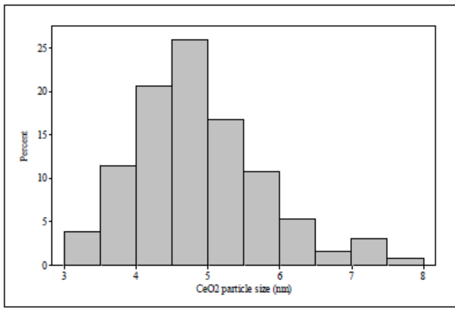
Pages 24-29

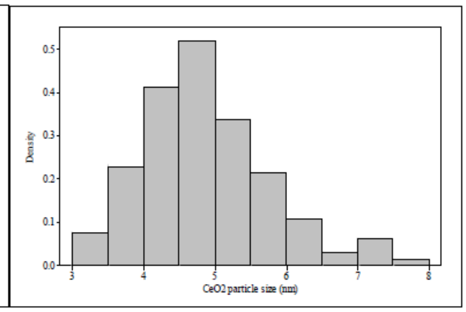
12. The sample size for this data set is $n = 5 + 15 + 27 + 34 + 22 + 14 + 7 + 2 + 4 + 1 = 131$.

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.





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.

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.

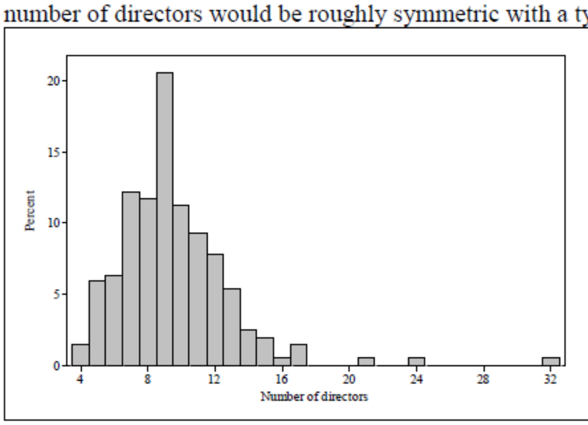
c.

```
 .  .   .   :..  : .: . . .    :          .          .           .
-+---------+---------+---------+---------+---------+-----
6.0       7.5       9.0      10.5      12.0      13.5
```

Cylinder strength (MPa)

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.



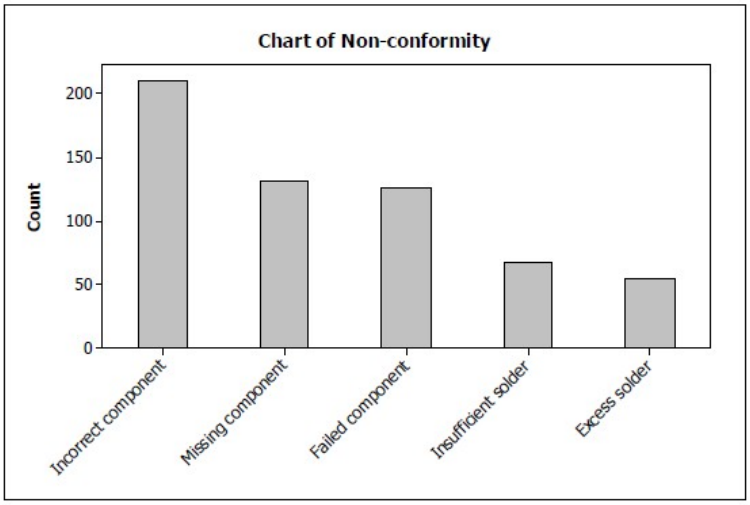
Note: One way to have Minitab automatically construct a histogram from grouped data such as this is to use Minitab's ability to enter multiple copies of the same number by typing, for example, 42(9) to enter 42 copies of the number 9. The frequency data in this exercise was entered using the following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

c. The sample size is 3 + 12 + … + 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.

30.



Pages 34-36

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) ≈ 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) ≈ 6.7%. The trimmed mean is below, though not far from, the mean and median of the entire sample.

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 sec)/(60 sec) = 6.18 min, while the median is 6.16 min.

102.23

92   119.26

40. $\tilde{x} = \cancel{60.8}$, $\bar{x}_{tr(25)} = 59.3083$, $\cancel{\bar{x}_{tr(10)}} = 58.3475$, $\cancel{\bar{x}} = 58.54$. All four measures of center have about the same value.

remove high & low 12:   remove 13:   weigted avg. ← Text book def.

