WEEK 7: CHAPTER 7, THE NORMAL DISTRIBUTION [AND NON-PARAMETRIC 1-SAMPLE TESTS]

TABLE OF CONTENTS

List of Figures

List of Tables

List of m.files

Assignment

Required reading

! Larsen, R. J. and M. L. Marx. 2006. An introduction to mathematical statistics and its applications, 4th edition. Prentice Hall, Upper Saddle River, NJ. 920 pp.

  "   Read All of Chapter 7
  "   Read 14.1 (p. 802-807)
  "   Read 14.3 (p. 810-820)

Understanding by Design Templates

Understanding By Design Stage I — Desired Results Week 7 (7/12-7/18)
LM Chapter 7 The Normal Distribution

G Established Goals
- Become proficient in calculating and interpreting confidence intervals
- Recognize whether and how to apply one-sample tests

U Understand
- The Student T distribution converges to the normal distribution
- The probability that a 95% confidence interval contains the true value of a parameter is never 95%.
- The expected value for the F distribution is 1 and the chi square distribution it is the degrees of freedom

Q Essential Questions
- Do (1-)% confidence intervals contain the true value of the parameter (1-)% of the time?
- What are degrees of freedom and how are they determined?
- What are three common ways of presenting the precision of an estimate?

K *Students will know how to define (in words or equations)*
- **Bonferroni adjustment**, **chi-square distribution**, **F distribution**, **Fisher's sign test**, **normal distribution**, **one-sample *t* test**, **sampling distributions**, **Student *t* distribution**, **Wilcoxon's sign rank test**

S *Students will be able to*
- Calculate confidence intervals for μ and 2 using the Student t and chi-square distributions, respectively
- Apply the one-sample *t* test, Sign test and Wilcoxon one-sample sign rank test to data and interpret the decision rules and p-values

Understanding by Design Stage II — Assessment Evidence Week 7 (7/12-7/18 M)

Chapter 7

- **Post one comment and one reply in the discussion section by 7/20/11 W 10 PM**
    - Why do we estimate a 95% confidence limit for μ and σ^2 instead of a 95% confidence limit for $\bar{X}$ and s^2, respectively?
- **HW 6 Problems due Wednesday 7/13/11 W 10 PM**
    - **Basic problems (4 problems 10 points)**
        - Problem 7.4.16 (p. 489) hint: use LMex070401_4th.m to check the assumptions of the one-sample Student T test.
        - Problem 7.4.18 Use LMex070403_4th.m as a model. Use a boxplot or histogram to assess whether the t-based CI can be used
        - Problem 7.5.14 Use data from Case Study 7.5.1 and program from Case Study 7.5.2. Use a left-tailed test.
        - Problem 14.2.6 (p 810, data from case study 1.2) Use case study 14.2.2 as a model
    - **Advanced problems (2.5 points each)**
        - **Problem 7.5.6** Use LMex070501_4th.m as a model
        - **Problem 7.5.16**
    - **Master problems (1 only, 5 points)** 7.5.4 Hint: use fsolve, chi2cdf and LMex070501_4th.m

Introduction

Synopsis

This handout will cover the following topics
! Definitions and theorems from the chapter
! Statistical tests
    " One sample t test: known $μ_o$, variance estimated from the sample
    " One sample variance test
    " Fisher's sign test
    " Wilcoxon signed rank test
! Examples & Case Studies from Chapter 7 applying these concepts.

Introduction

Definitions, Tests and Theorems

Testing μ $\underline{\mu}_o$ σ , (**known)**

In the previous chapter **Theorem 6.2.1** (**Larsen & Marx, 2006, p. 435**) introduced the 1-sample
Z test.

Theorem 6.2.1 Let Y_1, Y_2, … Y_n be a random sample of size n from a normal distribution where
is known. Let $z = \dfrac{\bar{y} - \mu_o}{\sigma/\sqrt{n}}$.

"   To test $H_o: \mu = \mu_o$ versus $H_1: \mu > \mu_o$ at the   level of significance, reject H_o if
           $z \geq z$.

"   To test $H_o: \mu = \mu_o$ versus $H_1: \mu < \mu_o$ at the   level of significance, reject H_o if
           $z \leq -z$.

"   To test $H_o: \mu = \mu_o$ versus $H_1: \mu \neq \mu_o$ at the   level of significance, reject H_o if z is
           either (1) $\leq -z_{/2}$ or (2) $\geq z_{/2}$

In most statistical testing,   is not known and must be estimated from the sample or samples. In
that case, the standard normal, or z, distribution is not appropriate. The usual distribution that is
used to account for the reduced precision when   must be estimated from a sample is Student's t
distribution, invented by William S. Gossett of the Guinness Brewing Company, who published

his famous t ratio under the pseudonym 'Student:' $t = \dfrac{\bar{y} - \mu_o}{s/\sqrt{n}}$. **Larsen & Marx (2006)**

introduce the use of Student's t distribution, introduced in Theorem 7.3.5, with their **Theorem
7.4.3.**

Theorem 7.3.5 Let Y_1, Y_2 , … , Y_n be a random sample from a normal distribution with mean μ
and standard deviation   . Then

$$ T_{n-1} = \frac{\bar{Y} - \mu}{S/\sqrt{n}} $$

has a Student t distribution with n-1 degrees of freedom.

The name 'degrees of freedom' was one of R.A. Fisher's many contributions to statistics. It
arose in a heated debate with Karl Pearson (father of Egon Pearson of Neyman-Pearson, the
inventors of critical values and confidence limits) over the appropriate chi-square distribution to
use for testing contingency tables. Fisher proposed a geometric explanation, describing the
degrees of freedom of spread of variation in a space describing a distribution. (Fisher won the
battle with Pearson, showing that the appropriate d.f. was (r-1)*(c-1), where r and c are the
number of rows and columns in a contingency table.) We've already used d.f. in conjunction

with fitting the chi-square distribution to Prussian horsemen deaths and Iraqi casualties. In those cases, one parameter, the Poisson parameter , was estimated from the data so the tests are performed using a ^{2} distribution with *n-1 df*, where n is the number of observations and the lost d.f. is due to the estimation of .

Current definitions of degrees of freedom rarely mention the geometric aspects. **Legendre & Legendre (1998)** have an apt discussion of the concept of degrees of freedom, shown below: **On confidence limits & Neyman-Pearson theory**

Degrees of freedom   Box 1.2

Statistical tests of significance often call upon the concept of degrees of freedom. A formal definition is the following: "The degrees of freedom of a model for expected values of random variables is the excess of the number of variables [observations] over the number of parameters in the model" (Kotz & Johnson, 1982).

In practical terms, the number of degrees of freedom associated with a statistic is equal to the number of its independent components, i.e. the total number of components used in the calculation minus the number of parameters one had to estimate from the data before computing the statistic. For example, the number of degrees of freedom associated with a variance is the number of observations minus one (noted $v = n - 1$): n components $(x_i - \bar{x})$ are used in the calculation, but one degree of freedom is lost because the mean of the statistical population is estimated from the sample data; this is a prerequisite before estimating the variance.

There is a different *t* distribution for each number of degrees of freedom. The same is true for the *F* and χ^2 families of distributions, for example. So, the number of degrees of freedom determines which statistical distribution, in these families (*t*, *F*, or χ^2), should be used as the reference for a given test of significance. Degrees of freedom are discussed again in Chapter 6 with respect to the analysis of contingency tables.

Figure 1. Degrees of Freedom discussed by **Legendre & Legendre's (1998)** Numerical Ecology.

Larsen & Marx (2006) is written firmly in the school of statistics proposed by Neyman & Pearson. In this approach to hypothesis testing, an   level is established in advance of performing the statistical test or even before the experiment or survey is undertaken, usually =0.05. This   is the probability of Type I error (Type I and Type II errors were Neyman-Pearson innovations), the probability of rejecting a true null hypothesis by chance. After analyzing the data, the test statistic is compared to the critical value and critical regions for the test statistic. If the test statistic, say Student's t, is in the critical region beyond the appropriate critical value, then the test was judged as providing significant evidence that the null hypothesis was false. The results would be reported as, "I rejected the null hypothesis that $\mu_1 = \mu_2$ at the 5%   -level." This approach to hypothesis testing made a great deal of sense when calculations had to be done by hand. A test could be performed and compared to tabulated values of the critical values of test

statistics at say the =0.1, 0.05, and 0.001 levels of significance. Now, however, most reputable journals insist that the actual *p* values for the test be included. Matlab, for example, will calculate the *p values* for all of the major test statistics, moreover, Matlab can perform the exact probability analysis for many of the major statistical tests and can perform randomization tests to produce *p values* to most desired levels of accuracy.

Ronald Fisher never accepted the Neyman-Pearson approach to hypothesis testing with critical values, insisting that the *p values* should be reported. Recall that the ***p-value*** associated with an observed test statistic is the probability of getting a value for that test statistic as extreme or more extreme than what was actually observed (relative to H_1) *given that H_o is true*. Indeed, **Mayo's (1996)** synthesis of frequentist hypothesis testing points out that Egon Pearson was less than enthused about the black-white, significant-nonsignificant dichotomy that the Neyman-Pearson approach entailed. **Mayo (1996)** argues that Neyman & Pearson's most influential contribution to scientific advance was the introduction of the Type II error and confidence intervals. Fisher's signficance levels could be used to reject null hypotheses, but Type II errors forced the investigator to specify alternate hypotheses. Confidence intervals, especially the 95% confidence intervals, allow an investigator to report the effect size and provide an estimate of the uncertainty of the effect size that allows a quick assessment of alternative hypotheses.

Confidence limits usually have the following form: Effect size ± {Z, t [more rarely ^{2} or F}* Standard Error. There is some confusion among practicing scientists about the difference between the standard deviation and the standard error. Part of that confusion is that there are many standard errors, the most common being 'the **standard error of the mean**,' which is the conventional {standard deviation of the sample}/$\sqrt{n}$. As **Larsen & Marx (2006, Section 4.3, pp. 292-299)** discuss and demonstrate with Example 4.3.2, the underlying population distribution may take a variety of forms — including uniform, Poisson, binomial, geometric — but as sample size increases, test statistics such as the sample mean and the difference in mean will tend towards the normal

distribution. **Ramsey & Schafer (2002)** present a nice graphic (shown at right) showing this effect for a 2-sample *t* test for the difference in means between a pair of populations. The underlying distributions are non-normal, but the difference inmeans has a distribution that is
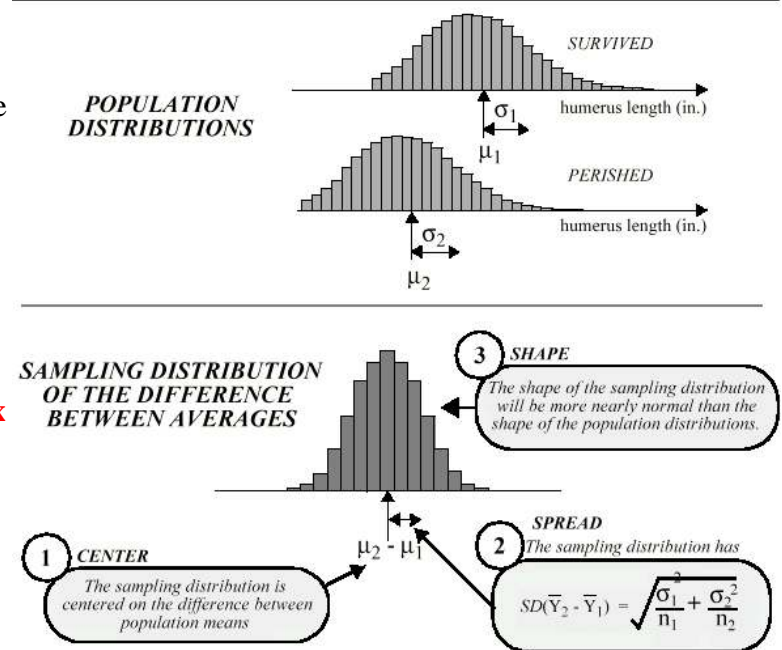
Figure 2. Demonstration of the effect of the central limit theorem on the distribution of differences in means from Ramsey & Schafer's Statistical Sleuth

more nearly normal. The standard deviation of the difference in means is also called the 'standard error' of the difference in means. It can be calculated from the standard deviations of the individual samples as shown in ②. Since the standard deviations are estimated from the samples, the appropriate multiplier for the 95% confidence interval around the observed difference in means would be the Student's t statistic with d.f. equal to $n+m - 2$, since two standard deviations are being estimated from the data.

The variance of the test statistic, such as the difference in means shown in Figure 2 will decrease, proportionate to $\sqrt{n}$. So, when reporting a test statistic, such as the observed difference in means, the appropriate measure of precision is either the standard error or the 95% confidence interval. In scientific meetings, it is not unusual to see the estimates of precision, or error, represented as ± 1 standard error. Indeed, that is how the MA Dept of Education reports the variability on MCAS scores. If the variance is known from theory, one can immediately recognize that the standard error must be multiplied by the critical value of the z statistic associated with the 97.5th percentile, or 1.96, to obtain

Figure 3. Effects of d.f. on the magnitude of Student's t statistic used to construct 95% confidence intervals. The Matlab program to calculate this (without labels),
df=1:12;alpha=.05;p=1-alpha/2;fyt = tinv(p,df);bar(df,fyt);grid

95% confidence intervals [I've often seen error bars on plots based on means with just 3 replicates. The appropriate multiplier is then not $Z_{0.975}$ or 1.96, but $t_{0.975, 2\,df}$ or 4.3! That could make a huge difference in determining whether the results might be due to chance alone. Figure 3 shows the effect of sample size, d.f., on Student's t statistic. An investigator showing error bars for means based on 2 replicates, with no indication that n=2, may be trying to deceive his audience, since those error bars would have to use a Student's t statistic of 12.7 to convert to 95% confidence intervals.

Interpreting confidence intervals

Ramsey & Schafer (2002) provide a superb graphic, Figure 4 below, showing how to interpret the error bars, presented as 95% confidence intervals in presentations.

Display 5.17   p. 133

Separate confidence intervals for two group means: are the means different?
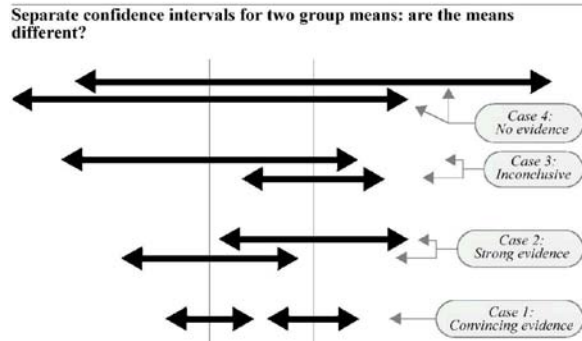
Figure 4. A rough guide to interpreting overlap in 95% confidence intervals, using the guidance that a p-value less than 0.05 is regarded as moderate to strong evidence of a difference. Note especially in Case 2 that the confidence limits can overlap and still produce a difference in sample statistics with p values less than 0.05. This interpretation of 'error bars' is not possible with standard errors if the sample sizes are not provided (and even with sample sizes provided, one would have to have a good memory of Student's t statistics for n<6-10.) [This is display 5.19 in the 2nd edition of Sleuth]

Confidence intervals, Standard Errors and Significant Figures for reporting results

Bevington & Robinson (1992) and **Taylor (1997)** are the two excellent 'how-to' guides on how to propagate errors and report errors in publications. Of the two, I prefer the more rigorous Bevington & Robinson, especially because it stresses Monte Carlo simulation, but Taylor may have the best cover photo of any recent book in statistics (see Figure 5). Neither book is in accord with current editorial practice in ecological, psychological or medication journals, because these authors recommend reporting standard errors, rather than 95% confidence intervals, in reporting results.

Both books argue that investigators should let the standard error (or confidence interval) dictate the level of precision reported for results. It is a sign of statistical ignorance to report a result as 7.51478 ± 0.5672. Note, in reporting results you should always state whether that estimate of error to the right of the ± is a
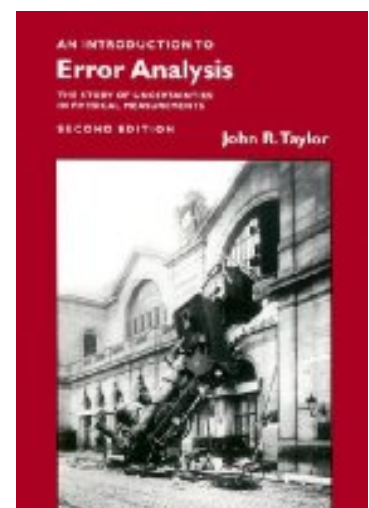
Figure 5. **Taylor (1997)** on error analysis.

http://images.amazon.com/ima
ges/P/093570275X.01._SX140
_SY225_SCLZZZZZZ_.jpg

standard deviation (a poor choice), a standard error (the most common in the physical sciences) or a half-95% confidence interval (the best choice in my opinion). This result could be reported as 7.5 ± 0.6. Bevington & Robinson argue that standard errors should be reported to 1 significant figure only, but I'll argue shortly that there are reasons for reporting 2 significant figures. As a follow-up recommendation, there are reasons for reporting 1 additional significant figure than can be justified by the data.

One of the most persuasive arguments for retaining 1 more significant figure than warranted by the data is revealed by the answer to the following question, "What is the normal human body temperature?" Of course, the answer is 98.6ºF, but that answer is wrong. John Alan Paulos, in his book, "A mathematician reads the newspaper" recounts the story of Wunderlich who carried out the most comprehensive survey of human temperatures. They are highly variable. I wish I could report his actual results, but I have been unable to find an appropriate citation, and Paulos doesn't provide citations for his story. Paulos reports that the actual data were reported to just two significant figures, say 37ºC $\pm$ 0.8. To Wunderlich, there was no reason to add a 3rd significant figure to his data. However, when people converted his 37ºC to Fahrenheit, they produced the now highly precise 98.6º F. According to Paulos, a report that I haven't been able to confirm, the best estimate of normal human temperature is 98.2º F. The conversion to Fahrenheit has given normal human temperature a diagnosis of a mild fever. Had Wunderlich presented that single extra, non-significant at =0.05, digit, the conversion to Fahrenheit would have been more accurate.

One sample t-test

Larsen & Marx (2006, p. 489-493) cover the one-sample Student's t test to test H_o: on $\mu = \mu_o$.

Theorem 7.4.2 Let $y_1, y_2, ..., y_n$ be a random sample of size n from a normal distribution where is unknown. Let $t = \dfrac{\bar{y} - \mu_o}{s/\sqrt{n}}$

        a.   To test H_o: $\mu = \mu_o$ versus H_1: $\mu > \mu_o$ at the   level of significance, reject H_o if $t \geq t_{,n-1}$.

        b.   To test H_o: $\mu = \mu_o$ versus H_1: $\mu < \mu_o$ at the   level of significance, reject H_o if t $t_{,n-1}$.

        c.   To test H_o: $\mu = \mu_o$ versus H_1: $\mu \neq \mu_o$ at the   level of significance, reject H_o if t is either (1) $-t_{/2,n-1}$ or $\geq t_{1-/2,n-1}$.

 For most common statistical procedures there will be several classes of statistical tests to test hypotheses:
!   The parametric test. In this case, the one-sample Student's t test
       "   It is available in Matlab's statistics toolbox as ttest.m
       "   I have programmed this test as Stud1sample.m.

!   The randomization test. There is a simple randomization test for the one-sample t test, invented by Fisher & discussed in **Manly (1991)**. In brief, Fisher calculated $\bar{X}$ -μ_o for the *n* data in the sample. Under the null hypothesis that μ=μ$_o$, the sum of positive and negative deviations should have an expected value of zero. Fisher randomized by randomly assigning the appropriate number of +'s and -'s to the observed deviations and determined how many sums of the deviations were equal to or more extreme than those observed.

!   Distribution-free tests.

"   All statistical tests have assumptions. Parametric tests, like the Student's *t* test, assume a parametric distribution for the errors. Student's one-sample and two-sample tests assume that the errors are independently and identically normally distributed. Now, I'd warrant that this is not strictly true for most applications of Student's *t* test. There are some assumptions that really matter and others that don't. If the *p-value* produced by a test if the assumptions are violated is about equal to the p-value if the assumption were met, we say that the test is robust to violations of that assumption. As noted by **Larsen & Marx (2006, p. 497)**, Student's *t* tests are remarkably robust to violations of the normality assumption. There are other assumptions of the t test that are more important than normality.

"   Nonparametric tests do not assume an underlying parametric distribution, like the normal or Student's *t* distributions, in deriving the p values. They are more properly named distribution-free tests. They **do** have assumptions though, as we'll see in the discussion of the two-sample t test in a few weeks.

