

WEEK 8: CHAPTER 9 & CHAPTER 14, TWO SAMPLE PROBLEMS (1 OF 2): INDEPENDENT T, TWO-SAMPLE BINOMIAL, F, AND WILCOXON RANK SUM TESTS

TABLE OF CONTENTS

	Page:
List of Figures.	2
List of Tables.	2
List of m.files.	2
Assignment.	3
Required reading.	3
Understanding by Design Templates.	3
Understanding By Design Stage I — Desired Results Week 8.	3
Understanding by Design Stage II — Assessment Evidence Week 8 (7/19-7/25 M).	3
Introduction.	4
The Independent Samples t Test.	5
The box plot.	6
Confidence Limits for the Difference in Means.	6
Welch’s <i>t</i> test.	7
Wilcoxon’s Rank Sum Test.	7
The Behrens-Fisher problem revisited for Wilcoxon’s rank-sum test.	9
Testing — The F test.	10
The two-sample binomial test.	11
Case Studies.	11
Case Study 9.2.1.	11
Case Study 9.2.2.	13
Case Study 9.3.1.	14
Case Study 9.4.1.	15
Case Study 9.4.2.	15
Case Study 9.5.1.	15
Case Study 9.5.2.	16
Case Study 9.5.3.	16
Case Study 14.3.4 Baseball Game Length.	17
Chapter Outlines.	18

Annotated outline (with Matlab scripts) for Larsen & Marx Chapter 9.	18
Annotated Outline (with Matlab scripts) for Larsen & Marx Chapter 14.	26
References.	45
Index.	46

List of Figures

Figure 1. Boxplots for the Case Study 9.2.1 data with Twain left and Snodgrass right. There is no indication of unequal spread.	12
Figure 2. Student's t distribution for 16 df showing the two tailed critical values for $\alpha=0.01$. Since 3.88 is greater than 2.9208, reject the null hypothesis at the $\alpha=0.01$ level.	12
Figure 3. Boxplots for the nonconfined (left) and solitary confinement (right) alpha waves.	14
Figure 4. The F distribution with 9 and 9 df. In order to reject the null hypothesis of equal variances, an F ratio of 0.25 or 4.03 would have to be observed. The observed F ratio of 1.7 is well within the two $\alpha = 0.05$ critical values.	15
Figure 5. Boxplots for enamel spectropenetration gradients for males (left) and females (right).	16
Figure 6. Boxplots for glacial flow rates for three-year photographs (left) and five-year photographs (right).	16
Figure 7. Boxplots for average game length for American (left) and National (right) league teams.	17

List of Tables

List of m.files

LMcs090201_4th.	18
LMcs090202_4th.	19
function [D,t,df,pvalue,CI,sp]=student2group(Xn,Yn,Xmean,Ymean,Xstd,Ystd,alpha).	20
LMcs090301_4th.	20
LMcs090401_4th.	22
function [D,phat,z,pvalue,CI,obsp]=binom2sample(x,n,y,m,alpha).	22
LMcs090402_4th.	23
LMcs090501_4th.	23
function [D,Studp,t,df,RandPermP,CIpD,CIf]=randp2sample(X,Y,Trials,UseT,alpha).	24
LMcs090502_4th.	26
LMcs090503_4th.	26
LMcs140201_4th.	26
LMcs140202_4th.	28
LMcs140203_4th.	29
LMcs140301_4th.	31
LMcs140302_4th.	32
LMcs140303_4th.	34
LMcs140304_4th.	36
function [pvalue,W,U,Wstar]=Wilcoxranksum(X,Y,Ex).	37
function pexuptail=Wilcoxrsexact(n,m,W,ranks);	38
function [T,R,ind]=ties(A).	41
LMcs140401_4th.	42
LMcs140501_4th.	43

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 Chapter 9, all
- " Read Chapter 14, Section 14.3

Understanding by Design Templates

Understanding By Design Stage I — Desired Results Week 8

LM Chapter 9 (Independent) Two-sample problems

G Established Goals

- Come to know the Student's t , Welch's t , Wilcoxon rank sum and permutation tests based on difference in mean and t ratios..

U Understand

- The major assumption of the independent tests is equal spread and that assumption applies to the Wilcoxon rank sum and permutation test

Q Essential Questions

- When do you use the independent samples t test and when the paired t test?
- What are the advantages of pairing?
- What assumptions matter for these tests?

K Students will know how to define (in words or equations)

- box plot, confidence limit for the difference in means, Levene's test, Mann-Whitney U test, pooled variance, power efficiency, Student's t test, **Satterthwaite df approximation, Welch's t test, Wilcoxon rank sum test**

S Students will be able to

- Analyze independent two-sample problems with Matlab
- Recognize when the assumptions of the independent samples t test have been violated
- Find and interpret the confidence limits for differences between means

Understanding by Design Stage II — Assessment Evidence Week 8 (7/19-7/25 M)

Chapter 9 (part) and Chapter (14) part

- **Post in the discussion section by 7/27 W 10 PM** by W
 - Run Rajeev Raizada's Matlab interactive 2-sample t-test.m simulation & note the effects of outliers, post a 1-paragraph summary of your conclusions
 - <http://www.dartmouth.edu/~raj/intro-stats.html>
 - http://www.dartmouth.edu/~raj/Matlab/Intro_stats/interactive_two_sample_t_test.m
- **HW 6 Problems due Wednesday 7/27/11 W 10 PM**
 - **Basic problems (4 problems 10 points)**
 - Problem 9.2.1 Alcoholic Authors. Use a box plot and decide whether an independent samples t test or Welch's t test is appropriate. Use Case Study 9.2.1 as a model.
 - Problem 9.4.2 using Case Studies 9.4.1 or 9.4.2 as a model
 - Problem 9.5.2 using Case Study 9.5.1 as a model
 - Problem 14.3.10 analyze Case Study 9.3.1 data using Wilcoxon's ranksum test using Case Study 14.3.4 as a model
 - **Advanced problems (2.5 points each)**
 - **Problem 1.** Perform an independent samples and Welch's t test on the Case Study 9.3.1 alpha wave data. Note any differences.
 - **Problem 2** Gallagher's exactpratio.m will Apply the exactpratio.m test to case study 9.2.1 data and compare the p value with the independent samples t test.
 - **Master problems (1 only, 5 points)** unequalvariances.m is Gallagher's program to generate random data with known differences in mean, standard deviation, skewness and kurtosis. Use this program to test whether the independent samples t test is robust to violations of equal variance. How is the result affected by different sample sizes. Write a brief report on your exploration of the program and its results.

Introduction

In Weeks 8 & 9, we'll cover two sample problems. In Week 8, we'll cover the independent samples t test (**Larsen & Marx 2006**, Chapter 9) and the Wilcoxon rank sum test which is the nonparametric analogue of the independent samples t test. Nonparametric tests are covered in Chapter 14 in Larsen & Marx. Also in Week 8, we'll introduce the **Welch's t test** which is not covered in **Larsen & Marx (2006)**, but Welch's test is appropriate when the two groups have different variances. A final class of two-sample tests will be based on random permutations. We'll either randomly permute or find all the permutations of two samples and test for differences in means between every pair of permutations. If all of the permutations have been identified, the test will be exact. As we'll see with Gallagher's Matlab program unequalvariances.m, random permutations don't solve perhaps the major problem for the t test: unequal variances between the groups.

In Week 9, we'll continue with the two sample tests starting with the paired t test, which **Larsen & Marx (2006)** cover in Chapter 13. Also in Week 9, we'll cover the nonparametric equivalents of the paired t test, which are the Wilcoxon signed rank test and Fisher's sign test.

The Independent Samples t Test

Theorem 9.2.1 provides the equations for the pooled estimate of the standard error for the difference in two means. There are three key assumptions for the use of the independent sample t test: 1) equal means, 2) both samples normally distributed, and 3) the errors are independently distributed. The best way to assess whether two samples have equal spread or variance is to use a box plot. There is a formal test, called Levene's test, for testing whether the equal

THEOREM 9.2.1. Let X_1, X_2, \dots, X_n be a random sample of size n from a normal distribution with mean μ_X and standard deviation σ and let Y_1, Y_2, \dots, Y_m be an independent random sample of size m from a normal distribution with mean μ_Y and standard deviation σ . Let S_X^2 and S_Y^2 be the two corresponding sample variances, and S_p^2 , the *pooled variance*, where

$$S_p^2 = \frac{(n-1)S_X^2 + (m-1)S_Y^2}{n+m-2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2}{n+m-2}$$

Then

$$T_{n+m-2} = \frac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{S_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

has a Student t distribution with $n + m - 2$ degrees of freedom.

