34. a) We know that 95% of the observations for a Normal model fall within 2 SDs of the mean. That corresponds to 23.84 − 2(3.56) = 16.72 mph and 23.84 + 2(3.56) = 30.96 mph.
b) The actual 97.5% and 2.5% tiles are 30.976 and 16.638, respectively. These are very close to the predicted values of 30.96 and 16.72 mph. The histogram is roughly unimodal and symmetric. It is very slightly right skewed and there is one outlier, but the Normal probability plot is quite straight. We should not be surprised that the approximation is good.

35. a) 2.5%
b) 2.5% of the receivers should gain less than −333 yards, but that's impossible, so the model doesn't fit well.
c) Data are strongly skewed to the right, not symmetric.

36. a) Median because the distribution is so skewed to the left.
b) IQR. Distribution is skewed.
c) 68% d) More than 75%
e) Normal model is not appropriate. Data are strongly skewed.

37. a) 12.2% b) 71.6% c) 23.3%

38. a) 89.4% b) 26.6% c) 20.4%

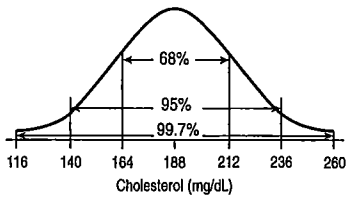
39. a) 1259.7 lb b) 1081.3 lb c) 1108 lb to 1196 lb

40. a) 126.3 b) 91.6 c) 79.5 to 120.5

41. a) 1130.7 lb b) 1347.4 lb c) 113.3 lb

42. a) 83.4 b) 132.9 c) 21.6

43. a)



b) 30.85% c) 17.00% d) 32 points e) 212.9 points

44. a) No, that's more than 3 SDs above the mean.
b) 21.2% c) 67.3%
d) Quartiles at 30,314 and 33,686 miles, so 3372 miles.
e) 27,623 miles

45. a) 11.1% b) (35.9, 40.5) inches c) 40.5 inches

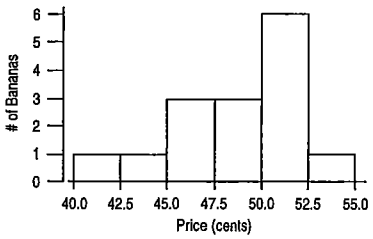
46. a) Based on the Normal model, we expect 95% to be between 96.8 and 99.6°F.
b) 28.4% c) 97.6°F

47. a) 5.3 grams b) 6.4 grams
c) Younger because SD is smaller.

48. a) 3.26 grams b) 75.70 grams c) 2.86 grams
d) The new tomatoes are more consistent in their weights.

PART I REVIEW

1. a)



b) Median 49 cents, IQR 6 cents.
c) The distribution is unimodal and left skewed. The center is near 50 cents; values range from 42 cents to 53 cents.

2. a) It is (rounded to 1 decimal place), but there's no reason it should be unless the number of women receiving each type of care was roughly the same.
b) Yes, but they do not prove that adequate prenatal care is important for pregnant women. The mortality rate is quite a bit lower for women with adequate care than for other women, but there may be a lurking variable.
c) Intensive care is given for emergency conditions. The data do *not* suggest that the care is the *cause* of the higher mortality.

3. a) If enough sopranos have a height of 65 inches, this can happen.
b) The distribution of heights for each voice part is roughly symmetric. The basses are slightly taller than the tenors. The sopranos and altos have about the same median height. Heights of basses and sopranos are more consistent than those of altos and tenors.

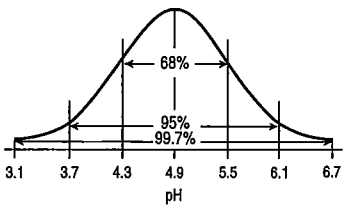
4. With only 3 cases, probably no display is appropriate.

5. a) It means their heights are also more variable.
b) The *z*-score for women to qualify is 2.40, compared with 1.75 for men, so it is harder for women to qualify.

6. a) The distribution is unimodal and skewed to the right. The mode is near 100, and values range from 95 to 140.
b) The mean will be larger than the median, since the distribution is right skewed.
c) Create a boxplot with quartiles at 97 and 105.5, and median at 100. The IQR is 8.5, so the upper fence is at (1.5 × 8.5) + 105.5 = 118.25. There are several outliers to the right. There are no outliers to the left because the minimum at 95 lies well within the left fence at 97 − (1.5 × 8.5) = 84.25.
d) No, the Normal model is not appropriate for these data. They are unimodal but not symmetric.

