11

Inference for Distributions

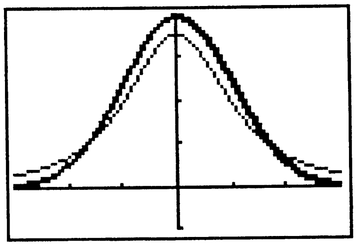
11.1 (a) $s/\sqrt{n} = 9.3/\sqrt{27} \doteq 1.7898$. (b) Since $s/\sqrt{3} = 0.01$, $s = (0.01)(\sqrt{3}) \doteq 0.0173$.

11.2 (a) 2.015. (b) 2.518.

11.3 (a) 2.145. (b) 0.688.

11.4 (a) df = 11, $t^* = 1.796$. (b) df = 29, $t^* = 2.045$. (c) df = 17, $t^* = 1.333$.

11.5 (c) The t_2 curve is a bit shorter at the peak and slightly higher in the tails (see TI-83 plot below). (d) The t_9 curve has moved toward coincidence with the standard normal curve. (e) The t_{30} curve cannot be distinguished from the standard normal curve. As the degrees of freedom increase, the t (df) curve approaches the standard normal density graph.



11.6 (a) 0.228.

(b), (c), and (d)

df	$P(t > 2)$	Absolute difference
2	.0917	.0689
10	.0367	.0139
30	.0273	.0045
50	.0255	.0027
100	.0241	.0013

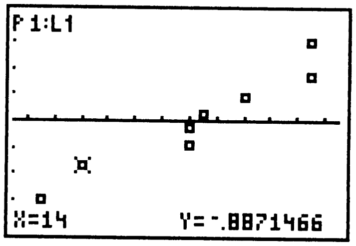
(e) As the degrees of freedom increases, the area to the right of 2 under the t_{df} distribution gets closer to the area under the standard normal curve to the right of 2.

11.7 (a) 14. (b) 1.82 is between 1.761 ($p = 0.05$) and 2.145 ($p = 0.025$). (c) The P-value is between 0.025 and 0.05 (in fact, $P = 0.0451$). (d) $t = 1.82$ is significant at $\alpha = 0.05$ but not at $\alpha = 0.01$.

11.8 (a) 24. (b) 1.12 is between 1.059 ($p = 0.15$) and 1.318 ($p = 0.10$). (c) The P-value is between 0.30 and 0.20 (in fact, $P = 0.2738$). (d) $t = 1.12$ is not significant at either $\alpha = 0.10$ or at $\alpha = 0.05$.

11.9 (a) Since the sample size is small ($n < 15$), the distribution of the CSB vitamin C data should be close to normal. We can check this using a stemplot and a normal probability plot.

1	14
2	2236
3	11



Since there are no outliers and the normal plot is reasonably linear, the assumption of normality seems justified despite the small number of observations.

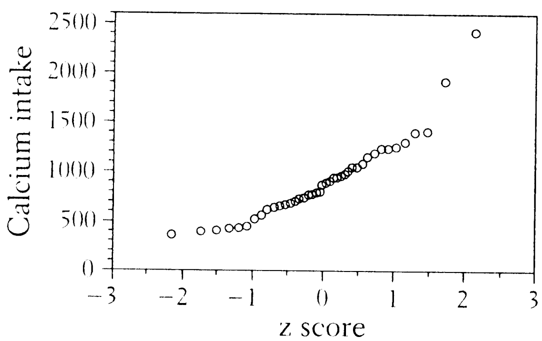
We also must assume that the eight observations represent an SRS from the population of all possible amounts of vitamin C in samples of CSB. Since the eight observations were taken from a production run, this seems like a reasonable assumption provided that the observations were taken at regular intervals.

(b) We will use the t-procedure. $\bar{x} = 22.50$, $s = 7.19$, df $= n - 1 = 7$. Using Table C with df $= 7$, we find $t^* = 2.365$. The 95% confidence interval is therefore $22.50 \pm (2.365)\ (7.19/\sqrt{8}) = 22.5 \pm 6.0$, or (16.5, 28.5).

(c) Letting $\mu =$ the mean vitamin C content per 100 g, we wish to test H_0: $\mu = 40$ vs. H_a: $\mu \neq 40$. The t test statistic is $t = \frac{22.5 - 40}{7.2/\sqrt{8}} = -6.88$ and the corresponding P-value (from software) is 0.0002. Clearly, this result is incompatible with a process mean of $\mu = 40$. We reject H_0 and conclude that the vitamin C content for this run does not conform to specifications (specifically, it is below the specifications).

11.10 (a) The stemplot shown below has stems in 1000s, split 5 ways. The data are right-skewed with a high outlier of 2433 (and possibly 1933). The normal plot shows these two outliers, but otherwise it is not strikingly different from a line.

0	3
0	4444455
0	666667777
0	88899999
1	0011
1	22223
1	44
1	
1	9
2	
2	
2	4



(b) $\bar{x} = 926$, $s = 427.2$, standard error $= 69.3$ (all in mg).

(c) Use of the t-procedure is justified here because the sample size is large ($n = 38 > 30$) and thus the distribution of $\bar{x}$ will be approximately normal by the central limit theorem. Using Table C with 30 degrees of freedom, we have $t^* = 2.042$. The approximate 95% confidence interval is then $926 \pm (2.042)(69.3)$, or 784.5 to 1067.5 mg; MINITAB reports 785.6 to 1066.5 mg.

(d) Without the outliers, the stemplot (see below) shows some details not previously apparent. The normal quantile plot is essentially the same as before (except that the two points that deviated greatly from the line are gone). $\bar{x} = 856.2$, $s = 306.7$, $SE_{\bar{x}} = 51.1$ (all in mg).