-      There are a variety of 1-sample nonparametric procedures

     #   The most widely used is the Wilcoxon signed rank test

     *   Available as Matlab's signrank.m

     #   Fisher's sign test, Matlab's signtest.m

For a 1-sample test using only n, the mean and standard deviation, I wrote student1group.m

Student1group.m
function [D,t,df,pvalue,CI]=student1group(Xn,Xmean, Xstd, M, alpha)
% Student't 1-sample t test with grouped data
% [D,t,df,pvalue,CI]=student1group(Xn,Xmean,Xstd,alpha)
% Input: Xn = Size of group = number of cases
%   Xmean = mean for groups
%   Xstd, Ystd= standard deviations for 2 groups
%   M=expected mean, 0 if not provided
%   alpha level for CI, optional, 95% if not specified
% Output: D=Xmean-M or Xmean-0;
%   t=Student's t statistic
%   df=degrees of freedom for t statistic
%   p value, 2-sided, for t Statisitc, with df degrees of freedom
%   CI for 1-alpha CI, [L U];95% CI if alpha not specified.
% Based on Theorem 7.4.2, page 490 in Larsen & Marx (2006)
% Introduction to Mathematical Statistics, 4th edition.

```
% uses Statistics toolbox
% Written by Eugene.Gallagher@umb.edu, Revised 11/14/10
% see also stud1sample, binom2sample, student2group, stud2sample
if nargin<4
    M=0;
    alpha=0.05;
elseif nargin <5
    alpha=0.05;
end
D=Xmean-M;
% Calculate sample standard deviation
t=D/(Xstd/sqrt(Xn));
% Use stixbox's pt.m for significance of t.
df=Xn-1;
if t>=0
    pvalue=2*(1-tcdf(t,df));
else
    pvalue=2*tcdf(t,df);
end
HalfCI=tinv(1-alpha/2,df)*Xstd*sqrt(1/Xn);
CI=[D-HalfCI D+HalfCI];
```

Parametric Examples & Case Studies

Larsen & Marx (2006, pp. 490-492) provide two examples of Student's one-sample t-test, Examples 7.4.2 and 7.4.3. Both can be solved directly in a pair of equations using Theorem 7.4.2.

Case Study 7.4.1

What is the effective range of a bat's echolocation system? Construct a 95% CI for the 11 observations in the table shown in Figure 6.

This is a simple call to Matlab's normfit or ttest. TTEST will perform a one sample t test on the data too (with zero as the default μ), but I just want to get access to the confidence intervals.
X=[62 52 68 23 34 45 27 42 83 56 40];
fprintf('The mean reaction distance is %4.2f cm\n',mean(X));
[H,P,CI,STATS]=ttest(X);
fprintf('The 95%% confidence interval is (%4.2f cm %4.2f cm)5\n',CI)
% or call normfit
[MUHAT,SIGMAHAT,MUCI,SIGMACI] = normfit(X,alpha)

Table 7.4.1

Catch number	Detection Distance (cm)
1	62
2	52
3	68
4	23
5	34
6	45
7	27
8	42
9	83
10	56
11	40

Figure 6. Table 7.4.1 Bat detection distances

fprintf('The 95%% confidence interval is (%6.4f cm %6.4f cm)\n',MUCI)

The mean reaction distance is 48.3 cm. The 95% confidence interval is (36.2 cm to 60.5 cm)

Example 7.41

The data in the following table have mean 2.6 and s 3.6. Is it appropriate to say the 95% CI for the sample mean is 0.9 and 4.3?

2.5	0.1	0.2	1.3
3.2	0.1	0.1	1.4
0.5	0.2	0.4	11.2
0.4	7.4	1.8	2.1
0.3	8.6	0.3	10.1

No it isn't because the sample deviates greatly from a normal population. The book uses a simple histogram to show that the data are not normally distributed. Figure 7 shows Matlab's version of that histogram. Clearly the sample distribution is strongly postively skewed.
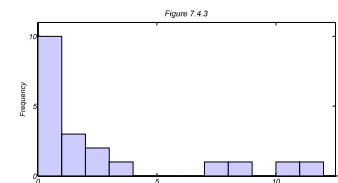
Figure 7. Histogram of the

The boxplot is a very nice method for assessing whether the *t* tools can be used. The term **t tools** was coined by **Ramsey & Schafer (2002)** to describe the set of statistical procedures that rely on the t distribution, including the creation of confidence limits and one-sample, two-sample and paired *t* tests. The boxplot of the Example 7.4.1 data is shown in Figure 8. There are four extreme outliers that could strongly affect the 95% confidence interval by greatly increasing the standard error of the sample.
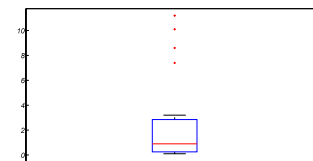
Figure 8. Boxplot of the Example 7.4.1 data with 4 outliers.

Matlab has a useful diagnostic function called normplot for plotting a sample vs. a normal cumulative frequency distribution given the samples mean and standard deviation. Figure 9 shows the normplot for the Example 7.4.1 data. When plotted on a normal probability scale, the observed data, converted to z scores, should fall along a straight line and Matlab's function provides that expected line. The observed data for Example 7.4.1 depart radically from a straight line, invalidating the use of the t tools for calculating a 95% confidence limit. The *t* tools are robust to normality, so the departures from linearity would have to be severe to invalidate the use of the t tools for calculating 95% confidence intervals.
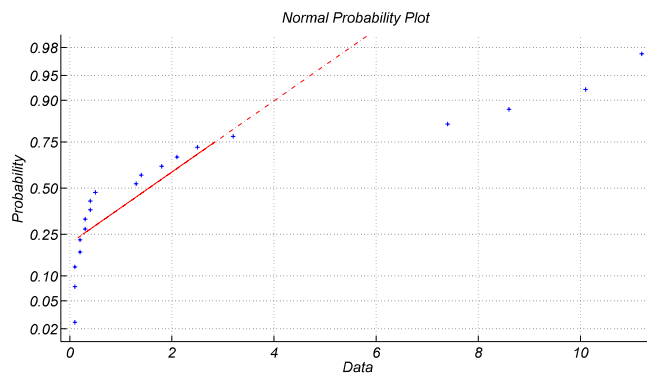
Figure 9. Normplot of the Example 7.4.1 data with the 4 outliers.

Environmental scientists often encounter data that don't meet the assumptions of the Student *t* test. The usual approach is to transform the data, a subject not covered in **Larsen & Marx (2006)**. For positively skewed data such as these, the log transform often works remarkably well. Figure 10 is the normal probability plot for the log-transformed data. Although a few data at the ends fall away from the line, there would be no problem
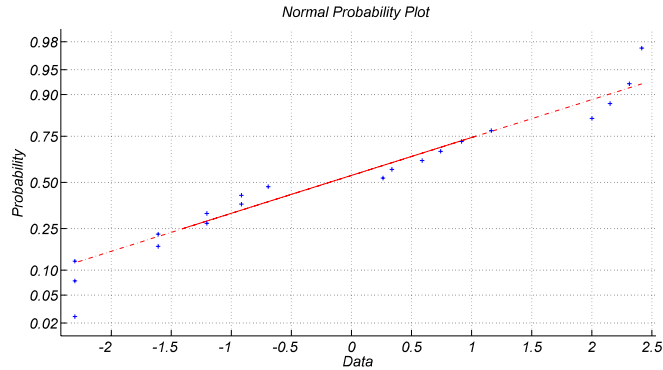
Figure 10. Normplot of the Example 7.4.1 data after log transformation. No evident problem with fit.

applying the t tools to the log-transformed data. While it would be inappropriate to calculate the 95% confidence interval based on the t tools in the natural scale, it is appropriate to calculate the mean and 95% confidence intervals of the log-transformed data. The mean and confidence limits can than be back transformed to the natural scale using antilogarithms to produce an estimate of the geometric mean and its 95% confidence interval. The geometric mean is 0.90 with 95% confidence intervals of (0.42 1.90). Because of the positive skew in the data, the geometric mean and its confidence interval is quite different from the arithmetic mean for these data and its confidence interval. The sample arithmetic mean is 2.61with calculated confidence interval of 0.92 and 4.30. Again, as noted in the text, the arithmetic mean and its confidence interval would be inappropriate for these data. As described below, the CDC routinely log transforms all pollutant data in its annual reports on the state of the nation's health and reports its results as back-transformed geometric means with appropriate confidence limits.

Example 7.4.2

Twelve students enrolled in a nursery had their blood lead levels measured. The average and standard deviation were 18.65 and 5.049 respectively. The expected lead level was 16 mg/l. Can\n it be concluded that children at this particular facility tend to have higher lead levels? At the = 0.05 level, is the increase from 16.0 to 18.65 statistically significant?

Results of analysis of LMex070402_4th.m

Matlab doesn't have a built-in function to perform the 1-sample t test using just a mean and standard deviation, but Gallagher wrote student1sample.m for that purpose. It produces the following results:
The 1-tailed p value = 0.04817 for t=*1*.8182
The blood lead difference of 2.65 mg/l had 95% CI =[-0.56 5.86]
Reject H*o*: Evidence that children had lead levels higher than 16.00
The difference from 'normal (=16 mg/l)' was 2.6 mg/l +/- 3.2 mg/l

Following the Neyman-Pearson model, **Larsen & Marx (2006, p. 490 bottom)** conclude, **"It would be correct to claim, then, that the $\bar{y}$ of 18.65 *does* represent a statistically significant**

increase in blood levels.'' Figure 11 shows how close the one-tailed t statistic is to the critical value of 1.7959. They plot almost on top of each other.

Figure 7.4.2

Figure 11. The Student t distribution for 11 df. Shown are is the 1-tailed upper 95% critical value of 1.7959. The test statistic of 1.82 is just to the right of this critical value, allowing Larsen & Marx (2006) to conclude that the results are statistically significant.

Note that the 95% CI does include zero. The program calculates this 95% CI in the conventional way with the 2-tailed critical value of the *t* distribution. The 1-tailed *p value* is less than 0.05 even though the two-tailed confidence interval includes the null hypothesis expectation of 0 difference in blood lead level. The difference in lead level in this sample of 12 from the expected value is calculated by Gallagher's program as 2.6 mg/l +/- 3.2 mg/l. This indicates no significant difference from the expected 16 mg/l **IF** the test was performed two-tailed, but the test was performed 1-tailed. But, can we really conclude that these data constitute a significant difference from the 16 mg/l 'normal' concentration. No.

Critique of Larsen & Marx's interpretation of Example 7.5.2

On the plus side, **Larsen & Marx (2006)** did mention, 'a random sample of 12 children.' Without mentioning random sampling, NO valid inferences to the larger population of 'children at Mighty Bear' could be made. A statistical test, like the one-sample *t* test, tells you only whether the results observed could have occurred by chance. The inference to a larger population from the 12 students sampled to the larger population of students, requires that a valid sampling procedure from the larger population be used. Fisher advocated simple random sampling. For example, use a random number generator, like Matlab's randperm, and if there are 50 students use the following procedure to select 12 for sampling:

i=randperm(50); % shuffles the integers 1 to 50;
Myrandomsample=i(1:12) % the SIMPLE RANDOM SAMPLE is the first 50.

Once the names associated with those 12 random numbers are selected, it is vital that they be sampled. If a student doesn't show up, you can't go to the next name on the list without introducing a potential bias in the sampling procedure. By bias, I mean that the estimate from the sample average of 12 may not estimate the true population mean. If students who don't show up routinely have higher (or lower) lead levels, say due to socioeconomic background and a potential increased chance of living in houses with lead paint, then the sample average will underestimate the true population mean. There is a distinct possibility that the estimator will be biased. **Larsen & Marx (2006, p. 381-385)** define & discuss 'unbiasedness.' In sampling from a population, a simple random sample will produce an unbiased estimate of μ, the population mean.

Now, one could think that you could do BETTER than simple random sampling, and indeed you can, but only with GREAT care in selecting an alternative to simple random sampling. One can produce a more efficient estimator, meaning an estimator with a lower standard error (**Larsen & Marx, 2006 p 396**). However, this requires a fair amount of prior information about the population, and adjustments for the sample average AFTER it is calculated. In stratified random sampling, the population is divided into 'strata' or groups which are strongly suspected to differ with respect to the response variable, such as lead concentration. Usually, samples are allocated proportionately to the strata with the highest variance. Don't use stratified random sampling unless you have clearly stated the statistical model that you are using, including preliminary information on the variances among strata, and you have a well-defined statistical model for calculating the expected values and means after you've collected the samples. Improper stratification and improper analysis of the data can lead to greatly biased estimates of population means and variances. Fisher argued strongly in favor of simple random sampling. If you were to stratify the sampling, such that the 3 major socioeconomic groups were sampled: African-Americans, Hispanic & White, with 4 blood samples each, you COULD NOT calculate the sample average and variance as if samples were taken using simple random sampling. One would have to analyze the groups separately, and the averages would have to be weighted by the relative proportions of students in each group. Calculating the variances of stratified samples is beyond the scope of EEOS601.

Ramsey & Schafer's (2002) Statistical Sleuth presents a simple BUT VERY IMPORTANT graphic showing the importance of random sampling, shown as Figure 12. The study design used in Larsen & Marx's Example 7.5.2 would be an example of the upper right box "Random samples are selected from existing distinct populations" Since the selection of units, 12 students in this case, was random,
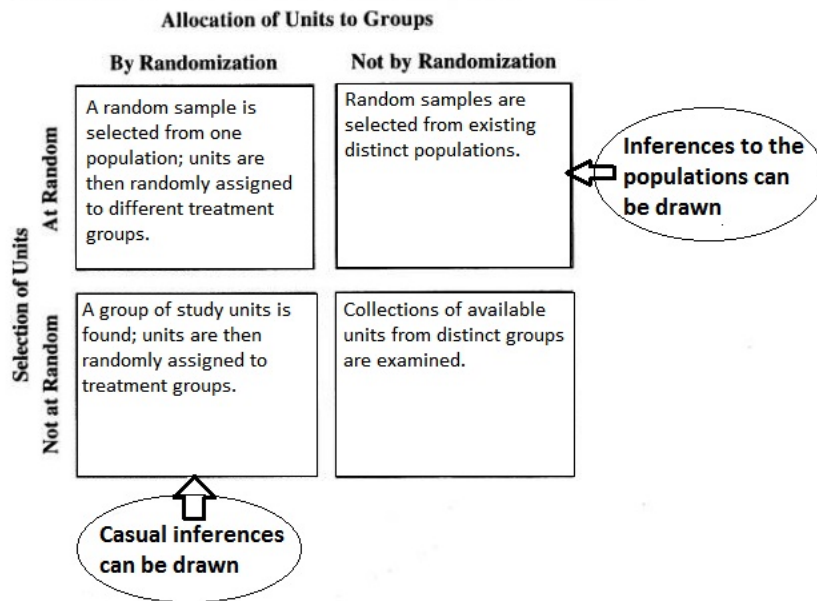
Figure 12. statistical inferences allowed by statistical tests.

"Inferences to the population can be drawn" If **Larsen & Marx (2006, p 490)** had not included that 1 phrase about random sampling, one could only conclude that the average lead content of these 12 students is higher than 16 mg/l, but you didn't need statistics to determine that.

Presentation of the results

In current practice, this result could be reported as offering moderate evidence in favor of an elevated lead level. One could report the 1-tailed (1-sided) *p value* as 0.048. Note that one has to present 3 significant figures to get the p value under that somewhat arbitrary, but conventional, =0.05. In this case, it would be misleading to present the results as merely, "This study showed a significantly enhanced lead level." The *p value* indicates that the lead level offers only moderate (some would say weak) evidence that the lead levels were higher than 16 mg/l. Now, this test was performed 1-tailed. If the investigator had no prior information about expected lead levels BEFORE CALCULATING THE AVERAGE, the result should have been reported with the two-tailed test, with a *p* value of 0.096. This would pose no difficulties for the current approach of reporting the *p values*, but would have changed the whole interpretation of the results using the Neyman-Pearson hypothesis testing using the decision rule at an   level of 0.05. This example provides a striking example of why journal editors and reviewers now insist on seeing *p* values and effect sizes.

What is the effect size for this study? Even if the test is performed 1-tailed (*i.e.,* the alternate hypothesis is that lead levels are HIGHER than 16 mg/l), the confidence limits are almost always reported using CI's calculated with t statistics for the two-tailed test. The effect size (the difference from 16 mg/l) would be reported as:

Difference from 16 mg/l = 2.6 mg/l $\pm$ 3.2 mg/l

This reported effect size offers little convincing evidence for elevated lead concentrations, and, of course, the 95% CI for the difference includes 0, as it must since the *p value* for the 2-sided test was p=0.096 (far higher than 0.05). If the investigator did not have valid *a priori* justification for suspecting enhanced lead levels in this nursery, he should not have used the 1-tailed test. Simply finding the mean is greater than expected is not a proper reason for using the 1-tailed test.

Was the correct statistical test used?

Much of the controversy in statistical analysis involves the extent to which inferences can be made to larger populations. However, occasionally one finds that the wrong test was used. Such is the case with Larsen & Marx's example 7.4.2. The one sample t-test, described in Section 7.4 assumes that μ_o, the population mean is a known parameter. Unfortunately, the normal lead concentration is not a known parameter in the human population. There are published averages and associated confidence limits, but it is not a known parameter and can't be treated as such. **Larsen & Marx (2006)** should NOT have used a 1-sample t test for these data. They should have saved this example for the two-sample t test in the next chapter.

The Center for Disease Control (CDC) has just published the results of the fourth national report of human exposure to environmental chemicals. These analyses are online at: **http://www.cdc.gov/exposurereport/pdf/FourthReport.pdf**

Larsen & Marx (2006, p. 490) report, *"Among children between the ages of 6 months and 5 years, blood levels of 16.0 mg/l are considered 'normal.'"* They apply the 1-sample t test as if

the 16 mg/l (identical to 1.6 µg/deciliter as the CDC reports their results) was measured without error, or is a known parameter. It isn't. The following table, Figure 13, shows the estimated values for lead in the US population. In 1999 the geometric mean concentration was 1.66 with a 95% confidence interval of 1.6 to 1.72 µg/dl based on 7970 individuals. One can forgive Larsen & Marx for using 16 mg/l, since that may have been an old lead value, but it isn't appropriate to treat the 16 mg/l as if it were measured without error. Recall that the observed average concentration for the 12 students was 1.865 ± 0.321 mg/dl. Using Figure 11 from Ramsey & Schafer above, the reported confidence limits offers only inconclusive evidence for elevated lead concentration (The 18.65 is beyond the upper confidence limit of 1.8 from the national survey, but the lower CI for the sample of 12 extends to 1.54 mg/dl, far below the 1.66 mg/l standard. There are insufficient data provided to do a formal 2-sample t test, but it is a certainty that the 2-sample test would reveal no differences (since the 1 sample test assuming 1.6 mg/dl had a 2-tailed pvalue=0.096). It is also certain that the two-sample t test, which included the standard error of the 1.6 mg/dl, would show a p value >0.05 since the 1-sample t test had a *p value* of 0.048 and that assumed that the 1.6 mg/dl had no error. As the statistician W. Edwards Deming emphasized, natural processes are often associated with variability. Only a small part of the error in blood lead levels is due to measurement error. Almost all is due to differences among individuals. The tables, most not shown below, also reveals striking differences in lead concentration within age, gender and ethnic groups.

Blood Lead

Geometric mean and selected percentiles of blood concentrations (in µg/dL) for the U.S. population from the National Health and Nutrition Examination Survey.