7. a) *Who*—People who live near State University
What—Age, attended college? Favorable opinion of State?
When—Not stated
Where—Region around State U.
Why—To report to the university's directors
How—Sampled and phoned 850 local residents
b) Age—Quantitative (years); attended college?—categorical; favorable opinion?—categorical.
c) The fact that the respondents know they are being interviewed by the university's staff may influence answers.

8. a)

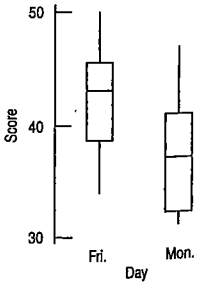


b) 3.3% c) 6.7%
d) pH 4.40 e) pH 5.89
f) Quartiles at 4.50 and 5.30, so IQR is 0.80.

9. a) These are categorical data, so mean and standard deviation are meaningless.
b) Not appropriate. Even if it fits well, the Normal model is meaningless for categorical data.

10. a) Stream name—categorical; substrate—categorical; pH—quantitative; temperature—quantitative (°C); BCI—quantitative (units missing).
b) Bar chart or pie chart.

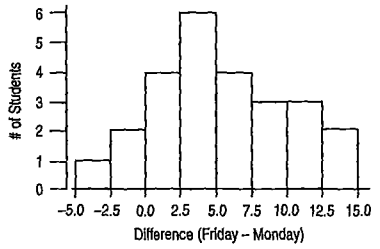
11. a)



b) The scores on Friday were higher by about 5 points on average. This is a drop of more than 10% off the average score and shows that students fared worse on Monday after prep

for the test on Friday. The spreads are about the same, but the scores on Monday are a bit skewed to the right.

c)

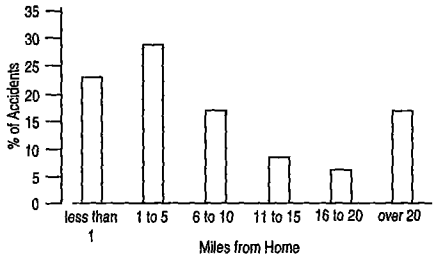


d) The changes (Friday–Monday) are unimodal and centered near 4 points, with a spread of about 5 (SD). They are fairly symmetric, but slightly skewed to the right. Only 3 students did better on Monday (had a negative difference).

12. No, you can't add these, since the groups are not disjoint.

13. a) Categorical
b) Go fish. All you need to do is match the denomination. The denominations are not ordered. (Answers will vary.)
c) Gin rummy. All cards are worth their value in points (face cards are 10 points). (Answers will vary.)

14. a)



b) We are given no information about how many miles are driven in each of these categories, so we have no idea how many accidents to *expect*. We also have no information about how many accidents were involved in compiling the data.

15. a) Annual mortality rate for males (quantitative) in deaths per 100,000 and water hardness (quantitative) in parts per million.
b) Calcium is skewed right, possibly bimodal. There looks to be a mode down near 12 ppm that is the center of a fairly tight symmetric distribution and another mode near 62.5 ppm that is the center of a much more spread out, symmetric (almost uniform) distribution. Mortality, however, appears unimodal and symmetric with the mode near 1500 deaths per 100,000.

16. a) Overall mean is $(34 \times 1631.59 + 27 \times 1388.85)/(34 + 27) = 1524.15$ deaths per 100,000.
b) Yes. Mortality for the Northern towns is generally higher than that for the South. Fully half of the towns in the South have mortality rates lower than the rates of *any* of the Northern towns. About 25% of Northern towns have rates higher than the rates of all of the Southern towns.

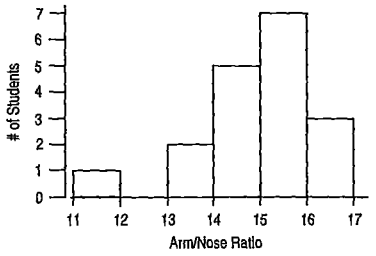
17. a) They are on different scales.
b) January's values are lower and more spread out.
c) Roughly symmetric but slightly skewed to the left. There are more low outliers than high ones. Center is around 40 degrees with an IQR of around 7.5 degrees.

