = \frac{56}{4-1} = 18.667 \qquad s = \sqrt{18.667} = 4.302$$

2.60 (a) We carry out all the necessary calculations in the following table. The mean is $\overline{x} = 9.5 / 5 = 1.9$

	x	$x - \overline{x}$	$(x - \overline{x})^2$
	2.5	.6	.36
	1.7	$-.2$	,04
	2.1	.2	.04
	1.5	$= .4$	.16
	1.7	$-.2$	.04
Total	9.5	0	.64

(b) The variance is

$$s^2 = \frac{.64}{5-1} = .16 \qquad s = \sqrt{.16} = .40$$

2.61 We carry out all the necessary calculations in the following table.

x		x^2
	6	36
	2	4
	4	16
Total	12	56

The variance is

$$s^2 = \frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) = \frac{1}{2}\left(56 - \frac{12^2}{3}\right) = 4$$

2.62 (a) We carry out all the necessary calculations in the following table.

	x	x^2
	100	10000
	45	2025
	60	3600
	130	16900
	30	900
Total	365	33425

The variance is

$$s^2 = \frac{1}{n-1}\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]$$

$$= \frac{1}{4}\left(33425 - \frac{365^2}{5}\right) = 1695$$

(b) The standard deviation $= \sqrt{1695} = 41.17$ minutes.

(c) In the alternate formula for s^2, only n changes.

$$s^2 = \frac{1}{5}\left(33425 - \frac{365^2}{6}\right) = 224.17$$

The value increased dramatically since the new value, 0, is very different from the other values of x.

2.63 (a) $s^2 = (3, -12^2/5)/4 = 1.30$

(b) $s^2 = (19 - (-7)^2/6)/5 = 2.167$

(c) $s^2 = (499 - 59^2/7)/6 = .286$

2.64 (a) $\overline{x} = \frac{\sum x_i}{n} = \frac{4789}{8} = 598.625$ and

$$s^2 = \frac{\sum (x_i - \overline{x})}{n-1} = \frac{71899.88}{7} = 10271.4$$

so $s = \sqrt{10271.4} = 101.3$ feet.

(b) $\overline{x} = \frac{\sum x_i}{n} = \frac{3961}{7} = 565.86$ and

$$s^2 = \frac{\sum (x_i - \overline{x})^2}{n-1} = \frac{11770.9}{6} = 1961.8$$

so $s = \sqrt{1961.8} = 44.3$ feet.

2.65

$$s = \frac{1}{13-1}\left(9726 - \frac{346^2}{13}\right) = 6.56 \text{ ft}$$

2.66 (a)

$$s^2 = \frac{1}{5-1}\left(142 - \frac{26^2}{5}\right) = 1.70$$

(b) $s = \sqrt{1.70} = 1.304$.

2.67 (a) $s^2 = (\,3,140,900 - 3810^2/15\,)\,/\,14 = 155,226$.

(b) $s = \sqrt{155,226} = 393.99$.

(b) $s^2 = (\,580,900 - 2210^2/14\,)\,/\,13 = 17,848.9$. The single very large value had greatly inflated the standard deviation.

2.68

(a) $s^2 = (\,2410 - 212^2/25\,)\,/\,24 = 25.51$.

(b) $s = \sqrt{25.51} = 5.05$.

2.69 (a) $\bar{x} = 1862/10 = 186.2$

(b) $s^2 = (\,353,796 - 1862^2/10)/9 = 787.96$.

(c) $s = \sqrt{787.96} = 28.07$.

2.70 (a) $\bar{x} = 62/50 = 1.24$ bags.

(b) $s^2 = (\,112 - 62^2\,/\,50)\,/\,49 = .71673$. Hence $s = .847$ bags.

2.71 (a) Median = 68.4.

(b) $\bar{x} = 478.4/7 = 68.343$.

(c) $s^2 = (\,32730.34 - 478.4^2/7)\,/\,6 = 5.853$. Hence $s = 2.419$.

(d)

```
x = c(66.9,66.2,71,68.6,65.4,68.4,71.9)
```

```
  mean(x)            sd(x)
68.34286           2.419268
```

2.72 (a) The output gives the standard deviation $s = 4.58$ so the sample variance $s^2 = (\,4.58\,)^2 = 20.98$

(b) The interquartile range = $Q_3 - Q_1 = 5 - 1 = 4$ hours. The middle 50 % of the hours students worked are contained in an interval of just 4 hours.

