

Complete Solutions Manual to Accompany

Probability and Statistics for Engineering and the Sciences

NINTH EDITION

Jay Devore

California Polytechnic State University,
San Luis Obispo, CA

Prepared by

Matthew A. Carlton

California Polytechnic State University, San Luis Obispo, CA



© 2016 Cengage Learning

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored, or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, Web distribution, information networks, or information storage and retrieval systems, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the publisher except as may be permitted by the license terms below.

For product information and technology assistance, contact us at
Cengage Learning Customer & Sales Support,
1-800-354-9706.

For permission to use material from this text or product, submit
all requests online at www.cengage.com/permissions
Further permissions questions can be emailed to
permissionrequest@cengage.com.

ISBN-13: 978-1-305-26061-0
ISBN-10: 1-305-26061-9

Cengage Learning
20 Channel Center Street
Fourth Floor
Boston, MA 02210
USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at:
www.cengage.com/global.

Cengage Learning products are represented in
Canada by Nelson Education, Ltd.

To learn more about Cengage Learning Solutions,
visit www.cengage.com.

Purchase any of our products at your local college
store or at our preferred online store
www.cengagebrain.com.

NOTE: UNDER NO CIRCUMSTANCES MAY THIS MATERIAL OR ANY PORTION THEREOF BE SOLD, LICENSED, AUCTIONED, OR OTHERWISE REDISTRIBUTED EXCEPT AS MAY BE PERMITTED BY THE LICENSE TERMS HEREIN.

READ IMPORTANT LICENSE INFORMATION

Dear Professor or Other Supplement Recipient:

Cengage Learning has provided you with this product (the "Supplement") for your review and, to the extent that you adopt the associated textbook for use in connection with your course (the "Course"), you and your students who purchase the textbook may use the Supplement as described below. Cengage Learning has established these use limitations in response to concerns raised by authors, professors, and other users regarding the pedagogical problems stemming from unlimited distribution of Supplements.

Cengage Learning hereby grants you a nontransferable license to use the Supplement in connection with the Course, subject to the following conditions. The Supplement is for your personal, noncommercial use only and may not be reproduced, posted electronically or distributed, except that portions of the Supplement may be provided to your students IN PRINT FORM ONLY in connection with your instruction of the Course, so long as such students are advised that they

may not copy or distribute any portion of the Supplement to any third party. You may not sell, license, auction, or otherwise redistribute the Supplement in any form. We ask that you take reasonable steps to protect the Supplement from unauthorized use, reproduction, or distribution. Your use of the Supplement indicates your acceptance of the conditions set forth in this Agreement. If you do not accept these conditions, you must return the Supplement unused within 30 days of receipt.

All rights (including without limitation, copyrights, patents, and trade secrets) in the Supplement are and will remain the sole and exclusive property of Cengage Learning and/or its licensors. The Supplement is furnished by Cengage Learning on an "as is" basis without any warranties, express or implied. This Agreement will be governed by and construed pursuant to the laws of the State of New York, without regard to such State's conflict of law rules.

Thank you for your assistance in helping to safeguard the integrity of the content contained in this Supplement. We trust you find the Supplement a useful teaching tool.

CONTENTS

Chapter 1	Overview and Descriptive Statistics	1
Chapter 2	Probability	48
Chapter 3	Discrete Random Variables and Probability Distributions	90
Chapter 4	Continuous Random Variables and Probability Distributions	126
Chapter 5	Joint Probability Distributions and Random Samples	177
Chapter 6	Point Estimation	206
Chapter 7	Statistical Intervals Based on a Single Sample	217
Chapter 8	Tests of Hypotheses Based on a Single Sample	234
Chapter 9	Inferences Based on Two Samples	255
Chapter 10	The Analysis of Variance	285
Chapter 11	Multifactor Analysis of Variance	299
Chapter 12	Simple Linear Regression and Correlation	330
Chapter 13	Nonlinear and Multiple Regression	368
Chapter 14	Goodness-of-Fit Tests and Categorical Data Analysis	406
Chapter 15	Distribution-Free Procedures	424
Chapter 16	Quality Control Methods	434

CHAPTER 1

Section 1.1

1.
 - a. *Los Angeles Times, Oberlin Tribune, Gainesville Sun, Washington Post*
 - b. Duke Energy, Clorox, Seagate, Neiman Marcus
 - c. Vince Correa, Catherine Miller, Michael Cutler, Ken Lee
 - d. 2.97, 3.56, 2.20, 2.97

2.
 - a. 29.1 yd, 28.3 yd, 24.7 yd, 31.0 yd
 - b. 432 pp, 196 pp, 184 pp, 321 pp
 - c. 2.1, 4.0, 3.2, 6.3
 - d. 0.07 g, 1.58 g, 7.1 g, 27.2 g

3.
 - a. How likely is it that more than half of the sampled computers will need or have needed warranty service? What is the expected number among the 100 that need warranty service? How likely is it that the number needing warranty service will exceed the expected number by more than 10?
 - b. Suppose that 15 of the 100 sampled needed warranty service. How confident can we be that the proportion of *all* such computers needing warranty service is between .08 and .22? Does the sample provide compelling evidence for concluding that more than 10% of all such computers need warranty service?

Chapter 1: Overview and Descriptive Statistics

- 4.
- a. Concrete populations: all living U.S. Citizens, all mutual funds marketed in the U.S., all books published in 1980
Hypothetical populations: all grade point averages for University of California undergraduates during the next academic year, page lengths for all books published during the next calendar year, batting averages for all major league players during the next baseball season
 - b. (Concrete) Probability: In a sample of 5 mutual funds, what is the chance that all 5 have rates of return which exceeded 10% last year?
Statistics: If previous year rates-of-return for 5 mutual funds were 9.6, 14.5, 8.3, 9.9 and 10.2, can we conclude that the average rate for all funds was below 10%?
(Hypothetical) Probability: In a sample of 10 books to be published next year, how likely is it that the average number of pages for the 10 is between 200 and 250?
Statistics: If the sample average number of pages for 10 books is 227, can we be highly confident that the average for all books is between 200 and 245?
- 5.
- a. No. All students taking a large statistics course who participate in an SI program of this sort.
 - b. The advantage to randomly allocating students to the two groups is that the two groups should then be fairly comparable before the study. If the two groups perform differently in the class, we might attribute this to the treatments (SI and control). If it were left to students to choose, stronger or more dedicated students might gravitate toward SI, confounding the results.
 - c. If all students were put in the treatment group, there would be no firm basis for assessing the effectiveness of SI (nothing to which the SI scores could reasonably be compared).
- 6.
- One could take a simple random sample of students from all students in the California State University system and ask each student in the sample to report the distance from their hometown to campus. Alternatively, the sample could be generated by taking a stratified random sample by taking a simple random sample from each of the 23 campuses and again asking each student in the sample to report the distance from their hometown to campus. Certain problems might arise with self reporting of distances, such as recording error or poor recall. This study is enumerative because there exists a finite, identifiable population of objects from which to sample.
- 7.
- One could generate a simple random sample of all single-family homes in the city, or a stratified random sample by taking a simple random sample from each of the 10 district neighborhoods. From each of the selected homes, values of all desired variables would be determined. This would be an enumerative study because there exists a finite, identifiable population of objects from which to sample.