Theorem 9.2.2 is a simple extension of Theorem 9.2.1 that defines the t statistic used to test for differences in the means in two samples.

Theorem 9.2.2. Let x_1, x_2, \dots, x_n and y_1, y_2, \dots, y_m be independent random samples from normal distributions with means μ_X and μ_Y , respectively, and with the same standard deviation σ . Let

$$t = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

- a. To test $H_0: \mu_X = \mu_Y$ versus $H_1: \mu_X > \mu_Y$ at the α level of significance, reject H_0 if $t \geq t_{\alpha, n+m-2}$.
- b. To test $H_0: \mu_X = \mu_Y$ versus $H_1: \mu_X < \mu_Y$ at the α level of significance, reject H_0 if $t \leq -t_{\alpha, n+m-2}$.
- c. To test $H_0: \mu_X = \mu_Y$ versus $H_1: \mu_X \neq \mu_Y$ at the α level of significance, reject H_0 if t is either $(1) \leq -t_{\alpha/2, n+m-2}$ or $(2) \geq t_{\alpha/2, n+m-2}$.

The box plot

The boxplot was invented by John Tukey as a quick way to plot data. The centerline of the boxplot is the median. The box itself contains, approximately, the interquartile range. In actuality, there are about 7 different ways to find estimates of the 1st and 3rd quartiles, but Tukey wanted a quick method. So, Tukey's boxes are found by finding the median first and then finding the median of the lower half and the median of the upper half of the data. These will be close to but not necessarily identical to the quartiles. The whiskers on the box plots extend to real data points that are within 1.5 interquartile ranges from the end of the box. Then there are outliers and extreme outliers marked by real data that are between 1.5 and 3 interquartile ranges and extreme outliers that are more than 3 interquartile ranges from the end of the box.

There are boxplots that are used in K-12 texts that use a different plotting convention. The whiskers extend to include all of the data. This is the boxplot that is tested in virtually every Massachusetts MCAS exam.

Matlab produces a notched boxplot. The notches are just a crude approximation for a confidence interval for the median. I don't recommend the use of notched boxplots, but Matlab often produces them as the default boxplot option in some routines such as ANOVA's.

Confidence Limits for the Difference in Means

Theorem 9.5.1 provides the equation to calculate the confidence interval for two means. The confidence interval is estimated from the pooled standard deviation shown above in Theorem 9.2.1 and a t statistic with $n+m-2$ *df*.

Theorem 9.5.1. *Let x_1, x_2, \dots, x_n and y_1, y_2, \dots, y_m be independent random samples drawn from normal distributions with means μ_X and μ_Y , respectively, and with the same standard deviation, σ . Let s_p denote the data's pooled standard deviation. A $100(1 - \alpha)\%$ confidence interval for $\mu_X - \mu_Y$ is given by*

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

Welch's t test

The independent samples t test should not be used if the spread in the two groups is markedly different, especially if the sample sizes are unequal. There are two approaches that are used if the spreads are unequal. If the group with the larger spread has a larger mean than a square root or logarithmic transformation can often equalize the variance. If these transformations don't work, others include the inverse transform can be used. If there is still unequal spread, then the **Welch's t test** should be used. The Welch's t test is not based on the assumption of equal variances in the group. A new formula for the standard error of the difference in mean is based on the pair of different standard deviations and their sample sizes as shown in the box at the right from **Ramsey & Schafer (2002)**. The exact sampling distribution of this t test are not known but an empirical approach to fitting p values has shown that if the degrees of freedom are reduced in proportion to the differences in variance, approximate p values can be attained. The df equation is called **Satterthwaite's approximation**, and it converges to the conventional Student's t df ($n+m-2$) when the variances in the two groups are equal. The problem of testing for differences in means with different variances is called the Behrens-Fisher problem and is one of the most vexing in statistics.

4.3.2 The Welch t -Test for Comparing Two Normal Populations with Unequal Spreads

Welch's t -test employs the individual sample standard deviations as separate estimates of their respective population standard deviations, rather than pooling to obtain a single estimate of a population standard deviation. The result is a different formula for the standard error of the difference in averages:

$$SE_W(\bar{Y}_2 - \bar{Y}_1) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

This becomes the denominator in the t -statistic for comparing the means of populations with different spreads. Even when the populations are normal, however, the exact sampling distribution of the Welch t -ratio is unknown. It can be approximated by a t -distribution with $d.f._W$ degrees of freedom, known as Satterthwaite's approximation:

$$d.f._W = \frac{[SE_W(\bar{Y}_2 - \bar{Y}_1)]^4}{\frac{[SE(\bar{Y}_2)]^4}{(n_2-1)} + \frac{[SE(\bar{Y}_1)]^4}{(n_1-1)}}$$

where

$$SE(\bar{Y}_1) = \frac{s_1}{\sqrt{n_1}} \quad \text{and} \quad SE(\bar{Y}_2) = \frac{s_2}{\sqrt{n_2}}$$

Matlab's `ttest2.m` calculates the independent samples t test as the default, but if method is specified as 'unequal,' then Welch's t test is calculated.

Wilcoxon's Rank Sum Test

Salsburg (2001) devotes Chapter 16 of his book on landmark statistical developments to the work of Frank Wilcoxon. The chapter is titled 'Doing away with parameters.' Wilcoxon was a chemist, not a statistician, working at American Cyanamid. He was applying Student's t tests for equality of means and was disturbed when single observations would drastically affect the results:

"But Wilcoxon was concerned about what often appeared to be a failure of these methods [Student's t tests and F tests]. He might run a series of experiments where it was obvious to him that treatments differed in effect. Sometimes the t -tests would declare significance, and sometimes they would not. It often happens, when running an experiment in chemical engineering, that the chemical reactor, where the reaction takes place, is not sufficiently warmed up at the beginning of

the sequence of experimental trials. It may happen that a particular enzyme begins to vary in its ability to react. The result is an experimental value that appears to be wrong. it is often a number that is much too large or much too small. Sometimes, it is possible to identify the cause of this outlying result. Sometimes, the result is an outlier, differing drastically from all the other results, but there is no obvious reason for this ...” (Salsburg 2001, p. 161-162)

Wilcoxon developed a procedure resistant to outliers which involved converting the data to ranks. If there were n cases for sample A and m cases for sample B, the A and B cases would be combined and the data converted to integers ranging from 1 to $n+m$. If there were ties, the tied cases would be assigned average ranks. The test statistic for Wilcoxon’s rank sum test is simply the sum of the ranks in the smaller group.

Wilcoxon worked out, using combinations, the probability of observing different rank sum statistics for groups of differing size. He submitted his results to *Biometrics*, assuming that a reviewer would point out that a statistician had already invented such a test. But, he was the first to develop a nonparametric procedure based on ranks. In 1947, Mann & Whitney proposed their 2-sample equivalent to Student’s t test, again based on ranks. Their approach was shown to produce results exactly the same as Wilcoxon’s rank sum procedure, and the Mann-Whitney U statistic can be readily calculated if one knows the Wilcoxon sum of ranks statistic, called W .

It was later determined that the Central Limit Theorem assures that the distribution of the Wilcoxon rank sum statistic is normally distributed for moderately large samples sizes. A variance formula was produced for the standard error of Wilcox’s W . Using this formula for the standard error, one can quickly calculate the p value for the test. There is a correction for the number of ties in the data. With increasing numbers of tied values, the variance of the W statistic is reduced.

Larsen & Marx introduce the rank sum statistic in chapter 14, but since it is appropriate for two-sample tests, we’ll cover it here. The rank sum statistic is calculated by ranking all of the items in both groups from smallest to largest, assigning average ranks to ties. The rank sum statistic, W , is the sum of the ranks in the smaller group. The variance of W is described in Theorem 14.3.4 (at right).

Theorem 14.3.4. Let X_1, X_2, \dots, X_n and $Y_{n+1}, Y_{n+2}, \dots, Y_{n+m}$ be two independent random samples from $f_x(x)$ and $f_y(y)$, respectively, where the two pdfs are the same except for a possible shift in location. Let r_i denote the rank of the i^{th} observation in the combined sample (where the smallest observation is assigned a rank of one and the largest observation, a rank of $n+m$). Let

$$W' = \sum_{i=1}^{n+m} r_i z_i$$

where z_i is 1 if the i^{th} observation comes from $f_x(x)$ and zero, otherwise. Then:

$$E(W') = \frac{n(n+m+1)}{2}$$

$$\text{Var}(W') = \frac{nm(n+m+1)}{12}$$

And

$$\frac{W' - n(n+m+1)/2}{\sqrt{nm(n+m+1)/12}}$$

Has approximately a standard normal pdf if $n > 10$ and $m > 10$

In the presence of ties, the variance of W is reduced and a different formula should be used (shown at right in a table from **Hollander & Wolfe 1999**).