18. Bimodal with modes around 50 and 80 minutes. Fairly symmetric around each mode.

19. a) Bimodal with modes near 2 and 4.5 minutes. Fairly symmetric around each mode.
b) Because there are two modes, which probably correspond to two different groups of eruptions, an average might not make sense.
c) The intervals between eruptions are longer for long eruptions. There is very little overlap. More than 75% of the short eruptions had intervals less than about an hour (62.5 minutes), while more than 75% of the long eruptions had intervals longer than about 75 minutes. Perhaps the interval could even be used to predict whether the next eruption will be long or short.

20. The chance of an accident is not the same for different age groups. The distribution of ages for drivers involved in fatal crashes is not the same as that for other drivers. So, the two variables are not independent.

21. a)

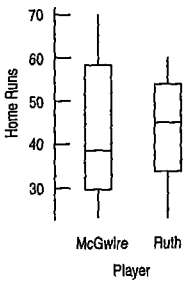


The distribution is left skewed with a center of about 15. It has an outlier between 11 and 12.
b) Even though the distribution is somewhat skewed, the mean and median are close. The mean is 15.0 and the SD is 1.25.
c) Yes. 11.8 is already an outlier. 9.3 is more than 4.5 SDs below the mean. It is a very low outlier.

22. a) About 38% b) 16%
c) Data are bimodal, and the Normal model is inappropriate.

23. If we look only at the overall statistics, it appears that the follow-up group is insured at a much lower rate than those not traced (11.1% of the time compared with 16.6%). But most of the follow-up group were black, who have a lower rate of being insured. When broken down by race, the follow-up group actually has a higher rate of being insured for both blacks and whites. So the overall statistic is misleading and is attributable to the difference in race makeup of the two groups.

24. a) 3, 25.5, 36, 50.5, 70
b) Because the IQR is so large, none are technically outliers, but the seasons with fewer than 20 home runs stand out as a separate group.
c)



d) Without the injured seasons, McGwire and Ruth's home run production distributions look similar. (Ruth's seasons as a pitcher were not included as well.) Ruth's median is a little higher, and he was a little more consistent (less spread), but McGwire had the two highest season totals.

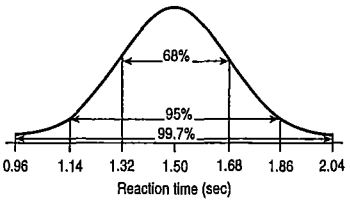
e)

	Stem and Leaf		
	0 | 7		
	5 | 6	0	
	82 | 5	449	
	92 | 4	1166679	
McGwire	9932 | 3	45	Ruth
	2 | 2	25	

7|0 = 70 home runs

f) Now we can see how much more consistent Ruth was. Most of Ruth's seasons had home run totals in the 40s or 50s. McGwire's seasons are much more spread out (not even including the three at the bottom).

25. a)



b) According to the model, reaction times are symmetric with center at 1.5 seconds. About 95% of all reaction times are between 1.14 and 1.86 seconds.

c) 8.2% d) 24.1%

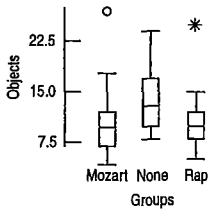
e) Quartiles are 1.38 and 1.62 seconds, so the IQR is 0.24 seconds.

f) The slowest 1/3 of all drivers have reaction times of 1.58 seconds or more.

26. a) *Who*—62 people
What—Number of objects correctly memorized
When—Not stated
Where—Not stated
Why—To see if music affects memorization ability
How—Randomized experiment

b) Type of music (categorical) and number of items remembered (quantitative).

c) Because we do not have all the data, we can't know exactly how the boxplots look, but we do know that the minimums are all within the fences and that two groups have at least one outlier on the high side.



d) All three distributions are right skewed. Mozart and rap had very similar distributions. The scores for the None groups are, if anything, slightly higher than those for the other two groups. It is clear that groups listening to music did not score higher than those who heard none.

27. a)

b) Mean 100.25, SD 25.54 pieces of mail.

c) The distribution is somewhat symmetric and unimodal, but the center is rather flat, almost uniform.

d) 64%. The Normal model seems to work reasonably well, since it predicts 68%.