	Survey years	Geometric mean (95% conf. interval)	50th	75th	90th	95th	Sample size
Total	99-00	1.66 (1.60-1.72)	1.60 (1.60-1.70)	2.50 (2.40-2.60)	3.80 (3.60-4.00)	5.00 (4.70-5.50)	7970
	01-02	1.45 (1.39-1.51)	1.40 (1.40-1.50)	2.20 (2.10-2.30)	3.40 (3.20-3.60)	4.50 (4.20-4.70)	8945
	03-04	1.43 (1.36-1.50)	1.40 (1.30-1.50)	2.10 (2.10-2.20)	3.20 (3.10-3.30)	4.20 (3.90-4.40)	8373
	05-06	1.29 (1.23-1.36)	1.27 (1.20-1.34)	2.01 (1.91-2.11)	3.05 (2.86-3.22)	3.91 (3.64-4.18)	8407
Age group							
1-5 years	99-00	2.23 (1.96-2.53)	2.20 (1.90-2.50)	3.40 (2.80-3.90)	4.90 (4.00-6.60)	7.00 (6.10-8.30)	723
	01-02	1.70 (1.55-1.87)	1.60 (1.50-1.80)	2.50 (2.20-2.90)	4.20 (3.50-5.20)	5.80 (4.70-6.90)	898
	03-04	1.77 (1.60-1.95)	1.70 (1.50-1.90)	2.50 (2.30-2.80)	3.90 (3.30-4.60)	5.10 (4.10-6.60)	911
	05-06	1.46 (1.36-1.57)	1.43 (1.34-1.55)	2.10 (1.97-2.20)	2.98 (2.72-3.32)	3.80 (3.49-4.54)	968
6-11 years	99-00	1.51 (1.36-1.66)	1.40 (1.30-1.60)	2.10 (1.80-2.50)	3.30 (2.80-3.80)	4.50 (3.40-6.20)	905
	01-02	1.25 (1.14-1.36)	1.20 (1.00-1.30)	1.70 (1.60-2.00)	2.80 (2.50-3.10)	3.70 (3.00-4.70)	1044
	03-04	1.25 (1.12-1.39)	1.20 (1.10-1.40)	1.80 (1.50-2.10)	2.60 (2.10-3.10)	3.30 (2.50-4.60)	856
	05-06	1.02 (.948-1.10)	.970 (.890-1.01)	1.40 (1.28-1.55)	2.06 (1.80-2.72)	3.00 (2.26-3.81)	934
12-19 years	99-00	1.10 (1.04-1.17)	1.10 (1.00-1.20)	1.50 (1.40-1.70)	2.30 (2.10-2.40)	2.90 (2.70-3.00)	2135
	01-02	.942 (.899-.986)	.900 (.900-1.00)	1.30 (1.20-1.40)	2.00 (1.90-2.10)	2.70 (2.40-2.90)	2231
	03-04	.946 (.878-1.02)	.900 (.800-1.00)	1.30 (1.20-1.40)	1.90 (1.70-2.10)	2.60 (2.20-3.00)	2081
	05-06	.797 (.746-.852)	.740 (.690-.790)	1.08 (.990-1.20)	1.69 (1.50-1.85)	2.23 (1.98-2.46)	1996

Figure 13. Just part of a table from the CDC's updated tables for the fourth report on exposure to chemicals in the environment.
http://www.cdc.gov/exposurereport/pdf/Update_Tables.pdf

Example 7.4.2 is probably composed of 'made-up' data, but it is interesting to note that the geometric mean lead concentration in the US for children 1-5 years in 1999-2000 was 2.23

µg/dl, with a 95% CI for the mean of 1.96 to 2.53 µg/dl. Given these data, a statistician would be making TWO errors in applying a 1-sample *t* test to test whether the observed 1.865 mg/dl lead was less than the national average (a 2-sample test could be used). Moreover, one should decide in advance whether the hypothesis should be performed 1-tailed or 2-tailed. If one calculates the average and then automatically sets the alternative hypothesis to minimize the *p value* (by choosing the appropriate alternative hypothesis) one will be doubling the chance of committing a Type I error.

What are geometric means?

The CDC reports lead concentrations as geometric means, which is a common way of reporting data that have been log transformed. **Larsen & Marx (2006, p. 385-386)** provide a definition of geometric mean that only a mathematician could appreciate. For the non-mathematician, when you calculate the average of log-transformed data and thenn back transform with the antilog, you get the geometric mean. If natural logs were used $e^{\bar{X}}$ is the geometric mean, of $\log_{10}(x)$ was used then $10^{\bar{X}}$ is the geometric mean.

Geometric means are widely used, and the reasons for their use have direct relevance to the use of Student's *t* tests. Student's *t* tests, least-squares regression, and parametric ANOVA all assume that the errors are 1) identically, 2) independently, and 3) normally distributed. The underlying data don't have to be normally distributed – indeed they rarely are. However, after a model is fit, even a model as simple as fitting an average to the data, the deviations from the model, the errors, should be independently, identically, normally distributed. In the older statistical literature, there were tests for normality proposed, but it is now known that the Student's *t* test is remarkably robust to violations of normality. By robust, I mean that the *p* values are not affected much by deviations from normality. Many of the tests are affected by the shape of the distribution. The normal distribution is symmetric, while many population parameters have a skewed distribution (defined on p. 200 in **Larsen & Marx (2006)**). Many statistics in nature have a right-skewed distribution, like the lognormal distribution. Pollutant concentrations in humans tend to be right or positively skewed, often following the log-normal distribution.

Figure 14 shows one of many right-skewed distributions that have a mean of 18.6 with standard deviation 5, similar to Larsen & Marx's Example 7.4.2.

Figure 14. An example of a lognormal distribution with mean 18.6 and standard deviation 5. 1000 cases were randomly generated by a Matlab m.file that generates lognormally distributed samples.

Twelve samples were drawn randomly from this distribution producing Figure 15. The single sample with 30 mg/l lead appears to be an outlier.

The boxplot for these positively skewed data is shown in Figure 16.

Figure 16. A Tukey boxplot of the 12 simulated lead data from the previous figure. The red centerline represents the median. The left & right box border are Tukey hinges, designed to contain 50% of the data. The width of the box is called the interhinge length. The whiskers —| extend to no more than 1.5 times the interhinge length (=the interquartile rang) and end with an observed datum. The values outside the whiskers but less than 3 interhinge lengths from the box are called outside values (sometimes called outliers). The data more than 3 interhinge lengths from the hinge are called extreme values (sometimes extreme outliers). These data had 1 extreme value at 31. The median was 16.5 and the lower hinge and upper hinges were at 14.9 and 20.6. The hinges approximate the lower and upper quartiles of a dataset.

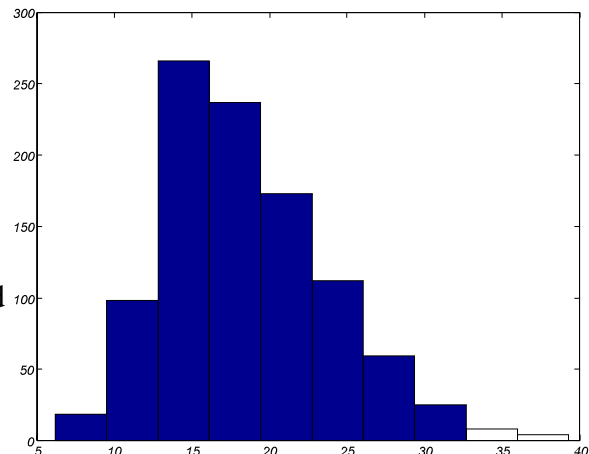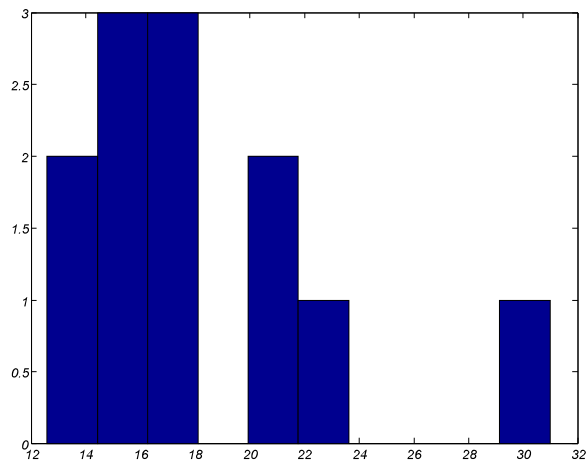**Figure 15**. An example of a lognormal distribution with mean 18.6 and standard deviation 5. Only 12 cases were randomly generated from this distribution, the same distribution used to generate the 1000 cases in Figure 14.

Now, one unfamiliar with statistics might think that a single case in Figure 17 with a blood level of 30 was an outlier. It would be a grave mistake to automatically delete points such apparent outliers as these without a thorough review of the data.

When data appear skewed like Figure 17, the standard procedure is to log transform the data. If there are zeros in the data, then a log(X+1) transform can be used. The data shown in Figure 16 were log transformed producing the boxplot shown in Figure 17. Several features are obvious. While the distribution is still skewed to the right somewhat, the single datum at 31 mg Pb/ml is no longer an extreme value. On the untransformed scale, that single datum could drastically affect both the mean and the standard deviation of the distribution. A log transform would be the appropriate transformation to perform on these data, prior to performing either a one-sample or two-sample Student's t test.
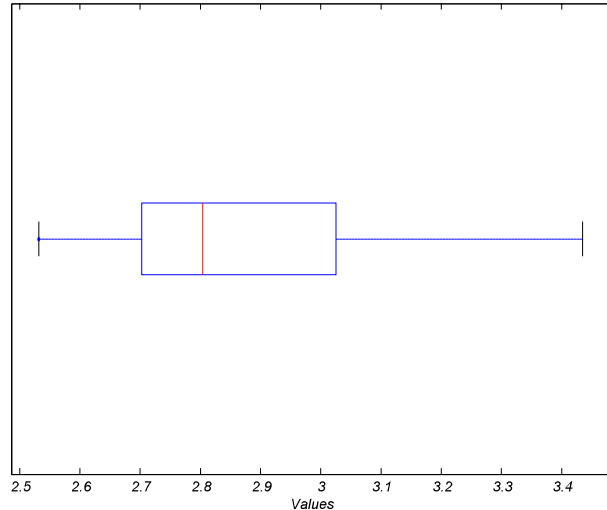
Figure 17. A Tukey boxplot of the 12 simulated lead data from the previous 2 figures after log transform

Geometric means Statistical tests are performed on the log transformed data when appropriate. The means and standard deviations are performed on the log transformed data. Any base logarithm could be used, natural logarithms (log(x) in Matlab) or log base 10 (log10(x)) are both used, producing identical results on the statistical tests. In reporting the results of these tests, report the p values AND report the back transformed mean. if $\bar{X}$ = mean(log(X)), then report the results as the back-transformed mean, $e^{\bar{X}}$, also known as the geometric mean [if y was the column of untransformed data, then exp(mean(log(y))) would find the geometric mean in Matlab]. The upper and lower 95% confidence intervals for the mean should be calculated on the transformed data and these confidence limits should be suitably back transformed, $\left[e^{Lower\ CI}, e^{Upper\ CI} \right]$. That is how the CDC has reported the geometric mean value and 95 confidence limits for each of the environmental chemicals measured in human blood (see Figure 7 above).

It may seem odd working on log-transformed data, but most of the data and analyses that I work with require some sort of transformation. The log transform is the most common. The boxplot is THE workhorse tool to judge the adequacy of the transformation, although there are more formal tests that could be used. The Student t tests are robust to small violations of normality, and a transformation that produces a more symmetric distribution is generally adequate for the t tests. Outlier detection and evaluation is a BIG issue. One should NOT delete outliers routinely.

Other common transformations used for data include the square root transform (often appropriate for Poisson distributed data, and the logit transform (log (p/(1-p)) for frequency data. In a famous paper, G.E.P. Box, son-in-law of Fisher, and Cox developed a method for analyzing a whole family of transforms, including the square root transform and log transform to

determine which is the best for producing a symmetric distribution of errors. R. E. Strauss distributes a free program called boxcox.m which will apply the boxcox transform on a set of data. His library of statistics m.files can be accessed at his Texas Tech web site:
http://www.biol.ttu.edu/Strauss/Matlab/Matlab.htm
The best description of the Box-Cox transform's use that I've seen is in **Draper & Smith (1998, p. 280-289)**, but one should be wary. With the very small sample size of 12 shown in the above figures, the Box-Cox method indicates that a very odd transform of 1/sqrt(y) is the most appropriate transform to produce symmetry.

Example 7.4.3, Page 491, an application of the 1-sample t test.

Twelve banks are surveyed to determine whether they are discriminating against inner city neighborhoods. The approval rate for rural residents is 62%. The percentage approval for inner city applicants is provided for three banks. LMex070403_4th.m produces the analyses reported in the book.

Comments on example 7.4.3

Wow, where should I begin in pointing out the problems with Example 7.4.3. First, and most important, a one-sample t-test is inappropriate here! One can not assume that the 62% approval rate for rural applicants is measured without error. One should have performed a test for a difference in proportions or a two-sample t test. The binomial test, described in Section 6.3, would have been a more appropriate test. The 62% approval rate for rural residence has a variance of n*p*(1-p), and this variance must be taken into account in assessing whether these 12 banks have proportional acceptance rates for inner city residents that differs from that of rural residents. Since the p value for the inappropriate 1-sample t test, which assumes that 62% approval for rural residents is a parameter, μ_o, with no associated error, was greater than 0.05, it is virtually a certainty that the *p value* for the binomial test, which assumes that p is estimated, not known, would be greater than 0.05. **Larsen & Marx (2006)**, in their example, do not provide the sample sizes for the binomial proportions for each bank, so it is impossible to do the binomial test for these data.

After testing all 12 banks, drawn from 3 affiliates, **Larsen & Marx (2006)** fail to reject the null hypothesis, reporting the t statistic and critical values in Table 7.4.3. It would have been clearer if they had adopted modern convention and reported the p value of **0.125**. However, they state, "The overall analysis of Table 7.5.3, though, may be too simplistic. Common sense would tell us to look at the three banks separately. What emerges then is an entirely different picture ..."

If the statistical analyst only decided to test the three separate bank types AFTER beginning the analysis of the data, then one CAN NOT use   =0.05 as the estimate of Type I error. We can use basic probability concepts to calculate the effects of testing 3 separate bank types, AFTER THE FACT, with each test being performed at alpha=0.05. What is the probability of seeing AT LEAST ONE significant result at   =0.05, if there were no true differences between any of the banks at the true proportion of 62%? It is:

   1- (1-alpha)* (1-alpha)*(1-alpha)=1-(1-alpha)^3=1-(1-alpha)^{Number of Tests}

The probability of Type I error is no longer 5%, but 14.2% (1-(1-0.05)^3). If there were 10 types of banks being tested, after the fact, the Probability of Type I error would rise to 40%! One can also solve this problem using the binomial pdf:

>> 1-binopdf(0,10,0.05)% 1-Prob of 0 successes in 10 trials with p=0.05
ans =   0.4013 % or a 40% chance of declaring at least 1 test significant,
                        % if none were truly different

In order to maintain a proper experiment-wise alpha level, or probability of Type I error, one must adjust the   level for the number of tests performed. The most widely used correction is the Bonferroni correction (not mentioned in Larsen & Marx). If you are performing tests that were not planned in advance, so called *a posteriori* tests, then alpha must be adjusted for the number of comparisons: level for tests = {overall   level (usually 0.05)}/{Number of tests}. This approximation works well, but it is just a shortcut for the following identity, solved for in LMex070403_4th.

        alpha=0.05;ntests=3;alphatest=1-exp(log(1-alpha)/ntests)

This indicates that each of the 3 tests in Larsen and Marx's (2006) Example 7.4.3 would have to be performed at an alpha level of 0.017 (slightly higher than 0.05/3) to ensure that P(TypeI error)=1-(1-alpha_test)^3, is indeed 0.05. The results of performing the 3 tests with the appropriate corrections are shown below, printed directly from the output of LMex070403_4th.m:

The Bank being analyzed is AU
The mean Percentage approved is 52.25% vs. 62.00% expected
For Bank AU, the difference is -9.75%, with 2-tailed p value=0.036
For bank AU, **reject** the null hypothesis at alpha=0.050
The 2 tailed critical values for the t statistic=4.33 and alpha=0.050, are -3.18 and 3.18

The Bank being analyzed is FT
The mean Percentage approved is 58.80% vs. 62.00% expected
For Bank FT, the difference is -3.20%, with 2-tailed p value=0.145
For bank FT, fail to reject the null hypothesis at alpha=0.050
The 2 tailed critical values for the t statistic=4.33 and alpha=0.050, are -2.78 and 2.78

The Bank being analyzed is TU
The mean Percentage approved is 67.00% vs. 62.00% expected
For Bank TU, the difference is 5.00%, with 2-tailed p value=0.049
For bank TU, **reject** the null hypothesis at alpha=0.050
The 2 tailed critical values for the t statistic=4.33 and alpha=0.050, are -4.30 and 4.30

Analyses using Bonferroni adjustment based on number of tests (3)
The Bonferroni adjusted alpha, based on 3 tests is 0.017.

The Bank being analyzed is AU
The mean Percentage approved is 52.25% vs. 62.00% expected
For Bank AU, the difference is -9.75%, with 2-tailed p value=0.036

For bank AU, **fail to reject** the null hypothesis at Bonferroni-adjusted alpha=0.017
The 2 tailed critical values for the t statistic=4.33 and alpha=0.017, are -4.83 and 4.83

The Bank being analyzed is FT
The mean Percentage approved is 58.80% vs. 62.00% expected
For Bank FT, the difference is -3.20%, with 2-tailed p value=0.145
For bank FT, **fail to reject** the null hypothesis at Bonferroni-adjusted alpha=0.017
The 2 tailed critical values for the t statistic=4.33 and alpha=0.017, are -3.94 and 3.94

The Bank being analyzed is TU
The mean Percentage approved is 67.00% vs. 62.00% expected
For Bank TU, the difference is 5.00%, with 2-tailed p value=0.049
For bank TU, **fail to reject** the null hypothesis at Bonferroni-adjusted alpha=0.017
The 2 tailed critical values for the t statistic=4.33 and alpha=0.017, are -7.58 and 7.58

Conclusion on LMex070403

Larsen & Marx used the wrong test, choosing a one-sample *t* test when a two-sample binomial test (or similar test for the same proportions) should have been used. Moreover, they failed to adjust the alpha level for the number of tests performed, AFTER the FACT, or *a posteriori*. When the alpha level for these tests is adjusted to maintain the overall probability of Type I error at 0.05, by using alpha_test=0.017, none of the three separate tests provide even moderate evidence for rejecting the null hypothesis that the proportion of loans approved was 62%. In Larsen & Marx's analysis, without an adjustment of the   levels, 2 of the 3 a posteriori or unplanned comparisons were significant.

Case Study 7.5.1

K-Ar dating was performed on 19 rocks believed to be at the same age. Assume that the procedure's estimated ages are normally distributed with (unknown) mean μ and (unknown) variance ^{2}. Construct a 95% confidence interval for   .

Now the easiest way to find the solution to this is to use Matlab's **NORMFIT**, which finds the maximum likelihood estimators for μ, and   and their 95% confidence limits.
[MUHAT,SIGMAHAT,MUCI,SIGMACI] = normfit(X)

The function produces the following output:
MUHAT =   276.8947; SIGMAHAT =27.0820
MUCI =   263.8416 289.9478
SIGMACI = 20.4635 40.0495

Table 7.5.1

Specimen	Estimated Age (millions of years)
1	249
2	254
3	243
4	268
5	253
6	269
7	287
8	241
9	273
10	306
11	303
12	280
13	260
14	256
15	278
16	344
17	304
18	283
19	310

Figure 18. K-Ar Ages of 19 rocks believed to be the same age.

In my Matlab script m.file LMcs070501_4th.m, I find the 95% CI using Theorem 7.5.1, which produces an identical result. So, the 95% confidence interval for sigma is 20.4 to 40.1 million years. I'm rounding up slightly to ensure that the reported 95% CI includes the computed confidence limits

Example 7.5.1

What is the minimum sample size such that the width of the 95% confidence interval for the variance is equal to or less than the variance? This is a straightforward application of Theorem 7.5.1 and can be solved in three lines in Matlab with Matlab's numeric equation solver fsolve:
```
alpha=0.05;
x = fsolve(@(n) (n-1)*(1/chi2inv(alpha/2,n-1)-1/chi2inv(1-alpha/2,n-1))-1,30)
fprintf(...
'The min n such that the %4.1f%% CI width for sigma^2 < sigma^2 is %2.0f\n',...
100*(1-alpha),ceil(x));;
```

FSOLVE is one of several Matlab equation solvers. It will find an estimate of a parameter or parameters needed to solve an equation. The equation or function should be written such that the solution is 0. Matlab iteratively searches parameter space near starting values that you provide. The above implementation of **fsolve** I've used what Matlab calls 'an anonymous function' by including the function to be solved in the same statement as the call to fsolve. The '30' in the call is an initial guess of n. With default parameters, fsolve will find the solution with any initial guess greater than 1 and about 200. For values much above 200, Matlab can't find a solution with its default number of iterations.

The solution to the problem is that n is 38.5622, but only integers are solutions, so a sample size of 39 is the minimum such that the width of the 95% confidence interval for ^{2} will be equal to or less than ^{2}.

Case Study 7.5.2

During the first quarter of 1994, Tennessee lenders were charging an average rate of 8.84% for a 30-year fixed mortgage. The standard deviation was 0.1%. Figure 19 shows the interest rates from 9 lenders on 1-year adjustable rates. Are these interest rates more variable than 30-year fixed rates?