If there are no ties, there are tables of the exact p values for the Wilcoxon rank sum test. For large sample sizes, a normal approximation is used. Matlab has an exact test for the rank sum test appropriate even with ties. In essence, this program analyzes every combination of the $n+m$ samples divided into two groups of size n and m . This program can take an exceptionally long time if the number of permutations exceeds a million or so. A million permutations might take a brief lunch break to finish depending on the speed of your computer and how fast you eat.

Ties	
If there are ties, give tied observations the average of the ranks for which those observations are competing. After computing W using average ranks, use procedures (4.4), (4.5) or (4.6) and refer the value of W to Table A.6. Now, however, the test is approximate rather than exact. (To get an exact test, even in the tied case, see Comment 5.)	
When applying the large-sample approximation, the following modification should be made. When there are ties, the null mean of W is unaffected, but the null variance is reduced to	
$\text{var}_0(W) = \frac{mn}{12} \left[m + n + 1 - \frac{\sum_{j=1}^g (t_j - 1)t_j(t_j + 1)}{(m + n)(m + n - 1)} \right], \quad (4.13)$	
or, equivalently,	
$\text{var}_0(W) = \frac{mn(N + 1)}{12} - \left\{ \frac{mn}{12N(N - 1)} \cdot \sum_{j=1}^g (t_j - 1)t_j(t_j + 1) \right\}. \quad (4.14)$	

Wilcoxon submitted his paper describing his new statistic fully expecting to be told in review that statisticians had already described the test. They hadn't. Mann & Whitney had been working on a test based on ranks and published their test. The Mann-Whitney U test calculates its test statistic with the following algorithm (**Hollander & Wolfe 1999**):

- " For the two groups X_i and Y_j with m & n cases, consider each of the $m \times n$ pairs
- " For each pair of values X_i and Y_j , observe which is smaller.
- " If the X_i value is smaller, score a 1 for that pair. If the Y_j value is smaller, score a 0 for that pair.
- " Mann & Whitney showed that in the case of no ties:
 - $T = U + [n(n+1)/2]$, where T is the sum of ranks from the Wilcoxon rank sum test; n is size of smaller group
 - $U = T - n*(n+1)/2$
 - Thus, the Wilcoxon & MW-U tests are exactly equivalent
- " When, X_i and Y_j are tied, score $\frac{1}{2}$

The Wilcoxon's test and the Mann-Whitney U test have different methods for analyzing samples, but they produce identical p values and the Mann-Whitney and Wilcoxon rank sum statistics can be quickly converted, so only one set of tables of p values is required.

The Wilcoxon rank sum test should be used if the data are just ordinal in scale or in the presence of a few outliers. The power efficiency of the rank sum test is generally quite high.

The Behrens-Fisher problem revisited for Wilcoxon's rank-sum test

The Wilcoxon rank sum test is distribution-free, not assumption free. **Hollander & Wolfe (1999, p. 120)** list the assumptions for Wilcoxon's rank-sum test, noting that "The significance level of the rank sum test is not preserved if the two populations differ in dispersion or shape. This is also the case for the normal theory two-sample t test." Thus, the equivalent of the equal-variance assumption applies to Wilcoxon's rank-sum test. The test also assumes independence among the X 's and Y 's and that the X 's are independent of the Y 's. This is similar to the

assumption of independently distributed errors for the Student's t test, derived under the assumption of randomly distributed errors.

Many people opt for the Wilcoxon rank sum test to avoid the problem of unequal variances among groups. **Zar (1999, p. 149)** notes that the Mann-Whitney U test assumes 'equal dispersion' along the measurement scale, but discounts its effects on the p -values, "Strictly speaking, Mann-Whitney statistics are affected by both differences between the two populations along the measurement scales [differences in median or central tendency] and differences between the dispersion and shapes of the two populations (e.g., Boneau, 1962), but the latter effect is generally small compared to the former." **Underwood (1997, p. 131)** describes the assumptions underlying the Mann-Whitney (=Wilcoxon rank sum test):

"This test makes no assumptions about normality of the distributions being sampled ... The crucial assumptions are that that data are sampled independently within and between populations, which is identical to the assumption of the t-test. Conover (1980, p. 222) described the other assumptions and pointed out that the probabilities associated with the test are based on the underlying theory that only if the samples are identically distributed will every arrangement of the data be equiprobable. Thus, the distributions being sampled must be identical except for their mean. In other words, the variances, skewnesses, etc., of the two populations being sampled must be the same. This is at least as restrictive as the assumption that the two populations are normally distributed. Even if it were only considered to be important for the variances of the two populations, this assumption still requires the two variances are equal ...

There are numerous examples in the ecological literature where someone has done analyses using the Mann-Whitney procedure, because the data did not have equal variances, so a test based on the t-test could not be used. This is obviously silly. If one test is not valid because its assumptions are not met, so is another!"

Testing $\sigma^2_X = \sigma^2_Y$ — The F test

Shortly, we'll be covering ANOVA in which differences in means and differences in variances are tested with F statistics. Case Study 9.3.1 applies the F test, which is available in simple form in Matlab's `vartest2.m`. This test is applied in Case Study 9.3.1 below.

The two-sample binomial test

Larsen & Marx Theorem 9.4.1 provides the large sample approximation for testing the difference in two proportions. This formula should not be used for small sample sizes. For small sample sizes, Fisher's exact hypergeometric probability test should be used. Actually, with Matlab, there is no restriction to using Fisher's exact test even with large sample sizes.

Theorem 9.4.1. Let x and y denote the numbers of successes observed in two independent sets of n and m Bernoulli trials, respectively, where p_X and p_Y are the true success probabilities associated with each set of trials. Let $p_e = \frac{x+y}{n+m}$ and define

$$z = \frac{\frac{x}{n} - \frac{y}{m}}{\sqrt{\frac{p_e(1-p_e)}{n} + \frac{p_e(1-p_e)}{m}}}$$

a. To test $H_0: p_X = p_Y$ versus $H_1: p_X > p_Y$ at the α level of significance, reject H_0 if $z \geq z_\alpha$.
 b. To test $H_0: p_X = p_Y$ versus $H_1: p_X < p_Y$ at the α level of significance, reject H_0 if $z \leq -z_\alpha$.
 c. To test $H_0: p_X = p_Y$ versus $H_1: p_X \neq p_Y$ at the α level of significance, reject H_0 if z is either (1) $\leq -z_{\alpha/2}$ or (2) $\geq z_{\alpha/2}$.

The two-sample binomial is still useful for providing 95% confidence intervals for differences in proportions. Larsen & Marx (2006) Theorem 9.5.3. (right)

Theorem 9.5.3. Let x and y denote the numbers of successes observed in two independent sets of n and m Bernoulli trials, respectively. If p_X and p_Y denote the true success probabilities, an approximate $100(1 - \alpha)\%$ confidence interval for $p_X - p_Y$ is given by

$$\left(\frac{x}{n} - \frac{y}{m} - z_{\alpha/2} \sqrt{\frac{\left(\frac{x}{n}\right)\left(1 - \frac{x}{n}\right)}{n} + \frac{\left(\frac{y}{m}\right)\left(1 - \frac{y}{m}\right)}{m}}, \frac{x}{n} - \frac{y}{m} + z_{\alpha/2} \sqrt{\frac{\left(\frac{x}{n}\right)\left(1 - \frac{x}{n}\right)}{n} + \frac{\left(\frac{y}{m}\right)\left(1 - \frac{y}{m}\right)}{m}} \right)$$

The two-sample binomial test is not available in Matlab, but I have programmed it as binom2sample.m. I apply it in Case Study 9.4.1 and 9.4.2 below.

Case Studies

Case Study 9.2.1

Introduction

A question of authorship has arisen over a series of ten essays written by "Quintus Curtius Snodgrass" in the *New Orleans Daily Crescent*. Some critics believe that Mark Twain was the author, so the proportion of three-letter words in the ten Snodgrass essays was compared with that proportion in eight Twain Essays written at about the same time.