(c) The value of standard deviation here is 4.58 hours so numbers like 8.2 suggest greater variation.

2.73 (a) The output gives the standard deviation $s = 15.47$ so the sample variance $s^2 = (\,15.47\,)^2 = 239.321$

(b) The interquartile range $= Q_3 - Q_1 = 131.00 - 106.00 = 25.00$. This means the middle half of data span an interval of length 25 ounces. hours.

(c) The value of standard deviation here is 15.47 ounces so numbers like 10.2 ounces correspond to smaller variation.

2.74 (i) For the observations 5, 9, 9, 8, 10, 7, we have $\overline{x} = 8$, $s^2 = 3.2$ and $s = 1.789$. Adding c = 4 to the observations x , we have 9, 13, 13, 12, 14, 11. The sample mean and variance of the new data set are

$$\text{mean of } (x + 4) = \frac{9+13+13+12+14+11}{6} = 12$$

$$\text{variance of } (x + 4) = \frac{9+1+1+0+4+1}{6-1} = \frac{16}{5} = 3.2$$

so the standard devation of the new data set for $x + 4$ is

$$\sqrt{3.2} = 1.789$$

. It is the same as for the original data.

(ii) Multiplying the original observations by $d = 2$ we get 10, 18, 18, 16, 20, 14. The sample mean and variance of the new data set are

$$\text{mean of } \; 2x = \frac{10+18+18+16+20+14}{6} = 16$$

$$\begin{aligned} \text{variance of } \; 2x &= \frac{36 + 4+4+1+0+16+4}{6-1} \\ &= \frac{4(9 + 1+1+0+4+1}{6-1}) \\ &= 4(3.2) = 12.8 \end{aligned}$$

so the standard devation of the new $2x$ data set is 4 times the standard deviation of the original data.

2.75 For the data in Exercise 2.22, In Exercise 2.47 we determined that $Q_1 = 135.5$ and $Q_3 = 166.5$. Hence

$$\text{Interquartile range } = Q_3 - Q_1 = 166.5 - 135.5 = 31.0$$

2.76 From the data set of Exercise 2.3, in Exercise 2.46 we determined that $Q_1 = 1$ and $Q_3 = 2.5$ so

$$\text{Interquartile range } = Q_3 - Q_1 = 2.5 - 1 = 1.5 \text{ days}$$

2.77 No. Typically, the middle half of a data set is much more concentrated than the sum of the two quarters, one in each tail. As an example, for the water quality data of Exercise 2.17, the range is $1600 - 10 = 1590$ because of one extremely large observation. From the quartiles determined in Exercise 2.49, the interquartile range is $340 - 60 = 280$. The range is six times larger than the interquartile range.

2.78 (a) $\bar{x} = 150.125$ and $2s = 49.354$ so $\bar{x} \pm 2s$ is the interval (100.771, 199.479). This interval contains 38 observations or proportion .95 of the observations. And $\bar{x} \pm 3s$ is the interval (76.094, 224.156) which contains proportion 1 of the observations.

(b) The empirical guideline suggests proportion .95 in the interval $\bar{x} \pm 2s$ and we observed .95. It suggests proportion .997 for the interval $\bar{x} \pm 3s$ and we observed 1.000. The agreement is excellent.

2.79 (a) $\bar{x} = 6.775$ and $s = \sqrt{19.4096} = 4.406$.

(b) The proportion of the observations are:

	$\bar{x} \pm s$	$\bar{x} \pm 2s$	$\bar{x} \pm 3s$
Interval:	(2.369, 11.181)	(−2.037, 15.587)	(−6.443, 19.993)
Proportion:	26/40 = 0.65	38/40 = 0.95	40/40 = 1.00
Guidelines:	0.68	0.95	0.997

(c) We observe a good agreement with the proportions suggested by the empirical guideline.

2.80 (a) $\bar{x} = 51.71/30 = 1.724$ and $s = \sqrt{(98.641 - 51.71^2/30)/29} = .5727$.