Table 7.5.3

Lender	Initial Rate on 1-Year Adjustables
AmSouth Mortgage	6.38%
Boatmen's National Mortgage	6.63
Cavalry Bank	6.88
First American National Bank	6.75
First Investment	6.13
First Republic	6.50
NationsBanc Mortgage	6.63
Union Planters	6.38
MortgageSouth Corp.	6.50

Figure 19. Interest rates on 1-year adjustables. Are they more variable than 30-year mortgages?

This is Matlab 3-liner:
```
X=[6.38;6.63;6.88;6.75;6.13;6.50;6.63;6.38;6.50];n=length(X);df=n-1;
sd=.1;alpha=0.05;
[H,P,CI,STATS] = vartest(X,sd^2,alpha,'both')
```

Programming Theorem 7.5.2 just takes 3 additional lines:

chi2=df*var(X)/sd^2
pvalue=(1-chi2cdf(chi2,df))*2
fprintf('The p-value for chi2 = %6.2f & %1.0f df is %5.2e\n',chi2,df,pvalue)

The variance for 30-y fixed mortgages is 0.01% and 1-year adjustable mortgages is 0.05%. Testing the hypothesis that the variances are equal, the p-value for observing a chi-squared statistic of 39.77 or more extreme if the null hypothesis is true with 8 df is 7.07×10^{-6}.

Nonparametric 1-sample tests

Fisher's Sign Test

Fisher invented the deceptively simple but very powerful sign test to test one-sample problems. It is available as Matlab's signtest, but it is easy enough to calculate using principles from early in the course. I'll use the data from example 7.4.3 to describe the test, which is described in the text in Section 14.2. If the true proportion is 62% then subtract this expected proportion from the observed proportion. The expected number of differences with positive signs should equal the expected differences with negative signs. This can be tested exactly with the binomial probability density function. Of the 12 banks surveyed, only 4 of the 12 had approval rates greater than 62%. What is the probability of observing that result, OR ONE MORE EXTREME, by chance. Using Matlab's binopdf.m, that is easy to calculate. One has to calculate the binomial pdf for 0, 1, 2, 3 and 4, given the null hypothesis that states that the probability of negative and positive deviations from the expected values should be 0.5 (equal likelihood of positive and negative deviations):

k=0:4;p1tailed=sum(binopdf(k,12,0.5))
p1tailed =   0.1938

In this case, the 2-tailed p value is exactly twice the one-tailed p value and is **0.3877.** This could also be obtained by finding the probabilities of those values as or more extreme than 3 on both sides of the expected value. A simple Matlab program to do this is:
Expected=0.5*12; Observed=4; EO=Expected-Observed;
k=0:12;i=find(abs(Expected-k)>=EO);p2tailed=sum(binopdf(k(i),12,0.5))

The decision would be 'fail to reject' H$_o$ at the   level of 0.05. The sign test offers no evidence to reject the null hypothesis that the true proportion is 62%. This could be solved in one step using the binomial cumulative distribution function as well:
p1tailed=binocdf(4,12,0.5).

Matlab has Fisher's sign test as part of the statistics toolbox. It could be called using the Example 7.4.3 data in one line:
[P,H,STATS] = signtest([59 65 69 53 60 53 58 64 46 67 51 59],62,'method','exact')

This signtest will also perform the large sample normal approximation, but in most cases Matlab can calculate the exact solution.

The sign test does not assume any underlying parametric distribution for the data, such as normally distributed errors. The relative efficiency of this test is 95% for n=6, declining to 63% for large n, IF THE ASSUMPTIONS OF THE T TEST ARE MET. The asymptotic relative efficiency of a statistic is a difficult concept, but there is a nice graphical display in Figure 6.4.5 in Larsen & Marx (2006, p. 450), shown as Figure 20. If the assumptions of the parametric test are not met, the power efficiency of the nonparametric test can be MUCH higher than the parametric test.
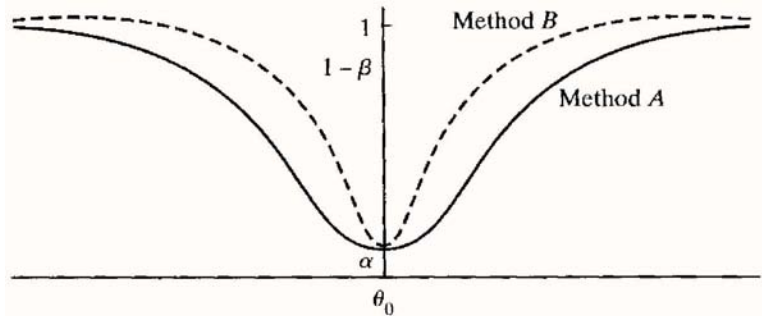
Figure 20. Power curves for two statistical tests, A & B. Method B has the higher asymptotic relative efficiency. The probability of Type II error ( ) is lower for B at most values for the alternate hypothesis, and the power (1-) is higher for Method B. The asymptotic relative efficiency reflects the reduction in sample size for the more powerful method to achieve the same power as the less powerful method. A relative power efficiency of 0.5 of test A relative to B, would mean that Method A would require twice as many samples to achieve the same power as method B.

One-sample Wilcoxon signed rank test

The sign test doesn't take into account the magnitude of the deviations from the expected value. There is a one-sample nonparametric test which DOES take into account the magnitude of the differences from the expected value. In the case of Example 7.4.3, the expected value is 62. The 12 banks had observed proportions shown below:

X= [59   65   69   53   60   53   58   64   46   67   51   59]

The Wilcoxon signed rank statistic subtracts the expected value from each observed value. It then ranks the observations from smallest to largest, without respect to sign, and assigns ranks to these values.

Observed differences:

-3   3   7   -9   -2   -9   -4   2   -16   5   -11   -3

Ranked by absolute value, there are 3 sets of tied ranks, and these are assigned average values.

4   4   8   9.5   1.5   9.5   6.0   1.5   12   7   11   4

Note, that there were 2 values tied with abs(obs-exp)=2, assigned an average rank of 1.5, 3 absolute differences of 3, assigned an average rank of 4 (=mean ([3 4 5]), and two differences of -9, assigned an average rank of 9.5.

The Wilcoxon signed rank statistic is the sum of positive ranks (could also be the sum of negative ranks too). I've programmed the test as Wilcoxsignrank, but Matlab has a nice version called signrank.m, called by:

[P,H,STATS] = signrank(X,62,'method','exact')

Matlab's solution is:

P =0.1577; H =0;STATS = signedrank: 20.5000

The signed rank statistic 20.5 is the sum of positive ranks (4+8+1.5+7)

There is an exact test available for the Wilcoxon signed rank statistic (not to be confused with the Wilcoxon rank sum statistic to be discussed in the 2-sample tests section). The approximate test, is based on the normal distribution and produces a 2-tailed pvalue of 0.1458.

Note that the p values for the Student t 1-sample test was p=0.125, the sign test was p=0.3877 and the Wilcoxon signed rank test was p=0.1577 . This is not surprising as the power efficiency of the nonparametric tests is often quite high compared to their parametric equivalents.

Nonparametric Case Studies

Case Study 14.2.1

Synovial fluid is the clear, viscid secretion that lubricates joints and tendons. Fourty-four patients with arthritis had the pH of they synovial fluid measured. In healthy patients the mean pH of synovial fluid is 7.39. At an   -level of 0.05, use an appropriate test to determine whether synovial pH can be used to test for the presence of arthritis.

I saved the data in vector D and then plotted them using Matlab's histc. I examined Figure 22, a histogram of the data with a superimposed normal probability plot. The data are poorly fit by the normal distribution with outliers on both tails of the distribution, ruling out most standard variance-normalizing transformations like the log transform. A non-parametric test, either a sign test or Wilcoxon signed rank test would be appropriate for these data.

Matlab's sign test is called in one line:

[p,h,stats] =
signtest(D,7.39,0.05,'method','exact');

Table 14.2.1

Subject	Synovial Fluid pH	Subject	Synovial Fluid pH
HW	7.02	BG	7.34
AD	7.35	GL	7.22
TK	7.32	BP	7.32
EP	7.33	NK	7.40
AF	7.15	LL	6.99
LW	7.26	KC	7.10
LT	7.25	FA	7.30
DR	7.35	ML	7.21
VU	7.38	CK	7.33
SP	7.20	LW	7.28
MM	7.31	ES	7.35
DF	7.24	DD	7.24
LM	7.34	SL	7.36
AW	7.32	RM	7.09
BB	7.34	AL	7.32
TL	7.14	BV	6.95
PM	7.20	WR	7.35
JG	7.41	HT	7.36
DH	7.77	ND	6.60
ER	7.12	SJ	7.29
DP	7.45	BA	7.31
FF	7.28		

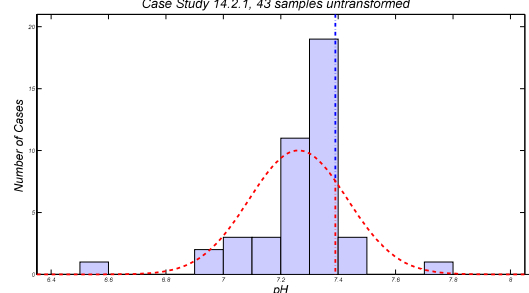**Figure 21**. Synovial pH data from 44

Figure 22. Synovial pH data from 44 patients with a superimposed fit to the normal distribution. I've also marked, pH 7.39, the pH of healthy patients.

Only 4 of the 44 pH measurements were less than 7.39 when the expected number would be 22 if the median pH of arthritic patients was 7.39. The two-sided p value for observing that result is $p=3.108 \times 10^{-8}$.

With these data, it would also be possible to perform a Wilcoxon signed rank test. The test is called by:
[P,H,STATS] = signrank(D,7.39,'alpha',0.05,'method','exact')

The signed rank test The signed rank test exact p value of 2.207×10^{-8}. The p-value for the one-sample t test with these data produced a two-tailed p-value of $p=1.36 \times 10^{-5}$. While the power efficiency of the parametric one-sample t test is greater than that of the two non-parametric equivalents when the assumptions of the test are met, such is not always the case if the assumptions are violated. With the non-normal data of Case Study 14.2.1, both nonparametric tests produced much lower *p values* than the parametric 1-sample t test.

Case Study 14.2.2

There are two different processes, freeze-drying and spray-drying, for producing instant coffee. There is interest in whether the caffeine concentration differs between these two treatments. The median amount of caffeine left by the freeze-drying method is 3.55 g caffeine per 100 g of dry matter. Listed in Figure 24 are the caffeine residues for eight brands of coffee produced by the spray-dried method.

Table 14.2.2

Brand	Caffeine Residue (gms/100gms dry weight)
A	4.8
B	4.0
C	3.8
D	4.3
E	3.9
F	4.6
G	3.1
H	3.7

Figure 23. Caffeine residues in 8 brands of spray-dried instant coffee.

Eight of the 7 coffees produced caffeine residues of 3.55 g per 100 g residue. How likely is that to have occurred by chance if the methods had similar amounts of caffeine per g of residue?

There are three appropriate tests for the null hypothesis of no difference in caffeine concentration: the 1-sample t test, the Wilcoxon signed rank test and Fisher's sign test. Figure 24 shows the distribution of the 8 samples relative to a fitted normal curve. The data are too few to provide much power in ruling out the normal distribution. All three tests would be appropriate, and the one-sample t test would probably have the highest power. For pedagogical purposes only, I'll run all three tests.
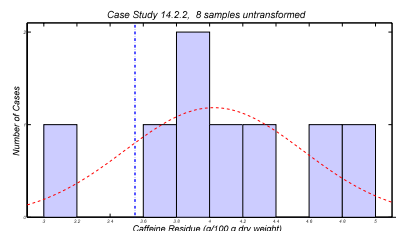

Figure 24. Caffeine residues in 8 brands of spray-dried instant coffee. Also shown as a vertical line is the freeze-dried coffee caffeine concentration of 3.55 g per 100 g residue.

The mean caffeine = 4.03 (g/100g residue) with 95% CI: [3.57 4.48], which doesn't include the 3.55 g caffeine per 100 g residue concentration found for the freeze-dried method. The one-sample *t* test has a 2-tailed p value of 0.041. The sign test statistic has an exact 2-tailed p value of 0.07031. he signed rank test statistic is 4.0 with exact 2-tailed p=0.05469. Now, in the formal Neyman-Pearson hypothesis testing procedure, if an   level of 0.05 had been set in advance, then one could reject the null hypothesis if one had chosen the t test in advance. But, one could not reject the null hypothesis at an   -level of 0.05 if the Wilcoxon signed rank test or sign test were used.

Case Study 14.3.1

Swell sharks are found off the coast of California south of Monterey Bay. Shown in Figure 25 are measurements of the total length and height of the first dorsal fin for 10 sharks caught off Santa Catalina. "It has been estimated on the basis of past data that the true average TL/HDl ratio for male swell sharks caught off the coast is 14.6. Is that figure consistent with the data shown in the table?

Table 14.3.2: Measurement Made on Ten Sharks Caught near Santa Catalina

Total Length (mm)	Height of First Dorsal Fin (mm)	TL / HDI
906	68	13.32
875	67	13.06
771	55	14.02
700	59	11.86
869	64	13.58
895	65	13.77
662	49	13.51
750	52	14.42
794	55	14.44
787	51	15.43

Figure 25. Lengths of 10 swell sharks caught off the coast of Santa Catalina Island.

The first step in performing an analysis is to quickly inspect the data to note any violation of assumptions. Figure 26 shows the distribution of shark length ratios. There are too few data to assess whether there are violations of the normality assumption of the 1-sample t-test. It is surely ok, but for pedagogical reasons, I'll analyze the data with all three one-sample tests. The one-sample t test, Fisher's sign test and the Wilcoxon signed rank tests. I'll use the exact versions of the latter two tests.

Figure 26. Lengths of 10 swell sharks caught off the coast of Santa Catalina Island.

The one-sample t test is the most effective test for analyzing data if its assumptions are met. In addition to being the most powerful test, it produces a direct way of reporting the effect size. The mean TL/HDl for Santa Catalina sharks is 13.74 with 95% CI: [13.06 14.42]. The 95% confidence interval does not include the mainland coastal value of 14.6. The one-sample t test has a 2-tailed p of 0.01889. The sign test statistic has an exact 2-tailed p value of 0.02148. The signed rank test statistic is 4.5 with exact 2-tailed p=0.01563.

These data support would provide modest evidence for the rejection of the null hypothesis that the Santa Catalina and coastal mainland sharks are members of the same population.

Case Study 14.3.2

Cyclazocine is a drug used to treat heroin addicts. Fourteen males were given a test to measure their psychological dependence on heroin. Figure 27 shows their Q scores after treatment. The mean Q score for addicts not given cyclazocine is 28. Is there evidence that cyclazocine is an effective treatment?

Table 14.3.4: Q-Scores of Heroin Addicts after Cyclazocine Therapy	
51	43
53	45
43	27
36	21
55	26
55	22
39	43

Figure 27. Q scores for 14 heroin addicts

The normality assumption is not met for the data whose distribution is plotted in Figure 28. The distribution is platykurtic, but it is reasonably symmetric. The one-sample t test would probably be app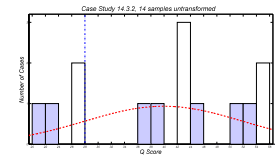ropriate. **Larsen & Marx (2006, p. 819)** argue that it would be better to analyze the data with a Wilcoxon one-sample signed rank test, which assumes only symmetry and not normality. For pedagogical purposes, I'll analyze the data with all three one-sample tests.

The mean Q score was 39.9 and since the test was performed one tailed, only the lower 95% CI of 34.3 is calculated. This confidence interval does not contain the non-addict Q score of 28. The one-sample t test 1-tailed p value is 0.001255. The Wilcoxon signed rank test statistic is 10.0 with exact 1-tailed p value of 0.002563. The signed rank test approximate 1-tailed p=0.003778 for z=-2.6713. I'd never consider reporting the approximate p

Figure 28. Distribution of the Q scores for 14 heroin addicts.

value if I'd calculated the exact p value, but I'm including it here because the text uses this case study as an example of the large sample normal approximation for the Wilcoxon signed rank test. The sign test statistic is 4.0 with exact 1-tailed p value of 0.08978. There is quite a difference in p values between the three tests. Even though the data aren't normally distributed, I would probably use the one-sample t test with these data because of the ease with which the effect size can be reported.

Annotated outline (with Matlab scripts) for Larsen & Marx Chapter 7

7   **The Normal distribution**
Francis Galton quote
     7.1   **INTRODUCTION**
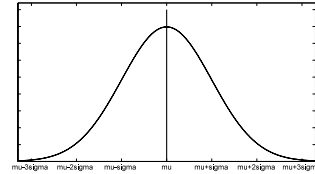
Figure 29. The normal distribution

Normal distribution:

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{1}{2}[(y-\mu)/\sigma]^2}, \quad -\infty < y < \infty$$

     7.2   Comparing $\dfrac{\overline{Y}-\mu}{\sigma/\sqrt{(n)}}$ and $\dfrac{\overline{Y}-\mu}{s/\sqrt{(n)}}$

[2] is seldom known; Are there probabilisitic differences between $\dfrac{\overline{Y}-\mu}{\sigma/\sqrt{(n)}}$ and $\dfrac{\overline{Y}-\mu}{s/\sqrt{(n)}}$? Willam

Sealey Gossett gets credit for discovering the difference.
Figure 30 (Figure 7.2.1a in the text) shows the normal distribution superimposed on 500 sample of size 4, in which the z ratios were calculated from the known population ($=1$)

Figure 30. Normal distributrion plotted with the histogram of 500 z ratios calculated with known ($=1$) from 500 samples of size 4.

Figure 31 (Figure 7.2.1b in the text) shows the normal distribution superimposed on 500 samples of size 4, in which the z ratios were calculated from *s* estimated from each sample.

```
% LMFig070201_4th.m
% Use as a model
LMFig040301_4th.m
% Page 471-472 in
% Larsen & Marx (2006)
Introduction to Mathematical
Statistics, 4th edition
% written by
Eugene.Gallagher@umb.edu,
3/12/11 6:50, revised 3/12/11
```


Figure 31. Normal distribution plotted with the histogram of 500 z ratios calculated with known (=1) from 500 samples of size 4. The histogram is poorly modeled by the normal distribution in the tails (the histogram is platykurtic relative to the normal)
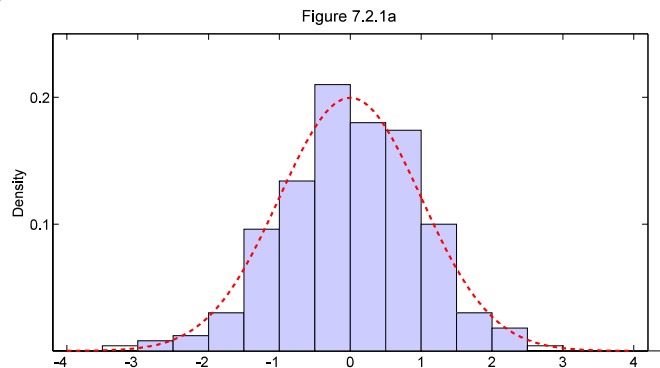
```
% Generate 400 samples from a standard normal distribution, size 4
clf;cla;hold off
K=menu('Choose an Option','Figure 7.2.1a','Figure7.2.1b')
Binsize=.5;
n=4;N=500;
if K==1
    R1 = normrnd(0,1,n,N);
    Z1=(mean(R1)-0)./(1./sqrt(n));
    tl='Figure 7.2.1a';
else
    R2 = normrnd(0,1,n,N);
    Z1=(mean(R2)-0)./(std(R2)./sqrt(n));
    tl='Figure 7.2.1b';
end
edges=-4:Binsize:4;
[N1,BIN1] = histc(Z1,edges);
bar(edges,N1/N,'histc');
axis([-4.2 4.2 0 0.25]);
set(get(gca,'Children'),'FaceColor',[.8 .8 1])
ylabel('Density','FontSize',20);
ax1=gca;set(ax1,'xtick',-4:4,'FontSize',20);
set(ax1,'ytick',[0.1 0.2],'FontSize',20)
figure(gcf);pause
% Superimpose the normal probability pdf:
% The more general equation(Legendre & Legendre, 1998 p. 147,
% or LM eq 7.1.1 is:
% f(y_j)=1/(sqrt(2*pi)*sigma_j)*exp(-1/2*((y_j-mu_j)/sigma_j)^2)
mu_j=0;
sigma_j=1; % sigmaj is the standard deviation
```
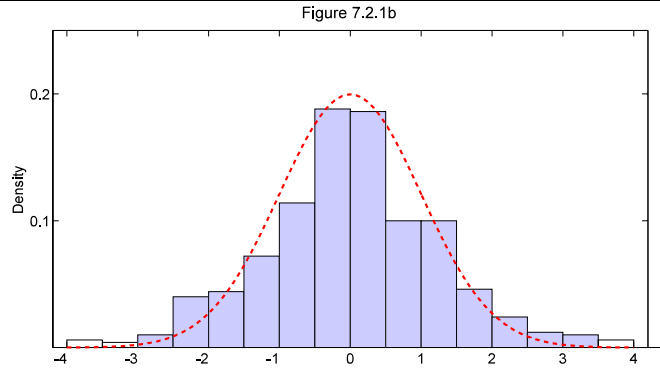
y_j=-4:0.01:4;
%fy_j=1/(sqrt(2*pi)*sigma_j)*exp(-1/2*((y_j-mu_j)./sigma_j).^2);
% or the identical function:
fy_j=normpdf(y_j,mu_j,sigma_j)*Binsize;
% fyj=1/(sqrt(2*pi)*sigmaj)*exp(-1/2*((y-muj)/sigmaj).^2);
% Plot using ax1 handle, saved above,to save this graph
% on top of the previous graph.
h1=line(y_j,fy_j,'Color','r','Parent',ax1,'Linewidth',3);
set(h1,'linestyle','--','color','r','linewidth',3)
xlabel("),ylabel('Density','FontSize',20)
title(tl,'FontSize',22)
figure(gcf);pause
hold off

7.3 DERIVING THE DISTRIBUTION OF $\dfrac{\bar{Y} - \mu}{\sigma/\sqrt{(n)}}$

Probability functions fall into two categories: those that model individual measurements (e.g., normal, binomial, Poisson, exponential, hypergeometric & uniform) and "There is a smaller set of probability distributions that model the behavior of *functions* based on a set of n random variables. These are called **sampling distributions**, and the functions they model are typically used for inference purposes." The normal belongs to both categories. Others sampling distributions include Student t distribution, chi square distribution and the F distribution. The t distribution will be derived as a special case of the F distribution, which is derived as a ratio of chi-square distributions.