Using 30 degrees of freedom, we have $856.2 \pm (2.042)(51.1)$, or 751.9 to 960.5 mg; Minitab reports 752.4 to 960.0 mg.

3	7
4	01346
5	47
6	25789
7	1478
8	008
9	04779
10	56
11	05
12	0556
13	2
14	22

11.11 $H_0: \mu = 1200$, $H_a: \mu < 1200$, where μ = mean daily calcium intake in mg. The t-procedure is justified for the same reason as that given in the previous exercise. The value of the t-statistic is $t = \frac{926 - 1200}{69.3} \approx -3.95$. With df = 37, we have $P = 0.00015$ (using unrounded standard error, we get $t = -3.90$ and $P = .0002$). We reject H_0 and conclude that the daily intake is significantly less than the RDA.

11.12 (a) μ is the difference between the population mean yields for Variety A plants and Variety B plants; that is, $\mu = \mu_A - \mu_B$. Another (equivalent) description is: μ is the mean difference between Variety A yields and Variety B yields.

(b) $H_0: \mu = 0$ vs. $H_a: \mu > 0$. With df = 9, we obtain $t = 1.295$, $P = 0.1137$. This is not enough evidence to reject H_0—the difference could be due to chance variation.

11.13 (a) $H_0: \mu = 0$ vs. $H_a: \mu > 0$, where μ is the mean improvement in score (posttest–pretest).

(b) Stemplot below. The stemplot of the differences, with stems split 5 ways, shows that the data are slightly left-skewed, with no outliers; the t test should be reliable.

(c) $\bar{x} = 1.450$; $SE_{\bar{x}} = 3.203/\sqrt{20} \doteq 0.716$, so $t \doteq 2.02$. With df = 19, we see that $0.025 < P < 0.05$; Minitab reports $P = 0.029$. This is significant at 5%, but not at 1%—we have some evidence that scores improve, but it is not overwhelming.

(d) Minitab gives 0.211 to 2.689; using $t^* = 1.729$ and the values of $\bar{x}$ and $SE_{\bar{x}}$ above, we obtain 1.45 ± 1.238, or 0.212 to 2.688.

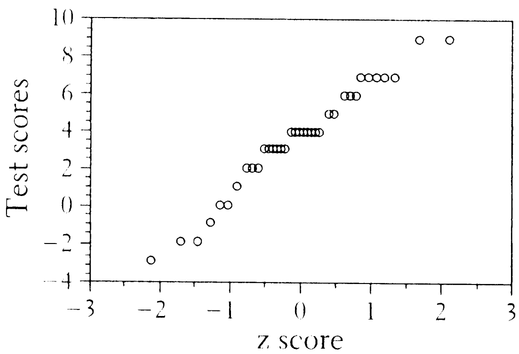
−0	54
−0	32
−0	11
0	11
0	2223333
0	4455
0	7

11.14 (a) Neither the subjects getting the capsules nor the individuals providing them with the capsules knew which capsules contained caffeine and which were placebos.

(b) The differences between "with caffeine" and "without caffeine" heart rates for the 11 subjects are 80, 22, 17, 131, −19, 3, 23, −1, 20, −51, and −3. For these data, $\bar{x} = 20.2$, $s = 48.75$. A 95% confidence interval for the mean difference μ is (−12.57, 52.931), and a test of H_0: $\mu = 0$ against H_a: $\mu \neq 0$ yields a P-value of approximately 0.2. Both statistical procedures suggest that there is no significant difference in heart rate.

11.15 (a) Methods of displaying will vary. Below is a stemplot where the digits are the stems, and all leaves are "0"—this is essentially the same as a histogram. The scores are slightly left-skewed. The normal quantile plot looks reasonably straight, except for the granularity of the data. (b) $\bar{x} = 3.618$, $s = 3.055$, $SE_{\bar{x}} = 0.524$. (c) Using df = 30, we have $t^* = 2.042$ and the interval is 2.548 to 4.688. Minitab reports 2.551 to 4.684.

```
-3 | 0
-2 | 00
-1 | 0
-0 | 0
 0 | 0
 1 | 0
 2 | 000
 3 | 00000
 4 | 0000000
 5 | 00
 6 | 000
 7 | 00000
 8 |
 9 | 00
```



11.16 Test H_0: $\mu = 0$ vs. H_a: $\mu > 0$, where μ is the mean improvement in scores. $\pm = (\bar{x} - \mu)/SE_{\bar{x}} = 3.618/0.524 \doteq 6.90$, which has $P < 0.0005$; we conclude that scores are higher.

11.17 (a) 1.54 to 1.80. (b) We are told the distribution is symmetric; because the scores range from 1 to 5, there is a limit to how much skewness there might be. In this situation, the assumption that the 17 Mexicans are an SRS from the population is the most crucial.

11.18 (a) H_0: $\mu = 0$ vs. H_a: $\mu \neq 0$. For each subject, randomly choose which test to administer first. Alternatively, randomly assign 11 subjects to the "ARSMA first" group, and the rest to the "BI first" group. (b) $t = 4.27$; the P-value is less than 0.001, so we reject H_0. (c) 0.1292 to 0.3746.

11.19 Letting μ = the mean dimension in millimeters, we wish to test H_0: $\mu = 224$ against H_a: $\mu \neq 224$. A stem-and-leaf diagram (see next page) reveals that the distribution of the data is slightly skewed to the right, but there are no apparent outliers, so the t procedure may be used. From software, $t = 0.1254$ and $P = 0.9019$, so there is very little evidence against H_0. We have no reason to believe that the mean dimension differs from 224 mm.

