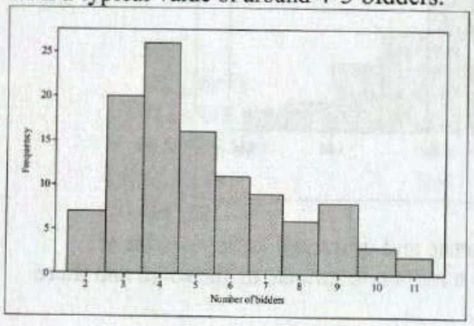
17. The sample size for this data set is $n = 7 + 20 + 26 + \ldots + 3 + 2 = 108$.

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.

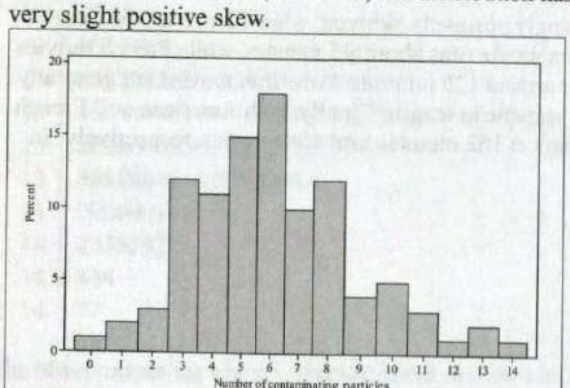


19.

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.



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).

35. The sample size is $n = 15$.

a. The sample mean is $\bar{x} = 18.55/15 = 1.237$ μg/g and the sample median is $\tilde{x}$ = the 8th ordered value = .56 μg/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 + … + 3.07)/13 = 1.162. Similarly, a 2/15 trimmed mean is the average of the middle 11 values: (.25 + … + 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 μg/g.

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.

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.

39.

a. $\Sigma x_i = 16.475$ so $\bar{x} = \dfrac{16.475}{16} = 1.0297$; $\tilde{x} = \dfrac{(1.007 + 1.011)}{2} = 1.009$

b. 1.394 can be decreased until it reaches 1.011 (i.e. by 1.394 – 1.011 = 0.383), the largest of the 2 middle values. If it is decreased by more than 0.383, the median will change.

41.

a. $x/n = 7/10 = .7$

b. $\overline{x} = .70$ = the sample proportion of successes

c. To have x/n equal .80 requires $x/25 = .80$ or $x = (.80)(25) = 20$. There are 7 successes (S) already, so another $20 - 7 = 13$ would be required.

43. The median and certain trimmed means can be calculated, while the mean cannot — the exact values of the "100+" observations are required to calculate the mean. $\tilde{x} = \frac{(57+79)}{2} = 68.0$,

$\overline{x}_{tr(20)} = 66.2$, $\overline{x}_{tr(30)} = 67.5$.

Section 1.4

45.

a. $\overline{x} = 115.58$. The deviations from the mean are $116.4 - 115.58 = .82$, $115.9 - 115.58 = .32$, $114.6 - 115.58 = -.98$, $115.2 - 115.58 = -.38$, and $115.8 - 115.58 = .22$. Notice that the deviations from the mean sum to zero, as they should.

b. $s^2 = [(.82)^2 + (.32)^2 + (-.98)^2 + (-.38)^2 + (.22)^2]/(5-1) = 1.928/4 = .482$, so $s = .694$.

c. $\Sigma x_i^2 = 66795.61$, so $s^2 = S_{xx}/(n-1) = \left(\Sigma x_i^2 - (\Sigma x_i)^2/n\right)/(n-1) =$ $(66795.61 - (577.9)^2/5)/4 = 1.928/4 = .482$.

d. The new sample values are: 16.4 15.9 14.6 15.2 15.8. While the new mean is 15.58, all the deviations are the same as in part (a), and the variance of the transformed data is identical to that of part (b).

47.

a. From software, $\tilde{x} = 14.7\%$ and $\overline{x} = 14.88\%$. The sample average alcohol content of these 10 wines was 14.88%. Half the wines have alcohol content below 14.7% and half are above 14.7% alcohol.

b. Working long-hand, $\Sigma(x_i - \overline{x})^2 = (14.8 - 14.88)^2 + \ldots + (15.0 - 14.88)^2 = 7.536$. The sample variance equals $s^2 = \Sigma(x_i - \overline{x})^2 = 7.536/(10-1) = 0.837$.

c. Subtracting 13 from each value will not affect the variance. The 10 new observations are 1.8, 1.5, 3.1, 1.2, 2.9, 0.7, 3.2, 1.6, 0.8, and 2.0. The sum and sum of squares of these 10 new numbers are $\Sigma y_i = 18.8$ and $\Sigma y_i^2 = 42.88$. Using the sample variance shortcut, we obtain $s^2 = [42.88 - (18.8)^2/10]/(10-1) = 7.536/9 = 0.837$ again.

