Homework 1 Solutions
Chapter 1: # 10; Chapter 2: # 9, 27
Chapter 2: #42ac, 51, 61, 75
Chapter 2: #81, 84, 99

Assigned Friday, September 24

1.10 **a.** This statement was based on an observational study. It would not be ethical to randomly assign women to breast feed their babies or not. The researchers observed who was breast-fed and who was not.
b. The better headline is "Link Found Between Breast-feeding and School Performance" because it does not imply that there is a cause-and-effect conclusion, while the other headline does. Because the study was observational, a cause-and-effect conclusion cannot be made. It is likely that children who were breast-fed as infants have other (confounding) factors in their lives that differ from children who were not, and those may influence school performance. For example, perhaps they are more likely to have mothers who don't work, perhaps they are more likely to be first-born children, and so on. We can say that a link was found, but cannot say that breast-feeding *leads* to better performance.

2.9 **a.** The two variables are treatment used (placebo or aspirin) and whether individual died from a heart attack or not.
b. Observational units are male physicians between 40 and 84 years old.
c. Sample size is $n = 22{,}071$.
d. Sample data. They used the data to make generalizations about a larger population.

2.27 **a.**

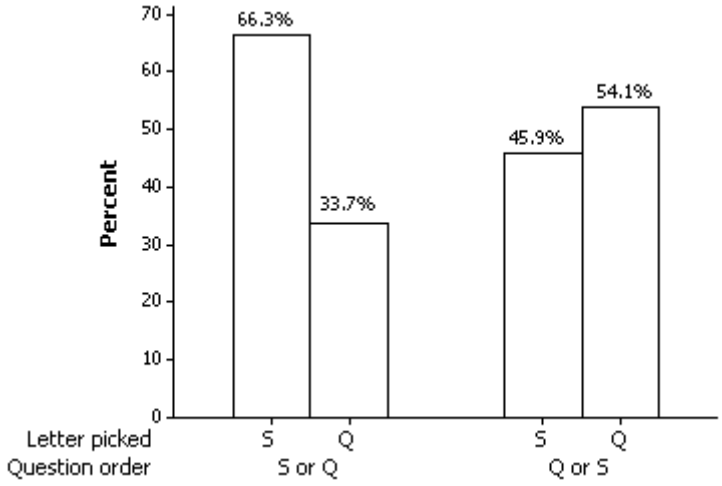
	Picked S	**Picked Q**	**Total**
S listed first	61	31	92
Q listed first	45	53	98
Total	106	84	190

b. Picked S $= (61/92) \times 100\% = 66.3\%$; Picked Q $= (31/92) \times 100\% = 33.7\%$;

c. Picked S $= (45/98) \times 100\% = 45.9\%$; Picked Q $= (53/98) \times 100\% = 54.1\%$;

d.

Figure for Exercise 2.27d



e. Parts (b) and (c) show that the percentage picking S was higher when S was listed first than when Q was listed first. It looks like the letter picked was influenced by the letter listed first.

Assigned Mon, Sept 27

2.42 **a.** In the figure shown here, two stems have been used for each possible "tens" place in the number (values under 10 are an exception because the lowest value is about 6 inches). We rounded the 1995 total of 24.5 inches up to 25. You could also use five stems for each "tens" place. And you do not need to put the values in order within each stem value.

Figure for Exercise 2.42a

```
|0| 689
|1| 11111122233444
|1| 566667777779
|2| 011
|2| 5555778889
|3| 0011
|3| 7
```

c. The data are skewed (but only slightly) to the right and appear to be bimodal. (This is especially clear in the histogram in the book.)

2.51 100 is a large value compared to the rest of the values; it causes the mean to increase, while not affecting the median.

2.61 There are $n = 47$ values so the median is 24^{th} value in the ordered data (23 will be smaller and 23 will be larger). The lower quartile is the median of the smallest 23 values and the upper quartile is the median of the largest 23 values. These will be, respectively, the 12^{th} lowest and the 12^{th} highest values in the data set. The five-number summary is:

	Rainfall (inches)		
Median		16.72	
Quartiles	12.05		25.37
Extremes	6.14		37.42

NOTE: If you use R Commander to create the five-number summary you get slightly different values for the quartiles: Q1 = 12.3 and Q3 = 25.215 because it uses a more complicated method.
The five-number summary above shows that the median annual rainfall for Davis, California is 16.72 inches. The middle ½ of the values are between 12.05 and 25.37 inches. The minimum is 6.14 inches and the maximum is 37.42 inches.

2.75 Most likely a mistake was made when the data were entered. For instance, it's possible that the value should have been 71 instead of 17. If possible, the instructor should correct the value (by looking again at the student's survey form). If the correct height is not available, the value 17 should be deleted from the dataset.

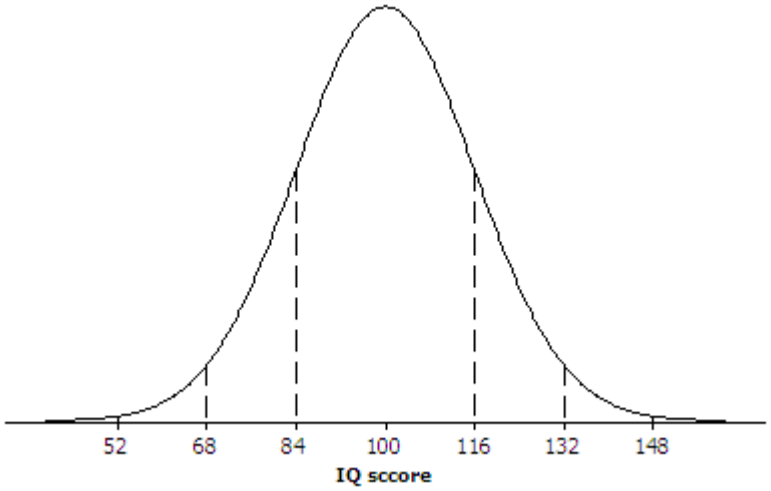
Assigned Wed, Sept 29

2.81 The interval is –6.1 to 22.7 hours. The interval includes negative values, which are impossible times. Thus, the interval based an assumption of a bell-shaped curve would not reflect reality.

2.84 **a.** About 68% fall in the interval Mean ± St. Dev, which is 100 ± 16, or 84 to 116.
About 95% fall in the interval Mean ± 2 St. Dev, which is 100 ± (2)(16), or 68 to 132.
About 99.7% fall in the interval Mean ± 3 St. Dev, which is 100 ± (3)(16), or 52 to 148.

b.

Figure for Exercise 2.84b



c. $s^2 = 16^2 = 256$ **(**Variance = squared standard deviation.)

2.99 You should be more satisfied if the standard deviation was 5. This would mean you scored 2 standard deviations above the mean and, if scores are bell-shaped, only about 2.5% of students are expected to score higher. If the standard deviation was 15, your *z*-score would be 10/15 or 2/3, indicating that you scored only 2/3 of a standard deviation above the mean.