11. A woman who has just moved to a new job in a new town discovers two routes to drive from her home to work. The first Monday, she flips a coin, deciding to take route A if it comes up heads and to take route B if it is tails. The following Monday, she will take the other route. The first Tuesday, she flips the coin again with the same plan. And so on for the first week. At the end of two weeks, she has traveled both routes five times and can compare their average commuting times. Why should she not use the t-tools for two independent samples? What should she use?

12. In which ways are the t-tools more robust for larger sample sizes than for smaller ones (i.e., robust with respect to normality, equal SDs, and/or independence)?

13. **Fish Oil.** Why is a log transformation inappropriate for the fish oil data in Exercise 1.12?

14. Will an outlier from a contaminating population be more consequential in small samples or large samples?

15. What would you suggest as an alternative estimate of the standard deviation of the difference in sample averages when it is clear that the two populations have different SDs? (Check the formula for the standard deviation of the sampling distribution of the difference in averages, in Display 2.6.)

16. A researcher has taken tissue cultures from 25 subjects. Each culture is divided in half, and a treatment is applied to one of the halves chosen at random. The other half is used as a control. After determining the percent change in the sizes of all culture sections, the researcher calculates the standard error for the treatment-minus-control difference using both the paired t-analysis and the two independent sample (Chapter 2) t-analysis. Finding that the paired t-analysis gives a slightly larger standard error (and gives only half the degrees of freedom), the researcher decides to use the results from the unpaired analysis. Is this legitimate?

17. Respiratory breathing capacity of individuals in houses with low levels of nitrogen dioxide was compared to the capacity of individuals in houses with high levels of nitrogen dioxide. From a sample of 200 houses of each type, breathing capacity was measured on 600 individuals from houses with low nitrogen dioxide and on 800 individuals from houses with high nitrogen dioxide. (a) What problem do you foresee in applying t-tools to these data? (b) Would comparing the average *household* breathing capacities avoid the problem?

18. **Trauma and Metabolic Expenditure.** The following data are metabolic expenditures for eight patients admitted to a hospital for reasons other than trauma and for seven patients admitted for multiple fractures (trauma). (Data from C. L. Long, et al., "Contribution of Skeletal Muscle Protein in Elevated Rates of Whole Body Protein Catabolism in Trauma Patients," *American Journal of Clinical Nutrition* 34 (1981): 1087–93.)

Metabolic Expenditures (kcal/kg/day)								
Nontrauma patients:	20.1	22.9	18.8	20.9	20.9	22.7	21.4	20.0
Trauma patients:	38.5	25.8	22.0	23.0	37.6	30.0	24.5	

(a) Is the difference in averages resistant? (*Hint*: What happens if 20.0 is replaced by 200?)
(b) Replacing each value with its rank, from the lowest to highest, in the combined sample gives

Metabolic Expenditures (kcal/kg/day)								
Nontrauma patients:	3	9	1	4.5	4.5	8	6	2
Trauma patients:	15	12	7	10	14	13	11	

Consider the average of the ranks for the trauma group minus the average of the ranks for the nontrauma group. Is this statistic resistant?

19. In each of the following data problems there is some potential violation of one of the independence assumptions. State whether there is a cluster effect or serial correlation, and whether the questionable assumption is the independence within groups or the independence between groups.

(a) Researchers interested in learning the effects of speed limits on traffic accidents recorded the number of accidents per year for each of 10 consecutive years on roads in a state with speed limits of 90 km/h. They also recorded the number of accidents for the next 7 years on the same roads after the speed limit had been increased to 110 km/hr. The two groups of measurements are the number of accidents per year for those years under study. (Notice that there is also a potential confounding variable here!)
(b) Researchers collected intelligence test scores on twins, one of whom was raised by the natural parents and one of whom was raised by foster parents. The data set consists of test scores for the two groups, boys raised by their natural parents and boys raised by foster parents.
(c) Researchers interested in investigating the effect of indoor pollution on respiratory health randomly select houses in a particular city. Each house is monitored for nitrogen dioxide concentration and categorized as being either high or low on the nitrogen dioxide scale. Each member of the household is measured for respiratory health in terms of breathing capacity. The data set consists of these measures of respiratory health for all individuals from houses with low nitrogen dioxide levels and all individuals from houses with high levels.

Computational Exercises

20. **Means, Medians, Logs, Ratios.** Consider the following tuitions and their natural logs for five colleges:

College	In-State	Out-of-State	Out/In Ratio	Log(In-State)	Log(Out-of-State)
A	$1,000	$ 3,000	3	6.9078	8.0064
B	$4,000	$ 8,000	2	8.2941	8.9872
C	$5,000	$ 30,000	6	8.5172	10.3090
D	$8,000	$ 32,000	4	8.9872	10.3735
E	$40,000	$ 40,000	1	10.5966	10.5966

(a) Find the average In-State tuition. Find the average log(In-State). Confirm that the log of the average is *not* the same as the average of the logs. (b) Find the median In-State tuition and the median of the logs of In-State tuitions. Verify that the log of the median *is* the same as the median of the logs. (c) Compute the median of the ratios. Compute the differences of logged tuitions—log(Out-of-State) minus log(In-State) and compute the median of these differences. Verify that the median of the differences (of log tuitions) is equal to the natural log of the median of ratios (aside from some minor rounding error).

21. **Umpire Life Lengths.** When an umpire collapsed and died soon after the beginning of the 1990 U.S. major league baseball season, there was speculation that the stress associated with that job poses a health risk. Researchers subsequently collected historical and current data on umpires to investigate their life expectancies (Cohen et al., " Life Expectancy of Major League Baseball Umpires," *The Physician and Sportsmedicine*, 28(5) (2000): 83–89). From an original list of 441 umpires, data were found for 227 who had died or had retired and were still living. Of these, dates of birth and death were available for 195. Display 3.10 shows several rows of a generated data set based on the study.