2239	01
2239	6688899
2240	002
2240	69
2241	02

11.20 (a) The mean and standard deviation of the set of differences are $\bar{x} \doteq -5.71 \mu\text{m/hr}$ and $s \doteq 10.56\ \mu\text{m/hr}$, so $s/\sqrt{14} \doteq 2.82\ \mu\text{m/hr}$.
(b) We test H_0: $\mu = 0$ versus H_a: $\mu < 0$ and find $t = -5.71/2.82 \doteq -2.02$. With df = 13, this means that $0.025 < P < 0.05$ (software reports 0.032). This is fairly strong evidence (significant at 5% but not at 1%) that altering the electric field reduces the healing rate.
(c) Take $t^* = 1.771$; the 90% confidence interval is $-5.71 \pm (1.771)(2.82) = -10.70$ to -0.72 μm/hr. The method used to produce this interval works (gives an interval that includes the true mean) 90% of the time.

Minitab output

```
Test of mu = 0.00 vs mu < 0.00

Variable      N      Mean     StDev    SE Mean        T     P-Value
HealRate     14     -5.71     10.56       2.82    -2.02       0.032
```

11.21 (a) H_0: $\mu = 0$ vs. H_a: $\mu > 0$. $t = 43.5$; the P-value is basically 0, so we reject H_0 and conclude that the new policy would increase credit card usage.
(b) \$312.14 to \$351.86.
(c) The sample size is very large, and we are told that we have an SRS. This means that outliers are the only potential snag, and there are none.
(d) Make the offer to an SRS of 200 customers, and choose another SRS of 200 as a control group. Compare the mean increase for the two groups.

11.22 (a) Approximately 2.403 (from Table C), or 2.405 (using software). (b) Using $t^* = 2.403$: Reject H_0 if $t > 2.403$, which means $\bar{x} > 36.70$. (c) The power against $\mu = 100$ is 0.99998—basically 1. A sample of size 50 should be quite adequate. (d) Type I error: concluding that there has been a mean increase of \$100 in the amount charged when in fact no such increase has taken place. Type II error: concluding that there has been no mean increase of \$100 when in fact such an increase has taken place. Since the bank wants to be quite certain of detecting an increase, Type II error is the more serious here.

11.23 (a) The power is 0.5287. (Reject H_0 if $t > 1.833$, i.e., if $\bar{x} > 0.4811$.) (b) The power is 0.9034. (Reject H_0 if $t > 1.711$, i.e., if $\bar{x} > 0.2840$.) (c) Type I error: concluding that there is a mean difference in yield of 0.5 lb/plant when in fact there is none. Type II error: concluding that no mean difference of 0.5 lb/plant exists when in fact one does. Type II error is more serious, because the tomato experts are interested in detecting a mean difference of 0.5 when it in fact exists, and thus would prefer to use a high-power test.

11.24 (a) 9. (b) $P = 0.0255$; it lies between 0.05 and 0.025.

11.25 We want $|t| > t^*$ for t^* = the upper $\alpha/2$ critical value of the t-distribution with $20-1 = 19$ degrees of freedom. From Table C, for $\alpha = 0.005$, this value is $t^* = 3.174$. All values t such that $|t| > 3.174$ will be statistically significant at level $\alpha = 0.005$.

11.26 We use the upper $\alpha/2$ critical value of the t-distribution with $15 - 1 = 14$ degrees of freedom. From Table C, for $\alpha = 0.02$, this value is $t^* = 2.624$.

11.27 Letting μ = the mean HAV angle in the indicated population of patients, we wish to construct a 95% confidence interval for μ. The t-procedure may be used (despite the outlier at 50) because n is large (close to 40). For these data, $\bar{x} = 25.42$, $s = 7.475$. Using 30 degrees of freedom, $t^* = 2.042$. The 95% confidence interval is $25.42 \pm (2.042)\frac{7.475}{\sqrt{38}} = 25.42 \pm 2.476$ or (22.944, 27.896). MINITAB gives an interval of (22.96, 27.88).

11.28 (a) Dropping the outlier at 50, we have $\bar{x} = 24.76$, $s = 6.34$. Using 30 degrees of freedom, $t^* = 2.042$ and the 95% confidence interval is $24.76 \pm (2.042)\frac{6.34}{\sqrt{37}} = 24.76 \pm 2.128$ or (22.632, 26.888). MINITAB gives an interval of (22.64, 26.87).

(b) The interval in part (a) is narrower than the interval in Problem 11.27. Removing the outlier decreased the standard deviation and consequently decreased the margin of error.

11.29 (a) Letting μ = the average change (Haiti − Factory) in mg/100g, the researchers wish to know if there is evidence to conclude that $\mu < 0$. Test H_0: $\mu = 0$ vs. H_a: $\mu < 0$. For the given data, $t = -4.96$ with df = 26, which corresponds to $P < 0.0005$. The mean difference is significantly less than 0, indicating that the average amount of vitamin C has decreased as a result of storage and shipment.

(b) For the change (Haiti − Factory) data, $\bar{x} = -5.33$, $s = 5.59$. For a 95% confidence interval for μ, the mean change, $t^* = 2.056$. The interval is $-5.33 \pm (2.056)\frac{5.59}{\sqrt{27}} = -5.33 \pm 2.212$ or (−7.542, −3.118). MINITAB yields an interval of (−7.54, −3.12).

(c) Letting μ = the mean vitamin C content of all bags shipped to Haiti, we wish to test H_0: $\mu = 40$ against H_a: $\mu \neq 40$. For the 27 Factory observations, $\bar{x} = 42.852$, $s = 4.793$. The value of the t-statistic is $t = 3.092$, which corresponds to a P-value of between 0.002 and 0.005 (two-sided case). There is strong evidence indicating that μ in fact differs from the target mean of 40 mg/100g.

