AP Statistics
Packet 11

Inference for Distributions
Inference for the Mean of a Population
Comparing Two Means

HW #19 1 – 4, 7 - 10

11.1 Writers in some fields often summarize data by giving $\bar{x}$ and its standard error rather than $\bar{x}$ and *s*. The standard error of the mean $\bar{x}$ is often abbreviated as SEM.

(a) A medical study finds that $\bar{x}$ = 114.9 and *s* = 9.3 for the seated systolic blood pressure of the 27 members of one treatment group. What is the standard error of the mean?

$$\frac{s}{\sqrt{n}} = \frac{9.3}{\sqrt{27}} = 1.7898$$

(b) Biologists studying the levels of several compounds in shrimp embryos reported their results in a table, with the note, "Values are means ± SEM for three independent samples." The table entry for the compound ATP was 0.84 ± 0.01. The researchers made three measurements of ATP, which had $\bar{x}$ = 0.84. What was the sample standard deviation *s* for these measurements?

$$\text{Since } \frac{s}{\sqrt{3}} = 0.01, \ s = 0.0173$$

11.2 What critical value t^* from Table C satisfies each of the following conditions?

(a) The *t* distribution with 5 degrees of freedom has probability 0.05 to the right of t^*.
2.015 (invT (.95, 5))

(b) The *t* distribution with 21 degrees of freedom has probability 0.99 to the left of t^*.
2.518 (invT (.99, 21))

11.3 What critical value t^* from Table C satisfies each of the following conditions?

(a) The one-sample *t* statistic from a sample of 15 observations has probability 0.025 to the right of t^*.
2.145 (invT (.975, 14))

(b) The one-sample *t* statistic from an SRS of 20 observations has probability 0.75 to the left of t^*.
0.688 (invT (.75, 19))

11.4 What critical value t^* from Table C should be used for a confidence interval for the mean of the population in each of the following situations?

(a) A 90% confidence interval based on n = 12 observations.
df = 11, t^* = 1.796

(a) A 95% confidence interval from an SRS of 30 observations.
df = 29, t^* = 2.045

(a) An 80% confidence interval from a sample of size 18.
df = 17, t^* = 1.333

11.7 The one-sample t statistic for testing

$$H_0: m = 0$$
$$H_A: m > 0$$

from a sample of n = 15 observations has the value t = 1.82.

(a) What are the degrees of freedom for this statistic? df = 14

(b) Give the two critical values t^* from Table C that bracket t. What are the right-tail probabilities p for these two values? 1.82 is between 1.761 (p = 0.05) and 2.145 (p = 0.025).

(c) Between what two values does the P-value of the test fall?
The P-value is between 0.025 and 0.05 (in fact, p = tcdf(1.92, ∞, 14) = 0.0451).

(d) Is the value t = 1.82 statistically significant at the 5% level? At the 1% level?
t = 1.82 is significant at α = 0.05 but not at α = 0.01.

11.8 The one-sample t statistic from a sample of n = 25 observations for the two-sided test of

$$H_0: m = 64$$
$$H_A: m \neq 64$$

has the value t = 1.12.
,
(a) What are the degrees of freedom for this statistic? df = 24

(b) Give the two critical values t^* from Table C that bracket t. What are the right-tail probabilities p for these two values? 1.12 is between 1.059 (p = 0.15) and 1.318 (p = 0.10).

(c) Between what two values does the P-value of the test fall? (Note that H_A is two-sided.)
The P-value is between 0.30 and 0.20 (in fact, p = 2tcdf(1.12, ∞, 24) = 0.2738).

(d) Is the value t = 1.12 statistically significant at the 10% level? At the 5% level?
t = 1.12 is not significant at either α = 0.10 or at α = 0.05.

11.9 **VITAMIN C CONTENT** In fiscal year 1996, the U.S. Agency for International Development provided 238,300 metric tons of corn soy blend (CSB) for development programs and emergency relief in countries throughout the world. CSB is a highly nutritious, low-cost fortified food that is partially precooked and can be incorporated into different food preparations by the recipients. As part of a study to evaluate appropriate vitamin C levels in this commodity, measurements were taken on sample of CSB produced in a factory.

The following data are the amounts of vitamin C, measured in milligrams per 100 grams (mg/100 g) of blend (dry basis), for a random sample of size 8 from a production run:

26	31	23	22	11	22	14	31

(a) The specifications for the CSB state that the mixture should produce a mean (μ) vitamin C content in the final product of 40 mg/100 g. Does the CSB produced in this production run conform to these specifications? Perform a significance test to answer this question.

P – name the parameter and the population

Let μ = the mean vitamin C content per 100 g

H – state the hypotheses

$H_0 : m = 40$

$H_A : m \neq 40$

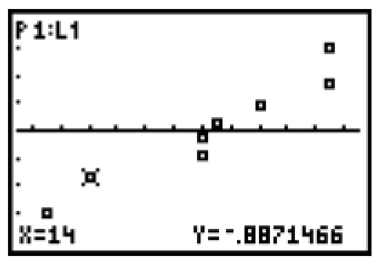
A – verify all assumptions

• SRS – not given, we must assume that the 8 observations represent an SRS from the population of all possible amounts of vitamin C in samples of CSB. Since the 8 observations were taken from a production run, this seems like a reasonable assumption provided that the observations were taken at regular intervals.

