DISPLAY 3.11 Partial listing of Connecticut skin cancer rates (per 100,000 people) from 1938 to 1972, with solar code (1 if there was higher than average sunspot activity and 2 if there was lower than average sunspot activity two years earlier)

Year	Rate	Code
1938	0.8	2
1939	1.3	1
1940	1.4	1
1941	1.2	1
. . .		
1972	4.8	1

DISPLAY 3.12 Proportions of pollen removed and visit durations (in seconds) by 35 bumblebee queens and 12 honeybee workers; partial listing.

Bee	Type	Removed	Duration
1	queen	0.07	2
2	queen	0.10	5
3	queen	0.11	7
4	queen	0.12	11
. . .			
45	worker	0.78	51
46	worker	0.74	64
47	worker	0.77	78

25. Agent Orange. With a statistical computer program, reanalyze the Agent Orange data of Display 3.3 with and without the two largest dioxin levels in the Vietnam veterans group. Verify the one-sided p-values in bubble 2 of Display 3.7.

26. Agent Orange. With a statistical computer package, reanalyze the Agent Orange data of Display 3.3 after taking a log transformation. Since the data set contains zeros—for which the log is undefined—try the transformation log(dioxin + .5). (a) Draw side-by-side box plots of the transformed variable. (b) Find a p-value from the t-test for comparing the two distributions. (c) Compute a 95% confidence interval for the difference in mean log measurements and interpret it on the original scale. (*Note*: Back-transforming does not provide an exact estimate of the ratio of medians since 0.5 was added to the dioxins, but it does provide an approximate one.)

27. Pollen Removal. As part of a study to investigate reproductive strategies in plants, biologists recorded the time spent at sources of pollen and the proportions of pollen removed by bumblebee queens and honeybee workers pollinating a species of lily. (Data from L. D. Harder and J. D. Thompson, "Evolutionary Options for Maximizing Pollen Dispersal of Animal-pollinated Plants," *American Naturalist* 133 (1989): 323–44.) Their data appear in Display 3.12.

(a) (i) Draw side-by-side box plots (or histograms) of the proportion of pollen removed by queens and workers. (ii) When the measurement is the proportion P of some amount, one useful transformation is $\log[P/(1-P)]$. This is the log of the ratio of the proportion removed to the proportion not removed. Draw side-by-side box plots or histograms on

Copyright 2012 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.

this transformed scale. (iii) Test whether the distribution of proportions removed is the same or different for the two groups, using the t-test on the transformed data.

(b) Draw side-by-side box plots of duration of visit on (i) the natural scale, (ii) the logarithmic scale, and (iii) the reciprocal scale. (iv) Which of the three scales seems most appropriate for use of the t-tools? (v) Compute a 95% confidence interval to describe the difference in means on the chosen scale. (vi) What are relative advantages of the three scales as far as interpretation goes? (vii) Based on your experience with this problem, comment on the difficulty in assessing equality of population standard deviations from small samples.

28. Bumpus's Data. Obtain p-values from the t-test to compare humerus lengths for sparrows that survived and those that perished (Exercise 2.21), with and without the smallest length in the perished group (length $=$ 0.659 inch). Do the conclusions depend on this one observation? What action should be taken if they do?

29. Cloud Seeding—Multiplicative vs. Additive Effects. On the computer, create a variable containing the rainfall amounts for only the unseeded days. (a) Create four new variables by adding 100, 200, 300, and 400 to each of the unseeded day rainfall amounts. Display a set of five box plots to illustrate what one might expect if the effect of seeding were additive. (b) Create four additional variables by multiplying each of the unseeded day rainfall amounts by 2, by 3, by 4, and by 5. Display a set of five box plots to illustrate what could be expected if the effect of seeding were multiplicative. (c) Which set of plots more closely resembles the actual data?

Data Problems

30. Education and Future Income. Display 3.13 shows the first five rows of a data set with annual incomes in 2005 of the subset of National Longitudinal Survey of Youth (NLSY79) subjects (described in Exercise 2.22) who had paying jobs in 2005 and who had completed either 12 or 16 years of education by the time of their interview in 2006. All the subjects in this sample were between 41 and 49 years of age in 2006. Analyze the data to describe the amount (or percent) by which the population distribution of incomes for those with 16 years of education exceeds the distribution for those with 12 years of education. (*Note*: The NLSY79 data set codes all incomes above $150,000 as $279,816. To make an exercise version that better matches the actual income distribution, those values have been replaced in the data set by computer-simulated values from a realistic distribution of incomes greater than $150,000.)

DISPLAY 3.13 Annual incomes in 2005 (in U.S. dollars) of 1,020 Americans who had 12 years of education and 406 who had 16 years of education by the time of their interview in 2006; "Subject" is a subject identification number; first 5 of 1,426 rows

Subject	Educ	Income2005
2	12	5,500
6	16	65,000
7	12	19,000
13	16	8,000
21	16	253,043

31. Education and Future Income II. The data file ex0331 contains a subset of the NLSY79 data set (see Exercise 30) with annual incomes of subjects with either 16 or more than 16 years of

Copyright 2012 Cengage Learning. All Rights Reserved. May not be copied, scanned, or duplicated, in whole or in part. Due to electronic rights, some third party content may be suppressed from the eBook and/or eChapter(s). Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. Cengage Learning reserves the right to remove additional content at any time if subsequent rights restrictions require it.