Minitab output

```
Test of mu = 40.000 vs mu not = 40.000

Variable     N     Mean    StDev   SE Mean      T       P
VITC        27   42.852    4.793     0.923   3.09  0.0047
```

11.30 (a) 109.97 to 119.87. (b) We assume that the 27 members of the placebo group can be viewed as an SRS of the population, and that the distribution of seated systolic BP in this population is normal, or at least not too nonnormal. Since the sample size is somewhat large, the procedure should be valid as long as the data show no outliers and no strong skewness.

11.31 (a) Randomly assign 12 (or 13) into a group that will use the right-hand knob first; the rest should use the left-hand knob first. Alternatively, for each student, randomly select which knob he or she should use first.

(b) μ is the mean difference between right-handed times and left-handed times; the null hypothesis is H_0: $\mu = 0$ (no difference). How the alternative is written depends on exactly how μ is defined. Let μ_R be the mean right-hand thread time for all right-handed people (or students), and μ_L be the mean left-hand thread time. As described above, we would most naturally write $\mu = \mu_R - \mu_L$; in this case, H_a: $\mu < 0$. Alternatively, we might define $\mu = \mu_L$

$-\mu_R$, so that H_a: $\mu > 0$. Either way, the null hypothesis says $\mu_R = \mu_L$ and the alternative is $\mu_R < \mu_L$.

(c) A plot of the differences shows no outliers or strong skewness. $\bar{x} = -13.32$ (or $+ 13.32$), $SE(\bar{x}) = 4.5872$, $t = \pm 2.9037$, and $P = 0.0039$. We reject H_0 in favor of H_a.

11.32 5.47 to 21.17 seconds. For our sample $\bar{x}_R \div \bar{x}_L = 88.7\%$; this suggests that right-handed students working on an assembly line with right-handed threads would complete their task in about 90% of the time that it would take them to complete the same task with left-handed threads.

11.33 (a) See below. (b) H_0: $\mu = 105$ vs. H_a: $\mu \neq 105$, $t = -0.3195$, $P = 0.7554$. We do not reject the null hypothesis—the mean detector reading could be 105.

```
 9 | 2
 9 | 578
10 | 024
10 | 55
11 | 1
11 | 9
12 | 2
```

11.34 We know the data for *all* presidents; we know about the whole population, not just a sample. (We might want to try to make statements about future presidents, but doing so from this data would be highly questionable; they can hardly be considered an SRS from the population.)

11.35 (a) 2.080. (b) Reject H_0 if $|t| \geq 2.080$, i.e., if $|\bar{x}| \geq 0.133$. (c) $P(|\bar{x}| \geq 0.133) = P(\bar{x} \leq -0.133$ or $\bar{x} \geq 0.133) = P(Z \leq -5.207$ or $Z \geq -1.047) = 0.852$.

11.36 (a) Yes, provided that the sample is representative of the population at large. (For example, the papers should not all come from the same school or class.) Since the population of 42,000 is much larger than the sample size of 25, the population size itself does not play much of a role in the variability of the sampling distribution of $\bar{x}$.

(b) Probably not. The 25 observations have a strongly right-skewed distribution, indicating that the scores on this question tend to be lower than the "average" value of 2.

(c) While the data are skewed, the absence of outliers and the moderately large sample size ($n = 25$) will permit us to use the t-procedure as an approximation. For these data, $\bar{x} = 1.04$ and $s = 1.14$. The critical value t^* (with 24 degrees of freedom) $= 2.064$, and the 95% confidence interval is $1.04 \pm (2.064)\frac{1.14}{\sqrt{25}} = 1.04 \pm 0.471$ or $(0.569, 1.511)$. MINITAB yields an interval of (0.571, 1.509).

Minitab output

Variable	N	Mean	StDev	SE Mean	95.0 % CI
APSCORE	25	1.040	1.136	0.227	(0.571, 1.509)

11.37 (a) (3)—two samples. (b) (2)—matched pairs.

11.38 (a) (1)—single sample. (b) (3)—two samples.

11.39 Both sample sizes are quite large, so we do not need to worry about the normality of the corresponding populations. Letting μ_1 = the mean Chapin Test score for males and μ_2 = the mean Chapin Test score for females, we wish to test H_0: $\mu_1 = \mu_2$ against H_a: $\mu_1 \neq \mu_2$. The two-sample t statistic for this test is $t = \frac{25.34 - 24.94}{\sqrt{\frac{(5.05)^2}{133} + \frac{(5.44)^2}{162}}} = 0.654$. df = 132, so use the $t(100)$ distribution in Table C: we find that $P > 0.50$. The data provide no evidence that males and females differ in social insight.

11.40 (a) If the loggers had known that a study would be done, they might have (consciously or subconsciously) cut down fewer trees than they typically would, in order to reduce the impact of logging.

(b) We test H_0: $\mu_1 = \mu_2$ versus H_a: $\mu_1 > \mu_2$. The means and standard deviations are given in the Minitab output below; we compute $SE = \sqrt{3.53^2/12 + 4.50^2/9} \doteq 1.813$ and $t = \frac{17.50 - 13.67}{1.813} \doteq 2.11$. With df = 8, we find $0.025 < P < 0.05$; this is significant at 5% but not at 1%.

(c) Use $t^* = 1.860$: $(17.50 - 13.67) \pm (1.860)(1.813) = 0.46$ to 7.20.

Minitab output

```
Twosample T for Species
Code       N      Mean     StDev     SE Mean
1         12     17.50      3.53         1.0
2          9     13.67      4.50         1.5
```

11.41 (a) The study was a randomized comparitive experiment. The large sample sizes will help ensure the accuracy of the t-procedures. Test H_0: $\mu_1 = \mu_2$ against H_a: $\mu_1 < \mu_2$ where μ_1 and μ_2 are the mean lengths of stay for normothermic (blanket) and hypothermic (non-blanket) patients. $t = \frac{12.1 - 14.7}{\sqrt{\frac{(4.4)^2}{104} + \frac{(6.5)^2}{96}}} = -3.285$. Using df = 80 in Table C, we find that $0.0005 < P < 0.01$. Heating blanketsdo seem to reduce the length of a patient's hospital stay.