Theorem 7.3.1 Let $U = \displaystyle\sum_{j=1}^{m} Z_j^2$, where $Z_1, Z_2, ..., Z_m$ are independent standard normal random variables. Then U has a gamma distribution with $r = \dfrac{m}{2}$ and $= \frac{1}{2}$. That is,

$$f_U(u) = \frac{1}{2^{\frac{m}{2}} \, \Gamma\left(\dfrac{m}{2}\right)} \, u^{\frac{m}{2}-1} \, e^{-\frac{u}{2}}, \; u \geq 0$$

The distribution of the sum of squares of independent standard normal random variables is sufficiently important that it gets its own name, the chi square distribution, despite the fact that it is just a special case of the gamma distribution.

Definition 7.3.1 The pdf of $U = \displaystyle\sum_{j=1}^{m} Z_j^2$ where $Z_1, Z_2, ..., Z_m$ are independent standard normal random variables, is called the **chi square distribution** with m degrees of freedom.

Theorem 7.3.2 Let Y1, Y2, ..., Yn be a random sample from a normal distribution with mean μ and variance ^{2}. Then

    a. S^2 and $\bar{Y}$ are independent

b.　$\dfrac{(n-1)\,S^2}{\sigma^2} = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(Y_i - \bar{Y})^2$　　has a chi square distribution with

Definition 7.3.2 Suppose that U and V are independent chi square random variables with n and m degrees of freedom respectively. A random variable of the form $\dfrac{V/m}{U/n}$ is said to have an **F distribution** with *m* and *n* degrees of freedom.

Theorem 7.3.3 pdf of the F distribution

      7.3.1　**F tables** The symbol $F_{p,m\,n}$ will be used to denote the 100pth percentile of the F distribution with m and n degrees of freedom.

Figure 32 shows the F distribution for 3 and 5 df.

```
% LMFig070301_4th.m
xmax=6.5;
r=0:0.01:xmax;
y=fpdf(r,3,5);
r95=finv(0.95,3,5);
plot(r,y,'--r','LineWidth',3)
axis([0 xmax 0 0.71])
ax1=gca;
xlabel('r','FontSize',20),
ylabel('Probability
Density','FontSize',20);
set(ax1,'xtick',[0:5 r95
6],'FontSize',18);
set(ax1,'ytick',0:0.2:0.7,'FontSize',14)
hold on;
xf=r95:0.01:xmax;
yf=fpdf(xf,3,5);
fill([r95 xf xmax]',[0 yf 0]',[.8 .8 1])
text(r95+0.05,0.05,'Area=0.05','FontSize',18);
title('Figure 7.3.1','FontSize',22)
figure(gcf);pause
hold off;
```
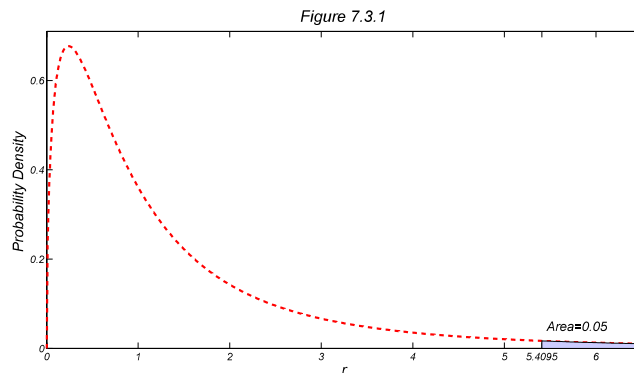

Figure 32. F distribution with 3 and 5 df, the 95% critical value of 5.4095 is also shown.

      7.3.2　**Using the *F* Distribution to Derive the pdf for t Ratios**

Definition 7.3.3 Let Z be a standard norml random variable and let U be a chi square random variable independent of Z with n degrees of freedom. The **Student t ratio** with n degrees of freedom is denoted T_n, where

$$T_n = \frac{Z}{\sqrt{\dfrac{U}{n}}}$$

Theorem 7.3.4 The pdf for a Student t random variable with n degrees of freedom is given by

$$f_{T_n}(t) = \frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n\,\pi}\,\Gamma\left(\dfrac{n}{2}\right)\left(1 + \dfrac{t^2}{n}\right)^{(n+1)/2}}, \quad -\infty < t < \infty$$

Theorem 7.3.5 Let Y_1, Y_2, ..., Y_n be a random sample from a normal distribution with mean μ and standard deviation . Then

$$T_{n-1} = \frac{\bar{Y} - \mu}{S/\sqrt{n}}$$

has a Student t distribution with *n-1* degrees of freedom.

$f_{Tn}(t)$ and f_Z (Z): How the two pdf's are related

Figure 33 shows the relation between the normal and t distributions Student's t curves are flatter.
% LMFig070302_4th.m
% Page 427 in Larsen & Marx, 2nd ed, page 479 in 4th edition
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% Pots of the z & t distribution.
% E. Gallagher, Eugene.Gallagher@umb.edu
% Revised: 2/5/04
% Plot the normal probability pdf:
% The normal probability equation is provided on p. 264 & 402 in L&M

Figure 33. Normal distribution and Student T distribution for 2 df (green) and 10 df (blue)

```
clf
muj=0;
sigmaj=1; % sigmaj is the standard deviation; 1 for a standardnormal
y=linspace(-4.2,4.2,100);
```

```
% fyj=1/(sqrt(2*pi)*sigmaj)*exp(-1/2*((y-muj)/sigmaj).^2);
fyj=normpdf(y,muj,sigmaj);
% Plot using ax1 handle, saved above,to save this graph
% on top of the previous graph.
h1=line(y,fyj,'Color','r','Linewidth',3);
set(h1,'linestyle','--','color','r','linewidth',3)
axis([-4 4 0 0.41]);
ax1=gca;
set(ax1,'Ytick',[0:0.2:0.4],'YColor','b')
set(ax1,'Xtick',[-4:4],'YColor','b')
title( 'Figure 7.3.2','Fontsize',22);
% Now plot the t distribution with 10 df;
df=10;
%ftn=LMTheorem070402(df);  % find the pdf for t with 10 df.
ftn=tpdf(y,10);
h2=line(y,ftn,'Color','b');
set(h2,'linestyle','--','color','b','linewidth',3);
df=2;
ftn=tpdf(y,2);
% ftn=LMTheorem070402(df);
h2=line(y,ftn,'Color','g');
set(h2,'linestyle','-.','color','g','linewidth',3);
% add labels
labs=['f_z(z)    ';
    'f_T_1_0(t)';
    'f_T_2(t)  '];
text(0.9,0.37,labs(1,:),'Color','r','Fontsize',18);
text(1.5,0.2,labs(2,:),'Color','b','Fontsize',18);
text(3,0.05,labs(3,:),'Color','g','Fontsize',18);
figure(gcf);pause
```

Questions p. 479 {none appropriate for EEOS601}

       7.4   **DRAWING INFERENCES ABOUT μ**

              7.4.1   *t* **Tables**

                        7.4.1.1 Use the symbol $t_{,n}$ to denote the 100(1-)th percentile of $f_{T_n}(t)$

Figure 35 Student's t distribution

Figure 34. Student T distribution for 3 df and 95th % cutoff at 2.534

% LMFig070402_4th.m
% Page 427 in Larsen & Marx, 2nd ed, page 479 in 4th edition
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% Pots of the z & t distribution.
% E. Gallagher, Eugene.Gallagher@umb.edu
% Revised: 2/5/04
% Plot the normal probability pdf:
% The normal probability equation is provided on p. 264 & 402 in L&M
clf
muj=0;
sigmaj=1; % sigmaj is the standard deviation; 1 for a standardnormal
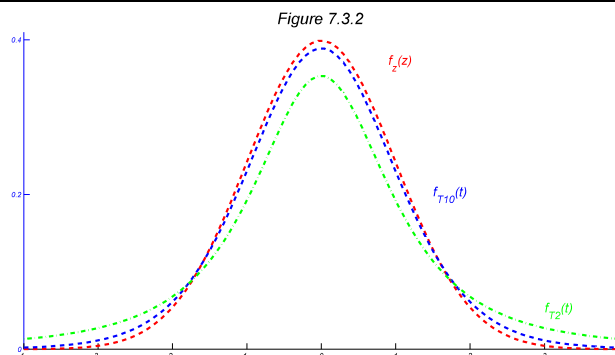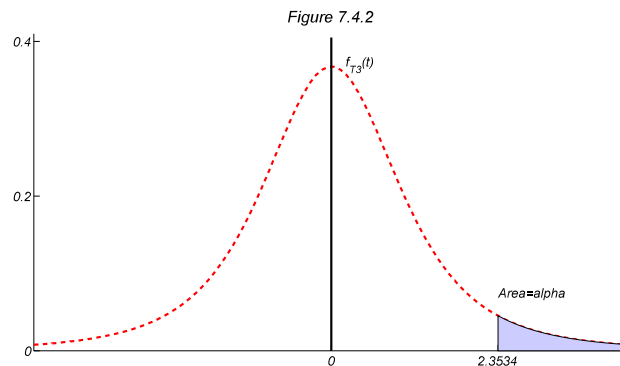y=linspace(-4.2,4.2,100);
fyj=tpdf(y,3);
r95=tinv(0.95,3);
% Plot using ax1 handle, saved above,to save this graph
% on top of the previous graph.
h1=line(y,fyj,'Color','r','Linewidth',3);
set(h1,'linestyle','--','color','r','linewidth',3)
axis([-4 4 0 0.41]);
ax1=gca;
set(ax1,'Xtick',[0 r95],'XColor','k','FontSize',18);
set(ax1,'Ytick',0:.2:.4,'FontSize',18);
title('Figure 7.4.2','Fontsize',22);
hold on
% add labels
labs=['f_T_3(t) '];
text(r95,0.075,'Area=alpha','Fontsize',18)
text(0.2,0.37,'f_T_3(t)','Fontsize',18)
xf=r95:0.01:xmax;
yf=tpdf(xf,3);
fill([r95 xf xmax]',[0 yf 0]',[.8 .8 1])
figure(gcf);pause
hold off;

7.4.2   Constructing a confidence interval for μ

Theorem 7.4.1 Let y_1, y_2, ..., y_n be a random sample of size n from a normal distribution with (unknown) mean μ. A 100(1-)% confidence interval for μ is the set of values

$$\left(\bar{y} - t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}}, \ \bar{y} + t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}} \right)$$

Case Study 9.4.1
See above

7.4.3   Fitting a normal distribution to data

Case Study 7.4.1
%LMcs070401_4th.m

```
% p 483-484 in
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% Written by Eugene.Gallagher@umb.edu 10/2010, revised 3/12/2011
X=[62 52 68 23 34 45 27 42 83 56 40];
fprintf('The mean reaction distance is %6.4f cm\n',mean(X));
alpha=0.05;
[H,P,CI,STATS]=ttest(X,0,alpha,'both');
fprintf('The 95%% confidence interval is (%6.4f cm %6.4f cm)\n',CI)
% or call normfit
[MUHAT,SIGMAHAT,MUCI,SIGMACI] = normfit(X,alpha)
fprintf('The 95%% confidence interval is (%6.4f cm %6.4f cm)\n',MUCI)
```

Example 7.4.1
```
%LMex070401_4th.m
X=[2.5 0.1 0.2 1.3
   3.2 0.1 0.1 1.4
0.5 0.2 0.4 11.2
0.4 7.4 1.8 2.1
0.3 8.6 0.3 10.1]; X=X(:);
[H,P,CI,STATS]=ttest(X);
fprintf('The sample arithmetic mean is %6.4f\n',mean(X))
fprintf('The 95%% confidence interval is (%6.4f %6.4f)\n',CI)
Binsize=1;
edges=0:Binsize:ceil(max(X));
[N1,BIN1] = histc(X,edges);
bar(edges,N1,'histc');
axis([0 12.5 0 11]);
set(get(gca,'Children'),'FaceColor',[.8 .8 1])
ylabel('Frequency','FontSize',20);
ax1=gca;set(ax1,'xtick',0:5:10,'FontSize',20);
set(ax1,'ytick',0:5:10,'FontSize',20)
title('Figure 7.4.3','FontSize',22)
figure(gcf);pause
% Matlab's boxplot and norm pdf are also very good at determining if
% samples have nonnormal distributions.
boxplot(X); pause
normplot(X);pause
% ----
% Optional. A log transform makes these data suitable for analysis with the
% t tools
lx=log(X);
boxplot(lx); pause
normplot(lx);pause
edges=-3:3;
[N1,BIN1] = histc(lx,edges);
bar(edges,N1,'histc');axis([-3.1 3.1 0 5.1]);
```

```
set(get(gca,'Children'),'FaceColor',[.8 .8 1])
ylabel('Frequency','FontSize',20);
ax1=gca;set(ax1,'xtick',-3:3,'FontSize',20);
set(ax1,'ytick',0:5:10,'FontSize',20);
title('Figure 7.4.3b','FontSize',22);figure(gcf)
[H,P,CI,STATS]=ttest(lx);
fprintf('\nThe geometric mean is %6.4f\n',exp(mean(lx)))
fprintf('The 95%% confidence interval is (%6.4f %6.4f)\n',exp(CI))
```

Questions p. 486-489

7.4.4   Testing H_o $\mu = \mu_o$ (The One-Sample t Test)

Theorem 7.4.2 Let $y_1, y_2, ..., y_n$ be a random sample of size n from a normal distribution where

is unknown. Let $t = \dfrac{\bar{y} - \mu_o}{s/\sqrt{n}}$

   a.   To test $H_o: \mu = \mu_o$ versus $H_1: \mu > \mu_o$ at the   level of significance, reject H_o
         if $t \geq t_{,n-1}$.
   b.   To test $H_o: \mu = \mu_o$ versus $H_1: \mu < \mu_o$ at the   level of significance, reject H_o
         if t $t_{,n-1}$.
   c.   To test $H_o: \mu = \mu_o$ versus $H_1: \mu \neq \mu_o$ at the   level of significance, reject H_o if t is
         either (1) $-t_{/2,n-1}$ or $\geq t_{1-/2,n-1}$.

```
Example 7.4.2
% LMex070402_4th.m
% Case 7.4.2 p. 490
% Example of one-sample t test in
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% Written by Eugene.Gallagher@umb.edu October 2010, revised 1/21/11
% http://alpha.es.umb.edu/faculty/edg/files/edgwebp.html
% Theorem 7.4.2: T=(Avg-mu_o)/(s/sqrt(n))
% The p value can for T can be determined
% using the wording indicating that a 1-tailed
% test is required: "Can it be concluded
% that children at this particular facility
% have a HIGHER risk of pica?
alpha=0.05;M=16
[D,t,df,pvalue,CI]=student1group(12,18.65, 5.049,M, alpha);
fprintf('The 1-tailed p value = %6.4g for t=%6.4f\n',pvalue/2,t);
fprintf(...
 'The blood lead difference of %4.2f mg/l had 95%% CI =[%4.2f %4.2f]\n',...
   D,CI(1),CI(2));
if pvalue/2<alpha
   fprintf(...
'Reject Ho: Evidence that children had lead levels higher than %4.2f\n',M);
fprintf(...
```

'The difference from "normal (=%2.0f mg/l)" was %3.1f mg/l +/- %3.1f mg/l\n',...
M,D,D-CI(1));
end

```
function [D,t,df,pvalue,CI]=student1group(Xn,Xmean, Xstd, M, alpha)
% Student't 1-sample t test with grouped data
% [D,t,df,pvalue,CI]=student1group(Xn,Xmean,Xstd,alpha)
% Input: Xn = Size of group = number of cases
%        Xmean = mean for groups
%        Xstd, Ystd= standard deviations for 2 groups
%        M=expected mean, 0 if not provided
%        alpha level for CI, optional, 95% if not specified
% Output: D=Xmean-M or Xmean-0;
%         t=Student's t statistic
%         df=degrees of freedom for t statistic
%         p value, 2-sided, for t Statisitc, with df degrees of freedom
%         CI for 1-alpha CI, [L U];95% CI if alpha not specified.
% Based on Theorem 7.4.2, page 490 in Larsen & Marx (2006)
% Introduction to Mathematical Statistics, 4th edition.
% uses Statistics toolbox
% Written by Eugene.Gallagher@umb.edu, Revised 11/14/10
% see also stud1sample, binom2sample, student2group, stud2sample
if nargin<4
   M=0;
   alpha=0.05;
elseif nargin <5
   alpha=0.05;
end
D=Xmean-M;
% Calculate sample standard deviation
t=D/(Xstd/sqrt(Xn));
% Use stixbox's pt.m for significance of t.
df=Xn-1;
if t>=0
   pvalue=2*(1-tcdf(t,df));
else
   pvalue=2*tcdf(t,df);
end
HalfCI=tinv(1-alpha/2,df)*Xstd*sqrt(1/Xn);
CI=[D-HalfCI D+HalfCI];
```

```
Example 7.4.3
% LMex070403_4th.m
% Application of the one sample t test using
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% Written by Eugene.Gallagher@umb.ed 11/14/2010, revised: 11/17/2010
% Percentage loan approval from different banks
```

```
X=[59 65 69 53 60 53 58 64 46 67 51 59]';
fprintf('The mean of X is %6.3f %%\n',mean(X));
meanHo=62;
fprintf('The true mean being tested is %6.3f%%\n',meanHo);
alpha=0.05;
% Matlab's 1-sample t test is:
[h,p,ci,stats] = ttest(X,meanHo)
% The book reports the critical values for the 2-tailed t
% statistic based on 11 df.
df=stats.df;
fprintf(...
'The 2 tailed critical values for D=0, %d df and alpha=0.05, are %4.2f and %4.2f\n',...
df,tinv(alpha/2,df),tinv(1-alpha/2,df));
fprintf('The observed mean was %6.3f with 95%% CI: [%6.3f %6.3f]\n',mean(X),ci);
fprintf(...
'The 2-tailed p of observing a difference of %6.3f by chance is %5.3f.\n',...
mean(X)-meanHo,p)
% Test the 3 banks separately, Table 7.5.4, page 492:
BANK=['AU';'TU';'TU';'FT';'FT';'AU';'FT';'FT';'AU';'TU';'AU';'FT'];
% Find how many different unique banks
UB=unique(BANK,'rows');
[r,c]=size(UB);
alpha=0.05;
for i=1:r
   j=strmatch(UB(i,:),BANK); % finds indices for all rows with the 1st unique bank name
fprintf('\nThe Bank being analyzed is %3s\n',UB(i,:));
x=X(j);
fprintf('The mean Percentage approved is %5.2f%% vs. %5.2f%%
expected\n',mean(x),meanHo);
% [D,t,df,pvalue,CI]=stud1sample(x,meanHo,alpha);
% Matlab's 1-sample t test is:
[h,pvalue,ci,stats] = ttest(x,meanHo,alpha);
D=mean(x)-meanHo;
% The book reports the critical values for the 2-tailed t
% statistic based on 11 df.
   fprintf(...
     'For Bank %s, the difference is %5.2f%%, with 2-tailed p value=%5.3f\n',...
     UB(i,:),D,pvalue);
   if pvalue<alpha
     fprintf('For bank %s, reject the null hypothesis at alpha=%5.3f\n',...
        UB(i,:),alpha);
   else
     fprintf('For bank %s, fail to reject the null hypothesis at alpha=%5.3f\n',UB(i,:),alpha);
   end
   df=stats.df;
```

```
    fprintf(...
    'The 2 tailed critical values for the t statistic=%4.2f and alpha=%5.3f, are %4.2f and
%4.2f\n',...
    t,alpha,tinv(alpha/2,df),tinv(1-alpha/2,df));
end
% HOWEVER, if the 3 banks were not idenitified in advance, then a Bonferroni correction
% to alpha should be used.
fprintf('\nAnalyses using Bonferroni adjustment based on number of tests (3)\n')
alpha=0.05;
alpha=0.05;ntests=r; % r is 3, the 3 types of banks
alpha=1-exp(log(1-alpha)/ntests);  % I'll overwrite alpha to avoid having to recode the if loop
fprintf('The Bonferroni adjusted alpha, based on %d tests is %5.3f.\n',ntests,alpha);
for i=1:r
    j=strmatch(UB(i,:),BANK); % finds indices for all rows with the 1st unique bank name
fprintf('\nThe Bank being analyzed is %3s\n',UB(i,:));
x=X(j);
fprintf('The mean Percentage approved is %5.2f%% vs. %5.2f%%
expected\n',mean(x),meanHo);
% [D,t,df,pvalue,CI]=stud1sample(x,meanHo,alpha);
% Matlab's 1-sample t test is:
[h,pvalue,ci,stats] = ttest(x,meanHo,alpha);
D=mean(x)-meanHo;
% The book reports the critical values for the 2-tailed t
% statistic based on 11 df.
    fprintf(...
      'For Bank %s, the difference is %5.2f%%, with 2-tailed p value=%5.3f\n',...
      UB(i,:),D,pvalue);
    if pvalue<alpha
      fprintf(...
'For bank %s, reject the null hypothesis at Bonferroni-adjusted alpha=%5.3f\n',...
        UB(i,:),alpha);
    else
      fprintf(...
'For bank %s, fail to reject the null hypothesis at Bonferroni-adjusted
alpha=%5.3f\n',UB(i,:),alpha);
    end
    df=stats.df;
    fprintf(...
    'The 2 tailed critical values for the t statistic=%4.2f and alpha=%5.3f, are %4.2f and
%4.2f\n',...
    t,alpha,tinv(alpha/2,df),tinv(1-alpha/2,df));
end
```