TABLE 9.2.1: Proportion of Three-Letter Words

Twain	Proportion	QCS	Proportion
Sergeant Fathom letter	0.225	Letter I	0.209
Madame Caprell letter	0.262	Letter II	0.205
Mark Twain letters in		Letter III	0.196
<i>Territorial Enterprise</i>		Letter IV	0.210
First letter	0.217	Letter V	0.202
Second letter	0.240	Letter VI	0.207
Third letter	0.230	Letter VII	0.224
Fourth letter	0.229	Letter VIII	0.223
First <i>Innocents Abroad</i> letter		Letter IX	0.220
First half	0.235	Letter X	0.201
Second half	0.217		

Table 9.2.1 shows the proportions of three-letter words found in eight Twain essays and in the ten Snodgrass essays. (Each of the Twain works was written at approximately the same time the Snodgrass essays appeared.)
 If $x_1 = 0.225, x_2 = 0.262, \dots, x_8 = 0.217$, and $y_1 = 0.209, y_2 = 0.205, \dots, y_{10} = 0.201$, then

$$\bar{x} = \frac{1.855}{8} = 0.2319 \quad \text{and} \quad \bar{y} = \frac{2.097}{10} = 0.2097$$

Does an independent samples t test provide evidence to reject the null hypotheses that the proportions are the same? The first step in any independent samples t test is to identify the alternate hypothesis, at least so far as whether it is one-tailed or two-tailed. Since we have no advanced or a priori knowledge about the the 3-letter word usage of either author, we'll perform our tests two-tailed.

I am absolutely confident that analysis by independent samples t test is appropriate. Moreover, it would be inappropriate to analyze the same data with multiple tests if the goal were to examine the p values to choose which test should be used. Purely for pedagogical reasons, I will perform the Welch's t test and Wilcoxon rank sum test on these data.

Results

First, we must analyze the equal spread assumption. Are the variances of the two groups about equal? The Tukey boxplot, shown in Figure 2, reveals about equal spread in the two groups. The t test is relatively robust to minor violations in the equal spread assumption if the group sizes are about equal but can produce wildly inflated Probability of Type I error, rejecting true null hypotheses more often than the nominal Type I level, if the smaller group has a larger variance.

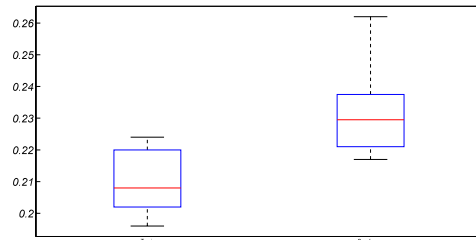


Figure 1. Boxplots for the Case Study 9.2.1 data with Twain left and Snodgrass right. There is no indication of unequal spread.

I analyzed the data with Matlab's `ttest2` with equal variance. The observed t statistic was 3.88, and the two-tailed probability of observing a t statistic of 3.88 with 16 df is 0.0013. The difference in means was 0.022 ± 0.012 (\pm half 95% confidence interval). The pooled sd for the difference was 0.012.

In Figure 9.2.1, shown as Figure 2, Larsen & Marx present the critical value method for evaluating the null hypothesis. Rather than looking up or calculating the p value for the observed statistic, 3.88 in this case with 16 df, one finds in advance the critical value at which one rejects the null hypothesis. In this case, Larsen & Marx chose an alpha level, which is the Probability of Type I error of 0.01. The value shown in their Figure, 2.9208, is calculated in Matlab using the inverse t distribution: $\text{tinv}(1-\alpha/2,16)$. Current practice in the age of fast computers and the internet is to report the actual p value rather than using the dichotomous reject, fail to reject decision rule.

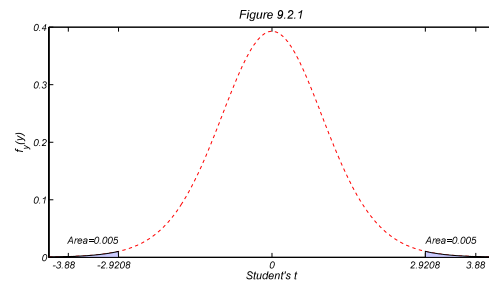


Figure 2. Student's t distribution for 16 df showing the two tailed critical values for $\alpha=0.01$. Since 3.88 is greater than 2.9208, reject the null hypothesis at the $\alpha=0.01$ level.

For pedagogical purposes, I analyzed the data with Matlab's `ttest2` with the unequal variance option. This performs the Welch's t test. The observed t statistic was 3.7, with the Satherthwaite approximation for the df of 11.7 (reduced from 16 because of the unequal variance). The two-tailed probability of observing a Welch's t statistic of 3.7 with 11.7 df is 0.0032. The difference in means was 0.022 ± 0.013 (\pm half 95% confidence interval). The separate standard deviations were 0.015 and 0.010.

Case Study 9.2.2

Introduction to 9.2.2

This case study involves two separate t tests. An instructor wanted to determine if acting more enthusiastic in class improved teacher ratings in another category, knowledge of the subject. So, using an unreplicated design, he taught a spring and fall class. He tried to deliver the same material in both classes. He selected the spring class to have the same

demographics as the fall class. [Just a note here. An experiment without replication is not an experiment at all. Don't follow this professor's methods].

In this problem, we are not given any of the raw data, only summaries. That is sufficient to perform the t tests, but not with the standard software. If this were an IBM PASW course, I'd have to write an m.file to solve the problem in syntax, but there already is such a file at Raynald Levesque's SPSS tools website. Matlab doesn't have a program for performing t tests with aggregated data such as that in Case Study 9.2.2, but I wrote one called student2group.m which solves the problem.

As usual in performing any statistical test, we must evaluate the assumptions of the test and set the a priori hypotheses. In this case, the a priori null and alternate hypotheses are that the professor's enthusiasm ratings remained unchanged vs. the 1-tailed alternative that students perceived his increased enthusiasm. For the second test, the null hypothesis is that his perceived knowledge remained the same vs. the alternative hypothesis that students in the enthusiastic class perceived him to be more knowledgeable.

Results for 9.2.2

The sample sizes were large and approximately equal and the standard deviation were approximately equal. There would be no reason to question the appropriateness of an independent samples t test.

Fall, x_i	Spring, y_i
$n = 229$	$m = 243$
$\bar{x} = 2.14$	$\bar{y} = 4.21$
$s_X = 0.94$	$s_Y = 0.83$

Let μ_X and μ_Y denote the true means associated with the two different teaching styles. There is no reason to think that increased enthusiasm on the part of the instructor would *decrease* the students' perception of enthusiasm, so it can be argued here that H_1 should be one sided. That is, we want to test

$H_0: \mu_X = \mu_Y$
 versus
 $H_1: \mu_X < \mu_Y$

Let $\alpha = 0.05$.

Fall, x_i	Spring, y_i
$n = 229$	$m = 243$
$\bar{x} = 3.61$	$\bar{y} = 4.05$
$s_X = 0.84$	$s_Y = 0.95$

Using my student2group.m, I first analyzed the enthusiasm hypothesis. The observed t statistic was -25.4 , and the one-tailed probability of observing a t statistic this low or lower with 470 df is 1.7×10^{-90} . I would never report a difference this small. Conventionally, unless the test is exact, one would report the p value as $< 10^{-6}$. The pooled standard deviation for the differences was 0.899 . The difference in means (Fall - Spring) was -2.07 ± 0.16 (\pm half 95% confidence interval), indicating greater perceived enthusiasm.

I then analyzed the knowledgeability hypothesis. The observed t statistic was -5.3 , and the one-tailed probability of observing a t statistic this low or lower with 470 df is 8.1×10^{-8} , which I would report as $p < 10^{-6}$. The pooled standard deviation for the differences was 0.898 . The difference in means (Fall - Spring) was -0.44 ± 0.16 (\pm half 95% confidence interval), indicating greater perceived knowledge in the spring class.

Case Study 9.3.1

Inmates were randomly assigned to solitary confinement or allowed to remain in their cells. After 7 days, their alpha waves were measured.

The distributions of these data are shown in Figure 3.

The variances of solitary and nonconfined groups were 0.357 and 0.211 , respectively with a ratio of variances of 1.70 . The probability of observing this ratio by chance equals the probability of observing an F statistic larger than 1.70 or smaller than 0.59 (i.e., $1/1.7$) with 9 and 9 df. That two-tailed probability is 0.44 . The 95% confidence interval for the ratio of 1.70 is 0.42 and 6.83 , which clearly includes the null hypothesis's expected ratio of 1.0 .