Chapter 1: Overview and Descriptive Statistics

- 8.
- a. Number observations equal $2 \times 2 \times 2 = 8$
 - b. This could be called an analytic study because the data would be collected on an existing process. There is no sampling frame.
- 9.
- a. There could be several explanations for the variability of the measurements. Among them could be measurement error (due to mechanical or technical changes across measurements), recording error, differences in weather conditions at time of measurements, etc.
 - b. No, because there is no sampling frame.

Section 1.2

- 10.
- a.

5		9	
6		33588	
7		00234677889	
8		127	
9		077	stem: ones
10		7	leaf: tenths
11		368	

A representative strength for these beams is around 7.8 MPa, but there is a reasonably large amount of variation around that representative value.

(What constitutes large or small variation usually depends on context, but variation is usually considered large when the range of the data – the difference between the largest and smallest value – is comparable to a representative value. Here, the range is $11.8 - 5.9 = 5.9$ MPa, which is similar in size to the representative value of 7.8 MPa. So, most researchers would call this a large amount of variation.)

- b. The data display is not perfectly symmetric around some middle/representative value. There is some positive skewness in this data.
- c. Outliers are data points that appear to be *very* different from the pack. Looking at the stem-and-leaf display in part (a), there appear to be no outliers in this data. (A later section gives a more precise definition of what constitutes an outlier.)
- d. From the stem-and-leaf display in part (a), there are 4 values greater than 10. Therefore, the proportion of data values that exceed 10 is $4/27 = .148$, or, about 15%.

Chapter 1: Overview and Descriptive Statistics

11.

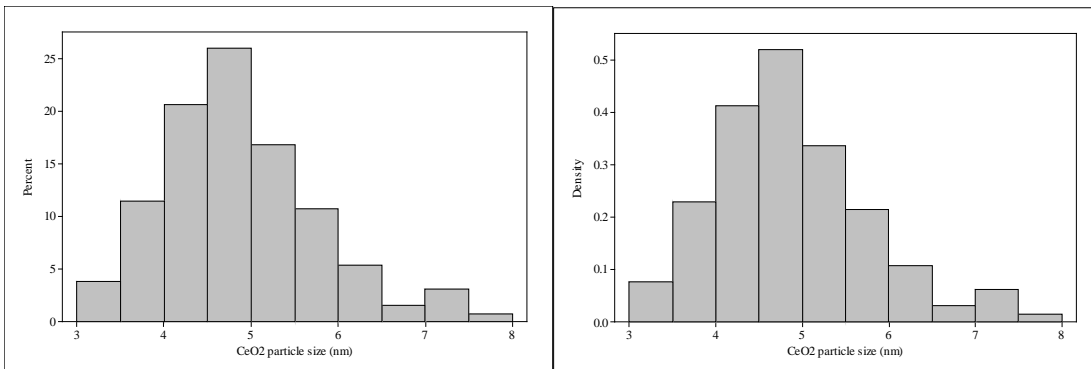
3L	1	
3H	56678	
4L	000112222234	
4H	5667888	stem: tenths
5L	144	leaf : hundredths
5H	58	
6L	2	
6H	6678	
7L		
7H	5	

The stem-and-leaf display shows that .45 is a good representative value for the data. In addition, the display is not symmetric and appears to be positively skewed. The range of the data is $.75 - .31 = .44$, which is comparable to the typical value of .45. This constitutes a reasonably large amount of variation in the data. The data value .75 is a possible outlier.

12.

The sample size for this data set is $n = 5 + 15 + 27 + 34 + 22 + 14 + 7 + 2 + 4 + 1 = 131$.

- a. The first four intervals correspond to observations less than 5, so the proportion of values less than 5 is $(5 + 15 + 27 + 34)/131 = 81/131 = .618$.
- b. The last four intervals correspond to observations at least 6, so the proportion of values at least 6 is $(7 + 2 + 4 + 1)/131 = 14/131 = .107$.
- c. & d. The relative (percent) frequency and density histograms appear below. The distribution of CeO_2 sizes is not symmetric, but rather positively skewed. Notice that the relative frequency and density histograms are essentially identical, other than the vertical axis labeling, because the bin widths are all the same.



Chapter 1: Overview and Descriptive Statistics

14.

a.

2	23	stem: 1.0
3	2344567789	leaf: .10
4	01356889	
5	00001114455666789	
6	0000122223344456667789999	
7	00012233455555668	
8	02233448	
9	012233335666788	
10	2344455688	
11	2335999	
12	37	
13	8	
14	36	
15	0035	
16		
17		
18	9	

- b. A representative is around 7.0.
- c. The data exhibit a moderate amount of variation (this is subjective).
- d. No, the data is skewed to the right, or positively skewed.
- e. The value 18.9 appears to be an outlier, being more than two stem units from the previous value.

15.

American		French
	8	1
755543211000	9	00234566
9432	10	2356
6630	11	1369
850	12	223558
8	13	7
	14	
	15	8
2	16	

American movie times are unimodal strongly positively skewed, while French movie times appear to be bimodal. A typical American movie runs about 95 minutes, while French movies are typically either around 95 minutes or around 125 minutes. American movies are generally shorter than French movies and are less variable in length. Finally, both American and French movies occasionally run very long (outliers at 162 minutes and 158 minutes, respectively, in the samples).