(b) The proportion of the observations are:

	$\bar{x} \pm s$	$\bar{x} \pm 2s$	$\bar{x} \pm 3s$
Interval:	(1.151, 2.296)	(.579, 2.869)	(.006, 3.442)
Proportion:	20/30 = 0.667	29/30 = 0.967	30/30 = 1.000
Guidelines:	0.68	0.95	0.997

(c) We observe quite good agreement with the proportions suggested by the empirical guideline.

2.81 (a) $\bar{x} = 2.6$ and $s = \sqrt{1.69655} = 1.303$.

(b) The proportion of the observations are:

	$\bar{x} \pm s$	$\bar{x} \pm 2s$	$\bar{x} \pm 3s$
Interval:	(1.297, 3.903)	(−.006, 5.206)	(−1.30, 6.509)
Proportion:	15 / 30= 0.50	30 / 30= 1.00	30 / 30 =1.00
Guidelines:	0.68	0.95	0.997

(c) The proportion based on one standard deviation is somewhat low. guideline.

2.82 (a) The z-values of 350 and 620 are

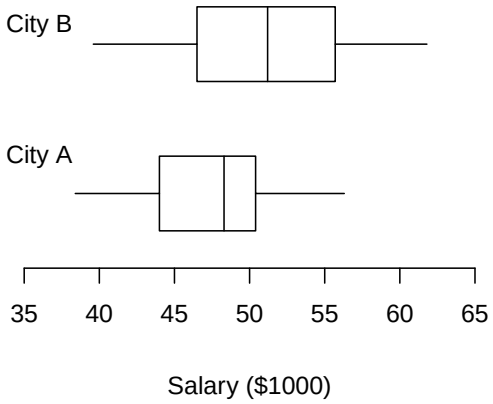
$$z = \frac{350 - 490}{120} = -1.167 \quad , \quad z = \frac{620 - 490}{120} = 1.083$$

(b) For the $z-$score of 2.4, the original score is obtained by solving the equation

$$z = 2.4 = \frac{x - 120}{50} \quad \text{so} \quad x = 210 + 50\,(2.4)\ 330$$

2.83 (a) $z = \dfrac{102 - 118.05}{15.47} = -\,1.037$ (b) $z = \dfrac{144 - 118.05}{15.47} = 1.677.$

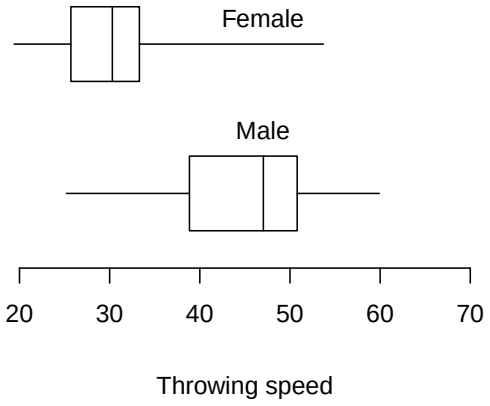
2.84 (a) and (b) The boxplots for salaries in both cities are



(c) There is a greater difference between the cities with respect to the higher salaries. For instance, any salary above the median in City B is greater than the 75th percentile in City A.

2.85 For males, the minimum and the maximum horizontal velocity of a thrown ball are 25.2 and 59.9 respectively. The quartiles are:
$Q_1 = (38.6 + 39.1)/2 = 38.85$, median $= (45.8 + 48.3)/2 = 47.05$,
$Q_3 = (49.9 + 51.7)/2 = 50.8$.
For females, the minimum and the maximum horizontal velocity of a thrown ball are 19.4 and 53.7 respectively. The quartiles are:
$Q_1 = 25.7$, median $= 30.3$, $Q_3 = 33.5$.
The boxplot of the male and female throwing speed are given at the top of the next page.

Comparing the two boxplots, we can see that males throw the ball faster than females.

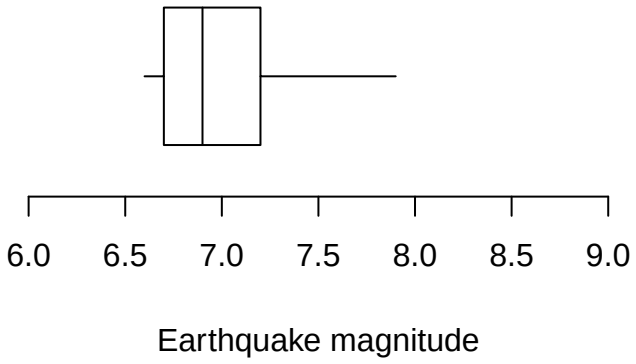


2.86 We give the boxplot obtained from the five number summary

(a)

min	Q_1	median	Q_3	max
6.6	6.7	6.9	7.2	7.9

(b)



2.87 (a) Using software we obtain $\overline{x} = 7.051$ and $s = .428$.