95.38   94.75   95,08

Pages 44-51

44.

a. The maximum and minimum values are 182.6 and 180.3, respectively, so the range is 182.6 – 180.3 = 2.3°C.

b. Note: If we apply the hint and subtract 180 from each observation, the mean will be 1.41, and the middle two columns will not change. The sum and sum of squares will change, but those effects will cancel and the answer below will stay the same.

	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	x_i^2
	180.5	–0.90833	0.82507	32580.3
	181.7	0.29167	0.08507	33014.9
	180.9	–0.50833	0.25840	32724.8
	181.6	0.19167	0.03674	32978.6
	182.6	1.19167	1.42007	33342.8
	181.6	0.19167	0.03674	32978.6
	181.3	–0.10833	0.01174	32869.7
	182.1	0.69167	0.47840	33160.4
	182.1	0.69167	0.47840	33160.4
	180.3	–1.10833	1.22840	32508.1
	181.7	0.29167	0.08507	33014.9
	180.5	–0.90833	0.82507	32580.3
sums:	2176.9	0	5.769167	394913.6
	$\bar{x}$ = 181.41			

$s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 / (n-1)$ = 5.769167/(12 – 1) = 0.52447.

c. $s = \sqrt{0.52447} = 0.724$.

d. $s^2 = \dfrac{\Sigma x^2 - (\Sigma x)^2 / n}{n-1} = \dfrac{394913.6 - (2176.9)^2 / 12}{11} = 0.52447$.

46.

a. Since all three distributions are somewhat skewed and two contain outliers (see **d**), medians are the more appropriate central measures. The medians are

varies → (may give means)

Cooler: 1.760°C Control: 1.900°C Warmer: 2.305°C

The median difference between air and soil temperature increases as the conditions of the minichambers transition from cooler to warmer ($1.76 < 1.9 < 2.305$).

b. With the aid of software, the standard deviations are

Cooler: 0.401°C Control: 0.531°C Warmer: 0.778°C

For the 15 observations under the "cooler" conditions, the typical deviation between an observed temperature difference and the <u>mean</u> temperature difference (1.760°C) is roughly 0.4°C. A similar interpretation applies to the other two standard deviations. We see that, according to the standard deviations, variability increases as the conditions of the minichambers transition from cooler to warmer ($0.401 < 0.531 < 0.778$).

c. Apply the definitions of lower fourth, upper fourth, and fourth spread to the sorted data within each condition.

1.43 1.90

Cooler: lower fourth = ~~(1.43 + 1.57)/2 = 1.50~~, upper fourth = ~~(1.88 + 1.90)/2 = 1.89~~,

.47 → f_s = ~~1.89 – 1.50 = 0.39°C~~

1.52 2.03

Control: lower fourth = ~~(1.52 + 1.78)/2 = 1.65~~, upper fourth = ~~(2.00 + 2.03)/2 = 2.015~~,

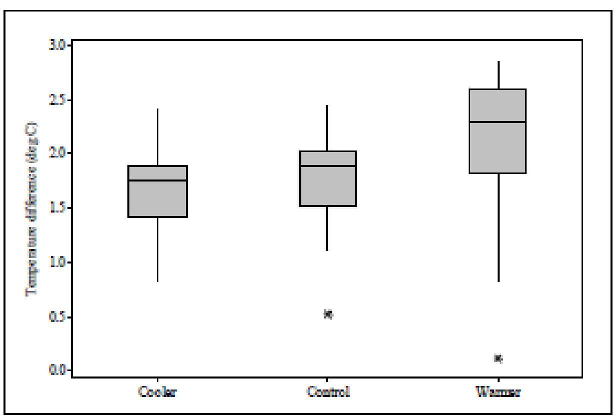
.51 → f_s = 2.015 – 1.65 ~~= 0.365°C~~

Warmer: lower fourth = 1.91, upper fourth = 2.60,

$f_s = 2.60 - 1.91 = 0.69$°C

The fourth spreads do <u>not</u> communicate the same message as the standard deviations did. The fourth spreads indicate that variability is quite similar under the cooler and control settings, while variability is much larger under the warmer setting. The disparity between the results of **b** and **c** can be partly attributed to the skewness and outliers in the data, which unduly affect the standard deviations.

d. As noted earlier, the temperature difference distributions are negatively skewed under all three conditions. The control and warmer data sets each have a single outlier. The boxplots confirm that median temperature difference increases as we transition from cooler to warmer, that cooler and control variability are similar, and that variability under the warmer condition is quite a bit larger.