TABLE 9.3.1: Alpha-Wave Frequencies (CPS)

Nonconfined, x_i	Solitary Confinement, y_i
10.7	9.6
10.7	10.4
10.4	9.7
10.9	10.3
10.5	9.2
10.3	9.3
9.6	9.9
11.1	9.5
11.2	9.0
10.4	10.9

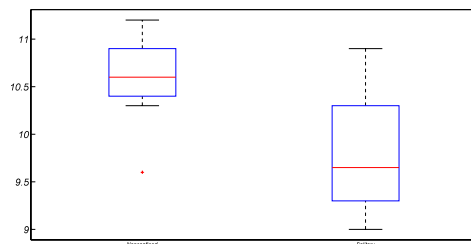


Figure 3. Boxplots for the nonconfined (left) and solitary confinement (right) alpha waves.

Figure 4 shows the F distribution with 9 and 9 df. The observed F ratio would have to exceed 4.026 in order to reject the null hypothesis with a two-tailed P value of 0.05.

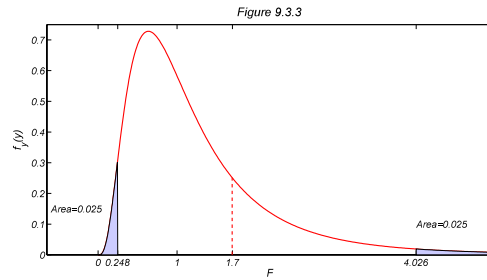


Figure 4. The F distribution with 9 and 9 df. In order to reject the null hypothesis of equal variances, an F ratio of 0.25 or 4.03 would have to be observed. The observed F ratio of 1.7 is well within the two $\alpha = 0.05$ critical values.

Case Study 9.4.1

The mitigation rate is the proportion of criminal cases in which the defendant qualifies for prison time but receives a greatly shortened term or no prison time at all. In Escambia County the mitigation rate was reduced from 1033 out of 1675 cases (61.7%) to 334 out of 660 cases (52.1%). Could this difference be attributed to chance?

The difference of 9.55% has a 95% confidence interval of 5.1% to 14.0%, which clearly doesn't include 0 so the the null hypothesis can be rejected at the $\alpha=0.05$ level. The actual p value for observing a z statistic of 4.22 is 2.4×10^{-5} . So, there is very strong evidence to reject the null hypothesis of the difference in mitigation rate being due to chance.

Case Study 9.4.2

The proportion of men and women reporting nightmares often or seldom is shown to the right. Using the two sample binomial test, could these differences in proportions (34.4% vs. 31.3%) be attributed to chance?

	Men	Women	Total
Nightmares often	55	60	115
Nightmares seldom	105	132	237
Totals	160	192	
% often:	34.4	31.3	

The difference in proportions was a mere 3.13%. If the null hypothesis were true, the two sample binomial tests whether these two proportions could have been drawn from a population with an observed proportion of 32.67%. The large sample approximation uses a z statistic of 0.6225 to calculate the probability of observing a difference this great or greater to be 0.5336. The difference has a 3.13% had a 95% confidence limit of -6.7 to 13.0% which clearly includes the null hypothesis's expected value of 0.

Case Study 9.5.1

The enamel spectropenetration gradient differs between men and women (see right). Test for this difference and find the 95% confidence limit for the difference.

Male, x_i	Female, y_i
4.9	4.8
5.4	5.3
5.0	3.7
5.5	4.1
5.4	5.6
6.6	4.0
6.3	3.6
4.3	5.0

The boxplots for males and females shown in Figure 5 indicates about equal spread, so the independent samples t test is appropriate.

The difference in gradient between men and women was 0.91 ± 0.80 (\pm half 95% CI). The two-tailed p for observing $t=2.43$ with 14 df is 0.030. There is strong evidence to reject the null hypothesis of equal gradients between men and women.

Case Study 9.5.2

The goal of this case study is to construct a 95% confidence interval for the ratio of variances for glacial flow rates with photographs taken every 3 years vs. every 5 years.

The unusual boxplot for these data is shown in Figure 6. The variances were 0.00022 and 0.00012 with ratio 1.87. The 95% CI for the observed ratio is [0.203 11.614], which includes 1. The probability of observing an $F_{6,4}$ statistic ≥ 1.87 under the null hypothesis of equal variances is 0.57. Therefore, I conclude that there is little evidence to reject the hypothesis that glacial flow rate variance in 3-year and 5-year photographs are the same.

Case Study 9.5.3

Lister performed amputations with and without carbolic acid disinfectant. In the 40 with carbolic acid 34 patients survived; in the 35 without carbolic acid, 19 patients survived. Could these differences be due to chance? Calculate a 95% confidence interval for the difference using the two-sample binomial.

Under the null hypothesis, the two sample binomial estimates the probability of survival as 71%, with 85% of the patients surviving with carbolic acid and 54% without. The large sample z statistic for the two-sample binomial test is 2.91 with two-tailed p value of 0.0036. The difference in proportions with 95% confidence interval is 30.7 ± 19.9 .

I've also used the user-contributed m.file fishertest.m to calculate the exact probability of observing the results shown in Table 9.5.3 or results more extreme. Fisher's test is based on the

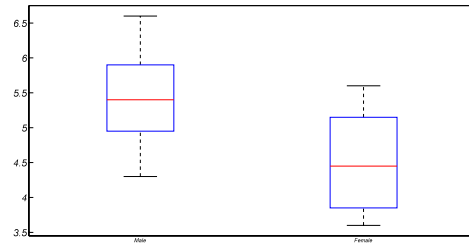


Figure 5. Boxplots for enamel spectropenetration gradients for males (left) and females (right).

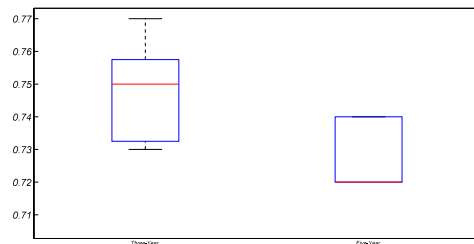


Figure 6. Boxplots for glacial flow rates for three-year photographs (left) and five-year photographs (right).

		<i>Carbolic acid used?</i>		
		No	Yes	Total
Patient lived?	Yes	19	34	53
	No	16	6	22
Total		35	40	

hypergeometric probability distribution and is the appropriate test if sample sizes are too small to warrant the use of the large-sample z approximation. In this case, the exact p value is 0.0037.

Case Study 14.3.4 Baseball Game Length

The National league doesn't have designated hitters but the American League does. Are the lengths of the games different in the National and American league? Shown at right are the average game completion times for the twenty six major league teams in 1992.

These data were tested with a Wilcoxon rank sum test. First, the boxplot shown in Figure 7 didn't reveal any issues with unequal spread.

The Wilcoxon rank sum statistic was 110.5. The large sample p value for this test statistic was $p=0.0084$, and the exact p value was $p=0.0065$. Note that the text lists the value of the test statistic as 240.5. This will produce an identical p value as 110.5. The sum of ranks of n and m groups is $n(n+m+1)$ or $13*27$ for these 26 major league teams. $240.5 = 13*27 - 110.5$. Larsen & Marx just summed the American League teams, whereas, the Matlab ranksum.m program summed the National league teams. Since it is the deviation from the average rank of $n(n+m+1)/2$ that serves as the numerator of the test statistic, both test statistics with identical deviations from the average will produce the same p value.

TABLE 14.3.7

Obs. #	Team	Time (min)	r_i	z_i	$r_i z_i$
1	Baltimore	177	21	1	21
2	Boston	177	21	1	21
3	California	165	7.5	1	7.5
4	Chicago (AL)	172	14.5	1	14.5
5	Cleveland	172	14.5	1	14.5
6	Detroit	179	24.5	1	24.5
7	Kansas City	163	5	1	5
8	Milwaukee	175	18	1	18
9	Minnesota	166	9.5	1	9.5
10	New York (AL)	182	26	1	26
11	Oakland	177	21	1	21
12	Seattle	168	12.5	1	12.5
13	Texas	179	24.5	1	24.5
14	Toronto	177	21	1	21
15	Atlanta	166	9.5	0	0
16	Chicago (NL)	154	1	0	0
17	Cincinnati	159	2	0	0
18	Houston	168	12.5	0	0
19	Los Angeles	174	16.5	0	0
20	Montreal	174	16.5	0	0
21	New York (NY)	177	21	0	0
22	Philadelphia	167	11	0	0
23	Pittsburg	165	7.5	0	0
24	San Diego	161	3.5	0	0
25	San Francisco	164	6	0	0
26	St. Louis	161	3.5	0	0

$w' = 240.5$

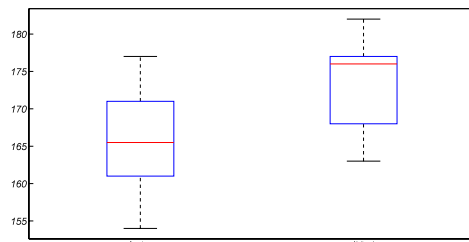


Figure 7. Boxplots for average game length for American (left) and National (right) league teams.