(b) Using df = 80, the critical value $t^* = 1.990$ and the confidence interval for $\mu_1 - \mu_2$ is $(12.1 - 14.7) \pm (1.990)\sqrt{\frac{(4.4)^2}{104} + \frac{(6.5)^2}{96}} = -2.6 \pm 1.575$ or $(-4.175, -1.025)$. Since the interval contains only negative values, it reinforces the result of the test in (a) that heating blankets reduce the lengths of stays. Specifically, we are 95% confident that the mean reduction is between (roughly) 1 and 4 days.

11.42 (a) Because the sample sizes are so large (and the sample sizes are almost the same), deviation from the assumptions have little effect.

(b) Using $t^* = 1.660$ from a $t(100)$ distribution, the interval is \$412.68 to \$635.58. Using $t^* = 1.6473$ from a $t(620)$ distribution (obtained with software), the interval is \$413.54 to \$634.72.

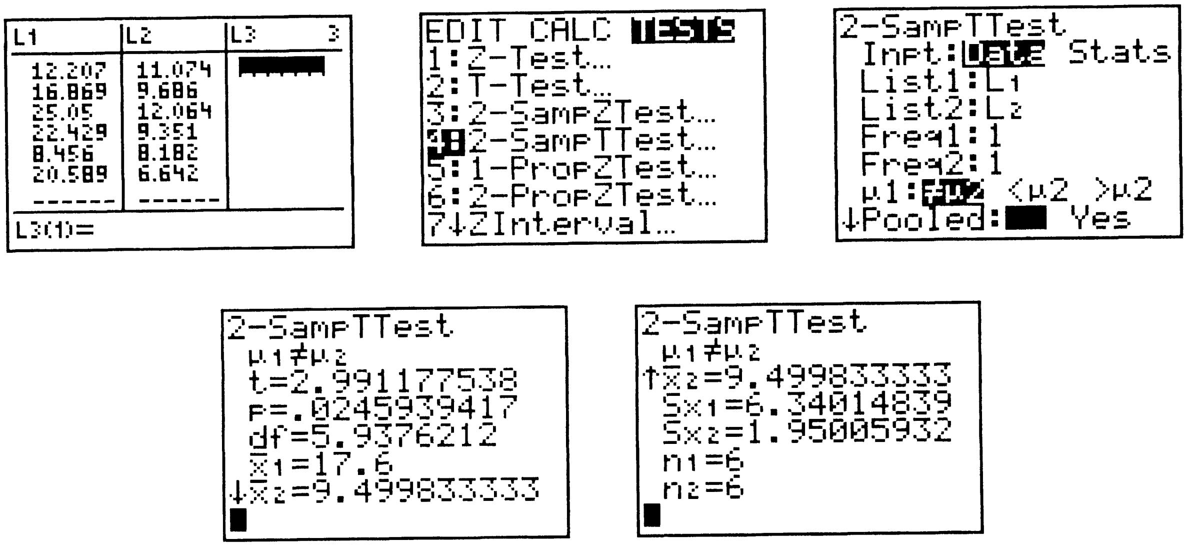
(c) The sample is not *really* random, but there is no reason to expect that the method used should introduce any bias into the sample.

(d) Students without employment were excluded, so the survey results can only (possibly) extend to *employed* undergraduates. Knowing the number of unreturned questionnaires would also be useful.

11.43 H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 > \mu_2$, where μ_1 and μ_2 are the mean number of beetles on untreated (control) plots and malathion-treated plots, respectively. $t = 5.8090$, which yields $P < 0.0001$ for a $t(12)$ distribution—this is significant at the 1% level.

11.44 Consider the left and right endpoint of the confidence interval. If these endpoints remain fixed, then as the degrees of freedom increase, the area in the tails (outside the confidence interval and under the t_{df} curve) decreases. See Exercise 11.5. To make up this difference in area, the endpoints of the intervals have to move toward the center of the distribution, giving up some area to the tails. Thus the confidence interval becomes narrower.

11.45 Here are the key calculator screens.



11.46 The test statistic is $t = 2.99$. With df = 5.9376, the P-value is .0246.

11.47 (a) H_0: $\mu_{\text{skilled}} = \mu_{\text{novice}}$ vs. H_a: $\mu_s > \mu_n$. (b) The t statistic we want is the "Unequal" value: $t = 3.1583$; its P-value is 0.0052. This is strong evidence against H_0. (c) Using $t^* = 1.895$ from a $t(7)$ distribution: 0.4691 to 1.876. Using $t^* = 1.8162$ from a $t(9.8)$ distribution (from software): 0.4982 to 1.8474.

11.48 H_0: $\mu_{\text{skilled}} = \mu_{\text{novice}}$ vs. H_a: $\mu_s \neq \mu_n$ (use a two-sided alternative since we have no preconceived idea of the direction of the difference). The t statistic we want is $t = 0.5143$; its P-value is 0.6165. There is no significant difference in weight between skilled and novice rowers.

11.49 (a) With the small sample sizes, and with no way to check the normality of the data, the t-procedures will be only approximately accurate. Test H_0: $\mu_1 = \mu_2$ against H_a: $\mu_1 \neq \mu_2$, where μ_1 and μ_2 are the mean number of trials to completion for the nicotine group and the saline (control) group. $t = \frac{111.9 - 75.6}{\sqrt{\frac{(50.28)^2}{10} + \frac{(27.512)^2}{10}}} = 2.003$. Using df = 9 in Table C, $0.05 < P < 0.1$.