Chapter 1: Overview and Descriptive Statistics

16.

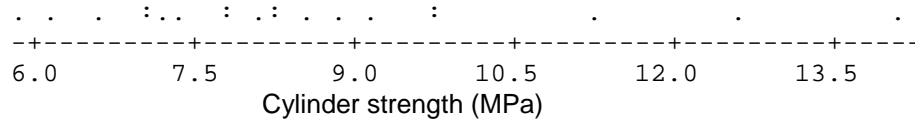
a.

Beams		Cylinders	
9	5	8	
88533	6	16	
98877643200	7	012488	
721	8	13359	stem: ones
770	9	278	leaf: tenths
7	10		
863	11	2	
	12	6	
	13		
	14	1	

The data appears to be slightly skewed to the right, or positively skewed. The value of 14.1 MPa appears to be an outlier. Three out of the twenty, or 15%, of the observations exceed 10 MPa.

b. The majority of observations are between 5 and 9 MPa for both beams and cylinders, with the modal class being 7.0-7.9 MPa. The observations for cylinders are more variable, or spread out, and the maximum value of the cylinder observations is higher.

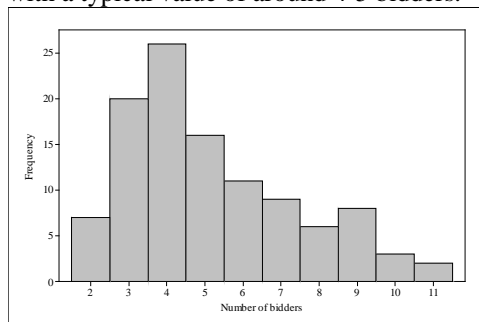
c.



17.

The sample size for this data set is $n = 7 + 20 + 26 + \dots + 3 + 2 = 108$.

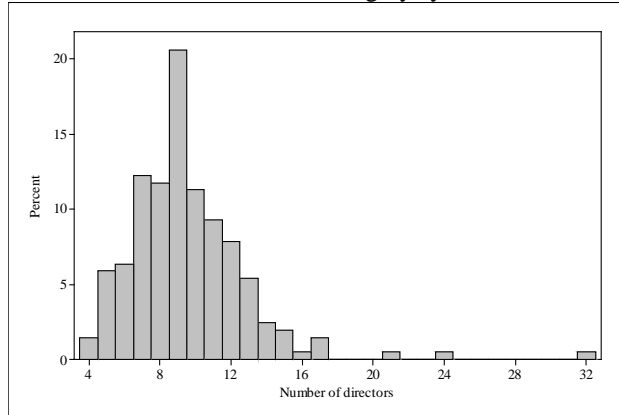
- a. “At most five bidders” means 2, 3, 4, or 5 bidders. The proportion of contracts that involved at most 5 bidders is $(7 + 20 + 26 + 16)/108 = 69/108 = .639$. Similarly, the proportion of contracts that involved at least 5 bidders (5 through 11) is equal to $(16 + 11 + 9 + 6 + 8 + 3 + 2)/108 = 55/108 = .509$.
- b. The number of contracts with between 5 and 10 bidders, inclusive, is $16 + 11 + 9 + 6 + 8 + 3 = 53$, so the proportion is $53/108 = .491$. “Strictly” between 5 and 10 means 6, 7, 8, or 9 bidders, for a proportion equal to $(11 + 9 + 6 + 8)/108 = 34/108 = .315$.
- c. The distribution of number of bidders is positively skewed, ranging from 2 to 11 bidders, with a typical value of around 4-5 bidders.



Chapter 1: Overview and Descriptive Statistics

18.

- a. The most interesting feature of the histogram is the heavy presence of three very large outliers (21, 24, and 32 directors). Absent these three corporations, the distribution of number of directors would be roughly symmetric with a typical value of around 9.



Note: One way to have Minitab automatically construct a histogram from grouped data such as this is to use Minitab's ability to enter multiple copies of the same number by typing, for example, 42(9) to enter 42 copies of the number 9. The frequency data in this exercise was entered using the following Minitab commands:

```
MTB > set c1
DATA> 3(4) 12(5) 13(6) 25(7) 24(8) 42(9) 23(10) 19(11) 16(12)
11(13) 5(14) 4(15) 1(16) 3(17) 1(21) 1(24) 1(32)
DATA> end
```

- b. The accompanying frequency distribution is nearly identical to the one in the textbook, except that the three largest values are compacted into the “ ≥ 18 ” category. If this were the originally-presented information, we could not create a histogram, because we would not know the upper boundary for the rectangle corresponding to the “ ≥ 18 ” category.

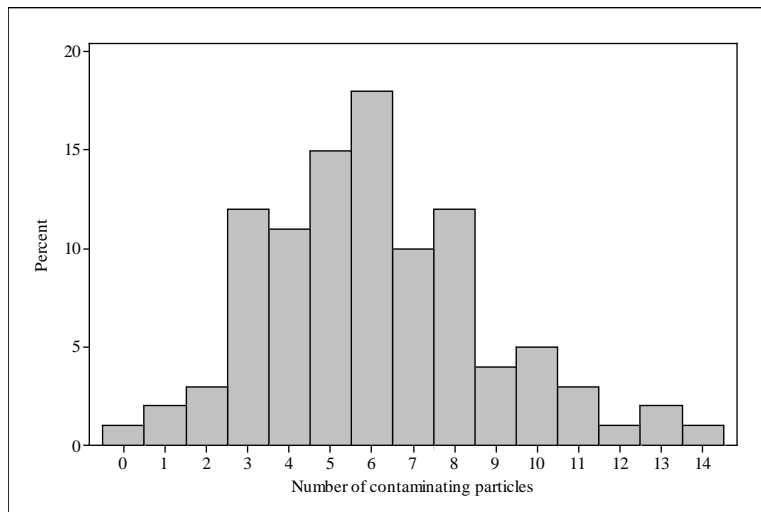
No. dir.	4	5	6	7	8	9	10	11
Freq.	3	12	13	25	24	42	23	19
No dir.	12	13	14	15	16	17	≥ 18	
Freq.	16	11	5	4	1	3	3	

- c. The sample size is $3 + 12 + \dots + 3 + 1 + 1 + 1 = 204$. So, the proportion of these corporations that have at most 10 directors is $(3 + 12 + 13 + 25 + 24 + 42 + 23)/204 = 142/204 = .696$.
- d. Similarly, the proportion of these corporations with more than 15 directors is $(1 + 3 + 1 + 1 + 1)/204 = 7/204 = .034$.