Chapter Outlines

Annotated outline (with Matlab scripts) for Larsen & Marx Chapter 9

9 Two-sample problems (Summer 2011 **Week 8** & **Week 9**)[p 553]

William Sealy Gosset (“Student”) (1876-1937)

9.1 INTRODUCTION

9.2 Testing $H_0 : \mu_x = \mu_y$ — The two sample t test

Theorem 9.2.1

Theorem 9.2.2

Case Study 9.2.1 Disputed Authorship

% LMcs090201_4th.m

% Case Study 9.2.1 p. 557 in

% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition

% Examples of two sample tests

% Written by Eugene.Gallagher@umb.edu in 2001, revised 1/21/11, 2/27/11

% <http://alpha.es.umb.edu/faculty/edg/files/edgwebp.html>

Twain=[.225 .262 .217 .240 .230 .229 .235 .217]';

Snodgrass=[.209 .205 .196 .210 .202 .207 .224 .223 .220 .201]';

% boxplot

[T,C]=size(Twain);[S,CS]=size(Snodgrass);

DATA=[Twain;Snodgrass];

G=[ones(T,1);zeros(S,1)];

boxplot(DATA,G,'labels',{'Twain','Snodgrass'})

figure(gcf),pause

fprintf('The difference in means (Twain - Snodgrass) was %6.4f\n',...

mean(Twain)-mean(Snodgrass));

[H,P,CI,STATS] = ttest2(Twain,Snodgrass,0.05,'both','equal');

fprintf(...

\nThe 2-tailed p for Student's t (= %4.2f, %d df) = %6.4f\n',...

STATS.tstat,STATS.df,P);

fprintf('The pooled sd for the difference is %6.4f\n',STATS.sd)

fprintf(...

'The lower and upper CI for the observed difference = %6.3f: [%6.3f %6.3f]\n',...

mean(Twain)-mean(Snodgrass),CI(1),CI(2));

fprintf('\n\nThe unequal variance, or Welch t test\n');

[H,P,CI,STATS] = ttest2(Twain,Snodgrass,0.05,'both','unequal');

fprintf(...

\n\nThe 2-tailed p for Welch's t (= %4.2f, %5.2f df) = %6.4f\n',...

STATS.tstat,STATS.df,P);

fprintf('The separate sd's for the samples were %6.4f and %6.4f\n',STATS.sd)

fprintf(...

'The lower and upper CI for the observed difference = %6.3f: [%6.3f %6.3f]\n',...

mean(Twain)-mean(Snodgrass),CI(1),CI(2));

```
% Plot Figure 9..2.1 using Figure 4.3.6 (LMex040307_4th.m) as a model.
df=16;
X=-4.25:0.01:4.25;
Y = tpdf(X,df);
plot(X,Y,'-r');
axis([-4.25 4.25 0 0.4]);title('Figure 9.2.1','FontSize',22);
ax1=gca;
xlabel('Student's t','FontSize',20),
ylabel('f_y(y)','FontSize',20);
ax1=gca;
set(ax1,'xtick',[-3.88 -2.9208 0 2.9208 3.88],'FontSize',18);
set(ax1,'ytick',0:.1:0.4,'FontSize',18)
hold on;
xfu=2.9208:0.01:4.25;yfu=tpdf(xfu,df);
fill([2.9208 xfu 4.25],[0 yfu 0],[.8 .8 1])
xfl=-4.25:0.005:-2.9208;yfl=tpdf(xfl,df);
fill([-4.25 xfl -2.9208],[0 yfl 0],[.8 .8 1])
text(-3.9,0.03, 'Area=0.005','FontSize',18);
text(2.9208,0.03,'Area=0.005','FontSize',18)
figure(gcf);pause
hold off;
```

Case Study 9.2.2 Teacher evaluations using only aggregate data

```
% LMcs090202_4th.m
% Table 9.2.2 & Table 9.2.3 data
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% Page 560-562. Case Study 9.2.2, 2-sample t-test using grouped data
% Written by Eugene.Gallagher@umb.edu
% http://alpha.es.umb.edu/faculty/edg/files/edgwebp.html
% Written & last revised 11/17/2010
Xn=229;Xmean=2.14;Xstd=0.94;Yn=243;Ymean=4.21;Ystd=0.83;alpha=0.05;
[D,t,df,pvalue,CI,sp]=student2group(Xn,Yn,Xmean,Ymean,Xstd,Ystd,alpha);
fprintf('The Student's t statistic was %6.4f with %3.0f df.\n',t,df)
fprintf('The pooled standard deviation was %6.4f\n',sp)
fprintf('The 1-tailed p for enthusiasm =%6.2g\n',pvalue/2);
fprintf('The difference is %4.3f with 95%% CI = [%5.3f %5.3f]\n',D, ...
    CI(1),CI(2));
% Table 9.2.3 data
Xn=229;Xmean=3.61;Xstd=0.84;Yn=243;Ymean=4.05;Ystd=0.95;alpha=0.05;
[D,t,df,pvalue,CI,sp]=student2group(Xn,Yn,Xmean,Ymean,Xstd,Ystd,alpha);
fprintf('The Student's t statistic was %6.4f with %3.0f df.\n',t,df)
fprintf('The pooled standard deviation was %6.4f\n',sp)
fprintf('The 1-tailed p for knowledgeable =%6.2g\n',pvalue/2);
fprintf('The difference is %4.3f with 95%% CI = [%5.3f %5.3f]\n',D, ...
    CI(1),CI(2));
```

```

function [D,t,df,pvalue,CI,sp]=student2group(Xn,Yn,Xmean,Ymean,Xstd,Ystd,alpha)
function [D,t,df,pvalue,CI,sp]=student2group(Xn,Yn,Xmean,Ymean,Xstd,Ystd,alpha)
% Student't 2-sample equal variance t test with grouped data
% [D,t,df,pvalue,CI,sp]=student2group(Xn,Yn,Xmean,Ymean,Xstd,Ystd,alpha)
% Input: Xn, Yn = Size of X and Y groups
%     Xmean, Ymean = means for 2 groups
%     Xstd, Ystd = standard deviations for 2 groups
%     alpha level for CI, optional, 95% if not specified
% Output: D=Xmean-Ymean;
%     t=Student's t statistic
%     df=degrees of freedom for t statistic
%     p value, 2-sided, for t Statistic, with df degrees of freedom
%     CI for 1-alpha CI, [L U];95% CI if alpha not specified.
%     sp pooled standard deviation for difference
% Based on Theorems 9.2.1 & 9.2.2, page 555 & 557 Larsen & Marx (2006)
% Introduction to Mathematical Statistics, 4th edition.
% uses Student's t from Statistics toolbox
% Written by Eugene.Gallagher@umb.edu, Revised 11/17/10, 2/27/11 (added
% sp to output)
% see also stud1sample, stud2sample,
n=Xn;
m=Yn;
D=Xmean-Ymean;
% Estimated pooled standard deviation (Theorem 9.2.1)
sp=sqrt( ((n-1)*Xstd^2+(m-1)*Ystd^2)...
        / (n+m-2) );
t=D/(sp*sqrt(1/n+1/m));
% Use Matlab's tcdf for p values
df=n+m-2;
if t>=0
    pvalue = 2*(1-tcdf(t,df));
else
    pvalue=2*tcdf(t,df);
end
if nargin<3;
    alpha=0.05;
end
% Use tinv.m from statistics toolbox
HalfCI=tinv(1-alpha/2,df)*sp*sqrt(1/n+1/m);
CI=[D-HalfCI D+HalfCI];