Using the TI-83, the more precise df = 13.946 and $P = 0.065$. There does not seem to be very strong evidence of a difference between the two means.

(b) Using df = 9, the conservative degrees of freedom, the critical value $t^* = 2.262$ and the 95% confidence interval for $\mu_1 - \mu_2$ is $(111.9 - 75.6) \pm (2.262)\sqrt{\frac{(50.28)^2}{10} + \frac{(27.512)^2}{10}} = 36.3 \pm 41$ or $(-4.7, 77.3)$. Using df = 13.946, the TI-83 gives the 95% confidence interval as $(-2.587, 75.187)$. The more precise df results in a slightly narrower interval.

11.50 (a) Two-sample t test. (b) Matched pairs t test. (c) Matched pairs t test. (d) Two-sample t test. (e) Matched pairs t test.

11.51 (a) Completed table below. The only values not given directly are the standard deviations, which are found by computing $s = \text{SEM}\sqrt{10}$. (b) Use df = 9 (the "smaller of 9 and 9").

Treatment	n	$\bar{x}$	s
IDX	10	116	17.71
Untreated	10	88.5	6.01

11.52 (a) The means and SEMs were given in Exercise 11.51 as 88.5 and 1.9 days (control) and 116 and 5.6 days (IDX), so $SE = \sqrt{1.9^2 + 5.6^2} \doteq 5.91$ and $t = (88.5 - 116)/SE \doteq -4.65$. With either df = 9 or df = 11.04, we have a significant result ($P < 0.001$ or $P < 0.0005$, respectively), so there is strong evidence that IDX prolongs life.

(b) The interval is $(116 - 88.5) \pm t^*(5.91)$, which equals either 14.13 to 40.87 days (df = 9, $t^* = 2.262$) or 14.49 to 40.51 days (df = 11, $t^* = 2.201$).

11.53 (a) Stemplots show little skewness, but one moderate outlier (85) for the control group on the right. Nonetheless, the t procedures should be fairly reliable since the total sample size is 44.

(b) H_0: $\mu_t = \mu_c$ vs. H_a: $\mu_t < \mu_c$; $t = 2.311$. Using $t(20)$ and $t(37.9)$ distributions, P equals 0.0158 and 0.0132, respectively; reject H_0.

(c) Randomization was not really possible, because existing classes were used—the researcher could not shuffle the students.

	1	079
4	2	068
3	3	377
9964333	4	1222368
98776432	5	3455
721	6	02
1	7	
	8	5

11.54 (a) The observations are "before-and-after" weights, so the pairs of observations will be highly correlated—it is the change in weight that we are interested in.

(b) We expect some variation in the weight change, and there may have been some loss due to chance, but the amount lost was so great that it is unlikely to occur merely by chance. In short, this weight-loss program seems to work.

(c) Table C shows that the P-value must be smaller than 0.0005; in fact, it is less than 0.00002.

11.55 (a) A stemplot of the differences (below) looks reasonably normal with no outliers, so the t procedures should be safe.

(b) For the differences, $SE = s/\sqrt{n} = 1071/\sqrt{20} \approx 239$. With $df = 19$, $t^* = 1.729$ and the interval is $-37 \pm (1.729)(239)$, or -451 to 376.5.

−2	0
−1	6
−1	1
−0	775
−0	433210
0	123
0	5
1	13
1	5
2	3

11.56 (a) A matched pairs t test should be used because the observations from Try 1 and Try 2 are dependent; each coached student generates a pair of observations.

(b) The t-procedures may be used because the sample size is very large. For μ = mean gain (Try 2 score − Try 1 score), we will test H_0: $\mu = 0$ vs. H_a: $\mu > 0$. The test statistic is

$t = \frac{29}{59/\sqrt{427}} = 10.157$. Using df = 100 in Table C, we obtain $P < 0.0005$. Using the TI-83, we get $P = 3.77 \times 10^{-22} \approx 0$. Coaching definitely seems to have improved the students' scores.

(c) (21.612, 36.388).

11.57 (a) Since the sample sizes are quite large, we should be safe applying the t-procedures. Test H_0: $\mu_1 = \mu_2$ against H_a: $\mu_1 > \mu_2$, where μ_1 and μ_2 are the mean gains for coached and uncoached students respectively, in a two-sample t test. $t = \frac{29 - 21}{\sqrt{\frac{(59)^2}{427} + \frac{(52)^2}{2733}}} = 2.646$.

Using df = 100 in Table C, we find that $0.0025 < P < 0.005$. The TI-83 yields $P = 0.0042$. There is evidence that coached students gained more on the average than uncoached students.

(b) Using the conservative df = 100, the critical value $t^* = 2.626$ and the confidence interval is $(29 - 21) \pm (2.626)\sqrt{\frac{(59)^2}{427} + \frac{(52)^2}{2733}} = 8 \pm 7.94$ or (0.06, 15.94). Using the TI-83, the more precise confidence interval is (0.184, 15.816). The intervals suggest that on the average, coached students gained somewhere between 0 and 15 points more than uncoached students.

(c) The average gain vis-a-vis uncoached students is rather small, so it may well be that coaching courses are *not* worth the time and expense.

11.58 Since the same students are taking the test over again, the effect of coaching will be confounded with the "experience factor," i.e., the tendency of a student to improve his/her score on Try 2 because of familiarity with the format and types of questions asked on the test. It is impossible to separate the two effects, so a "cause-and-effect" relationship cannot be inferred.

11.59 E.g., The difference between average female (55.5) and male (57.9) self-concept scores was so small that it can be attributed to chance variation in the samples ($t = -0.83$, df = 62.8, $P = 0.4110$). In other words, based on this sample, we have no evidence that mean self-concept scores differ by gender.