Chapter 1: Overview and Descriptive Statistics

19.

- a. From this frequency distribution, the proportion of wafers that contained at least one particle is $(100-1)/100 = .99$, or 99%. Note that it is much easier to subtract 1 (which is the number of wafers that contain 0 particles) from 100 than it would be to add all the frequencies for 1, 2, 3, ... particles. In a similar fashion, the proportion containing at least 5 particles is $(100 - 1-2-3-12-11)/100 = 71/100 = .71$, or, 71%.
- b. The proportion containing between 5 and 10 particles is $(15+18+10+12+4+5)/100 = 64/100 = .64$, or 64%. The proportion that contain strictly between 5 and 10 (meaning strictly *more* than 5 and strictly *less* than 10) is $(18+10+12+4)/100 = 44/100 = .44$, or 44%.
- c. The following histogram was constructed using Minitab. The histogram is *almost* symmetric and unimodal; however, the distribution has a few smaller modes and has a very slight positive skew.



20.

- a. The following stem-and-leaf display was constructed:

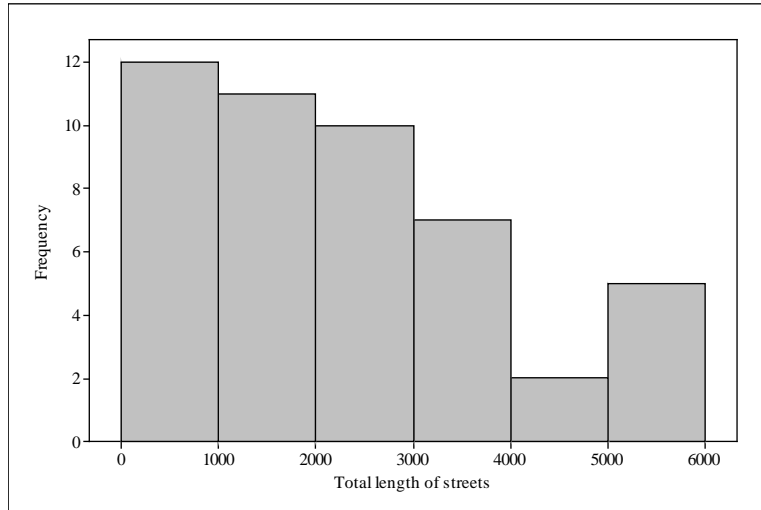
```

0|123334555599
1|00122234688      stem: thousands
2|1112344477      leaf: hundreds
3|0113338
4|37
5|23778
    
```

A typical data value is somewhere in the low 2000's. The display is bimodal (the stem at 5 would be considered a mode, the stem at 0 another) and has a positive skew.

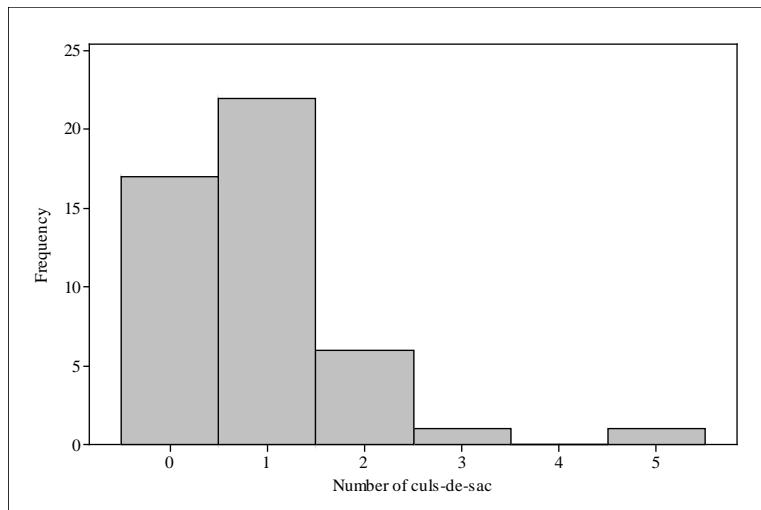
Chapter 1: Overview and Descriptive Statistics

- b. A histogram of this data, using classes boundaries of 0, 1000, 2000, ..., 6000 is shown below. The proportion of subdivisions with total length less than 2000 is $(12+11)/47 = .489$, or 48.9%. Between 2000 and 4000, the proportion is $(10+7)/47 = .362$, or 36.2%. The histogram shows the same general shape as depicted by the stem-and-leaf in part (a).



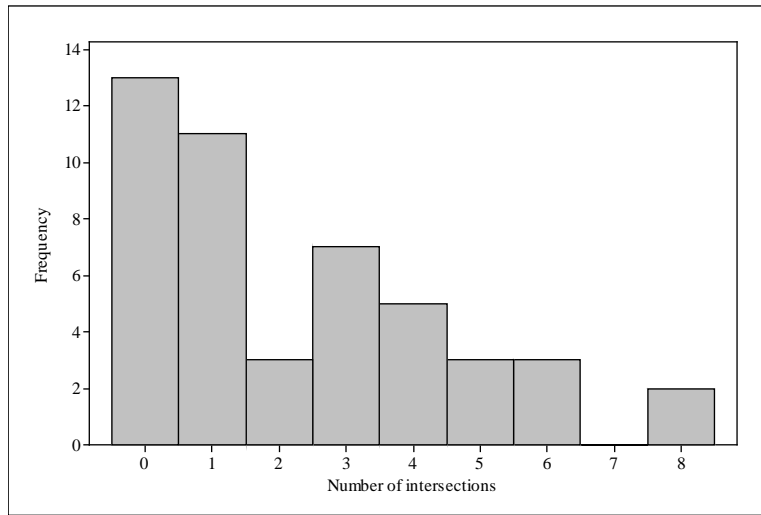
21.

- a. A histogram of the y data appears below. From this histogram, the number of subdivisions having no cul-de-sacs (i.e., $y = 0$) is $17/47 = .362$, or 36.2%. The proportion having at least one cul-de-sac ($y \geq 1$) is $(47 - 17)/47 = 30/47 = .638$, or 63.8%. Note that subtracting the number of cul-de-sacs with $y = 0$ from the total, 47, is an easy way to find the number of subdivisions with $y \geq 1$.