49.

a. $\Sigma x_i = 2.75 + \cdots + 3.01 = 56.80$, $\Sigma x_i^2 = 2.75^2 + \cdots + 3.01^2 = 197.8040$

b. $$s^2 = \frac{197.8040 - (56.80)^2/17}{16} = \frac{8.0252}{16} = .5016, \quad s = .708$$

51.

a. From software, $s^2 = 1264.77$ min^2 and $s = 35.56$ min. Working by hand, $\Sigma x = 2563$ and $\Sigma x^2 = 368501$, so

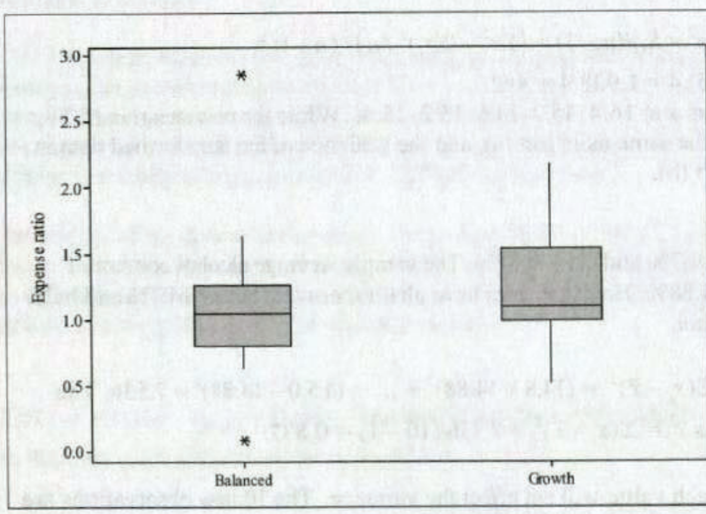
$$s^2 = \frac{368501-(2563)^2/19}{19-1} = 1264.766 \text{ and } s = \sqrt{1264.766} = 35.564$$

b. If y = time in hours, then $y = cx$ where $c = \frac{1}{60}$. So, $s_y^2 = c^2 s_x^2 = \left(\frac{1}{60}\right)^2 1264.766 = .351\text{ hr}^2$ and $s_y = cs_x = \left(\frac{1}{60}\right)35.564 = .593\text{ hr}$.

53.

a. Using software, for the sample of balanced funds we have $\overline{x} = 1.121, \tilde{x} = 1.050, s = 0.536$; for the sample of growth funds we have $\overline{x} = 1.244, \tilde{x} = 1.100, s = 0.448$.

b. The distribution of expense ratios for this sample of balanced funds is fairly symmetric, while the distribution for growth funds is positively skewed. These balanced and growth mutual funds have similar median expense ratios (1.05% and 1.10%, respectively), but expense ratios are generally higher for growth funds. The lone exception is a balanced fund with a 2.86% expense ratio. (There is also one unusually low expense ratio in the sample of balanced funds, at 0.09%.)



55.

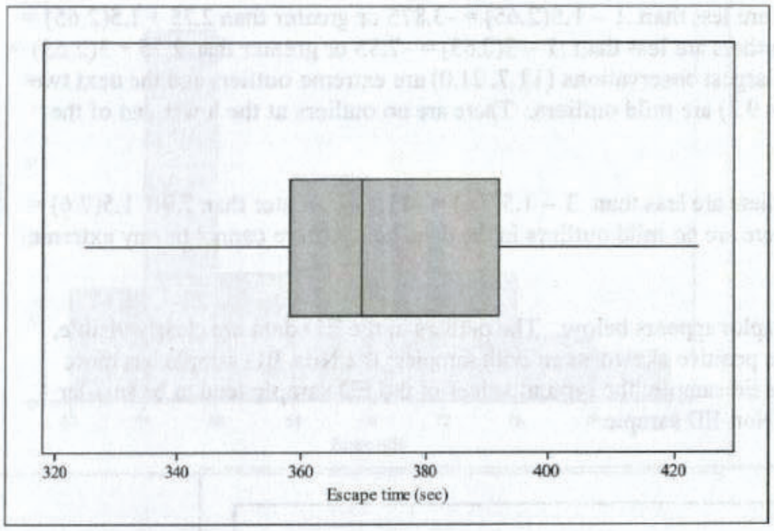
a. Lower half of the data set: 325 325 334 339 356 356 359 359 363 364 364 366 369, whose median, and therefore the lower fourth, is 359 (the 7th observation in the sorted list).

Upper half of the data set: 370 373 373 374 375 389 392 393 394 397 402 403 424, whose median, and therefore the upper fourth is 392.

So, $f_s = 392 - 359 = 33$.

b. inner fences: $359 - 1.5(33) = 309.5$, $392 + 1.5(33) = 441.5$
To be a mild outlier, an observation must be below 309.5 or above 441.5. There are none in this data set. Clearly, then, there are also no extreme outliers.

c. A boxplot of this data appears below. The distribution of escape times is roughly symmetric with no outliers. Notice the box plot "hides" the fact that the distribution contains two gaps, which can be seen in the stem-and-leaf display.



d. Not until the value $x = 424$ is lowered below the upper fourth value of 392 would there be any change in the value of the upper fourth (and, thus, of the fourth spread). That is, the value $x = 424$ could not be decreased by more than $424 - 392 = 32$ seconds.

57.

a. $f_s = 216.8 - 196.0 = 20.8$
inner fences: $196 - 1.5(20.8) = 164.6$, $216.8 + 1.5(20.8) = 248$
outer fences: $196 - 3(20.8) = 133.6$, $216.8 + 3(20.8) = 279.2$
Of the observations listed, 125.8 is an extreme low outlier and 250.2 is a mild high outlier.

b. A boxplot of this data appears below. There is a bit of positive skew to the data but, except for the two outliers identified in part (a), the variation in the data is relatively small.