Questions p. 492-493

7.4.5   Testing H_o $\mu = \mu_o$ When the Normality Assumption is not Met

Is the validity of the *t* test compromised if the normality assumption is not met?

Figure 7.4.6
```
% LMFig070406_4th.m
% Pages 495-498 in
% Larsen & Marx (2006) Introduction to Mathematical Statistics,
4th edition
% Written by Eugene.Gallagher@umb.edu 3/12/11, revised
3/13/11
% Use 7.2.1b as a model
K=menu('Pick a Figure','7.46a Uniform','7.46b exponential, n6',...
   '7.46c Poisson','7.46d Exponential n15');
n=6;N=100;
Binsize=1;
tx='t ratio (n=6)';
ymax=0.42;
if K==1
   % Plot the distribution figure
   y=[0 1];fy=[1 1];
   plot(y,fy,'--k','LineWidth',3)
   axis([0 1.05 0 1.1]);
   set(gca,'ytick',[0 1],'xtick',[0 1],'FontSize',20)
   xlabel('y','FontSize',20);ylabel('Probability
Density','FontSize',20);
   text(0.5,0.92,'f_y(y) = 1','FontSize',18);
   title('Figure 7.4.6: Uniform Distribution');figure(gcf);pause
   % settings for the Monte Carlo simulated data:
   tl='Figure 7.4.6a: Uniform';
   R=rand(n,N);
   xmin=-6.2;xmax=6.2;
   mu=0.5;
elseif K==2
   % Plot the distribution figure
   YMAX=6.2;y=[0:0.01:YMAX];fy=exppdf(y,1);
   plot(y,fy,'--k','LineWidth',3)
   axis([0 YMAX 0 1.1]);
   set(gca,'ytick',[0:.5:1],'xtick',[0:2:6],'FontSize',20)
   xlabel('y','FontSize',20);ylabel('Probability
Density','FontSize',20);
   text(1.5,0.35,'f_y(y) = e^{-y}','FontSize',18);
   title('Figure 7.4.6: Exponential Distribution');figure(gcf);pause
   % Settings for the Monte Carlo Simulated data
   tl='Figure 7.4.6b: Exponential';
   R=exprnd(1,n,N);
   xmin=-7.2;xmax=4.2;
   mu=1;
elseif K==3
```
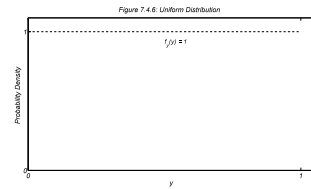
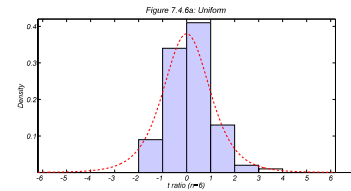

Figure 35. Uniform distribution
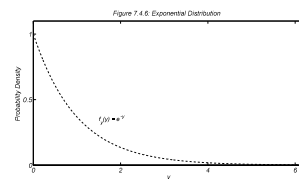

Figure 36. Mean (n=6) from uniform distribution as t ratios.

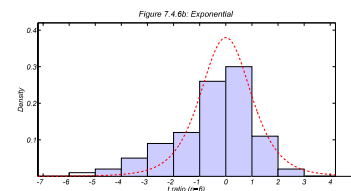**Figure 37**. Exponential distribution ($\mu = 1$)

Figure 38. Mean (n=6) from exponential distribution ($\mu=1$) as t ratios.

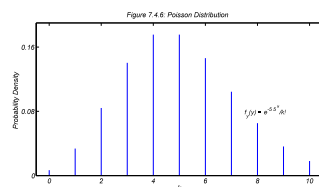**Figure 39**. Poisson distribution, =5

```
   % Plot the distribution figure
   % use Figure 5.1.1 as a model
   lambda=5;k5=0:10;pxk5=poisspdf(k5,lambda);
   h2=stem(k5,pxk5,'Marker','none');
   xlabel('k','FontSize',20)
   axis([-.5 10.5 0 0.19])
   set(h2,'LineWidth',3)
   figure(gcf);pause
   set(gca,'ytick',[0:.08:.16],'xtick',[0:2:10],'FontSize',20)
   xlabel('k','FontSize',20);ylabel('Probability
Density','FontSize',20);
   text(7.5,0.08,'f_y(y) = {e^{-5 5^k}}/k!','FontSize',18);
   title('Figure 7.4.6: Poisson Distribution');figure(gcf);pause
   % Settings for the Monte Carlo Simulated data
   tl='Figure 7.4.6c: Poisson';
   lambda=5;
   R=poissrnd(lambda,n,N);
   xmin=-5.2;xmax=4.2;
   mu=lambda;
elseif K==4
   % Plot the distribution figure same as before
   YMAX=6.2;y=[0:0.01:YMAX];fy=exppdf(y,1);
   plot(y,fy,'--k','LineWidth',3)
   axis([0 YMAX 0 1.1]);
   set(gca,'ytick',[0:.5:1],'xtick',[0:2:6],'FontSize',20)
   xlabel('y','FontSize',20);ylabel('Probability Density','FontSize',20);
   text(1.5,0.35,'f_y(y) = e^{-y}','FontSize',18);
   title('Figure 7.4.6: Exponential Distribution');figure(gcf);pause
   % Settings for the Monte Carlo Simulated data
   n=15;
   tx='t ratio (n=15)';
   tl='Figure 7.4.6d';
   R=exprnd(1,n,N);
   xmin=-4.2;xmax=4.2;
   mu=1;
end
edges=ceil(xmin):Binsize:floor(xmax);
tdf=n-1;
V=[xmin xmax 0 ymax];
   mr=mean(R)';
   s=std(R)';
   mu=repmat(mu,N,1);
   Z1=(mr-mu)./(s./sqrt(n));
[N1,BIN1] = histc(Z1,edges);
bar(edges,N1/N,'histc');
```
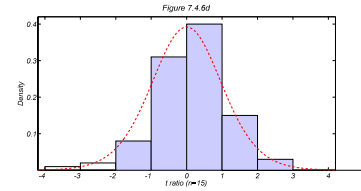

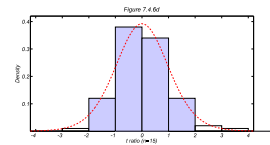**Figure 40**. Mean (n=6) from Poisson distribution, =5, as t ratios.

Figure 41. Mean (n=15) from exponential distribution, μ=1, as t ratios.

```
axis(V)
set(get(gca,'Children'),'FaceColor',[.8 .8 1])
ylabel('Density','FontSize',20);
ax1=gca;set(ax1,'xtick',[ceil(xmin):floor(xmax)],'FontSize',20);
set(ax1,'ytick',[0.1:0.1:0.6],'FontSize',20)
hold on
figure(gcf);pause
y_j=xmin:0.01:xmax;
fy_j=tpdf(y_j,tdf)*Binsize;
h1=line(y_j,fy_j);
set(h1,'linestyle','--','color','r','linewidth',3)
xlabel(tx,'FontSize',20),ylabel('Density','FontSize',20)
title(tl,'FontSize',22)
figure(gcf);pause
hold off
```

Questions p. 498-499 None really appropriate for Matlab programming

7.5 DRAWING INFERENCES ABOUT σ^2

Shift attention to scale parameter σ^2

Figure 42 shows the chi square distribution for 5 df.

7.5.1 Chi Square Tables

7.5.2 Constructing Confidence Intervals for σ^2

Theorem 7.5.1 Confidence interval for σ^2

Figure 42. Chi square distribution for 5 df

```
%LMFig070501_4th.m
% Based on LMFig060205_4th. Page 500 in
% Larsen & Marx (2006)
Introduction to Mathematical Statistics, 4th edition
% Written by Eugene.Gallagher@umb.edu 3/13/11
df=5;chi05=chi2inv(0.05,df);
chi99=chi2inv(0.99,df);
zmax=17.1;
z=0:.01:zmax;
fzz=chi2pdf(z,df);
plot(z,fzz,'linestyle','--','color','r','linewidth',3)
ylabel('Probability Density','FontSize',20)
xlabel('z','FontSize',20)
axis([0 zmax 0 0.17])
set(gca,'Ytick',[0:.05:0.15],'FontSize',18)
set(gca,'Xtick',[chi05 4:4:12 chi99],'FontSize',18)
ax=axis;
```
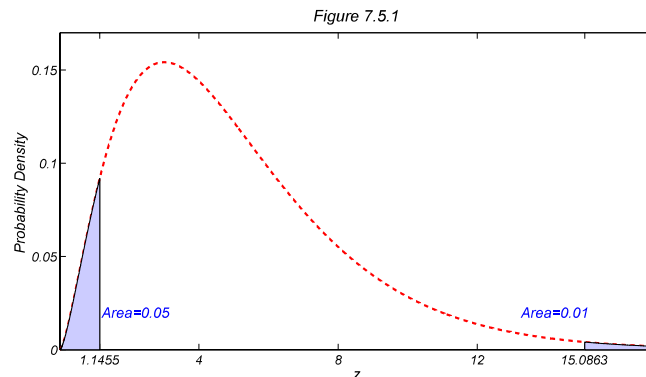
ax1=gca; % save the handle of the graph
title('Figure 7.5.1','FontSize',22)
hold on
fz=chi2pdf([chi05 chi99],df);
plot([chi05 chi05;chi99 chi99]',[0 fz(1);0 fz(2)]','-k','linewidth',1)
% Fill in the upper tail with fill
y2=chi99:.001:ax(2);
fy2=chi2pdf(y2,df);
fymax=chi2pdf(ax(2),df);
fill([chi99 y2 ax(2) ax(2)],[0 fy2 fymax 0],[.8 .8 1])
% Fill in the lower tail with fill
y3=ax(1):.001:chi05;
fy3=chi2pdf(y3,df);
% fymin=chi2pdf(ax(1),df);
fymin=0;
fill([ax(1) ax(1) y3 chi05],[0 fymin fy3 0],[.8 .8 1])
t=sprintf('Area=0.05');
t1='Area=0.01';
text(chi05+.05,.02,t,'Color','b','FontSize',20);
text(13.25,.02,t1,'Color','b','FontSize',20);
figure(gcf)
hold off

Case Study 7.5.1
% LMcs070501_4th.m
% Page 502-503 in
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% Written by Eugene.Gallagher@umb.edu, 3/13/11, revised:
% Will solve the easy and easiest way.
X=[249;254;243;268;253;269;287;241;273;306;303;280;260;256;278;344;304;283;310];
[MUHAT,SIGMAHAT,MUCI,SIGMACI] = normfit(X);df=length(X)-1;
fprintf('The 95%% CI for sigma is %4.2f and %4.2f million years\n',SIGMACI)
alpha=0.05;chi2low=chi2inv(alpha/2,df);
chi2up=chi2inv(1-alpha/2,df);
s2=var(X);s=std(X);
fprintf('The variance is %5.1f and sd is %5.1f million years.\n',s2,s);
Low95s=sqrt(df*s2/chi2up);Up95s=sqrt(df*s2/chi2low);
fprintf('The 95%% CI for sigma is %4.2f and %4.2f million years\n',Low95s,Up95s)

Example 7.5.1
Solve with fsolve
% LMex070501_4th.m
% Page 503 in
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% Written by Eugene.Gallagher@umb.edu 3/13/11
% What is the minimum n such that the width of the 95% confidence interval
% for sigma^2 is no greater than sigma^2;

Width=(n-1)*s2*(1/chi2inv(alpha/2,n-1)-1/chi2inv(1-alpha/2,n-1));
% So
alpha=0.05;
x = fsolve(@(n) (n-1)*(1/chi2inv(alpha/2,n-1)-1/chi2inv(1-alpha/2,n-1))-1,30)
% x=38.562, so
fprintf(...
'The min n such that 95%% CI width for sigma^2 < sigma^2 is %2.0f\n',...
ceil(x));

7.5.3 H_o: $\sigma^2 = \sigma^2_o$

Theorem 7.5.2

Case Study 7.5.2
% LMcs070502_4th.m
% Page 505-506 in
% Larsen & Marx (2006)
Introduction to Mathematical
Statistics, 4th edition
% Written by
Eugene.Gallagher@umb.edu
X=[6.38;6.63;6.88;6.75;6.13;6.50;6.63;6.38;6.50];n=length(X);df=n-1;
sd=.1;alpha=0.05;
fprintf(...
'The variance for 30-y fixed
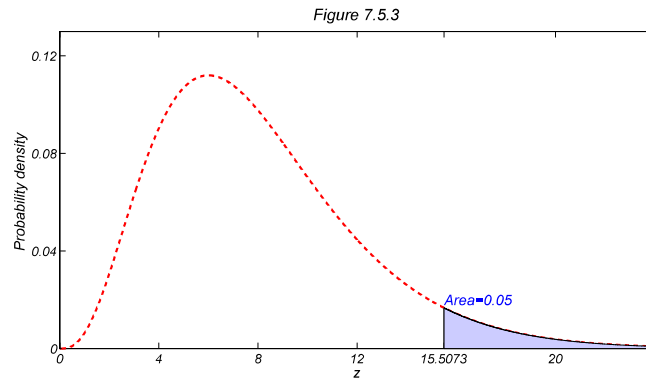
Figure 43. Chi square distribution for 8 df

mortgages = %5.3f%% and 1-year adjustables = %5.3f%%\n',...
sd^2,var(X))
[H,P,CI,STATS] = vartest(X,sd^2,alpha,'both')
if H==1
   fprintf('Reject the null hypothesis\n')
else
   fprintf('Fail to reject the null hypothesis\n')
end
% Or using theorem 7.5.2
chi2=df*var(X)/sd^2
pvalue1tail=(1-chi2cdf(chi2,df));
pvalue2tail=pvalue1tail*2;
fprintf('The two-tailed p-value for chi2 = %6.2f & %1.0f df is %5.2e\n',chi2,df,pvalue2tail)
% Plot the chi square distribution using Figure 7.5.1 as a model
chi95=chi2inv(0.95,df);
zmax=24;
z=0:.01:zmax;
fzz=chi2pdf(z,df);
plot(z,fzz,'linestyle','--','color','r','linewidth',3)
ylabel('Probability density','FontSize',20)
xlabel('z','FontSize',20)
axis([0 zmax 0 0.13])

```
set(gca,'Ytick',[0:.04:0.12],'FontSize',18)
set(gca,'Xtick',[0:4:12 chi95 20],'FontSize',18)
ax=axis;
ax1=gca; % save the handle of the graph
title('Figure 7.5.3','FontSize',22)
hold on
fz=chi2pdf(chi95,df);
plot([chi95 chi95]',[0 fz]','-k','linewidth',1)
% Fill in the upper tail with fill
y2=chi95:.001:ax(2);
fy2=chi2pdf(y2,df);
fymax=chi2pdf(ax(2),df);
fill([chi95 y2 ax(2) ax(2)],[0 fy2 fymax 0],[.8 .8 1])
t='Area=0.05';
text(chi95,.02,t,'Color','b','FontSize',20);
figure(gcf)
hold off
```

       7.6   Taking a second look at statistics ("Bad" Estimators)

14   **Nonparametric statistics**

       14.1   Introduction
       14.2   The Sign Test

Theorem 14.2.1

Case Study 14.2.1

```
% LMcs140201_4th.m
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% page 804. A case study solved by the sign test
% Written by Eugene.Gallagher@umb.edu 11/16/10 Revised 11/16/10, 3/1/11
%
D=[7.02 7.35 7.32 7.33 7.15 7.26 7.25 7.35 7.38 7.20 7.31 7.24 7.34 ...
   7.32 7.34 7.14 7.20 7.41 7.77 7.12 7.45 7.28 7.34 7.22 7.32 7.4 ...
   6.99 7.1 7.3 7.21 7.33 7.28 7.35 7.24 7.36 7.09 7.32 6.95 7.35 ...
   7.36 6.6 7.29 7.31];
[p,h,stats] = signtest(D,7.39,0.05,'method','exact');
fprintf('\nThe sign test exact p=%6.4g\n',p);
[p,h,stats] = signtest(D,7.39,'method','approximate');
fprintf('The sign test approximate p=%6.4g;z=%6.4f\n',p,stats.zval);
[H,P,CI,STATS] = ttest(D,7.39);
fprintf('The one-sample t test 2-tailed p=%6.4g\n',P);
fprintf('The mean pH = %4.2f with 95%% CI: [%4.2f %4.2f]\n',mean(D),...
   CI(1),CI(2));
[P,H,STATS] = signrank(D,7.39,'alpha',0.05,'method','exact');
fprintf('The sign rank test exact p=%6.4g\n',P);
```

```
[P,H,STATS] = signrank(D,7.39,'alpha',0.05,'method','approximate');
fprintf('The sign rank test approximate p=%6.4g\n',P);
% Plot histogram and check for symmetry
binsize=.1;  % Needed in order to properly scale the normal pdf
edges=6.4:binsize:8;
% hist(D);
[N,BIN] = histc(D,edges);
bar(edges,N,'histc')
axis([6.35 8.05 0 21])
set(get(gca,'Children'),'FaceColor',[.8 .8 1]);
xlabel('pH','FontSize',20);
ylabel('Number of Cases','FontSize',20);
ax1=gca;
set(ax1,'Ytick',[0:5:25],'Xtick',[6.4:0.2:8])
figure(gcf);pause
% Superimpose the normal probability pdf on a histogram of differences.
% The normal probability equation is provided on p. 293
% This is for mean 0, and unit standard
% deviation. The more general equation (Legendre & Legendre, 1998 p. 147) is:
% f(y_j)=1/(sqrt(2*pi)*sigma_j)*exp(-1/2*((y_j-mu_j)/sigma_j)^2);
n=length(D);
mu_j=mean(D);
sigma_j=std(D); % sigmaj is the standard deviation; = 1 after Z transform
y_j=6.35:0.01:8.05;
fy_j=1/(sqrt(2*pi)*sigma_j)*exp(-1/2*((y_j-mu_j)./sigma_j).^2);
fy_j=n*binsize*fy_j;
cutoff=7.39;
fy_cutoff=n*binsize/(sqrt(2*pi)*sigma_j)*exp(-1/2*((cutoff-mu_j)./sigma_j).^2);
% will properly scale the height of the pdf
% fyj=1/(sqrt(2*pi)*sigmaj)*exp(-1/2*((y-muj)/sigmaj).^2);
% Plot using ax1 handle, saved above,to save this graph
% on top of the previous graph.
v=axis;
h1=line(y_j,fy_j ,'Color','r','Parent',ax1);
set(h1,'linestyle','--','color','r','linewidth',2)
h1=line([cutoff cutoff]',[0 fy_cutoff]','Color','b','Parent',ax1);
set(h1,'linestyle','-.','color','b','linewidth',3)
h2=line([cutoff cutoff]',[fy_cutoff v(4)]','Color','b','Parent',ax1);
set(h2,'linestyle','-.','color','b','linewidth',3)
s=sprintf('Case Study 14.2.1, %2.0f samples untransformed',n);
title(s,'FontSize',22)
figure(gcf);pause
```