• 8 < 10% of all possible observations

• 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 normal probability plot.

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



N – name the test

One sample t test for means

T – calculate the test statistic (Show your work)

$with\ \bar{x} = 22.50, s = 7.19, df = 7$

$$t = \frac{22.50 - 40}{7.2/\sqrt{8}} = -6.88$$

O – obtain the p-value

p-value = 2P(t < -6.88) = 2tcdf($-\infty$, -6.88, 7) = 0.00024

M – make decision

Because the p-value is SO small, we reject H_0. We conclude that the vitamin C content for this run does not conform to specifications.

(b) Use your calculator to find a 95% confidence interval for μ.

t^* = invT(0.025, 7) = -2.365

The 95% confidence interval is $22.50 \pm (2.365)\dfrac{7.19}{\sqrt{8}}$ = 22.5 ± 6.0, or (16.5, 28.5).

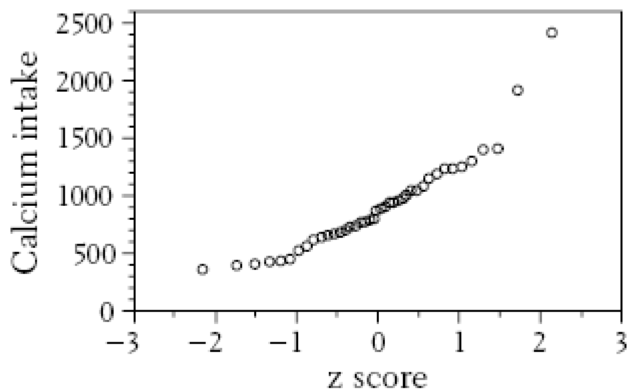
11.10 **HEALTHY BONES, I** Here are estimates of the daily intakes of calcium (in milligrams) for 38 women between the ages of 18 and 24 years who participated in a study of women's bone health:

808	882	1062	970	909	802	374	416	784	997
651	716	438	1420	1425	948	1050	976	572	403
626	774	1253	549	1325	446	465	1269	671	696
1156	684	1933	748	1203	2433	1255	1100		

(a) Display the data using a stemplot and make a normal probability plot. Describe the distribution of calcium intakes for these women.

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) Calculate the mean, standard deviation, and the standard error.

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

(c) Use your calculator to find a 95% confidence interval for the mean.

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 ± (2.042) (69.3), or 784.5 to 1067.5 mg

(d) Eliminate the two largest values and recompute the 95 % confidence interval.

Without the outliers, $\bar{x}$ = 856.2, s = 306.7, standard error = 51.1 (all in mg).
Using 30 degrees of freedom, we have 856.2 ± (2.042) (51.1), or 751.9 to 960.5 mg

HW #20   12, 13, 15 - 17, 19

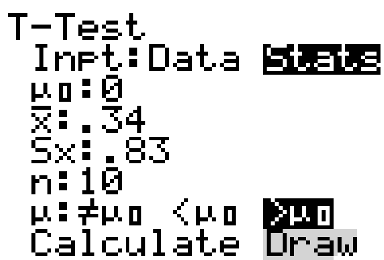
11.12 **GROWING TOMATOES** An agricultural field trial compares the yield of two varieties of tomatoes for commercial use. The researchers divide in half each of 10 small plots of land in different locations and plant each tomato variety on one half of each plot. After harvest, they compare the yields in pounds per plant at each location. The 10 differences (Variety A – Variety B) give $\bar{x} = 0.34$ and $s = 0.83$. Is there convincing evidence that Variety A has the higher mean yield?

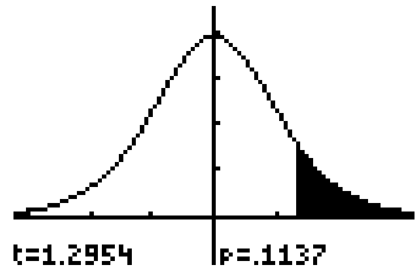
(a) Describe in words what the parameter μ is in this setting.

μ is the difference between the population mean yields for Variety A plants and Variety B plants; that is, $m = m_A - m_B$

(b) Use your calculator to perform a hypothesis test to answer the question.

$H_0: m = 0$, $H_A: m > 0$. With df = 9, we obtain $t = 1.295$, $p = 0.1137$. Because the **p**-value is so large, we do not have enough evidence to reject H_0 - the observed difference could be due to chance variation.