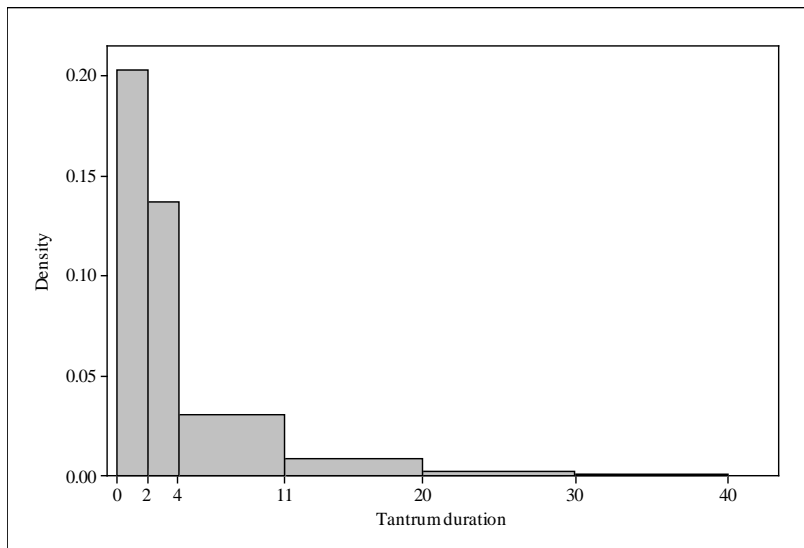


Chapter 1: Overview and Descriptive Statistics

- b. A histogram of the z data appears below. From this histogram, the number of subdivisions with at most 5 intersections (i.e., $z \leq 5$) is $42/47 = .894$, or 89.4%. The proportion having fewer than 5 intersections (i.e., $z < 5$) is $39/47 = .830$, or 83.0%.



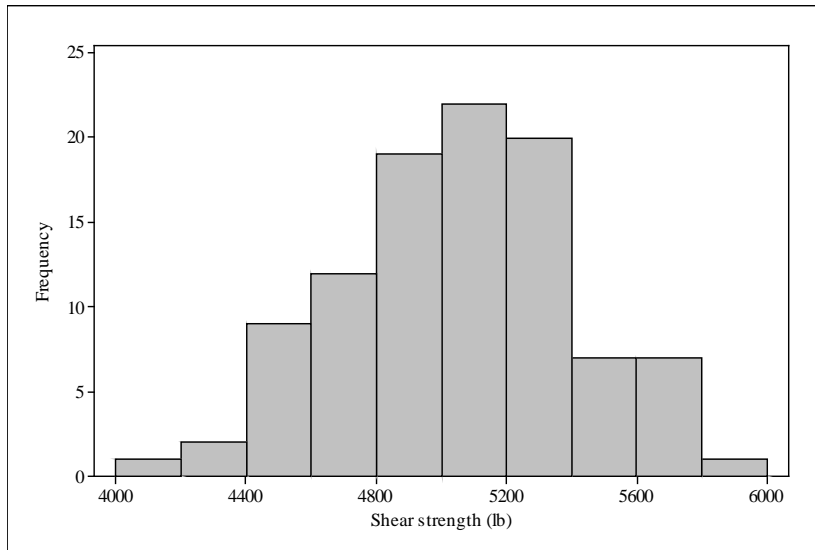
22. A very large percentage of the data values are greater than 0, which indicates that most, but not all, runners do slow down at the end of the race. The histogram is also positively skewed, which means that some runners slow down a *lot* compared to the others. A typical value for this data would be in the neighborhood of 200 seconds. The proportion of the runners who ran the last 5 km faster than they did the first 5 km is very small, about 1% or so.
23. Note: since the class intervals have unequal length, we must use a *density scale*.



The distribution of tantrum durations is unimodal and heavily positively skewed. Most tantrums last between 0 and 11 minutes, but a few last more than half an hour! With such heavy skewness, it's difficult to give a representative value.

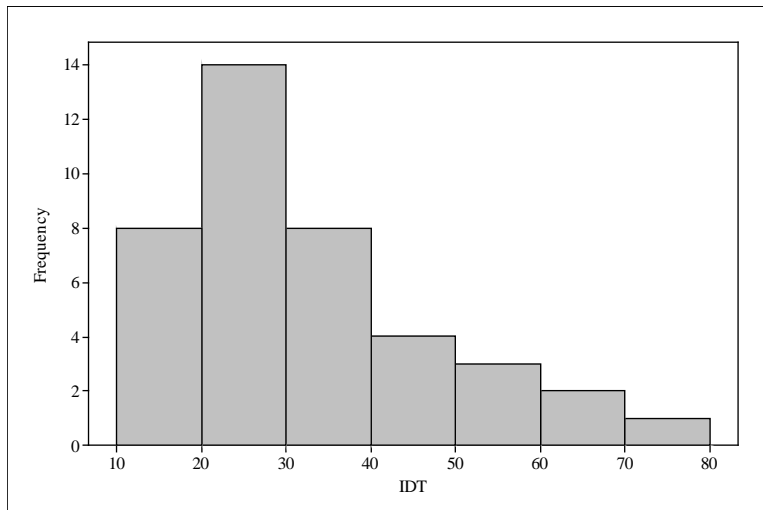
Chapter 1: Overview and Descriptive Statistics

24. The distribution of shear strengths is roughly symmetric and bell-shaped, centered at about 5000 lbs and ranging from about 4000 to 6000 lbs.



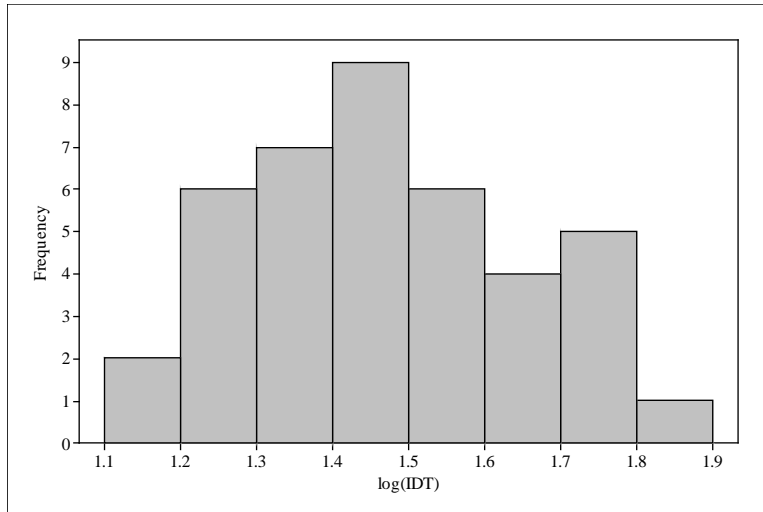
25. The transformation creates a much more symmetric, mound-shaped histogram.