(b) The limits become

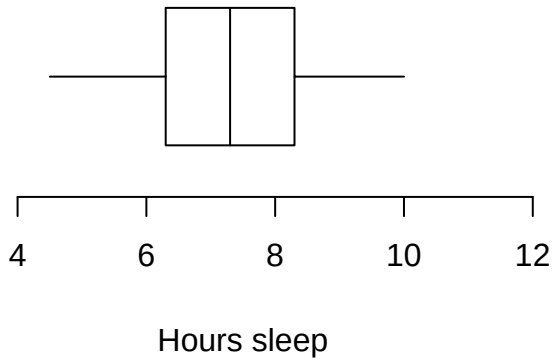
$$7.051 - 2\,(.428) = 6.195 \qquad 7.907 = 7.051 + ,2\,(.428)$$

Here 35 of the 35 observations lie in the interval. A histogram shows two peaks and that is far from the case where the empirical rule applies.

2.88 (a) Using the ordered data set from Example 6, we have

min	Q_1	Q_2	Q_3	max
4.5	6.3	7.3	8.3	10

(b)



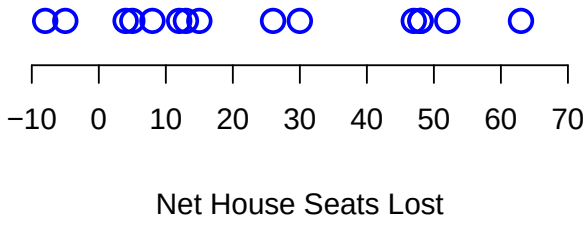
2.89 From Exercise 2.38, we know that $\overline{x} = 1.925$ and

$$s = \frac{249 - (77)^2/40}{40 - 1} = 2.584$$

so that $s = \sqrt{2.584} = 1.607$

2.90 (a) and (b) Using software we obtain $\overline{x} = 23.87$ and $s = 22.76$ seats lost.

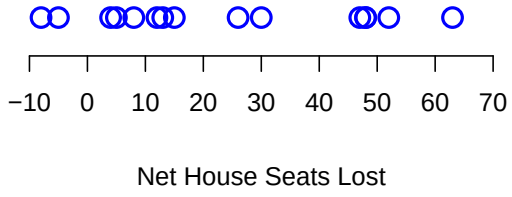
(c) The dot plot, with some double points because of ties, is



(d) The most striking feature is that a gain in seats only occurred twice.

2.91 (a) The ordered data are

$$-8 \quad -5 \quad 4 \quad 5 \quad 8 \quad 11 \quad 12 \quad 13 \quad 16 \quad 26 \quad 30 \quad 43 \quad 47 \quad 52 \quad 55 \quad 63$$

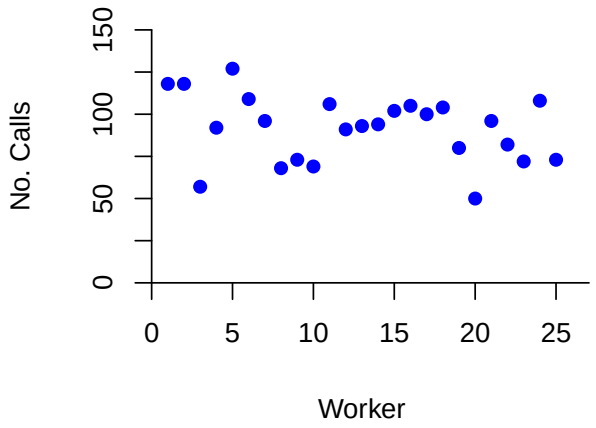


median $= (13 + 16)/2 = 14.5$ seats lost.

(b) The maximum number of seats lost, 63, occurred when B. Obama was President. The minimum number, -8 or a gain, occurred during G. W. Bush's second term as President.