11.13 **SPANISH TEACHERS WORKSHOP** The National Endowment for the Humanities sponsors summer institutes to improve the skills of high school language teachers. One institute hosted 20 Spanish teachers for four weeks. At the beginning of the period, the teachers took the Modern Language Association's listening test of understanding of spoken Spanish. After four weeks of immersion in Spanish in and out of class, they took the listening test again. (The actual spoken Spanish in the two tests was different, so that simply taking the first test should not improve the score on the second test.) The table below gives the pretest and posttest scores. The maximum possible score on the test is 36.

TABLE 11.4 MLA listening scores for 20 Spanish teachers

Subject	Pretest	Posttest	Subject	Pretest	Posttest
1	30	29	11	30	32
2	28	30	12	29	28
3	31	32	13	31	34
4	26	30	14	29	32
5	20	16	15	34	32
6	30	25	16	20	27
7	34	31	17	26	28
8	15	18	18	25	29
9	28	33	19	31	32
10	20	25	20	29	32

Source: Data provided by Joseph A. Wipf, Department of Foreign Languages and Literatures, Purdue University.

(a) Perform a hypothesis test to determine if attending the institute improves listening skills.

P – name the parameter and the population

m_d is the mean improvement in score (posttest–pretest).

H – state the hypotheses

$H_0 : m_d = 0$

$H_A : m_d > 0$

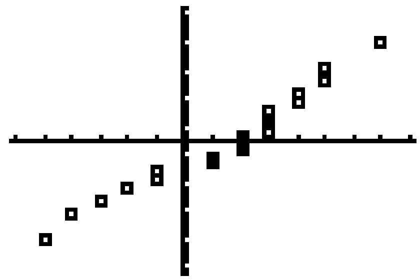
A – verify all assumptions

SRS – we can assume this

20 teachers < 10% of all Spanish teachers

n is small, but NPP looks fairly linear, with no outliers

So the *t* test should be reliable.



N – name the test

matched pairs *t* –test

T – calculate the test statistic (Show your work)

$$t = \frac{1.450 - 0}{3.203/\sqrt{20}} = 2.025, \ df = 19$$

O – obtain the *p*-value

p-value = P(*t* > 2.025) = tcdf(2.025, ∞, 19) = 0.0286

M – make decision

The *p*-value is small, less than 0.05, so we reject the null hypothesis. We have some evidence that scores improve, but it is not overwhelming.

(b) Can you reject H_0 at the 5% significance level? yes At the 1% significance level? no

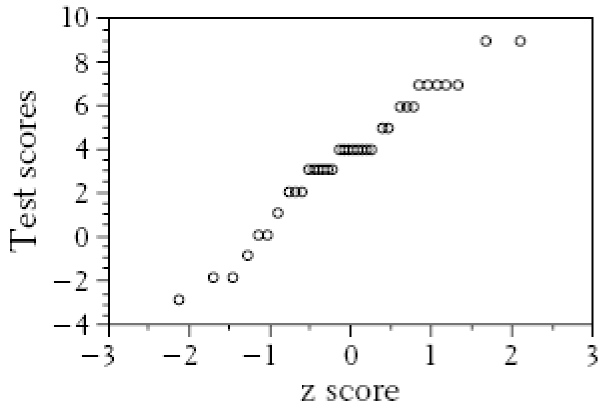
(c) Use your calculator to find a 90% confidence interval for the mean increase in listening score due to attending the summer institute. t^* = invT(.05, 19) = 1.729

CI = 1.45 ± 1.238, or 0.212 to 2.688.

11.15 **DOES PLAYING THE PIANO MAKE YOU SMARTER?** Do piano lessons improve the spatial-temporal reasoning of preschool children? Neurobiological arguments suggest that this may be true. A study designed to test this hypothesis measured the spatial-temporal reasoning of 34 preschool children before and after six months of piano lessons? (The study also included children who took computer lessons and a control groups; but we are not concerned with those here.) The changes in the reasoning scores are:

2	5	7	-2	2	7	4	1	0	7	3	4	3	4	9	4	5
2	9	6	0	3	6	-1	3	4	6	7	-2	7	-3	3	4	4

(a) Display the data and summarize the distribution.



The NPP looks reasonably straight, except for the granularity of the data.

(b) Find the mean, standard deviation, and the standard error of the mean.

$\bar{x} = 3.618,\ s = 3.055,\ SE_{\bar{x}} = 0.524$

(c) Use your calculator to find a 95% confidence interval for the mean improvement in reasoning.

With 33 df, we have $t^* = \text{invT}(.025, 33) = -2.035$

CI is 2.552 to 4.684

11.16 **PIANO PLAYING, II** Refer to the preceding exercise. Use your calculator to test the null hypothesis that there is no improvement versus the alternative suggested by the neurobiological arguments. What do you conclude?

$H_0: m = 0,\ H_A: m > 0$ where μ is the mean improvement in scores.

With df = 33, we obtain $t = \dfrac{3.618 - 0}{3.055/\sqrt{34}} = 6.906$, $p < 0.0005$. Because the p-value is so small we reject the null hypothesis and conclude that the scores are higher for students who study the piano.