14.2.1 A Small-Sample Sign Test, Use the exact binomial

Case Study 14.2.2
% LMcs140202_4th.m

```
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% page 806. A case study solved by the sign test
% Written by Eugene.Gallagher@umb.edu 11/16/10 Revised 11/16/10, 3/1/11
%
D=[4.8 4.0 3.8 4.3 3.9 4.6 3.1 3.7];
expected=3.55;
[H,P,CI,STATS] = ttest(D,expected,0.05,'both');
fprintf('\nThe one-sample t test 2-tailed p=%6.4g\n',P);
fprintf(...
'The mean caffeine = %4.2f (g/100g residue) with 95%% CI: [%4.2f %4.2f]\n',...
   mean(D), CI(1),CI(2));
[p,h,stats] = signtest(D,expected,0.05,'method','exact');
fprintf('The sign test statistic is %4.1f with exact 2-tailed p=%6.4g\n',stats.sign,p);
[p,h,stats] = signtest(D,expected,'method','approximate');
fprintf('The sign test z=%5.3f with approximate 2-tailed p=%6.4g\n',stats.zval,p);
[P,H,STATS] = signrank(D,expected,'alpha',0.05,'method','exact');
fprintf(...
'The sign rank test statistic is %4.1f with exact 2-tailed p=%6.4g\n',...
 STATS.signedrank,P);
[P,H,STATS] = signrank(D,expected,'alpha',0.05,'method','approximate');
fprintf('The sign rank test approximate 2-tailed p=%6.4g for z=%6.4f\n',P,STATS.zval);


% Plot histogram and check for symmetry
binsize=.2;  % Needed in order to properly scale the normal pdf
edges=3:binsize:5;
% hist(D);
[N,BIN] = histc(D,edges);
bar(edges,N,'histc')
axis([2.9 5.1 0 2.1])
set(get(gca,'Children'),'FaceColor',[.8 .8 1]);
xlabel('Caffeine Residue (g/100 g dry weight)','FontSize',20);
ylabel('Number of Cases','FontSize',20);
ax1=gca;
set(ax1,'Ytick',[0:2],'Xtick',[3:0.2:5])
figure(gcf);pause
% Superimpose the normal probability pdf on a histogram of differences.
% The normal probability equation is provided on p. 293
% This is for mean 0, and unit standard
% deviation. The more general equation (Legendre & Legendre, 1998 p. 147) is:
% f(y_j)=1/(sqrt(2*pi)*sigma_j)*exp(-1/2*((y_j-mu_j/sigma_j)^2);
n=length(D);
mu_j=mean(D);
sigma_j=std(D); % sigmaj is the standard deviation; = 1 after Z transform
y_j=2.9:0.01:5.1;
fy_j=1/(sqrt(2*pi)*sigma_j)*exp(-1/2*((y_j-mu_j)./sigma_j).^2);
```

```
fy_j=n*binsize*fy_j;
cutoff=expected;
fy_cutoff=n*binsize/(sqrt(2*pi)*sigma_j)*exp(-1/2*((cutoff-mu_j)./sigma_j).^2);
% will properly scale the height of the pdf
% fyj=1/(sqrt(2*pi)*sigmaj)*exp(-1/2*((y-muj)/sigmaj).^2);
% Plot using ax1 handle, saved above,to save this graph
% on top of the previous graph.
v=axis;
h1=line(y_j,fy_j ,'Color','r','Parent',ax1);
set(h1,'linestyle','--','color','r','linewidth',2)
h1=line([cutoff cutoff]',[0 fy_cutoff]','Color','b','Parent',ax1);
set(h1,'linestyle','-.','color','b','linewidth',3)
h2=line([cutoff cutoff]',[fy_cutoff v(4)]','Color','b','Parent',ax1);
set(h2,'linestyle','-.','color','b','linewidth',3)
s=sprintf('Case Study 14.2.2, %2.0f samples untransformed',n);
title(s,'FontSize',22)
figure(gcf);pause
```

14.2.2 Using the Sign Test for Paired Data (p. 807)

```
Case Study 14.2.3
% LMcs140203_4th.m
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% page 807. A case study solved by the sign test
% Written by Eugene.Gallagher@umb.edu 11/16/10 Revised 11/16/10, 3/1/11
%
D=[15 13;12 8;12 12.5;14 12;13 12;13 12.5;13 12.5;12 14;12.5 12;12 11;
   12.5 10];
[p,h,stats] = signtest(D(:,2),D(:,1),0.05,'method','exact');
fprintf('\nThe sign test statistic is %4.1f with exact 1-tailed p=%6.4g\n',stats.sign,p/2);
[p,h,stats] = signtest(D(:,2),D(:,1),'method','approximate');
fprintf('The sign test z=%5.3f with approximate 1-tailed p=%6.4g\n',stats.zval,p/2);
% It is a 1-tailed p test to the left since the expectation is that mean
% circulation time is reduced by 4 months of cyclandelate.
[H,P,CI,STATS] = ttest(D(:,2),D(:,1),0.05,'left');
fprintf('The t statistic was %5.3f with %2.0f df.\n',STATS.tstat,STATS.df)
fprintf('The paired t test 1-tailed p=%6.4g\n',P);
fprintf(...
'The mean circulation time = %4.2f secs with 95%% CI: [%4.2f %4.2f]\n',...
   mean(D(:,2)-D(:,1)), CI(1),CI(2));
[P,H,STATS] = signrank(D(:,2),D(:,1),'alpha',0.05,'method','exact');
fprintf(...
'The sign rank test statistic is %4.1f with exact 1-tailed p=%6.4g\n',...
 STATS.signedrank,P/2);
[P,H,STATS] = signrank(D(:,2),D(:,1),'alpha',0.05,'method','approximate');
```

fprintf('The sign rank test approximate 1-tailed p=%6.4g for z=%6.4f\n',P/2,STATS.zval);

% Plot histogram and check for symmetry
binsize=.5; % Needed in order to properly scale the normal pdf
edges=-4:binsize:2.5;
% hist(D);
[N,BIN] = histc(D(:,2)-D(:,1),edges);
bar(edges,N,'histc')
axis([-4.1 2.6 0 3.3])
set(get(gca,'Children'),'FaceColor',[.8 .8 1]);
xlabel('"After-Before" Mean Circulation Time (secs)','FontSize',20);
ylabel('Number of Cases','FontSize',20);
ax1=gca;
set(ax1,'Ytick',[0:3],'Xtick',[-4:0.5:2.5])
figure(gcf);pause
% Superimpose the normal probability pdf on a histogram of differences.
% The normal probability equation is provided on p. 293
% This is for mean 0, and unit standard
% deviation. The more general equation (Legendre & Legendre, 1998 p. 147) is:
% f(y_j)=1/(sqrt(2*pi)*sigma_j)*exp(-1/2*((y_j-mu_j/sigma_j)^2);
[n,c]=size(D);
mu_j=mean(D(:,2)-D(:,1));
sigma_j=std(D(:,2)-D(:,1)); % sigmaj is the standard deviation; = 1 after Z transform
y_j=-4.1:0.01:2.6;
fy_j=1/(sqrt(2*pi)*sigma_j)*exp(-1/2*((y_j-mu_j)./sigma_j).^2);
fy_j=n*binsize*fy_j;
cutoff=0;
fy_cutoff=n*binsize/(sqrt(2*pi)*sigma_j)*exp(-1/2*((cutoff-mu_j)./sigma_j).^2);
% will properly scale the height of the pdf
% fyj=1/(sqrt(2*pi)*sigmaj)*exp(-1/2*((y-muj)/sigmaj).^2);
% Plot using ax1 handle, saved above,to save this graph
% on top of the previous graph.
v=axis;
h1=line(y_j,fy_j ,'Color','r','Parent',ax1);
set(h1,'linestyle','--','color','r','linewidth',2)
h1=line([cutoff cutoff]',[0 fy_cutoff]','Color','b','Parent',ax1);
set(h1,'linestyle','-.','color','b','linewidth',3)
h2=line([cutoff cutoff]',[fy_cutoff v(4)]','Color','b','Parent',ax1);
set(h2,'linestyle','-.','color','b','linewidth',3)
s=sprintf('Case Study 14.2.2, %2.0f samples untransformed',n);
title(s,'FontSize',22)
figure(gcf);pause

Questions p 809-810

      14.3   **WILCOXON TESTS**
          14.3.1 **Testing** H_o: $\mu = \mu_o$

Theorem 14.3.1

> 14.3.2 Calculating $p_w(w)$
>
> 14.3.3 Tables of the cdf, $F_W(w)$

Case Study 14.3.1 Swell sharks

```
% LMcs140301_4th.m
% Case Study 14.3.1 from
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% page 815. A case study using Wilcoxon signed rank test
% Written by Eugene.Gallagher@umb.edu 11/16/10 Revised 11/16/10, 3/1/11
%
D=[13.32 13.06 14.02 11.86 13.58 13.77 13.51 14.42 14.44 15.43];
expected=14.6;
[P,H,STATS] = signrank(D,expected,'alpha',0.05,'method','exact');
fprintf(...
\nThe signed rank test statistic is %4.1f with exact 2-tailed p=%6.4g\n',...
 STATS.signedrank,P);
[P,H,STATS] = signrank(D,expected,'alpha',0.05,'method','approximate');
fprintf('The signed rank test approximate 2-tailed p=%6.4g for z=%6.4f\n',P,STATS.zval);
[H,P,CI,STATS] = ttest(D,expected,0.05,'both');
fprintf('The one-sample t test 2-tailed p=%6.4g\n',P);
fprintf(...
'The mean TL/HDl = %4.2f with 95%% CI: [%4.2f %4.2f]\n',...
   mean(D), CI(1),CI(2));
[p,h,stats] = signtest(D,expected,0.05,'method','exact');
fprintf('The sign test statistic is %4.1f with exact 2-tailed p=%6.4g\n',stats.sign,p);
[p,h,stats] = signtest(D,expected,'method','approximate');
fprintf('The sign test z=%5.3f with approximate 2-tailed p=%6.4g\n',stats.zval,p);

% Plot histogram and check for symmetry
binsize=0.25;  % Needed in order to properly scale the normal pdf
edges=11:binsize:16;
% hist(D);
[N,BIN] = histc(D,edges);
bar(edges,N,'histc')
axis([10.9 16.1 0 max(N)+0.2])
set(get(gca,'Children'),'FaceColor',[.8 .8 1]);
xlabel('TL/HDl','FontSize',20);
ylabel('Number of Cases','FontSize',20);
ax1=gca;
set(ax1,'Ytick',[0:max(N)],'Xtick',edges)
figure(gcf);pause
% Superimpose the normal probability pdf on a histogram of differences.
% The normal probability equation is provided on p. 293
% This is for mean 0, and unit standard
```

% deviation. The more general equation (Legendre & Legendre, 1998 p. 147) is:
% f(y_j)=1/(sqrt(2*pi)*sigma_j)*exp(-1/2*((y_j-mu_j/sigma_j)^2);
n=length(D);
mu_j=mean(D);
sigma_j=std(D); % sigmaj is the standard deviation; = 1 after Z transform
y_j=10.9:0.01:16.1;
fy_j=1/(sqrt(2*pi)*sigma_j)*exp(-1/2*((y_j-mu_j)./sigma_j).^2);
fy_j=n*binsize*fy_j;
cutoff=expected;
fy_cutoff=n*binsize/(sqrt(2*pi)*sigma_j)*exp(-1/2*((cutoff-mu_j)./sigma_j).^2);
% will properly scale the height of the pdf
% fyj=1/(sqrt(2*pi)*sigmaj)*exp(-1/2*((y-muj)/sigmaj).^2);
% Plot using ax1 handle, saved above,to save this graph
% on top of the previous graph.
v=axis;
h1=line(y_j,fy_j ,'Color','r','Parent',ax1);
set(h1,'linestyle','--','color','r','linewidth',2)
h1=line([cutoff cutoff]',[0 fy_cutoff]','Color','b','Parent',ax1);
set(h1,'linestyle','-.','color','b','linewidth',3)
h2=line([cutoff cutoff]',[fy_cutoff v(4)]','Color','b','Parent',ax1);
set(h2,'linestyle','-.','color','b','linewidth',3)
s=sprintf('Case Study 14.3.1, %2.0f samples untransformed',n);
title(s,'FontSize',22)
figure(gcf);pause

Questions p 816-817
14.3.4 **A large sample Wilcoxon signed rank test**
Theorem 14.3.2
Theorem 14.3.3

Case Study 14.3.2 Heroine addiction
% LMcs140302_4th.m
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% page 819. A case study using Wilcoxon signed rank test
% Written by Eugene.Gallagher@umb.edu 11/16/10 Revised 12/12/10, 3/1/11
%
D=[51 53 43 36 55 55 39 43 45 27 21 26 22 43];
expected=28;
[P,H,STATS] = signrank(D,expected,'alpha',0.05,'method','exact');
fprintf(...
\nThe signed rank test statistic is %4.1f with exact 1-tailed p=%6.4g\n',...
 STATS.signedrank,P/2);
[P,H,STATS] = signrank(D,expected,'alpha',0.05,'method','approximate');
fprintf('The signed rank test approximate 1-tailed p=%6.4g for z=%6.4f\n',P/2,STATS.zval);
[H,P,CI,STATS] = ttest(D,expected,0.05,'right');
fprintf('The one-sample t test 1-tailed p=%6.4g\n',P);
fprintf(...

```
'The mean Q Score = %4.2f with 95%% CI: [%4.2f %4.2f]\n',...
    mean(D), CI(1),CI(2));
[p,h,stats] = signtest(D,expected,0.05,'method','exact');
fprintf('The sign test statistic is %4.1f with exact 1-tailed p=%6.4g\n',stats.sign,p/2);
[p,h,stats] = signtest(D,expected,'method','approximate');
fprintf('The sign test z=%5.3f with approximate 1-tailed p=%6.4g\n',stats.zval,p/2);


% Plot histogram and check for symmetry
binsize=2;  % Needed in order to properly scale the normal pdf
edges=20:binsize:56;
% hist(D);
[N,BIN] = histc(D,edges);
bar(edges,N,'histc')
axis([19.5 56.5 0 max(N)+0.2])
set(get(gca,'Children'),'FaceColor',[.8 .8 1]);
xlabel('Q Score','FontSize',20);
ylabel('Number of Cases','FontSize',20);
ax1=gca;
set(ax1,'Ytick',[0:max(N)],'Xtick',edges)
figure(gcf);pause
% Superimpose the normal probability pdf on a histogram of differences.
% The normal probability equation is provided on p. 293
% This is for mean 0, and unit standard
% deviation. The more general equation (Legendre & Legendre, 1998 p. 147) is:
% f(y_j)=1/(sqrt(2*pi)*sigma_j)*exp(-1/2*((y_j-mu_j)/sigma_j)^2);
n=length(D);
mu_j=mean(D);
sigma_j=std(D); % sigmaj is the standard deviation; = 1 after Z transform
y_j=19.5:0.01:56.5;
fy_j=1/(sqrt(2*pi)*sigma_j)*exp(-1/2*((y_j-mu_j)./sigma_j).^2);
fy_j=n*binsize*fy_j;
cutoff=expected;
fy_cutoff=n*binsize/(sqrt(2*pi)*sigma_j)*exp(-1/2*((cutoff-mu_j)./sigma_j).^2);
% will properly scale the height of the pdf
% fyj=1/(sqrt(2*pi)*sigmaj)*exp(-1/2*((y-muj)/sigmaj).^2);
% Plot using ax1 handle, saved above,to save this graph
% on top of the previous graph.
v=axis;
h1=line(y_j,fy_j ,'Color','r','Parent',ax1);
set(h1,'linestyle','--','color','r','linewidth',2)
h1=line([cutoff cutoff]',[0 fy_cutoff]','Color','b','Parent',ax1);
set(h1,'linestyle','-.','color','b','linewidth',3)
h2=line([cutoff cutoff]',[fy_cutoff v(4)]','Color','b','Parent',ax1);
set(h2,'linestyle','-.','color','b','linewidth',3)
s=sprintf('Case Study 14.3.2, %2.0f samples untransformed',n);
```

```
title(s,'FontSize',22)
figure(gcf);pause
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% page 819. A case study using Wilcoxon signed rank test
% Written by Eugene.Gallagher@umb.edu 11/16/10 Revised 12/12/10
%
D=[51 53 43 36 55 55 39 43 45 27 21 26 22 43];
hist(D-28);figure(gcf);pause
hist(log(D)-log(28));figure(gcf);pause
M=28;
[H,P,CI,STATS] = ttest(D,M,0.05,'right');
fprintf('\nThe paired t test 1-tailed p=%6.4g\n',P);
fprintf('The mean Q score = %4.2f with 95%% CI: [%4.2f %4.2f]\n',...
    mean(D), CI(1),CI(2));
[H,P,CI,STATS] = ttest(log(D),log(M),0.05,'right');
fprintf('\nThe paired t test of log transform 1-tailed p=%6.4g\n',P);
[p,h,stats] = signtest(D,M,0.05,'method','exact');
fprintf('The sign test exact 1-tailed p=%6.4g\n',p/2);
[p,h,stats] = signtest(D,M,'method','approximate');
fprintf('The sign test approximate 1-tailed p=%6.4g\n',p/2);
[P,H,STATS] = signrank(D,M,'alpha',0.05,'method','exact');
fprintf('The sign rank test exact 1-tailed p=%6.4g\n',P/2);
[P,H,STATS] = signrank(D,M,'alpha',0.05,'method','approximate');
fprintf('The sign rank test approximate 1-tailed p=%6.4g\n',P/2);
```