Histogram of original data:



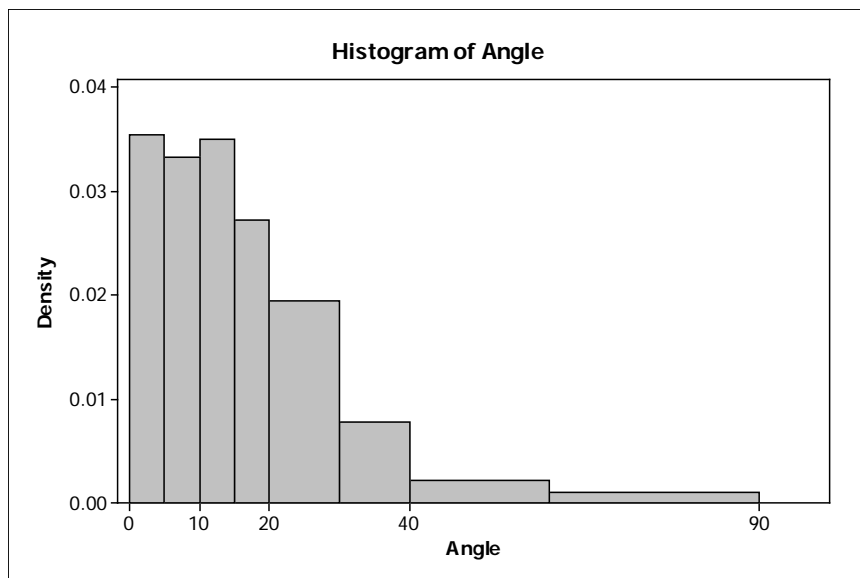
Chapter 1: Overview and Descriptive Statistics

Histogram of transformed data:



26.

- Yes: the proportion of sampled angles smaller than 15° is $.177 + .166 + .175 = .518$.
- The proportion of sampled angles at least 30° is $.078 + .044 + .030 = .152$.
- The proportion of angles between 10° and 25° is roughly $.175 + .136 + (.194)/2 = .408$.
- The distribution of misorientation angles is heavily positively skewed. Though angles can range from 0° to 90° , nearly 85% of all angles are less than 30° . Without more precise information, we cannot tell if the data contain outliers.

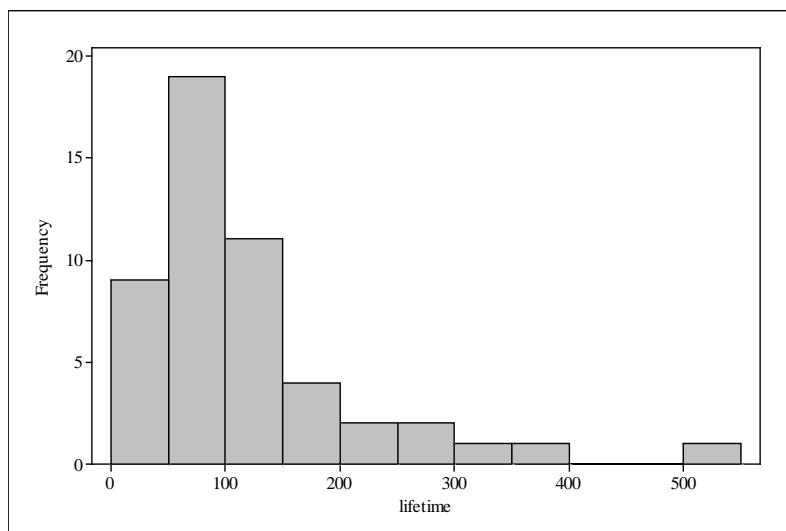


Chapter 1: Overview and Descriptive Statistics

27.

- a. The endpoints of the class intervals overlap. For example, the value 50 falls in both of the intervals 0–50 and 50–100.
- b. The lifetime distribution is positively skewed. A representative value is around 100. There is a great deal of variability in lifetimes and several possible candidates for outliers.

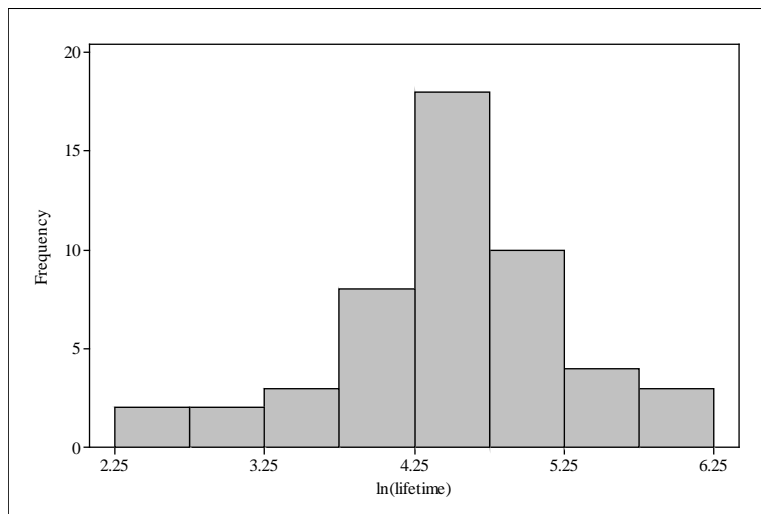
Class Interval	Frequency	Relative Frequency
0–< 50	9	0.18
50–<100	19	0.38
100–<150	11	0.22
150–<200	4	0.08
200–<250	2	0.04
250–<300	2	0.04
300–<350	1	0.02
350–<400	1	0.02
400–<450	0	0.00
450–<500	0	0.00
500–<550	1	0.02
	<hr/>	<hr/>
	50	1.00



Chapter 1: Overview and Descriptive Statistics

- c. There is much more symmetry in the distribution of the transformed values than in the values themselves, and less variability. There are no longer gaps or obvious outliers.