11.17 **MEASURING ACCULTURATION** The Acculturation Rating Scale for Mexican Americans (ARSMA) measures the extent to which Mexican Americans have adopted Anglo/English culture. During the development of ARSMA, the test was given to a group of 17 Mexicans. Their scores, from a possible range of 1.00 to 5.00, had a symmetric distribution with $\bar{x} = 1.67$ and $s = 0.25$. Because low scores should indicate a Mexican cultural orientation, these results helped to establish the validity of the test.

(a) Use your calculator to find a 95% confidence interval for the mean ARSMA scores of Mexicans.

1.54 to 1.80

(b) What conditions does your confidence interval require? Which of these conditions is most important in this case?

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.

11.19 **AUTO CRANKSHAFTS** Here are measurements (in millimeters) of a critical dimension for 16 auto engine crankshafts:

224.120	224.001	224.017	223.982	223.989	223.961
223.960	224.089	223.987	223.976	223.902	223.980
224.098	224.057	223.913	223.999		

The dimension is supposed to be 224 mm and the variability of the manufacturing process in unknown. Is there evidence that the mean dimension is not 224 mm? Give appropriate statistical evidence to support your conclusion.

P – name the parameter and the population

Let μ = the mean dimension in millimeters

H – state the hypotheses

$H_0: m = 224$

$H_A: m \neq 224$

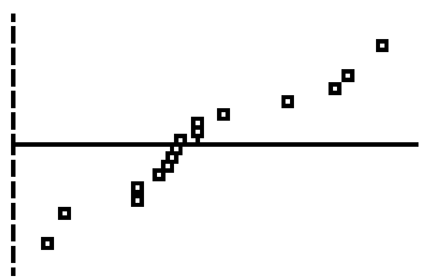
A – verify all assumptions

SRS, must be assumed

16 < 10% all possible measurements

NPP is fairly linear, with no outliers

So the *t* procedure should be reliable.



N – name the test

One sample *t* test for means

T – calculate the test statistic (Show your work)

$$t = \frac{224.0019375 - 224}{.0618/\sqrt{16}} = 0.1254; \quad df = 15$$

O – obtain the *p*-value

p-value = 2P(t > .1254) = 0.9019

M – make decision

Because the *p*-value is so large, we fail to reject H_0. We have no reason to believe that the mean dimension differs from 224 mm.

HW #21   29, 34, 37 - 42

11.29 **VITAMIN C CONTENT** The researchers studying vitamin C in CSB were also interested in a similar commodity called wheat soy blend.(WSB). A major concern was the possibility that some of the vitamin C content would be destroyed as a result of storage and shipment of the commodity to its final destination. The researchers specially marked a collection of bags at the factory and took a sample from each of these to determine the vitamin C content. Five months later in Haiti they found the specially marked bags and took samples. The data consist of two vitamin C measures for each bag, one at the time of production in the factory and the other five months later in Haiti. The units are mg/100 g. Here are the data:

Factory	Haiti	Factory	Haiti	Factory	Haiti
44	40	45	38	39	43
50	37	32	40	52	38
48	39	47	35	45	38
44	35	40	38	37	38
42	35	38	34	38	41
47	41	41	35	44	40
49	37	43	37	43	35
50	37	40	34	39	38
39	34	37	40	44	36

(a) Examine the question of interest to these researchers.

P – name the parameter and the population

Let m_d = the average change (Haiti − Factory) in mg/100g

H – state the hypotheses

$H_0 : m_d = 0$

$H_A : m_d < 0$

A – verify all assumptions

SRS, given

27 < 10% of all possible measurements

n = 27, we can proceed with *t* test

N – name the test

matched pairs *t* test

T – calculate the test statistic (Show your work)

$$t = \frac{-5.33 - 0}{5.59/\sqrt{27}} = -4.96; \quad df = 26$$

O – obtain the *p*-value

p-value = P(t < -4.96) < 0.0005

M – make decision

Because the *p*-value is so small, we reject H_0. 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) Estimate the loss in vitamin C content over the five-month period. Use a 95% confidence level.

$t^* = 2.056$, $\quad CI = -5.33 \pm 2.056\dfrac{5.59}{\sqrt{27}} = -5.33 \pm 2.212$

$(-7.452, -3.118)$

(c) Do these data provide evidence that the mean vitamin C content of all of the bags of WSB shipped to Haiti differs from the target value of 40mg/100g?

Letting μ = the mean vitamin C content of all bags shipped to Haiti, we wish to test $H_0: m = 40,\ H_A: m \neq 40$. 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 between 0.002 and 0.005 (two-sided case). With such a small ***p***-value we reject H_0. There is strong evidence indicating that μ in fact differs from the target mean of 40 mg/100g.

11.34 On page 19 (Table 1.4) we see the ages of U.S. presidents when they took office. It does not make sense to use the ***t*** procedures (or any other statistical procedures) to give a 95% confidence interval for the mean age of the presidents. Explain why not.

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.)

11.37 **WHICH DATA DESIGN?** The following situations require inference about a mean or means. Identify each as (1) single sample, (2) matched pairs, or (3) two samples. The procedures of Section 11.1 apply to cases (1) or (2). We are about to learn procedures for (3).