```

Questions p. 563-568

9.3 TESTING $H_0: \sigma_x^2 = \sigma_y^2$ — THE F TEST

Theorem 9.3.1

Case Study 9.3.1

% LMcs090301_4th.m

```
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th ed
% Case Study 9.3.1, an example of the 2-sample Variance test
% Written by Eugene.Gallagher@umb.edu
%
alpha=0.05;
Nonconfined=[10.7 10.7 10.4 10.9 10.5 10.3 9.6 11.1 11.2 10.4];
Solitary=[9.6 10.4 9.7 10.3 9.2 9.3 9.9 9.5 9.0 10.9];
fprintf(...
  'The variance of solitary and nonconfined groups were %6.3f and %6.3f\n',...
    var(Solitary),var(Nonconfined));
fprintf('The ratio of variances was %6.4f\n',var(Solitary)/var(Nonconfined))
boxplot([Solitary;Nonconfined],[ones(10,1);zeros(10,1)],...
  'labels',{'Nonconfined','Solitary'});figure(gcf);pause
[H,P,CI,STATS] = vartest2(Solitary,Nonconfined,alpha,'both');
df1=STATS.df1;df2=STATS.df2;
fprintf('The F statistic with %2.0f, %2.0f df is %6.4f\n',...
  STATS.df1,STATS.df2,STATS.fstat)
fprintf('The p value is %6.4f with CI: %6.4f %6.4f.\n',P,CI)
% Plot Figure 9.3.3 using Figures 4.3.6 (LMex040307_4th.m) & Figure 9.2.1
% as model.
X=0:0.01:5;
Y = fpdf(X,df1,df2);
Ytest=fpdf(STATS.fstat,df1,df2);
plot(X,Y,'r','LineWidth',2);
axis([-0.65 5 0 0.75]);
title('Figure 9.3.3','FontSize',22);
ax1=gca;
xlabel('F','FontSize',20),
ylabel('f_y(y)','FontSize',20);
ax1=gca;
alpha=0.05;Fcritl=finv(alpha/2,df1,df2);Fcritu=finv(1-alpha/2,df1,df2);
set(ax1,'xtick',[0 0.248 1 1.70 Fcritu],'FontSize',18);
set(ax1,'ytick',0:.1:0.7,'FontSize',18)
hold on;
plot([STATS.fstat STATS.fstat],[0 Ytest]','-r','Linewidth',2)
xfu=Fcritu:0.01:5;yfu=fpdf(xfu,df1,df2);
fill([Fcritu xfu 5],[0 yfu 0],[.8 .8 1])
xfl=0:0.005:Fcritl;yfl=fpdf(xfl,df1,df2);
fill([0 xfl Fcritl],[0 yfl 0],[.8 .8 1])

text(-.6, 0.15,'Area=0.025','FontSize',18);
text(Fcritu,0.1,'Area=0.025','FontSize',18)
figure(gcf);pause
hold off;
```

9.4 Binomial data: testing $H_0 : p_x = p_y$

Case Study 9.4.1

9.4.1 Applying the generalized likelihood ratio criterion

Theorem 9.4.1

Case Study 9.4.1

```
% LMcs090401_4th.m
% Application of 2-sample binomial test, Theorem 9.4.1 in
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% Written by Eugene.Gallagher@umb.edu
% Calls Gallagher's binom2sample.m
% Written Nov. 2010, revised 12/12/2010, 2/27/2011
[D,phat,z,pvalue,CI,obsp]=binom2sample(1033,1675,344,660,0.05);
fprintf('The expected joint probability is %6.4f\n',phat)
fprintf('The z statistic is %6.4f with two-tailed p = %6.4g\n',z,pvalue)
fprintf('The observed proportions were %6.4f and %6.4f\n',obsp)
fprintf(...)
The difference in proportions is %6.4f with confidence interval: %6.4f %6.4f\n',D,CI)
function [D,phat,z,pvalue,CI,obsp]=binom2sample(x,n,y,m,alpha)
% 2-sample binomial test.
% [D,phat,z,pvalue,CI]=binom2sample(x,n,y,m,alpha)
% Input: Sample 1: x successes in n trials
%       Sample 2: y successes in m trials
%       alpha level for CI, optional, 95% if not specified
% Output: D=x/n-y/m;
%         phat= expected p value under the no-difference null hypothesis
%         z=test statistic
%         p value, 2-sided, for z statistic
%         CI (1-alpha) percentile for D;95% CI if alpha not specified.
%         obsp=obsrved proportions = [x/n y/m]
% Based on Theorem 9.4.1, page 506 Larsen & Marx (2001, 3rd edition) and
% Page 578 in Larsen & Marx (2006, 4th edition) Introduction to
% Mathematical Statistics and for CI's Theorem 9.5.3 (p. 514) in Larsen &
% Marx (2001, 3rd edition), Page 587 in Larsen & Marx (2006, 4th edition)
% Written by E. Gallagher; revised 12/19/10
if x>n | y>m
    error('Successes can't exceed number of trials')
end
D=x/n-y/m;
phat=(x+y)/(n+m);
z=D/sqrt(phat*(1-phat)/n+phat*(1-phat)/m);
% Use Matlab stats toolbox normcdf for significance of z,
% the standard normal variate
% could use statbox normp.m or Gallagher's zprob.m or stibox pnorm.
if z>=0
    pvalue=2*(1-normcdf(z));
```

```

else
    pvalue=2*normcdf(z);
end
if nargin<5;
    alpha=0.05;
end
% Use Matlab's norminv.
% Theorem 9.5.3 (p. 514);
% n.b., the CI uses a different estimator for the standard deviation
% to find CI's for the difference D
HalfCI=norminv(1-alpha/2)*sqrt(((x/n)*(1-x/n)/n) +((y/m)*(1-y/m)/m));
CI=[D-HalfCI D+HalfCI];
obsp=[x/n y/m];

```

Case Study 9.4.2

```

% LMcs090402_4th.m
% Written by Eugene.Gallagher@umb.edu
% Calls Gallagher's binom2sample.m
% Written October 2010, revised 2/27/11
[D,phat,z,pvalue,CI,obsp]=binom2sample(55,160,60,192,0.05);
fprintf('The expected joint probability is %6.4f\n',phat)
fprintf('The z statistic is %6.4f with two-tailed p = %6.4f\n',z,pvalue)
fprintf('The observed proportions were %6.4f and %6.4f\n',obsp)
fprintf(...
'The difference in proportions is %6.4f with confidence interval: %6.4f %6.4f\n',D,CI)

```

Questions p. 580-582

9.5 CONFIDENCE INTERVALS FOR THE TWO-SAMPLE PROBLEM

Theorem 9.5.1

Case Study 9.5.1

```

% LMcs090501_4th.m
% Larsen & Marx (2006,p. 583) Introduction to Mathematical Statistics, 4th
% Edition. Case Study 9.5.1 An example of CI's for a two-sample proportion
% Examples of two sample tests
% Written by Eugene.Gallagher@umb.edu 11/14/10, revised 11/17/10
Male =[4.9 5.4 5.0 5.5 5.4 6.6 6.3 4.3]';
Female=[4.8 5.3 3.7 4.1 5.6 4.0 3.6 5.0]';
% boxplot
[M,CM]=size(Male);[F,CF]=size(Female);
DATA=[Male;Female];
G=[ones(M,1);zeros(F,1)];
boxplot(DATA,G)
[H,P,CI,STATS] = ttest2(Male,Female,0.05,'both','equal');
fprintf(...
\nThe 2-tailed p for Student's t (= %4.2f, %d df) = %6.4f\n',...
STATS.tstat,STATS.df,P);
fprintf(...