<u>Class Interval</u>	<u>Frequency</u>	<u>Relative Frequency</u>
2.25–<2.75	2	0.04
2.75–<3.25	2	0.04
3.25–<3.75	3	0.06
3.75–<4.25	8	0.16
4.25–<4.75	18	0.36
4.75–<5.25	10	0.20
5.25–<5.75	4	0.08
5.75–<6.25	3	0.06

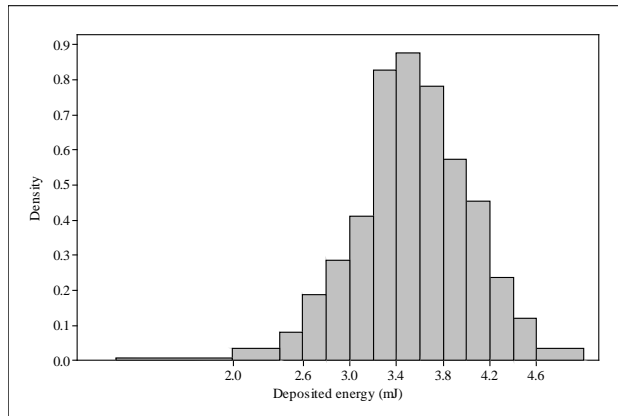


- d. The proportion of lifetime observations in this sample that are less than 100 is $.18 + .38 = .56$, and the proportion that is at least 200 is $.04 + .04 + .02 + .02 + .02 = .14$.

28. The sample size for this data set is $n = 804$.
- $(5 + 11 + 13 + 30 + 46)/804 = 105/804 = .131$.
 - $(73 + 38 + 19 + 11)/804 = 141/804 = .175$.
 - The number of trials resulting in deposited energy of 3.6 mJ or more is $126 + 92 + 73 + 38 + 19 + 11 = 359$. Additionally, 141 trials resulted in deposited energy within the interval 3.4–<3.6. If we assume that roughly half of these were in the interval 3.5–<3.6 (since 3.5 is the midpoint), then our estimated frequency is $359 + (141)/2 = 429.5$, for a rough proportion equal to $429.5/804 = .534$.

Chapter 1: Overview and Descriptive Statistics

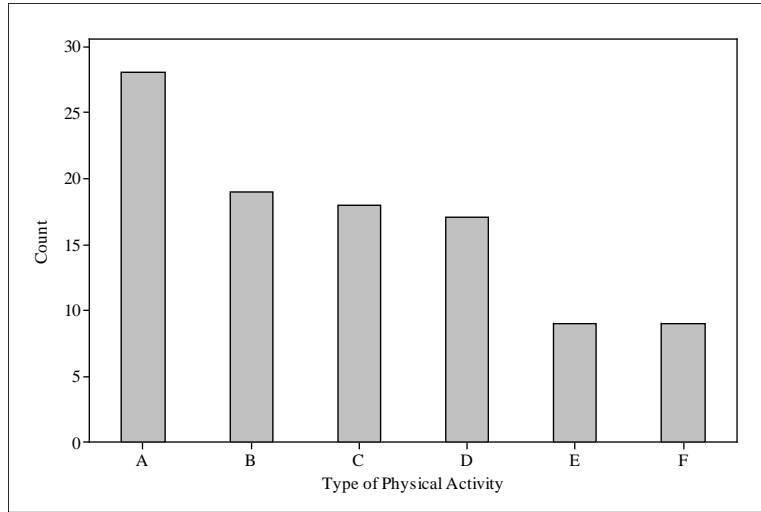
- d. The deposited energy distribution is roughly symmetric or perhaps slightly negatively skewed (there is a somewhat long left tail). Notice that the histogram must be made on a density scale, since the interval widths are not all the same.



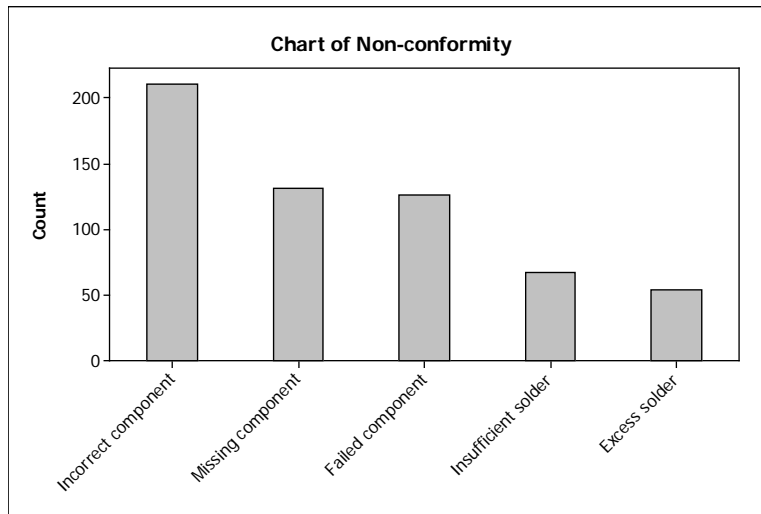
Chapter 1: Overview and Descriptive Statistics

29.

Physical Activity	Frequency	Relative Frequency
A	28	.28
B	19	.19
C	18	.18
D	17	.17
E	9	.09
F	9	.09
	100	1.00



30.



Chapter 1: Overview and Descriptive Statistics

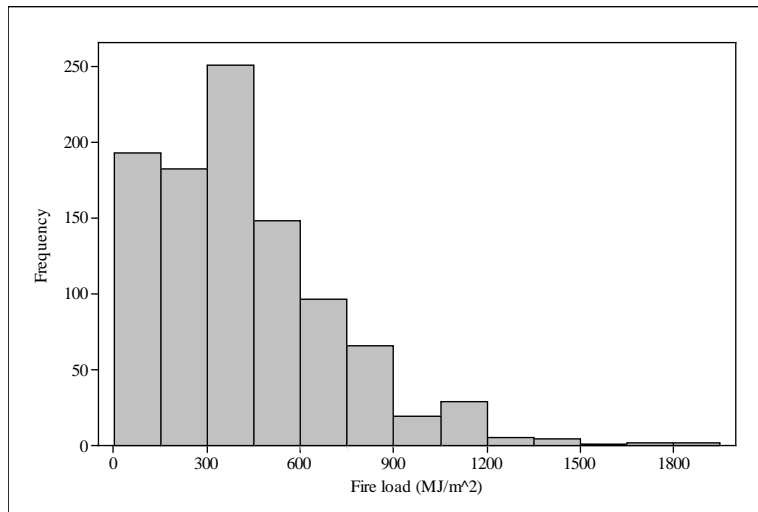
31.