(a) An education researcher wants to learn whether it is more effective to put questions before or after introducing a new concept in an elementary school mathematics text. He prepares two text segments that teach the concept, one with motivating question before and the other with review questions after. He uses each text segment to teach a separate group of children. The researcher compares the scores of the groups on a test over the material.

Two samples

(b) Another researcher approaches the same issue differently. She prepares text segments of two unrelated topics. Each segment comes in two versions, one with question before and the other with question after. The subjects are a single group of children. Each child studies both topics, one (chosen at random) with questions before and the other with questions after. The researcher compares test scores for each child on the two topics to see which topic he or she learned better.

Matched pairs

11.38 **WHICH DATA DESIGN?** The following situations require inference about a mean or means. Identify each as (1) single sample, (2) matched pairs, or (3) two samples. The procedures of Section 11.1 apply to cases (1) or (2). We are about to learn procedures for (3).

(a) To check a new analytical method, a chemist obtains a reference specimen of known concentration from the National Institute of Standards and Technology. She then makes 20 measurements of the concentration of this specimen with the new method and checks for bias by comparing the mean result with the known concentration.

Single sample

(b) Another chemist is checking the same new method. He has no reference specimen, but a familiar analytic method is available. He wants to know if the new and old methods agree. He takes a specimen of unknown concentration and measures the concentration 10 times with the new method and 10 times with the old method.

Two samples

11.39 **SOCIAL INSIGHT AMONG MEN AND WOMEN** The Chapin Social Insight Test is a psychological test designed to measure how accurately a person appraises other people. The possible scores on the test range from 0 to 41. During the development of the Chapin test, it was given to several different groups of people. Here are the results for male and female college students majoring in the liberal arts:

Group	Sex	n	$\bar{x}$	s
1	Male	133	25.34	5.05
2	Female	162	24.94	5.44

Do these data support the contention that female and male students differ in average social insight? Perform a significance test to help you answer this question.

- Let m_1 = the mean Chapin Test score for males and m_2 = the mean Chapin Test scores for females
- $H_0: m_1 - m_2 = 0$
  $H_A: m_1 - m_2 \neq 0$
- Assume both samples are SRS
  Reasonable to assume the samples are independent
  162 < 10% of females taking the test, 133 < 10% of males taking the test
  Both samples sizes are quite large, so we do not need to worry about the normality of the corresponding populations.
- Two sample *t* test
- $t = \dfrac{(25.34 - 24.94) - 0}{\sqrt{\dfrac{5.05^2}{133} + \dfrac{5.44^2}{162}}} = 0.654;\quad df = 288.58\,(\mathit{from\ calc})$
- p-value = 2P(t > .654) = 0.514
- The p-value is very large, so we fail to reject H_0. The data provide no evidence that males and females differ in social insight.

11.40 **THE EFFECT OF LOGGING** How badly does logging damage tropical rainforests? One study compared forest plots in Borneo that had never been logged with similar plots nearby that had been logged 8 years earlier. The study found that the effects of logging were somewhat less severe than expected. Here are the data on the number of tree species in 12 unlogged plots and 9 logged plots:

Unlogged:	22	18	22	20	15	21	13	13	19	13	19	15
Logged:	17	4	18	14	18	15	15	10	12			

(a) The study report says, "Loggers were unaware that the effects of logging would be assessed." Why is this important? The study report also explains why the plots can be considered to be randomly assigned.

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) Does logging significantly reduce the mean number of species in a plot after 8 years? Use your calculator to perform an appropriate test to support your conclusion.

We use a two sample *t* test of $H_0: m_1 - m_2 = 0,\ H_A: m_1 - m_2 > 0$

The *t* statistic = 2.11, with 14.8 df.

The ***p***-value = 0.026

This ***p***-value is fairly small, at the 5% level we would reject H_0. There is evidence to show that the mean number of tree species in unlogged plots is greater than the mean number of species in logged plots.

(c) Give a 90% confidence interval for the difference in mean number of species between unlogged and logged plots.

Use t^* = 1.860; CI is (17.50 − 13.67) ± (1.860)(1.813) = 0.46 to 7.20. We have evidence to show that there are about ½ to 7 more tree species in unlogged plots than in logged plots.

11.41 **SURGERY IN A BLANKET** When patients undergo surgery, the operating room is kept cool so that the physicians in heavy gowns will not be overheated. The patient may pay the price for the surgeon's comfort. The exposure to cold, in addition to impairment of temperature regulation caused by anesthesia and altered distribution of body heat, may result in mild hypothermia (approximately 2° C below the normal core body temperature.) As a result of the hypothermia, patients may have an increased susceptibility to wound infections or even heart attacks. In 1996, researchers in Austria investigated whether maintaining a patient's body temperature close to normal by heating the patient during surgery decreases wound infection rates. Patients were assigned at random to two groups: the normothermic group (patients' core temperatures were maintained at near normal 36.5 ° C with heating blankets) and the hypothermic group (patients' core temperatures were allowed to decrease to about 34.5 ° C). If keeping patients warm during surgery reduces the chance of infection, then patients in the normothermic group should have shorter hospital stays than those in the hypothermic group.

