1. You read in the report of a psychology experiment: "Separate analyses for our two groups of 12 participants revealed no overall placebo effect for our student group ($\bar{x} = 0.08, s = 0.37, t(11) = 0.49$) and a significant effect for our non-student group ($\bar{x} = 0.35, s = 0.37, t(11) = 3.25, p < 0.01$)." The null hypothesis is that the mean effect is zero. What are the correct values of the two t statistics based on the means and standard deviations? Compare each correct t-value with the critical values in Table C. What can you say about the two-sided P-value in each case?

2. An NHANES report gives data for 654 women aged 20 to 29 years. The mean BMI of these 654 women was $\bar{x} = 26.8$. On the basis of this sample, we want to estimate the mean BMI μ in the population of all 20.6 million women in this age group. Assume that the population standard deviation is known to be $\sigma = 7.5$.

(a) Give a 95% z confidence interval using the above information.

(b) In fact, the sample data had a sample standard deviation $s = 7.42$. What is the 95% t confidence interval for the mean BMI of all young women?

3. The Trial Urban District Assessment (TUDA) is a government-sponsored study of student achievement in large urban school districts. TUDA gives a reading test scored from 0 to 500. A score of 243 is a "basic" reading level and a score of 281 is "proficient." Scores for a random sample of 3000 eighth-graders in Atlanta had $\bar{x} = 250$ with standard error 1.0.

(a) We don't have the 3000 individual scores, but use of the t procedures is surely safe. Why?

(b) Give a 99% confidence interval for the mean score of all Atlanta eighth-graders. (Be careful: the report gives the standard error of $\bar{x}$, not the standard deviation s.)

(c) Urban children often perform below the basic level. Is there good evidence that the mean for all Atlanta eighth-graders is less than the basic level?

4. The placebo effect is particularly strong in patients with Parkinson's disease. To understand the workings of the placebo effect, scientists measure activity at a key point in the brain when patients receive a placebo that they think is an active drug and also when no treatment is given. The same 6 patients are measured both with and without the placebo, at different times.

(a) Explain why the proper procedure to compare the mean response to placebo with control (no treatment) is a matched pairs t test.

(b) The 6 differences (treatment minus control) had $\bar{x} = -0.326$ and $s = 0.181$. Is there significant evidence of a difference between treatment and control?

5. Our bodies have a natural electrical field that is known to help wounds heal. Does changing the field strength slow healing? A series of experiments with newts investigated this question. In one experiment, the two hind limbs of 12 newts were assigned at random to either experimental or control groups. This is a matched pairs design. The electrical field in the experimental limbs was reduced to zero by applying a voltage. The control limbs were left alone. Here are the rates at which new cells closed a razor cut in each limb, in micrometers per hour:

Newt	1	2	3	4	5	6	7	8	9	10	11	12
Control Limb	36	41	39	42	44	39	39	56	33	20	49	30
Experimental Limb	28	31	27	33	33	38	45	25	28	33	47	23

(a) Make a stemplot of the differences between limbs of the same newt (control limb minus experimental limb). There is a high outlier.

(b) A good way to judge the effect of an outlier is to do your analysis twice, once with the outlier and a second time without it. Carry out two t tests to see if the mean healing rate is significantly lower in the experimental limbs, one test including all 12 newts and another that omits the outlier. What are the test statistics and their P-values? Does the outlier have a strong influence on your conclusion?

6. Hallux abducto valgus (call it HAV) is a deformation of the big toe that often requires surgery. Doctors used X-rays to measure the angle (in degrees) of deformity in 38 consecutive patients under the age of 21 who came to a medical center for surgery to correct HAV. The angle is a measure of the seriousness of the deformity. Here are the data:

$$
\begin{array}{ccccccccccccc}
28 & 32 & 25 & 34 & 38 & 26 & 25 & 18 & 30 & 26 & 28 & 13 & 20 \\
21 & 17 & 16 & 21 & 23 & 14 & 32 & 25 & 21 & 22 & 20 & 18 & 26 \\
16 & 30 & 30 & 20 & 50 & 25 & 26 & 28 & 31 & 38 & 32 & 21 &
\end{array}
$$

A good way to judge the effect of an outlier is to do your analysis twice, once with the outlier and a second time without it. The data above follow a Normal distribution quite closely except for one patient with HAV angle 50 degrees, a high outlier.

(a) It is reasonable to regard these patients as a random sample of young patients who require HAV surgery?

(b) Find the 95% confidence interval for the population mean based on the 38 patients who remain BEFORE you drop the outlier.

(c) Find the 95% confidence interval for the population mean based on the 37 patients who remain AFTER you drop the outlier. Compare your interval in (b) with this interval. What is the most important effect of removing the outlier?

7. Researchers claim that women speak significantly more words per day than men. One estimate is that a woman uses about 20,000 words per day while a man uses about 7,000. To investigate such claims, one study used a special device to record the conversations of male and female university students over a four-day period. From these recordings, the daily word count of the 20 men in the study was determined. Here are their daily word counts:

$$
\begin{array}{ccccc}
28408 & 10084 & 15931 & 21688 & 37786 \\
10575 & 12880 & 11071 & 17799 & 13182 \\
8918 & 6495 & 8153 & 7015 & 4429 \\
10054 & 3998 & 12639 & 10974 & 5255
\end{array}
$$

(a) Examine the data. Is it reasonable to use the t procedures? (Assume these men are an SRS of all male students at this university.)

(b) If your conclusion in part (a) is "Yes," do the data give convincing evidence that the mean number of words per day of men at this university differs from 7,000?

8. Velvetleaf is a particularly annoying weed in corn fields. It produces lots of seeds, and the seeds wait in the soil for years until conditions are right. How many seeds do velvetleaf plants produce? Here are counts from 28 plants that came up in a corn field when no herbicide was used:

$$
\begin{array}{cccccccccc}
2450 & 721 & 2228 & 2504 & 2114 & 1110 & 2137 & 8015 & 863 & 1136 \\
2819 & 1911 & 2101 & 363 & 5973 & 1050 & 1961 & 1809 & 1623 & 1051 \\
130 & 1531 & 218 & 880 & 2008 & 1711 & 1716 & 164 & &
\end{array}
$$

We would like to give a confidence interval for the mean number of seeds produced by velvetleaf plants. Alas, the t interval can't be safely used for these data. Why not?

9. Cola makers test new recipes for loss of sweetness during storage. Trained tasters rate the sweetness before and after storage. Here are the sweetness losses (sweetness before storage minus sweetness after storage) found by 10 tasters for one new cola recipe:

$$
2.0 \quad 0.4 \quad 0.7 \quad 2.0 \quad -0.4 \quad 2.2 \quad -1.3 \quad 1.2 \quad 1.1 \quad 2.3
$$

Take the data from these 10 carefully trained tasters as an SRS from a large population of all trained tasters.

(a) Use these data to see if there is good evidence that the cola lost sweetness.

(b) It is not uncommon to see the t procedures used for data like these. However, you should regard the results as only rough approximations. Why?