28. a) 50.7% b) 34.9%
c) 13.3% d) 6.7%

29. a) *Who*—100 health food store customers
What—Have you taken a cold remedy?, and Effectiveness (scale 1 to 10)
When—Not stated
Where—Not stated
Why—Promotion of herbal medicine
How—In-person interviews

b) Have you taken a cold remedy?—categorical. Effectiveness—categorical or ordinal.

c) No. Customers are not necessarily representative, and the Council had an interest in promoting the herbal remedy.

30. a) Math/Sci—25.6%, Ag—41.7%, Humanities—19.3%, Other—13.5%.

b) Math/Sci—30.1%, Ag—46.0%, Humanities—13.3%, Other—10.6%.

c) Math/Sci—20.3%, Ag—39.1%, Humanities—24.6%, Other—15.9%.

d) No; it appears that oldest children are more likely than second-born children to major in Math or Science (30% to 20%), while second-born children are more likely than first-born to major in the Humanities (25% to 13%)

31. a) 38 cars

b) Possibly because the distribution is skewed to the right.

c) Center—median is 148.5 cubic inches. Spread—IQR is 126 cubic inches.

d) No. It's bigger than average, but smaller than more than 25% of cars. The upper quartile is at 231 inches.

e) No. 1.5 IQR is 189, and 105 − 189 is negative, so there can't be any low outliers. 231 + 189 = 420. There aren't any cars with engines bigger than this, since the maximum has to be at most 105 (the lower quartile) + 275 (the range) = 380.

f) Because the distribution is skewed to the right, this is probably not a good approximation.

g) Mean, median, range, quartiles, IQR, and SD all get multiplied by 16.4.

32. a) Fairly symmetric, almost uniform except for the right tail.

b) 47 horses

c) IQR is 47, so fences are at 7.5 and 195.5. Since the range is only 90, we know there can't be any observations lower than 35 (upper quartile − range) or greater than 168 (lower quartile + range), so no.

d) Distribution is very roughly uniform, not unimodal, so the Normal modal might not be reasonable.

e) 22 of 38, or about 58%.

f) Mean, median, and quartiles would increase by 10. SD, IQR, and range would not change.

33. a) 30.4%

b) If this were a random sample of all voters, yes.

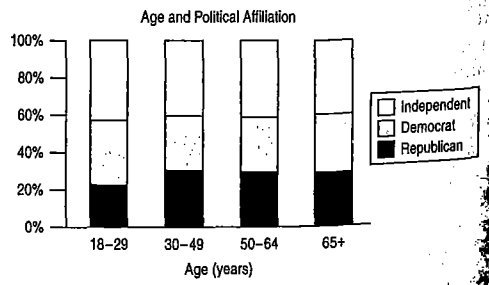
c) 36.6% d) 8.8%

e) 23.1% f) 47.0%

34. Chief executives have a mean salary less than the median, so the distribution is likely to be skewed to the left. General and operations managers have a mean salary larger than the median, so their distribution is likely to be skewed to the right.

35. a) Republican—16,535, Democrat—17,183, Other— 20,666; or Republican—30.4%, Democrat—31.6%, Other—38.0%.

b)



c) Among voters over 30, political affiliation appears to be largely unrelated to age. However there is some evidence that younger voters are less likely to be Republican

d) Voters who identified themselves as "Other" seem to be generally younger than Democrats or Republicans.

36. a) *Who*—Years from 1994 to 2003
What—Bicycle fatalities
When—1994–2003
Where—United States

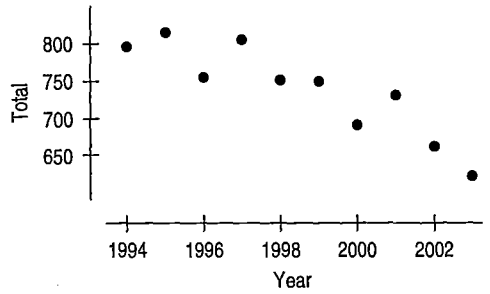
Why—To study bicycle helmet safety
How—Bicycle Helmet Safety Institute Report

b)

Stem	Leaf
6	1
6	68
7	2
7	5569
8	12

6/1 = 610 – 619 fatalities

c)



d) The stem-and-leaf display shows the distribution is skewed to the left. It also provides some idea about the center and spread of the annual fatalities.