Here are summary statistics on length of hospital stay for the two treatment groups.

Group	n	$\overline{x}$	s
Normothermic	104	12.1	4.4
Hypothermic	96	14.7	6.5

(a) Do these data provide evidence that the use of warming blankets reduces the length of a patient's hospital stay?

- Let m_1 = the mean lengths of stay for normothermic (blanket) and m_2 = the mean lengths of stay for hypothermic (non-blanket)
- $H_0: m_1 - m_2 = 0$
  $H_A: m_1 - m_2 < 0$
- Assume both samples are SRS
  Reasonable to assume the samples are independent
  104 < 10% of all patients in normothermic group 96 < 10% of all patients in hypothermic group
  Both samples sizes are quite large, so we do not need to worry about the normality of the corresponding populations.
- Two sample *t* test
- $t = \dfrac{(12.1 - 14.7) - 0}{\sqrt{\dfrac{4.4^2}{104} + \dfrac{6.5^2}{9.6}}} = -3.285; \quad df = 165.12\,(\textit{from calc})$
- *p*-value = P(t < -3.285) < 0.0005
- The ***p***-value is very small, so we reject H_0. The data provide evidence that blankets do seem to reduce the length of a patient's hospital stay.

(b) Use your calculator to construct a 95% confidence interval for the difference between the means for length of stay in the hospital for the normothermic and hypothermic groups. What does this interval tell you about the effect of the treatment?

The 95% confidence interval for $m_1 - m_2$ is (-4.162, -1.038). 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 **PAYING FOR COLLEGE** College financial aid offices expect students to use summer earnings to help pay for college. But how large are these earnings? One college studied this question by asking a sample of students how much they earned. Omitting students who were not employed, there were 1296 responses. Here are the data in summary form:

Groups	n	$\bar{x}$	s
Male	675	\$1884.52	\$1368.37
Female	621	\$1360.39	\$1037.46

(a) The distribution of earnings is strongly skewed to the right. Nevertheless, use of *t* procedures is justified. Why?

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

(b) Give a 90% confidence interval for the difference between the mean summer earnings of male and female students.

Using t^* = 1.6473 from a t(620) distribution (from calc), the interval is \$413.54 to \$634.72.

(c) Once the sample size was decided, the sample was chosen by taking every 20th name from an alphabetical list of all undergraduates. Is it reasonable to consider the samples as SRSs chosen from the male and female undergraduate populations?

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) What other information about the study would you request before accepting the results as describing all undergraduates?

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.

HW #23 44, 47, 49, 51, 53, 59, 63, 68

11.44 Example 11.13 demonstrates that if all other statistics stay the same, a higher number of degrees of freedom will produce a narrower (and hence more precise) confidence interval. Briefly explain why this is so.

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 *tcdf* curve) decreases. 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.

11.47 **COMPETITIVE ROWERS** What aspects of rowing technique distinguish between novice and skilled competitive rowers? Researchers compared two groups of female competitive rowers: a group of skilled rowers and a group of novices. The researchers measured many mechanical aspects of rowing style as the subjects rowed on a Stanford Rowing Ergometer. One important variable is the angular velocity of the knee, which describes the rate at which the knee joint opens as the legs push the body back on the sliding seat. The data show no outliers or strong skewness. Here is the SAS computer output:

TTEST PROCEDURE

Variable: KNEE

Group	N	Mean	Std Dev	Std Error
SKILLED	10	4.18283335	0.47905935	0.15149187
NOVICE	8	3.01000000	0.95894830	0.33903942

Variances	T	DF	Prob > | T |
Unequal	3.1583	9.8	0.0104
Equal	3.3918	16.0	0.0037

(a) The researchers believed that the knee velocity would be higher for skilled rowers. State H_0 and H_A.

$H_0: m_{skilled} - m_{novice} = 0$

$H_A: m_s - m_n > 0$

(b) What is the value of the two-sample *t* statistic and its ***P***-value? (Note that SAS provides two-sided ***P***-values. If you need a <u>one-sided</u> ***P***-value, divide the two-sided value by 2.) What do you conclude?

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. We believe the knee velocity is different for skilled an novice rowers.

(c) Use your calculator to find a 90% confidence interval for the mean difference between the knee velocities of skilled and novice female rowers.

Using $t^* = 1.8162$ from a $t(9.8)$ distribution, CI is 0.4982 to 1.8474. Evidence shows that the angular velocity of skilled rowers is between about ½ and 2 units greater than for novice rowers.

11.49 **NICOTINE AND GUINEA PIGS** Many studies have shown that smoking during pregnancy adversely affects the baby's health. Researchers investigated the behavior of guinea pig offspring whose mothers had been randomly assigned to receive either a normal saline or nicotine injection throughout the pregnancy. Each group consisted of 15 randomly chosen male and female guinea pigs. At 85 days of age, 10 subjects from each group were randomly chosen to run a maze and choose a black door rather than a white door at the end of the maze. The number of trials it took each guinea pig to complete the task successfully with no more than one mistake in two consecutive days was recorded.