11.60 (a) $t^* = 2.364$, the value for a $t(100)$ distribution (since values for a $t(99)$ distribution are not given). (b) Reject H_0 when $\bar{x}_1 - \bar{x}_2 \geq 2.6746$. (c) Power: $P(Z \geq -2.0554) = 0.9801$. (d) Type I error: detecting a lowering of blood pressure when none exists. Type II error: failing to detect a lowering of blood pressure when such an effect does exist. Since the clinical study is primarily interested in detecting a "beneficial" effect if it in fact exists, Type II error should be considered the more serious.

11.61 (a) H_0: $\mu_A = \mu_B$ vs. H_a: $\mu_A \neq \mu_B$; $t = (\bar{x}_A - \bar{x}_B)/\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}$.

(b) For a $t(349)$ distribution, $t^* = 1.967$; using a $t(100)$ distribution, take $t^* = 1.984$.

(c) We reject H_0 when $|\bar{x}_A - \bar{x}_B| \geq 59.48$ (using $t^* = 1.967$). To find the power against $|\mu_A - \mu_B| = 100$, we choose *either* $\mu_A - \mu_B = 100$ or $\mu_A - \mu_B = -100$ (the probability is the same either way). Taking the former, we compute: $P[(\bar{x}_A - \bar{x}_B) \leq -59.48$ or $(\bar{x}_A - \bar{x}_B) \geq 59.48]$ $= P(Z \leq -5.274$ or $Z \geq -1.340) = 0.9099$. Repeating these computations with $t^* = 1.984$ gives power 0.9071.

(d) Type I error: detecting a difference of \$100 in mean amount charged when no such difference exists. Type II error: failing to detect a difference of \$100 in mean amount charged when the difference does exist. Type II error should be of more concern to the bank because if there is a significant difference in the mean amount charged, the bank needs to know about it and adjust the handling of credit-card accounts appropriately.

11.62 (a) This is a two-sample t test—the two groups of women are (presumably) independent. (b) Use a $t(44)$ distribution. (c) The sample sizes are large enough that nonnormality has little effect on the reliability of the procedure.

11.63 (a) (See stemplot below.) The distribution looks reasonably symmetric; other than the low (9.4 ft) and high (22.8 ft) outliers, it appears to be nearly normal. The mean is $\bar{x} \doteq 15.59$ ft and the standard deviation is $s = 2.550$ ft.

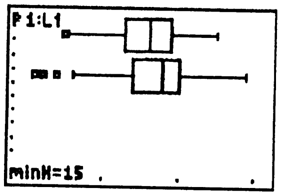
(b) Using df = 40 from Table C, $t^* = 2.021$, so the interval is $\bar{x} \pm t^*s/\sqrt{44} \doteq 15.59 \pm 0.78 = 14.81$ to 16.37 ft. Using software, we can find $t^* = 2.0167$ for df = 43, which also (after rounding) gives 15.59 ± 0.78 ft. Since 20 ft does not fall in (or even near) this interval, we reject this claim.

(c) We need to know what population we are examining: Were these all full-grown sharks? Were they all male? (I.e., is μ the mean adult male shark length? Or something else?)

9	4
10	
11	
12	12346
13	22225668
14	3679
15	237788
16	122446788
17	688
18	23677
19	17
20	
21	
22	8

11.64 (a) Matched pairs t test. (b) Two-sample t test. (c) Two-sample t test. (d) Matched pairs t test. (e) Matched pairs t test.

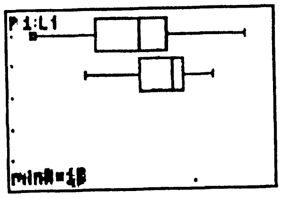
11.65 (a) Side-by-side boxplots of the data (see below) reveal rough symmetry for the air-filled football and strong left-skewness for the helium-filled football.



A two-sample t test of H_0: $\mu_A = \mu_H$ vs. H_a: $\mu_A < \mu_H$ yields $t = -0.37$, $P = 0.356$. We fail to reject H_0; there is no significant difference in the lengths of the kicks for the two balls.

(b) Without the outliers (i.e., the values 11, 12, 14, and 14 in the helium data set), $t = -1.931$ and $P = 0.0287$. Now we might conclude that there is a significant difference in the mean distance traveled by air-filled and helium-filled footballs.

(c) A time plot for each of the two balls shows an increasing trend, but there are also occasional short kicks that deviate from the trend. You could compare the lengths of the first 20 kicks and the remaining 19 kicks with each ball. For the air-filled ball, side-by-side boxplots of the first 20 kicks and the remaining 19 kicks (see below) suggest that the kicker may have improved. A two-sample t test for the two sets of "air-filled data" yields $t = -1.736$ and $P = 0.046$.



11.66 (a) The "unscented first" and "scented first" groups were separate, not matched pairs.
(b) The stemplot (below) and the means ($\bar{x}_1 = 54.38$ sec and $\bar{x}_2 = 45.21$ sec) suggest that unscented-first times are longer; that is, those subjects were slower. The unscented-first data show some hint of nonnormality, but the t procedures should be safe.
(c) We test H_0: $\mu_1 = \mu_2$ versus H_a: $\mu_1 > \mu_2$. The standard deviations are $s_1 \doteq 17.49$ and $s_2 \doteq 8.348$ sec, so $SE \doteq 5.897$ and $t \doteq 1.55$. With either df = 9 or df = 14.6, this is not significant ($0.05 < P < 0.10$), so we don't have strong enough evidence to conclude that there was a learning effect.

Unscented 1st		Scented 1st
710	3	267
3	4	3378
43	5	138
8540	6	
	7	
7	8	

11.67 (a) Using a $t(1361)$ distribution, you get \$1016.56 to \$1069.44, using a $t(2669.1)$ distribution, you get almost the same interval: \$1016.58 to \$1069.42. (b) Skewness will have little effect because the sample sizes are very large.