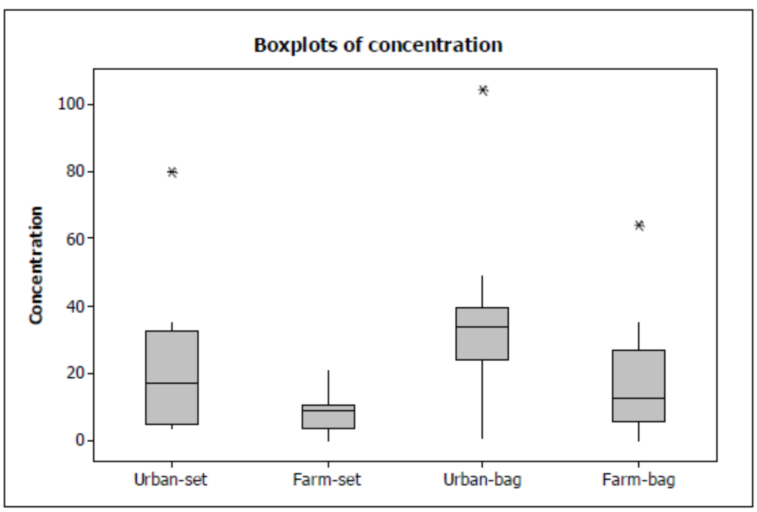


48.

a. Using the sums provided for urban homes, $S_{xx} = 10{,}079 - (237.0)^2/11 = 4972.73$, so $s = \sqrt{\dfrac{4972.73}{11-1}} = 22.3$ EU/mg. Similarly for farm homes, $S_{xx} = 518.836$ and $s = 6.09$ EU/mg. The endotoxin concentration in an urban home "typically" deviates from the average of 21.55 by about 22.3 EU/mg. The endotoxin concentration in a farm home "typically" deviates from the average of 8.56 by about 6.09 EU/mg. (These interpretations are very loose, especially since the distributions are not symmetric.) In any case, the variability in endotoxin concentration is far greater in urban homes than in farm homes.

b. The upper and lower fourths of the urban data are ~~28.0~~ 33 and ~~5.5~~ 5, respectively, for a fourth spread of ~~22.5~~ 28 EU/mg. The upper and lower fourths of the farm data are ~~10.1~~ 11 and 4, respectively, for a fourth spread of ~~6.1~~ 7 EU/mg. Again, we see that the variability in endotoxin concentration is much greater for urban homes than for farm homes.

c. Consider the box plots below. The endotoxin concentration in urban homes generally exceeds that in farm homes, whether measured by settled dust or bag dust. The endotoxin concentration in bag dust generally exceeds that of settled dust, both in urban homes and in farm homes. Settled dust in farm homes shows far less variability than any other scenario.



Q$_1$'s & Q$_3$'s a bit off

52. Let d denote the fifth deviation. Then $.3 + .9 + 1.0 + 1.3 + d = 0$ or $3.5 + d = 0$, so $d = -3.5$. One sample for which these are the deviations is $x_1 = 3.8$, $x_2 = 4.4$, $x_3 = 4.5$, $x_4 = 4.8$, $x_5 = 0$. (These were obtained by adding 3.5 to each deviation; adding any other number will produce a different sample with the desired property.)

58. The most noticeable feature of the comparative boxplots is that machine 2's sample values have considerably more variation than does machine 1's sample values. However, a typical value, as measured by the median, seems to be about the same for the two machines. The only outlier that exists is from machine 1.

76. The measures that are sensitive to outliers are: the mean and the midrange. The mean is sensitive because all values are used in computing it. The midrange is sensitive because it uses only the most extreme values in its computation.

The median, the trimmed mean, and the midfourth are not sensitive to outliers.

The median is the most resistant to outliers because it uses only the middle value (or values) in its computation.

The trimmed mean is somewhat resistant to outliers. The larger the trimming percentage, the more resistant the trimmed mean becomes.

The midfourth, which uses the quartiles, is reasonably resistant to outliers because both quartiles are resistant to outliers.

78.

a. Since the constant $\bar{x}$ is subtracted from each x value to obtain each y value, and addition or subtraction of a constant doesn't affect variability, $s_y^2 = s_x^2$ and $s_y = s_x$.

b. Let $c = 1/s$, where s is the sample standard deviation of the x's (and also, by part (a), of the y's). Then $z_i = cy_i \Rightarrow s_z^2 = c^2 s_y^2 = (1/s)^2 s^2 = 1$ and $s_z = 1$. That is, the "standardized" quantities $z_1, \ldots, z_n$ have a sample variance and standard deviation of 1.