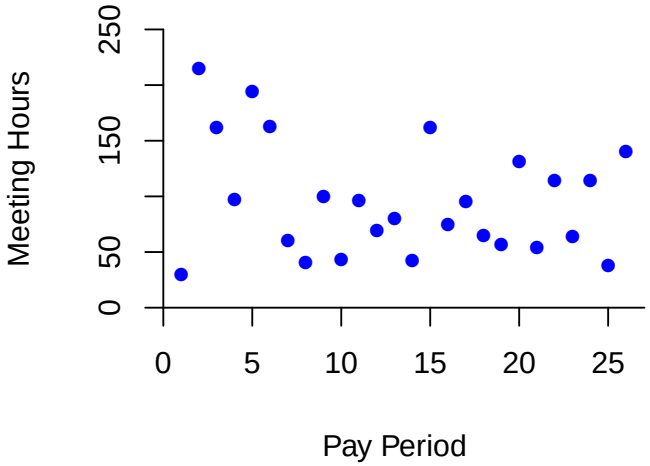
(c) range $= 63 - (-8) = 71$.

2.92



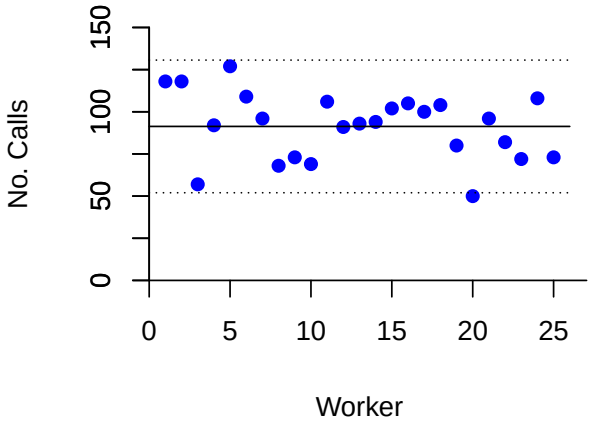
The process appears to be in statistical control. The pattern is nearly a horizontal band with one possible low value.

2.93



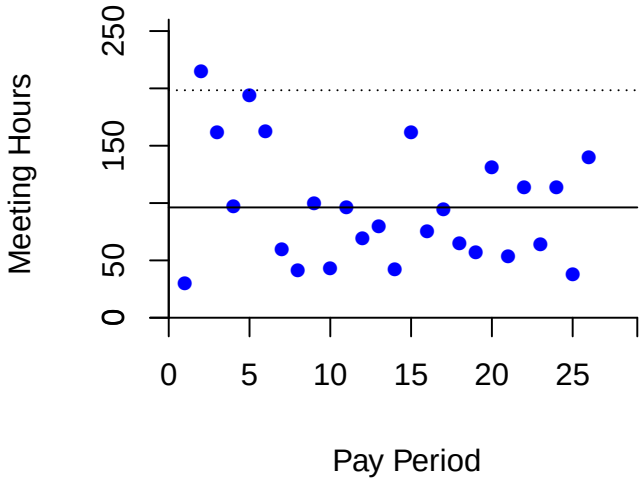
The value 215 from the second pay period looks high and 194 from the fifth period is possibly high.

2.94 We calculate $\bar{x} = 2283/25 = 91.32$ and $s = \sqrt{9281.44/24} = 19.67$ so the upper limit is $\bar{x}+2s = 130.66$ and the lower limit is $\bar{x}-2s = 51.98$.



Only the value 50 calls for worker 20 is out of control

2.95 We calculate $\bar{x} = 2501/26 = 96.2$ and $s = \sqrt{65254/25} = 51.1$ so the upper limit is $\bar{x}+2s = 198.4$ and the lower limit is $\bar{x}-2s = -6.0$ which we take as 0.

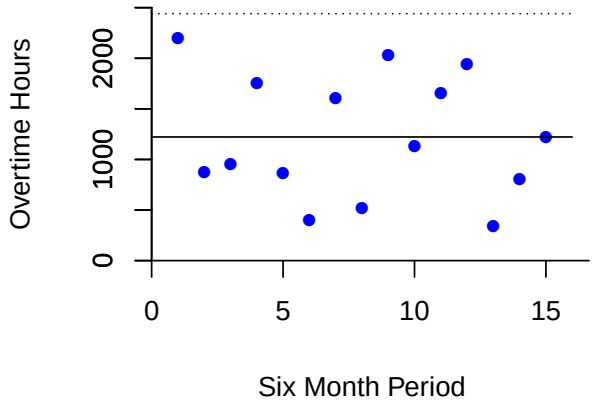


Only the value 215 from the second pay period is out of control.