11.68 (a) First compute each subject's improvement ("after" minus "before"). We test H_0: $\mu_1 = \mu_2$ versus H_a: $\mu_1 > \mu_2$. The means and standard deviations are given in the Minitab output below; we compute $SE = \sqrt{3.17^2/10 + 3.69^2/8} \doteq 1.645$ and $t = \frac{11.40-8.25}{1.645} \doteq 1.91$. With df = 7, we find $0.025 < P < 0.05$; this is significant at 5% (and also at 10%). (Note that Minitab uses the more accurate df = 13, rather than the conservative approach.)
(b) Use $t^* = 1.895$: $(11.40-8.25) \pm (1.895)(1.645) = 0.03$ to 6.27. (Minitab's result is based on df = 13, so is slightly narrower than the conservative interval.)

Minitab output

```
Two sample T for Diff
Code        N       Mean      StDev     SE Mean
1          10      11.40       3.17         1.0
2           8       8.25       3.69         1.3

90% C.I. for mu 1 - mu 2: (0.2, 6.1)
T-Test mu 1 = mu 2 (vs >): T = 1.91 P=0.039 DF = 13
```

11.69 We can compare heart rates between the groups by conducting two-sample t tests of the form H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 \neq \mu_2$. There are three separate tests to conduct: control group (C) vs. friend group (F), control group (C) vs. pet group (P), and friend group (F) vs. pet group (P). Below are the MINITAB outputs for these three cases. (In each case, the precise value of df is being used.)

```
Two sample T for CONTROL vs FRIEND

            N      Mean     StDev    SE Mean
CONTROL    15     106.5      10.1        2.6
FRIEND     15     117.7      13.0        3.4

95% CI for mu CONTROL - mu FRIEND: (-19.9, -2.4)
T-Test mu CONTROL = mu FRIEND (vs not =): T = -2.62 P = 0.015
DF = 26
```

There is evidence of a significant difference between the mean heart rates of the control group and the friend group. Specifically, since the 95% confidence interval contains only negative values, the mean for the control group appears to be significantly *lower* than the mean for the friend group (difference in means = 11.2).

```
Two sample T for CONTROL vs PET

            N      Mean     StDev    SE Mean
CONTROL    15     106.5      10.1        2.6
PET        15      78.5      15.4        4.0

95% CI for mu CONTROL - mu PET: (18.2, 37.8)
T-Test mu CONTROL = mu PET (vs not =): T = 5.90 P = 0.0000 DF = 24
```

There is very strong evidence of a significant difference between the control group and the pet group. In this case, since the 95% confidence interval contains only positive values, the mean for the control group is significantly *higher* than the mean for the pet group (difference in means = 28.0).

```
Two sample T for FRIEND vs PET

            N      Mean     StDev    SE Mean
FRIEND     15     117.7      13.0        3.4
PET        15      78.5      15.4        4.0

95% CI for mu FRIEND - mu PET: (28.5, 49.8)
T-Test mu FRIEND = mu PET (vs not =): T = 7.52 P = 0.0000 DF = 27
```

There is very strong evidence of a significant difference between the friend group and the pet group. Since the 95% confidence interval contains only positive values, the mean for the friend group is significantly *higher* than the mean for the pet group (difference in means = 39.2).

Combining the results of the tests, it appears that the friend group has the highest heart rate, the control group has the second highest, and the pet group has the lowest.

11.70 No—you have information about all Indiana counties (not just a sample).

11.71 The stemplot looks fairly symmetric; 4.88 is perhaps a moderate low outlier, but is not too far from the other observations. Our estimate is the mean, $\bar{x} = 5.4479$. The standard error of the mean is 0.0410; the margin of error depends on the confidence level chosen. Here are three possibilities:

Confidence level	Confidence interval	Margin of error
90%	(5.3781,5.5177)	0.0698
95%	(5.3639,5.5320)	0.0840
99%	(5.3345,5.5613)	0.1134

48	8
49	
50	7
51	0
52	6799
53	04469
54	2467
55	03578
56	12358
57	59
58	5

11.72 (a) H_0: $\mu_1 = \mu_2$ vs. H_a: $\mu_1 > \mu_2$; $t = 1.1738$, so $P = 0.1265$ (using $t(22)$) or 0.123453 (using $t(43.3)$). Not enough evidence to reject H_0. (b) −14.57 to 52.57 (using df = 22), or −13.64 to 51.64 (using df = 43.3). (c) 165.53 to 220.47. (d) We are assuming that we have two SRSs from each population, and that underlying distributions are normal. It is unlikely that we have random samples from either population, especially among pets.

11.73 (a) Stemplots (see below) show that both distributions are skewed right. There is one high outlier in the "Active" group. The summary statistics are:

	n	$\bar{x}$	s
Active	24	24.4167	6.31022
Passive	24	17.8750	4.02506

To test H_0: $\mu_A = \mu_P$ versus H_a: $\mu_A > \mu_P$, we find that $SE \doteq 1.5278$ and $t \doteq 4.28$. With either df = 23 or df = 39.1, $P < 0.0005$, so there is strong evidence that active learning results in more correct identifications. (b) With $t^* = 1.714$ from a $t(23)$ distribution, the 90% confidence interval is $24.4167 \pm (1.714)(6.31022/\sqrt{24}) = 22.2$ to 26.6 Blissymbols. (Note: This is a one-sample question.)

Active		Passive
	1	223
5	1	45555
76	1	66777
	1	889
111100	2	00111
332	2	
4444	2	5
7	2	66
9888	2	
1	3	
	3	
5	3	
	3	
	3	
	4	
	4	
4	4	