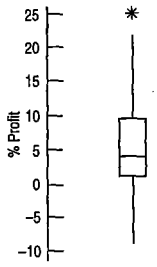
e) The number of bicycle fatalities has tended to decrease over the 10-year period.

f) In the 10-year period from 1994 to 2003, reported bicycle fatalities decreased fairly steadily from about 800 per year to around 620 a year.

37. a) 0.43 hours. b) 1.4 hours.
c) 0.89 hours (or 53.4 minutes).
d) Survey results vary, and the mean and the SD may have changed.

38. a) −9, 1, 4, 9, 25

b)



c) Mean 4.72% of sales, SD 7.55% of sales.

d) Fairly symmetric and unimodal, centered around 4% of sales. 50% of the companies report % profit between 1% and 9%. There are two outliers at 22% and 25% of sales.

CHAPTER 7

1. a) Weight in ounces: explanatory; Weight in grams: response. (Could be other way around.) To predict the weight in grams based on ounces. Scatterplot: positive, straight, strong (perfectly linear relationship).

b) Circumference: explanatory. Weight: response. To predict the weight based on the circumference. Scatterplot: positive, linear, moderately strong.

c) Shoe size: explanatory; GPA: response. To try to predict GPA from shoe size. Scatterplot: no direction, no form, very weak.

d) Miles driven: explanatory; Gallons remaining: response. To predict the gallons remaining in the tank based on the miles driven since filling up. Scatterplot: negative, straight, moderate.

2. a) Price: explanatory; Number sold: response. To predict the number sold based on the price. Scatterplot: negative, straight, moderate.

b) Depth: explanatory; Water pressure: response. To predict water pressure based on depth. Scatterplot: positive, straight, strong.

c) Visibility: explanatory; Depth: response. To predict depth based on visibility (although predicting visibility based on depth is also possible). Scatterplot: negative, straight(?), weak to moderate.

d) Weight: explanatory; Reading score: response. To predict reading test scores based on weight. Scatterplot: positive, possibly straight, moderate.

3. a) Altitude: explanatory; Temperature: response. (Other way around possible as well.) To predict the temperature based on the altitude. Scatterplot: negative, possibly straight, weak to moderate.

b) Ice cream cone sales: explanatory. Air-conditioner sales: response—although the other direction would work as well. To predict one from the other. Scatterplot: positive, straight, moderate.

c) Age: explanatory; Grip strength: response. To predict the grip strength based on age. Scatterplot: curved down, moderate. Very young and elderly would have grip strength less than that of adults.

d) Reaction time: explanatory; Blood alcohol level: response. To predict blood alcohol level from reaction time test. (Other way around is possible.) Scatterplot: positive, nonlinear, moderately strong.

4. a) Time: explanatory; Cost: response. To predict cost based on time. Scatterplot: positive, straight, strong.

b) Time delay: explanatory; Distance: response. To predict the distance from the lightning based on the time delay of the thunder. Scatterplot: positive, straight, strong.

c) Brightness: explanatory; Distance: response. To predict distance based on apparent brightness. Scatterplot: negative, curved, moderate.

d) Weight of car: explanatory; Age of owner: response. To predict the age of the owner based on the weight of the car. (Or other way around.) Scatterplot: no direction, no shape, very weak.

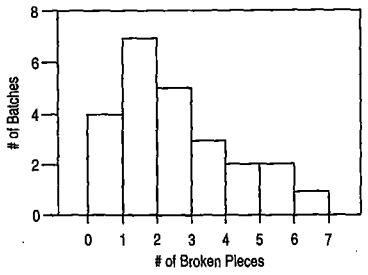
5. a) None b) 3 and 4 c) 2, 3, and 4
d) 1 and 2 e) 3 and possibly 1

6. a) 1 b) 4 c) 2 and 4 d) 3 e) 2 and 4

7. There seems to be a very weak—or possibly no—relation between brain size and performance IQ.

8. Nonlinear form. Moderately strong. The rate has not been constant; the rate of increase from the beginning to about 1950 is steeper than from 1950 to the present. One winner in the early 1890s was quite slow.

9. a)



b) Unimodal, skewed to the right. The skew.

c) The positive, somewhat linear relation between batch number and broken pieces.

10. a)