2.96 We re-calculate without the outlier 5326.

$$\bar{x} = \frac{18329}{15} = 1221.9 \quad \text{and} \quad s = \sqrt{\frac{5195138}{14}} = 609.16$$

so the upper limit is $\bar{x}+2s = 2440.2$ and the lower limit is $\bar{x}-2s = 3.6$. All of the points are within the control limits.



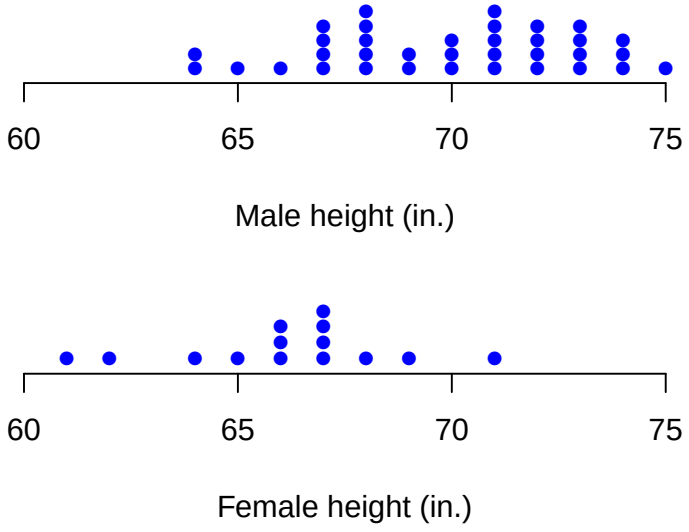
2.97 (a) The frequency table of "intended major" of the students is:

Intended major	Frequency	Relative Frequency
Biological Science	18	0.367
Humanities	4	0.082
Physical Sciences	9	0.184
Social Science	18	0.367
Total	49	1.000

(b) The frequency table of "year in college" of the students is:

Year	Frequency	Relative Frequency
1	4	0.082
2	10	0.204
3	20	0.408
4	15	0.306
Total	49	1.000

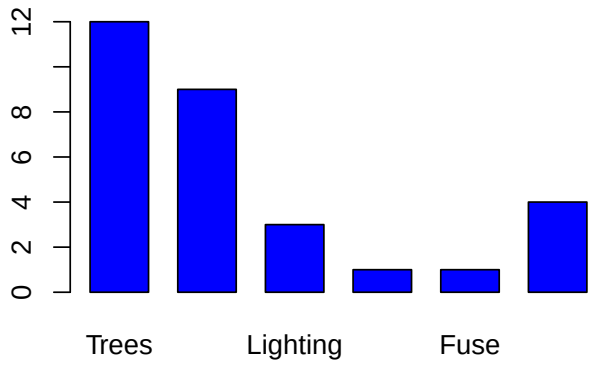
2.98 The dot diagrams of heights for the male and female students are



2.99 The frequency table of the causes for power outage is:

Frequency Table for Causes of Outrage	
Cause	Frequency
Trees and limbs	12
Animals	9
Lighting	3
Wind storm	1
Fuse	1
Unknown	4

The Pareto chart for the cause of the outage is



2.100 (a) Yes. The exact number of lunches is the sum of the frequencies of the first three classes.

(b) Yes. The exact number of lunches is the sum of the frequencies of the last three classes.

(c) No.

2.101 The sample mean and sample standard deviation are:

$$\overline{x} = \frac{143 + 131 + 101 + 143 + 111}{5} = 125.8 \quad s = \sqrt{\frac{1452.8}{5 - 1}} = 19.1 \text{ mm}$$

2.102 (a) The mean, 227.4, is one measure of center tendency and the median, 232.5, is another. These values may be interpreted as follows. On average, the 20 grizzly bears weigh 227.4 pounds apiece. Half of the grizzly bears sampled weighed at least 232.5 pounds while half weighed at most 232.5 pounds.

(b) The sample standard deviation is 82.7 pounds.