Class	Frequency	Cum. Freq.	Cum. Rel. Freq.
0.0–<4.0	2	2	0.050
4.0–<8.0	14	16	0.400
8.0–<12.0	11	27	0.675
12.0–<16.0	8	35	0.875
16.0–<20.0	4	39	0.975
20.0–<24.0	0	39	0.975
24.0–<28.0	1	40	1.000

32.

- a. Cumulative percents must be restored to relative frequencies. Then the histogram may be constructed (see below). The relative frequency distribution is almost unimodal and exhibits a large positive skew. The typical middle value is somewhere between 400 and 450, although the skewness makes it difficult to pinpoint more exactly than this.

Class	Rel. Freq.	Class	Rel. Freq.
0–< 150	.193	900–<1050	.019
150–< 300	.183	1050–<1200	.029
300–< 450	.251	1200–<1350	.005
450–< 600	.148	1350–<1500	.004
600–< 750	.097	1500–<1650	.001
750–< 900	.066	1650–<1800	.002
		1800–<1950	.002



- b. The proportion of the fire loads less than 600 is $.193 + .183 + .251 + .148 = .775$. The proportion of loads that are at least 1200 is $.005 + .004 + .001 + .002 + .002 = .014$.
- c. The proportion of loads between 600 and 1200 is $1 - .775 - .014 = .211$.

Section 1.3

33.

- a. Using software, $\bar{x} = 640.5$ (\$640,500) and $\tilde{x} = 582.5$ (\$582,500). The average sale price for a home in this sample was \$640,500. Half the sales were for less than \$582,500, while half were for more than \$582,500.
- b. Changing that one value lowers the sample mean to 610.5 (\$610,500) but has no effect on the sample median.
- c. After removing the two largest and two smallest values, $\bar{x}_{tr(20)} = 591.2$ (\$591,200).
- d. A 10% trimmed mean from removing just the highest and lowest values is $\bar{x}_{tr(10)} = 596.3$. To form a 15% trimmed mean, take the average of the 10% and 20% trimmed means to get $\bar{x}_{tr(15)} = (591.2 + 596.3)/2 = 593.75$ (\$593,750).

34.

- a. For urban homes, $\bar{x} = 21.55$ EU/mg; for farm homes, $\bar{x} = 8.56$ EU/mg. The average endotoxin concentration in urban homes is more than double the average endotoxin concentration in farm homes.
- b. For urban homes, $\tilde{x} = 17.00$ EU/mg; for farm homes, $\tilde{x} = 8.90$ EU/mg. The median endotoxin concentration in urban homes is nearly double the median endotoxin concentration in farm homes. The mean and median endotoxin concentration for urban homes are so different because the few large values, especially the extreme value of 80.0, raise the mean but not the median.
- c. For urban homes, deleting the smallest ($x = 4.0$) and largest ($x = 80.0$) values gives a trimmed mean of $\bar{x}_{tr} = 153/9 = 17$ EU/mg. The corresponding trimming percentage is $100(1/11) \approx 9.1\%$. The trimmed mean is less than the mean of the entire sample, since the sample was positively skewed. Coincidentally, the median and trimmed mean are equal.

For farm homes, deleting the smallest ($x = 0.3$) and largest ($x = 21.0$) values gives a trimmed mean of $\bar{x}_{tr} = 107.1/13 = 8.24$ EU/mg. The corresponding trimming percentage is $100(1/15) \approx 6.7\%$. The trimmed mean is below, though not far from, the mean and median of the entire sample.

35.

- The sample size is $n = 15$.
- a. The sample mean is $\bar{x} = 18.55/15 = 1.237$ $\mu\text{g/g}$ and the sample median is \tilde{x} = the 8th ordered value = .56 $\mu\text{g/g}$. These values are very different due to the heavy positive skewness in the data.
- b. A 1/15 trimmed mean is obtained by removing the largest and smallest values and averaging the remaining 13 numbers: $(.22 + \dots + 3.07)/13 = 1.162$. Similarly, a 2/15 trimmed mean is the average of the middle 11 values: $(.25 + \dots + 2.25)/11 = 1.074$. Since the average of 1/15 and 2/15 is .1 (10%), a 10% trimmed mean is given by the midpoint of these two trimmed means: $(1.162 + 1.074)/2 = 1.118$ $\mu\text{g/g}$.

Chapter 1: Overview and Descriptive Statistics

- c. The median of the data set will remain .56 so long as that's the 8th ordered observation. Hence, the value .20 could be increased to as high as .56 without changing the fact that the 8th ordered observation is .56. Equivalently, .20 could be increased by as much as .36 without affecting the value of the sample median.

36.

- a. A stem-and leaf display of this data appears below:

32	55	stem: ones
33	49	leaf: tenths
34		
35	6699	
36	34469	
37	03345	
38	9	
39	2347	
40	23	
41		
42	4	

The display is reasonably symmetric, so the mean and median will be close.

- b. The sample mean is $\bar{x} = 9638/26 = 370.7$ sec, while the sample median is $\tilde{x} = (369+370)/2 = 369.50$ sec.
- c. The largest value (currently 424) could be increased by any amount. Doing so will not change the fact that the middle two observations are 369 and 370, and hence, the median will not change. However, the value $x = 424$ cannot be changed to a number less than 370 (a change of $424 - 370 = 54$) since that will change the middle two values.
- d. Expressed in minutes, the mean is $(370.7 \text{ sec})/(60 \text{ sec}) = 6.18$ min, while the median is 6.16 min.

37. $\bar{x} = 12.01$, $\tilde{x} = 11.35$, $\bar{x}_{tr(10)} = 11.46$. The median or the trimmed mean would be better choices than the mean because of the outlier 21.9.