Here are the summary statistics on number of trials to successful completion for the nicotine group and the control (normal saline) group:

Groups	n	$\bar{x}$	s
Nicotine	10	111.9	50.28
Control	10	75.6	27.512

(a) Is there a significant difference in the mean number of trials recorded between the treatment group and the control group?

- Let m_n = the mean number of trials to completion for the nicotine group and m_c = the mean number of trials to completion for the control group.
- $H_0 : m_n - m_c = 0$
  $H_A : m_n - m_c \neq 0$
- Assume both samples are SRS
  Reasonable to assume the samples are independent
  10, 10 < 10% of all possible guinea pig offspring
  With the small sample sizes, and with no way to check the normality of the data, the *t* procedures will be only approximately accurate.
- Two sample ***t*** test
- $t = \dfrac{(111.9 - 75.6) - 0}{\sqrt{\dfrac{50.28^2}{10} + \dfrac{27.512^2}{10}}} = 2.003;\quad df = 13.95\,(\textit{from calc})$
- *p*-value = 2P(*t* > 2.003) = 0.065
- The ***p***-value is greater than 5%, so we fail to reject H_0. There does not seem to be very strong evidence of a difference between the two means.

(b) Calculate two 95% confidence intervals for the difference in the mean number of trials required to complete the task for the treatment group and the control groups. Use the conservative number of degrees of freedom for the first interval. For the second interval, use the more precise df given by the formula on page 659. Comment on what you notice.

Conservative df = 9; t^* = 2.262 the

95% CI for $m_n - m_c$ is $(111.9 - 75.6) \pm 2.262\sqrt{\dfrac{50.28^2}{10} + \dfrac{27.512^2}{10}}$

$(-4.7, 77.3)$

Calculator df = 113.95; the 95% CI for $m_n - m_c$ is (-2.587, 75.187)

The more precise df results in a slightly narrower interval

11.51 **TREATING SCRAPIE IN HAMSTERS** Scrapie is a degenerate disease of the nervous system. A study of the substance IDX as a treatment for scrapie used as subjects 20 infected hamsters. Ten, chosen at random, were injected with IDX. The other 10 were untreated. The researchers recorded how long each hamster lived. They reported, "Thus, although all infected control hamsters had died by 94 days after infection (mean ± SEM = 88.5 ± 1.9 days), IDX-treated hamsters lived up to 128 days (mean ± SEM = 116 ± 5.6 days).

(a) Fill in the values in this summary table:

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

(b) What degrees of freedom would you use in the conservative two-sample t procedures to compare the two treatments? Use df = 9, the smaller of 9 and 9.

11.53 **TEACHING READING** An educator believes that new reading activities in the classroom will help elementary school pupils improve their reading ability. She arranges for a third grade class of 21 students to follow these activities for an 8-week period. A control classroom of 23 third graders follows the same curriculum without the activities. At the end of the 8 weeks, all students are given the Degree of Reading Power (DRP) test, which measures the aspects of reading ability that the treatment is designed to improve. Here are the data:

Treatment					Control				
24	43	58	71	43	42	43	55	26	62
49	61	44	67	49	37	33	41	19	54
53	56	59	52	62	20	85	46	10	17
54	57	33	46	43	60	53	42	37	42
57					55	28	48		

(a) Examine the data with a graph. Are there strong outliers or skewness that could prevent us of the t procedures?

A back-to-back stemplot shows 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.

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

(b) Is there good evidence that the new activities improve the mean DRP score? Carry out a test and report your conclusions.

- Let m_t = the mean reading ability score of students in the treatment group and m_c = the mean reading ability score of students in the control group.

- $H_0 : m_t - m_c = 0$
  $H_A : m_t - m_c > 0$

- Assume both samples are SRS
  Reasonable to assume the samples are independent
  21 < 10% of all possible 3rd grade students using new reading techniques
  23 < 10% of all 3rd graders following standard curriculum
  Graphs of the data are explored in (a)

- Two sample *t* test

- $t = 2.311$; $df = 37.9$ (*from calc*)

- *P*-value = P(*t* > 2.311) = 0.0132

- The *p*-value is very small, so we reject H_0. The data provide evidence that students in the treatment group score higher on the reading ability test than students in the control group.

(c) Although this study is an experiment, its design is not ideal because it had to be done in a school without disrupting classes. What aspect of good experimental design is missing?

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

11.59 **STUDENTS' SELF-CONCEPT** Here is the SAS output for a study of the self-concept of seventh-grade students. The variable SC is the score on the Piers-Harris Self Concept Scale. The analysis was done to see if male and female students differ in mean self-concept score.