14.3.5 **Testing** H_o: $\mu_D = 0$ (Paired data)

```
Case Study 14.3.3
% LMcs140303_4th.m
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% page 821. A case study solved by the sign and Wilcoxon signed rank
% test
% Written by Eugene.Gallagher@umb.edu 11/16/10 Revised 3/1/11
%
D=[4.67 4.36;3.5 3.64;3.5 4;3.88 3.26;3.94 4.06;4.88 4.58;4 3.52
    4.4 3.66;4.41 4.43;4.11 4.28;3.45 4.25;4.29 4;4.25 5;4.18 3.85
    4.65 4.18];
expected=0;
[P,H,STATS] = signrank(D(:,1),D(:,2),'alpha',0.05,'method','exact');
fprintf(...
\nThe signed rank test statistic is %4.1f with exact 2-tailed p=%6.4g\n',...
 STATS.signedrank,P);
[P,H,STATS] = signrank(D(:,1),D(:,2),'alpha',0.05,'method','approximate');
fprintf('The signed rank test approximate 2-tailed p=%6.4g for z=%6.4f\n',P,STATS.zval);
[H,P,CI,STATS] = ttest(D(:,1),D(:,2),0.05,'both');
fprintf('The paired t test 2-tailed p=%6.4g\n',P);
```

```
fprintf(...
'The mean difference in ratings (In-Class - Online) = %4.2f with 95%% CI: [%4.2f %4.2f]\n',...
    mean(D(:,1)-D(:,2)), CI(1),CI(2));
[p,h,stats] = signtest(D(:,1),D(:,2),0.05,'method','exact');
fprintf('The sign test statistic is %4.1f with exact 2-tailed p=%6.4g\n',stats.sign,p);
[p,h,stats] = signtest(D(:,1),D(:,2),'method','approximate');
fprintf('The sign test z=%5.3f with approximate 2-tailed p=%6.4g\n',stats.zval,p);


% Plot histogram and check for symmetry
binsize=.2;  % Needed in order to properly scale the normal pdf
edges=-.8:binsize:.8;
% hist(D);
[N,BIN] = histc(D(:,1)-D(:,2),edges);
bar(edges,N,'histc')
axis([-.82 .82 0 max(N)+0.2])
set(get(gca,'Children'),'FaceColor',[.8 .8 1]);
xlabel('Difference in evaluations, In-Class - Online','FontSize',20);
ylabel('Number of Cases','FontSize',20);
ax1=gca;
set(ax1,'Ytick',[0:max(N)],'Xtick',edges)
figure(gcf);pause
% Superimpose the normal probability pdf on a histogram of differences.
% The normal probability equation is provided on p. 293
% This is for mean 0, and unit standard
% deviation. The more general equation (Legendre & Legendre, 1998 p. 147) is:
% f(y_j)=1/(sqrt(2*pi)*sigma_j)*exp(-1/2*((y_j-mu_j)/sigma_j)^2);
n=length(D(:,1)-D(:,2));
mu_j=mean(D(:,1)-D(:,2));
sigma_j=std(D(:,1)-D(:,2)); % sigmaj is the standard deviation; = 1 after Z transform
y_j=-.82:0.01:.82;
fy_j=1/(sqrt(2*pi)*sigma_j)*exp(-1/2*((y_j-mu_j)./sigma_j).^2);
fy_j=n*binsize*fy_j;
cutoff=expected;
fy_cutoff=n*binsize/(sqrt(2*pi)*sigma_j)*exp(-1/2*((cutoff-mu_j)./sigma_j).^2);
% will properly scale the height of the pdf
% fyj=1/(sqrt(2*pi)*sigmaj)*exp(-1/2*((y-muj)/sigmaj).^2);
% Plot using ax1 handle, saved above,to save this graph
% on top of the previous graph.
v=axis;
h1=line(y_j,fy_j ,'Color','r','Parent',ax1);
set(h1,'linestyle','--','color','r','linewidth',2)
h1=line([cutoff cutoff]',[0 fy_cutoff]','Color','b','Parent',ax1);
set(h1,'linestyle','-.','color','b','linewidth',3)
h2=line([cutoff cutoff]',[fy_cutoff v(4)]','Color','b','Parent',ax1);
set(h2,'linestyle','-.','color','b','linewidth',3)
```

```
s=sprintf('Case Study 14.3.3, %2.0f samples untransformed',n);
title(s,'FontSize',22)
figure(gcf);pause
```

14.3.6 **Testing** $H_o: \mu_X = \mu_Y$ (The Wilcoxon Rank Sum Test)
Theorem 14.3.4

Case Study 14.3.4
```
% LMcs140304_4th.m
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% Written by Eugene.Gallagher@umb.edu; written 11/16/10; revised 11/23/10
% Calls Matlab's ranksum.m and Gallagher's Wilcoxranksum.m

AL=[177 177 165 172 172 179 163 175 166 182 177 168 179 177]';
NL=[166 154 159 168 174 174 177 167 165 161 164 161]';
boxplot([AL;NL],[ones(length(AL),1);zeros(length(NL),1)]);figure(gcf)
[P,H,STATS] = ranksum(AL,NL,'alpha',0.05,'method','exact');
fprintf(...
   '\n\nUsing Matlab''s ranksum, exact p=%6.4f, Rank sum = %4.1f\n',P,...
   STATS.ranksum)
if H==1
   fprintf('Reject Ho\n\n')
else
   fprintf('Fail to reject Ho\n\n')
end
[pvalue,W,U]=Wilcoxranksum(AL,NL,1);
fprintf('Using Gallagher''s Wilcoxranksum, exact p=%6.4f;\n', P)
fprintf('Wilcoxon''s W = %4.1f; Mann-Whitney U=%4.1f;\n',W,U)
[P,H,STATS] = ranksum(AL,NL,'alpha',0.05,'method','approximate');
fprintf('\nUsing Matlab''s ranksum, large sample p=%6.4f;\n',P)
fprintf('Rank sum = %4.1f; z-value=%5.2f\n',STATS.ranksum,STATS.zval)
if H==1
   fprintf('Reject Ho\n\n')
else
   fprintf('Fail to reject Ho\n\n')
end
[pvalue,W,U,Wstar]=Wilcoxranksum(AL,NL,0);
fprintf('Using Gallagher''s Wilcoxranksum, large sample p=%6.4f;\n',P)
fprintf('Wilcoxon''s W = %4.1f; Mann-Whitney U=%4.1f; z-value=%5.2f\n',...
   W,U,Wstar)
function [pvalue,W,U,Wstar]=Wilcoxranksum(X,Y,Ex)
% Wilcoxon rank-sum test
% [pvalue,W,U,Wstar]=Wilcoxranksum(X,Y,Ex)
% Tests the null hypothesis that X & Y have the same pdf.
% Input: X,Y two samples,Ex~=0 indicates do an exact test.
% Output: pvalue: pvalue, 2-sided p value for large sample approximation N(0,1) distribution
```

```
%        W=Wilcoxon rank sum statistic
%        U=Mann-Whitney U statistic
%        Wstar=z value for asymptotic large sample approximation
% Calls Wilcoxrsexact
% Written by Eugene.Gallagher@umb.edu
% Revised 11/14/10

X=X(:);Y=Y(:);
n=length(X);
m=length(Y);
% Rank the X&Y values from smallest to largest, assigning average ranks to ties.
[T,R,ind]=ties([X;Y]);T=T';  % calls Gallagher's ties.m
% Find sum of ranks of the smaller sample;
if n<m;
   W=sum(R(1:n));
else
   W=sum(R(n+1:n+m));
   n=m;  % Expected value & variance equastions assume n is the size of the smaller group.
   m=length(X);
end
U=W-n*(n+1)/2; % Mann-Whitney U statistic
largesample=logical(1);
if nargin>2
   if Ex~=0
      largesample=logical(0);
   end
end
if nargin>2 & ~largesample
  ncomb=nchoosek(n+m,n);
  if ncomb>1e6
    t=sprintf(...
      '%d combinations, T=%d min (1e6 combs take 1 min on p4)\n',...
      ncomb,round(ncomb/1e6));
    toomany=menu(t,'Stop','Continue');
    if toomany==1
      largesample=logical(1);fprintf('Large sample approximation for 2-tailed p\n');
    end
  end
  if ~largesample
    pexuptail=wilcoxrsexact(n,m,W,R);
    if pexuptail<=0.5
      pvalue=2*pexuptail;
    else
      pvalue=2*(1-pexuptail);
    end
```

```
  end
end
if largesample
   % Large sample approximation;% Hollander & Wolfe p. 108
   EoW=(n*(m+n+1))/2;
   % Calculate the variance of W, without ties and with ties.
   if isempty(T)  % Size of tied groups from ties.m
      VaroW=(m*n*(m+n+1))/12;
   else
      VaroW=(m*n)/12*(m+n+1-(sum((T-1).*T.*(T+1)))/((m+n)*(m+n-1)));
   end
   Wstar=(W-(n*(m+n+1)/2))/sqrt(VaroW);  % Without ties, tends to an asymptotic N(0,1)
distribution.
   % Find the 2-tailedprobability of Wstar from the standard normal distributioin
   pvalue=erfc(abs(Wstar)/sqrt(2));
   % Note that the exact p values are tabulated, and an exact test, even in the presence of ties
   % can be performed, see pp. 113-116 in Hollander & Wolfe.
end
```

```
function pexuptail=Wilcoxrsexact(n,m,W,ranks);
% Exact upper tail p values for Wilcoxon Rank Sum statistic
% function pexuptail=Wilcoxrsexact(n,m,W,ranks);
% Borrows shamelessly from Strauss's combvals.m
% Note that Matlab's nchoosek will also generate the list
% of combinations. This program doesn't generate the full
% matrix of combinations, but calculates the test stat only.
% Input: n size of smaller group
%        m size of larger group
%        W Wilcoxon signed rank statistic
%        ranks, actual ranks of n+m items if there are ties present.
% Written by E. Gallagher, Eugene.Gallagher@umb.edu
% Help file for Strauss' combvals:
% COMBVALS: Generates the combinations of n integers taken r at a time.  The
%        number of such combinations is given by function nc=combin().
%    Usage: c = combvals(n,r)
%        n = number of integers (1:n) to be combined.
%        r = number to be taken at a time (0 < r <= n).
%        --------------------------------------------------------
%        c = [nc x r] matrix of combinations.

% Based on ACM Algorithm 94, J. Kurtzberg, Comm. ACM, June 1962.
% RE Strauss, 12/18/98

% An exact conditional distribution with ties follows Hollander & Wolfe p. 115
if nargin<4
   ranks=1:n+m;
```

```
      notiedr=logical(1);
else
   if length(ranks)<n+m
      error(...
         sprintf(...
         'Number of ranks (%d) doesn''t match n+m (%d)\n',...
         length(ranks),n+m));
   end
   ranks=sort(ranks);
   notiedr=logical(0);  % could do a check to see if there really are ties with ties.m
end
ranks=ranks(:);
fudranks=flipud(ranks);
N=n+m;
r = n;
ncomb = nchoosek(N,r);  % Matlab's built-in combination function.
if W>=n*(n+m+1)-W;
   uppertail=logical(1);
else
   W=n*(n+m+1)-W;
   uppertail=logical(0);
end
if W>sum(fudranks(1:n))
   if uppertail
      error('W impossibly large')
   else
      error('W impossibly small')
   end
elseif W==sum(fudranks(1:n)) & notiedr
   if uppertail
      pexuptail=1/ncomb;
   else
      pexuptail=(ncomb-1)/ncomb;
   end
   return
end
  % Strauss's combval lists combinations in c in lexicographic
  % order, thus the critical values for sum(C) are larger than
  % observed W. We can speed up the process by using
  % Wstar=min(W,n*(m+n+1)-W) and exiting loop when Wstar fails
  % to be less than critical value
  if ncomb>1e6
    t=sprintf(...
       '%d combinations, T=%d min (1e6 combs take 1 min on p4)\n',...
       ncomb,round(ncomb/1e6));
```

```
      toomany=menu(t,'Stop','Continue');
      if toomany==1
         return
      end
   end
   % c = zeros(ncomb,r);     % Don't need to store values.
   Tally=0;
   j = zeros(1,r);

   for i = 1:ncomb
     b = 1;
     endflag = 0;
     while(~endflag)
       if (j(b)>=b)
         a = j(b)-b-1;
         for l = 1:b
           j(l) = l+a;
         end;
         endflag = 1;
       else
         if (b==r)
           for b = 1:r
             j(b) = N-r-1+b;
           end;
           endflag = 1;
         end;
         b = b+1;
       end;
     end;
     % c(i,:) = N-j(r:-1:1);
     c=N-j(r:-1:1);
     if sum(ranks(c))>=W
        Tally=Tally+1;
     end
   end;
   pexuptail=Tally/ncomb;
   if ~uppertail
      pexuptail=1-pexuptail;
   end
```

```
function [T,R,ind]=ties(A)
% format: [T,R,ind]=ties(A)
% a function to return a row vector of tied groups, T,
% Ranks R (including average ranks) and indices of tied elements
% needed to calculate variance of S using Kendall's
% variance formula & Spearman's r.
```

```
% input: A is a row or column vector
% T: a row vector containing number of members of tied groups
% T=0 if there are no tied groups
% sum(T) is equal to the number of tied elements.
% each element of T equals the number in each tied group
% tied groups are sorted in ascending order.
% Examples: A=[1 2 3];[T,R,ind]=ties(A)=>  T=0,R=[1 2 3],ind=[]
%         A=[1 2 3 1];        T=2,R=[1.5 3 4 1.5],ind=[1 4]
%         A=[2 1 2 3 1 2];     T=[2 3],R=[4 1.5 4 6 1.5 4],
%                     ind=[5 2 3 1 6]
%         A=[2 1 2 3 3 1 2];   T=[2 3 2],R=[4 1.5 4 6.5 6.5 1.5 4]
%                     ind=[6 2 3 1 7 4 5]
% R (Row vec)=numerical rankings of A with ave. ranks for ties
% ind: indices of tied elements, sorted by rank; sorted tied elements=A(ind);
% ties.m is used in Kendall.m as T=ties(A), and Spear.m
% written by E. Gallagher, Environmental Sciences Program
% UMASS/Boston, Email: Eugene.Gallagher@umb.edu
% written: 6/16/93, revised 6/17/93
[r,c]=size(A);
if r>c
   A=A';              % change to row vector
end
[Asort,k]=sort(A);
iota=1:length(A);iota=iota';
R(k)=iota;
index=[k' iota];
ind=[];
CDA=[~diff(Asort) 0];
min1=min(find(CDA==1));
if isempty(min1)
   T=0;
   return
end
i=0;
[rw,cl]=size(CDA);
T=zeros(size(rw,cl));
while ~isempty(min1)
   min0=min(find(CDA==0));
   if min0<min1
      CDA(min0:min1-1)=[];
      index(min0:min1-1,:)=[];
   else
    i=i+1;
    T(i)=min0-min1+1;
    CDA(min1:min0)=[];
```

```
    ind=[ind index(min1:min0,1)'];
    R(1,index(min1:min0))=ones(1,T(i))*sum(index(min1:min0,2))/T(i);
    index(min1:min0,:)=[];
  end
  min1=min(find(CDA==1));
end
T(find(T==0))=[];
```

Questions p 825-826
      14.4   The **KRUSKAL-WALLIS TEST**
Theorem 14.4.1

Case Study 14.4.1 Draft lottery
% LMcs140401_4th.m
% Case Study 14.4.1
% 1969 draft lottery
% From Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th ed
% Written by Eugene.Gallagher@umb.edu 12/7/2010
% Are the data random?
DATA=[1 305 086 108 032 330 249 093 111 225 359 019 129
2 159 144 029 271 298 228 350 045 161 125 034 328
3 251 297 267 083 040 301 115 261 049 244 348 157
4 215 210 275 081 276 020 279 145 232 202 266 165
5 101 214 293 269 364 028 188 054 082 024 310 056
6 224 347 139 253 155 110 327 114 006 087 076 010
7 306 091 122 147 035 085 050 168 008 234 051 012
8 199 181 213 312 321 366 013 048 184 283 097 105
9 194 338 317 219 197 335 277 106 263 342 080 043
10 325 216 323 218 065 206 284 021 071 220 282 041
11 329 150 136 014 037 134 248 324 158 237 046 039
12 221 068 300 346 133 272 015 142 242 072 066 314
13 318 152 259 124 295 069 042 307 175 138 126 163
14 238 004 354 231 178 356 331 198 001 294 127 026
15 017 089 169 273 130 180 322 102 113 171 131 320
16 121 212 166 148 055 274 120 044 207 254 107 096
17 235 189 033 260 112 073 098 154 255 288 143 304
18 140 292 332 090 278 341 190 141 246 005 146 128
19 058 025 200 336 075 104 227 311 177 241 203 240
20 280 302 239 345 183 360 187 344 063 192 185 135
21 186 363 334 062 250 060 027 291 204 243 156 070
22 337 290 265 316 326 247 153 339 160 117 009 053
23 118 057 256 252 319 109 172 116 119 201 182 162
24 059 236 258 002 031 358 023 036 195 196 230 095
25 052 179 343 351 361 137 067 286 149 176 132 084
26 092 365 170 340 357 022 303 245 018 007 309 173
27 355 205 268 074 296 064 289 352 233 264 047 078
```

28 077 299 223 262 308 222 088 167 257 094 281 123
29 349 285 362 191 226 353 270 061 151 229 099 016
30 164 NaN 217 208 103 209 287 333 315 038 174 003
31 211 NaN 030 NaN 313 NaN 193 011 NaN 079 NaN 100];
DATA=DATA(:,2:13);
y=DATA(:); % convert the data into columns; drop the NaN elements
group=repmat(1:12,31,1);group=group(:);i=~isnan(y);y=y(i);group=group(i);
[p,table,stats] = kruskalwallis(y,group)
multcompare(stats)
% As described on page 829, test the 1st vs. 2nd 6 months.
g=group;g(group<=6)=1;g(group>6)=2;
[p2,table2,stats2] = kruskalwallis(y,g)

Questions p 830-832
       14.5    **THE FRIEDMAN TEST**
**Theorem 14.5.1**

**Case Study 14.5.1**
% LMcs140501_4th.m
% Case Study 14.5.1
% Base running example from Hollander & Wolfe
% From Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th ed
% Written by Eugene.Gallagher@umb.edu 12/7/2010
%
DATA=[5.5 5.55
   5.7 5.75
   5.6 5.5
   5.5 5.4
   5.85 5.7
   5.55 5.6
   5.4 5.35
   5.5 5.35
   5.15 5
   5.8 5.7
   5.2 5.1
   5.55 5.45
   5.35 5.45
   5 4.95
   5.5 5.4
   5.55 5.5
   5.55 5.35
   5.5 5.55
   5.45 5.25
   5.6 5.4
   5.65 5.55
   6.3 6.25];
plot(DATA');

```
ax1=gca;
set(ax1,'Xtick',[1 2])
set(ax1,'XtickLabel',{'Narrow-Angle','Wide-Angle'})

figure(gcf);pause
[P,TABLE,STATS]=friedman(DATA);
```

14.6 TESTING FOR RANDOMNESS

Case Study 14.6.1
```
% LMcs140601_4th.m
% Uses the resampling toolbox function runs.m
DATA=...
[61 53 58 51 52 34 45 52 46 52 37 39 50 38 55 59 57 64 73 46 48 47 40 35 40]';
n=length(DATA);
[H,P,STATS]=runstest(diff(DATA)>0); % This is not the same runs test a
                    % Larsen and Marx. Matlab's runs test
                    % considers the number of positive and
                    % negative runs, but L&M's test just
                    % considers the total N (25) in
                    % calculating its test statistic. Thus,
                    % L&M's test assumes no trend.
% Theorem 14.6.1:
EW=(2*n-1)/3;
VarW=(16*n-29)/90;
Z=(STATS.nruns-EW)/sqrt(VarW)
if Z>0
   p=1-normcdf(Z);
else
   p=normcdf(Z);
end
fprintf(...
  'With Matlab''s runs test, P(%2.0f runs with %2.0f cases) is %5.3f\n',...
   STATS.nruns,n,P)
fprintf(...
  'With Larsen & Marx''s runs test P(%2.0f runs with %2.0f cases) = %5.3f\n',...
   STATS.nruns,n,p)

% Although undocumented, Matlab is probably using the Wald-Wolfowitz runs
% test; When I can get access to my stats books with the exact version
% of the test, I'll check.
```

Questions p. 838-841

References

Bevington, P. R. and D. K. Robinson. 1992. Data reduction and error analysis for the physical sciences. McGraw-Hill, Boston. 328 pp. [**11**]

Draper, N. R. and H. Smith. 1998. Applied Regression Analysis, Third Edition. John Wiley & Sons, New York. 706 pp. & data diskette. [**24**]

Larsen, R. J. and M. L. Marx. 2006. An introduction to mathematical statistics and its applications, 4th edition. Prentice Hall, Upper Saddle River, NJ. 920 pp. [**7**, **8**, **9**, **12**, **13**, **14**, **16**, **17**, **18**, **19**, **21**, **24**, **33**]

Legendre, P. and L. Legendre. 1998. Numerical Ecology, 2nd English Edition, Elsevier, Amsterdam. 853 pp. [**8**]

Manly, B. F. J. 1991. Randomization, bootstrap, and Monte Carlo methods in biology. Chapman and Hall, London. [**12**]

Mayo, D. G. 1996. Error and the growth of experimental knowledge. University of Chicago Press, Chicago & London. 493 pp. [**9**]

Nahin, P. J. 2000. Duelling idiots and other probability puzzlers. Princeton University Press, Princeton & Oxford. 269 pp. [?]

Ramsey, F. L. and D. W. Schafer. 2002. The statistical sleuth: a course in methods of data analysis, Second edition Duxbury Press, Belmont CA, 742 pp & data diskette. [**10**, **15**, **18**]

Salsburg, ,D. 2001. The lady tasting tea: How statistics revolutionized science in the twentieth century. W. H. Freeman & Co., New York. 340 pp. [?]

Taylor, J. R. 1997. An introduction to error analysis: the study of uncertainties in physical measurements, 2nd edition. University Science Books, Sausalito CA 327 pp. [**11**, **15**, **16**, **17**]

Weir, B. S. 1990. Genetic data analysis. Sinauer Associates, Sunderland MA. [?]

Zar, J. H. 1999. Biostatistical analysis, Fourth Edition. Prentice-Hall, Upper Saddle River, N.J. 663 pp. plus appendices. [?]

Index