TTEST PROCEDURE

Variable: SC

SEX	N	Mean	Std Dev	Std Error
F	31	55.51612903	12.69611743	2.28029001
M	47	57.91489362	12.26488410	1.78901722

Variances	T	DF	Prob > | T |
Unequal	-0.8276	62.8	0.4110
Equal	-0.8336	76.0	0.4071

Write a sentence or two summarizing the comparison of females and males, as if you were preparing a report for publication.

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.

11.63 **SHARKS** Great white sharks are big and hungry. Here are the lengths in feet of 44 great whites.

18.7	12.3	18.6	16.4	15.7	18.3	14.6	15.8	14.9	17.6	12.1
16.4	16.7	17.8	16.2	12.6	17.8	13.8	12.2	15.2	14.7	12.4
13.2	15.8	14.3	16.6	9.4	18.2	13.2	13.6	15.3	16.1	13.5
19.1	16.2	22.8	16.8	13.6	13.2	15.7	19.7	18.7	13.2	16.8

(a) Examine these data for shape, center, spread, and outliers. The distribution is reasonably normal except for one outlier in each direction. Because these are not extreme and preserve the symmetry of the distribution, use of the t procedures is safe with 44 observations.

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 $\bar{x} = 15.59$ and the standard deviation is $s = 2.550$ ft.

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

(b) Give a 95% confidence interval for the mean length of great white sharks. Based on this interval, is there significant evidence at the 5% level to reject the claim "Great white sharks average 20 feet in length"?

Using our calc, we can find $t^* = 2.0167$ for df = 43, which gives CI = 15.59 ± 0.78 ft.
Since 20 ft does not fall in (or even near) this interval, we reject this claim.

(c) It isn't clear exactly what parameter μ you estimated in (b). What information do you need to say what μ is?

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?)

11.68 **EACH DAY I AM GETTING BETTER IN MATH** A "subliminal" message is below our threshold of awareness but may nonetheless influence us. Can subliminal messages help students learn math? A group of students who had failed the mathematics part of the City University of New York Skills Assessment Test agreed to participate in a study to find out.

All received a daily subliminal message, flashed on a screen too rapidly to be consciously read. The treatment group of 10 students (chosen at random) was exposed to "Each day I am getting better in math." The control group of 8 students was exposed to a neutral message, "People are walking on the street." All students participated in a summer program designed to raise their math skills, and all took the assessment test again as the end of the program. The table below gives data on the subjects' scores before and after the program.

TABLE 11.8 Mathematics skills scores before and after a subliminal message

Treatment Group		Control Group	
Pre-test	Post-test	Pre-test	Post-test
18	24	18	29
18	25	24	29
21	33	20	24
18	29	18	26
18	33	24	38
20	36	22	27
23	34	15	22
23	36	19	31
21	34		
17	27		

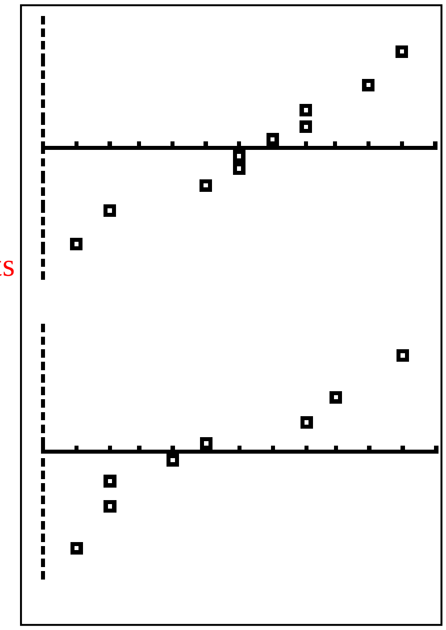
Source: Data provided by Warren Page, New York City Technical College, from a study done by John Hudesman.

(a) Is there good evidence that the treatment brought about a greater improvement in math scores than the neutral message? State hypotheses, carry out a test, and state your conclusion. Is your result significant at the 5% level? At the 10% level?

- Let m_1 = the mean improvement ("after" minus "before") in treatment group and m_2 = the mean improvement ("after" minus "before") in control group.

- $H_0: m_1 - m_2 = 0$
  $H_A: m_1 - m_2 > 0$

- SRS, given
  Reasonable to assume independence of the groups
  10 < 10% of possible students, 8 < 10% of possible students
  NPP of the observed differences are fairly linear,
  Despite the small sizes, the *t* procedures should be reliable.



- Two sample *t* test

- $t = \dfrac{(11.4 - 8.25) - 0}{\sqrt{\dfrac{3.17^2}{10} + \dfrac{3.69^2}{8}}} = 1.913;\ \ df = 13.9\,(\textit{from calc})$

- *p*-value = P(*t* > 1.913) = 0.038

- The ***p***-value is small, so we reject H_0 at both the 5% and 10% significance levels.
  The data provide evidence that students in the treatment group show greater improvement than students in the control group.

(b) Use your calculator to find a 90% confidence interval for the mean difference in gains between treatment and control.

The 90% CI is (0.025, 6.05). Again, since the CI does not include zero, the data show that the students in the treatment group show greater improvement than students in the control group.