```

```
'The lower and upper CI for the observed difference=%6.3f: [%6.3f %6.3f]\n',...
  mean(Male)-mean(Female),CI(1),CI(2));
% Analyze with Wilcoxon ranks sum statistic, the major non-parametric alternative
% Student's t test. First with the normal approximation:
[Wilcoxpvvalue,W,U]=Wilcoxranksum(Male,Female);
fprintf(...
  '\nThe 2-tailed p for Wilcoxon"s rank sum with large sample normal approx=%6.4f\n',...
  Wilcoxpvvalue);
% Do exact test with Wilcoxon ranksum
[Wilcoxexactp,W,U]=wilcoxranksum(Male,Female,1);
fprintf(...
  'The 2-tailed exact p for Wilcoxon"s rank sum =%6.4f\n',...
  Wilcoxexactp);
[D,Studp,t,df,RandPermP,CIpD,CIr]=randp2sample(Male,Female,1e4,1,0.05);
fprintf(...
  'The 2-tailed p using random permutations of t =%6.4f\n',...
  RandPermP);
```

```
function [D,Studp,t,df,RandPermP,CIpD,CIr]=randp2sample(X,Y,Trials,UseT,alpha);
% Test for difference between 2 means using random permutations
% format [D,Studp,t,df,RandPermP,CIpD,CIr]=randp2sample(X,Y,Trials,UseT,alpha);
% Example: [D,Studp,t,df,RandPermP,CIp,CIr]=randp2sample(X,Y);
% Input: X, Y column or row vectors with cases as rows, variables as columns.
% Trials, number of random permutations [Optional]
% if nargin>2 & nargin<3, Trials=1e4
% UseT, ==1, Use t statistic for random permutation tests
% ==0, Use abs(D from randperm)>abs(D) for random permutation tests
% Note, these two methods produce nearly identical results.
% Output: D Difference in means
% Studp parametric p value for difference based on Student's t, two-tailed
% T Student's t statistic, n+m-2 df
% df for 2-sample, equal mean t test, equal variance assumed
% RandPermP 2-tailed p test of D=0, based on Random permutations
% CIpD Parametric 1-alpha% CI for D based on Student's t
% CIr 1-alpha% CI for difference of X & Y,
% note that this CI is for a difference of 0. It is not the CI for
% the observed difference
% Reference: Legendre & Legendre, Manly
% Written by Eugene.Gallagher@umb.edu
if nargin>4 & nargin<3
  Trials=1e4;
  UseT=1;
  alpha=0.05;
elseif nargin<5
  alpha=0.05;
end
```



```

X=X(:);
n=length(X);
Y=Y(:);
m=length(Y);
[D,t,df,Studp,CIpD]=stud2sample(X,Y,alpha);
df=n+m-2;
XY=[X;Y];
if nargin>3
    UseT=logical(UseT); % Simple 1's and 0's are evaluated logically, but this is
                        % more formal, and perhaps faster.
else
    UseT=logical(0);
end
if nargin>3
    Tally=0; % Create a matrix of all zeros to hold results

    if nargin>6
        TallyD=zeros(Trials,1);
    end
    for i=1:Trials
        j=randperm(n+m);
        RandD=mean(XY(j(1:n)))-mean(XY(j(n+1:n+m)));
        if nargin>6
            TallyD(i)=RandD;
        end
        Sp=sqrt( ((n-1)*var(XY(j(1:n))))+(m-1)*var(XY(j(n+1:n+m))))...
            / (n+m-2);
        randt=RandD/(Sp*sqrt(1/n+1/m));
        if UseT % Use t statistic to determine p value
            if abs(randt)+eps >= abs(t)
                Tally=Tally+1;
            end
        else
            if abs(RandD)+eps>=abs(D) % Two-sided test using absolute value of D
                Tally=Tally+1;
            end
        end
    end
end
end
RandPermP=Tally/Trials;
if nargin>6
    TallyD=sort(TallyD);
    CIr=zeros(1,2);
    CIr(1)=TallyD(floor(alpha/2*Trials));
    CIr(2)=TallyD(ceil((1-alpha/2)*Trials));
end

```

end

Theorem 9.5.2

Case Study 9.5.2

```
% LMcs090502_4th.m
% Larsen & Marx (2006,p. 585) Introduction to Mathematical Statistics, 4th
% Edition. Case Study 9.5.1 An example of CI's for a ratio of variances
% Examples of two sample tests
% Written by Eugene.Gallagher@umb.edu 11/14/10, revised 11/17/10
X=[0.73 0.76 0.75 0.77 0.73 0.75 0.74]';
Y=[0.72 0.74 0.74 0.72 0.72]';
boxplot([Y;X],[ones(length(Y),1);zeros(length(X),1)]);figure(gcf)
[H,P,CI,STATS] = vartest2(X,Y,0.05,'both')
```

Theorem 9.5.3

Case Study 9.5.3

```
%LMcs090503_4th.m
% Written by Eugene.Gallagher@umb.edu
% Calls Gallagher's binom2sample.m, which
% Written November 2010, Revised 12/12/10
[D,phat,z,pvalue,CI]=binom2sample(34,40,19,35,0.05)
```

Questions p 588-591

9.6 Taking a second look at statistics (choosing samples)

Appendix 9.A.1 A derivation of the two-sample t test (A proof of Theorem 9.2.2)

Appendix 9.A.2 Minitab applications

Appendix 9.A.2 **Power calculations for a two-sample t test**

Annotated Outline (with Matlab scripts) for Larsen & Marx Chapter 14

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
%  $fy_j=1/(\sqrt{2*\pi}*\sigma_j)*\exp(-1/2*((y-\mu_j)/\sigma_j).^2)$ ;
% 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
%  $fy_j=1/(\sqrt{2*\pi}*\sigma_j)*\exp(-1/2*((y-\mu_j)/\sigma_j).^2)$ ;
% 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('\n\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
%  $fy_j=1/(\sqrt{2*\pi}*\sigma_j)*\exp(-1/2*((y-\mu_j)/\sigma_j).^2)$ ;
% 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 $H_0: \mu = \mu_0$

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/HDI','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); % sigma_j 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
%  $fy_j=1/(\sqrt{2*\pi}*\sigma_j)*\exp(-1/2*((y-\mu_j)/\sigma_j).^2)$ ;
% 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); % sigma_j 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
%  $fy_j=1/(\sqrt{2*\pi}*\sigma_j)*\exp(-1/2*((y-\mu_j)/\sigma_j).^2)$ ;
% 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 $H_0: \mu_D = 0$ (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
%  $fy_j=1/(\sqrt{2*\pi}*\sigma_j)*\exp(-1/2*((y-\mu_j)/\sigma_j).^2)$ ;
% 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 $H_0: \mu_X = \mu_Y$ (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 Wilcoxsexact
% 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 equations 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
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-tailed probability of Wstar from the standard normal distribution
  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) = 1+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

14.7 Taking a second look at statistics (comparing parametric and nonparametric
procedures

Appendix 14.A.1 Minitab applications

References

Hollander, M. and D. A. Wolfe. 1999. Nonparametric Statistical Methods, 2nd edition. John Wiley & Sons, New York. 787 p. [*An extremely valuable reference book that supplements brief discussion of distribution-free statistics in Ramsey & Schafer 2002*]{9}

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. {4, 5}

Ramsey, F. L. and D. W. Schafer. 2002. The statistical sleuth: a course in methods of data analysis, 2nd Edition. Duxbury Press, Belmont CA, 742 pp & data CD. {7, 45}

Salsburg, D. 2001. The lady tasting tea: how statistics revolutionized science in the twentieth century. W. H. Freeman & Co., New York. 340 pp. [*This is a wonderful little book, with superb New Yorker style articles on Fisher, Wilcoxon, Deming and other giants of 20th century statistics*] {7, 8}

Index

alpha level.	12, 20, 22
alternative hypothesis.	13
ANOVA.	6, 10
Behrens-Fisher.	7, 9
Boxplot.	6, 12, 16-18, 21, 23, 26, 36
Central limit theorem.	8
combinations.	8, 37-40
Confidence interval.	6, 12, 14-16, 22, 23
critical value.	12, 40
degrees of freedom.	3, 6, 7, 12, 14-16, 18-25, 29, 31, 44
Distributions	
F.	15
Efficiency.	3, 9
Expected value.	15, 37
Experiment.	7, 8, 13
Fisher.	7, 9, 45
Fisher's exact test.	11
Generalized likelihood ratio.	22
Gosset.	18
independence.	9
Kruskal-Wallis.	42
kurtosis.	3
Legendre.	24, 27, 29, 30, 32, 33, 35
level of significance.	9
Levene's test.	3, 5
likelihood.	22
Likelihood ratio.	22
Mann-Whitney U Test.	3, 9, 10
Matlab.	3, 4, 6, 7, 9-13, 17, 18, 20, 22, 23, 26, 36, 38, 39, 44, 45
Median.	6, 10
nonparametric.	4, 5, 8, 26, 45
normality.	10
null hypothesis.	12, 13, 15, 16, 22, 37
P-Value.	3, 8, 12, 14-17, 20-22, 24, 25, 37
paired data.	29, 34
population.	15
Power.	3, 9, 26
Probability.	1, 8, 11, 12, 14-17, 22, 23, 27, 28, 30, 32, 33, 35
Runs.	44, 45
sample.	1, 3-5, 7-9, 11, 13, 15-28, 31-33, 36-38
Sign test.	5, 26-29, 31, 33-35
Skewness.	3

Standard deviation.	3, 6, 13, 14, 19, 20, 23, 27, 29, 30, 32, 33, 35
Standard error.	5, 7, 8
Statistic.	1, 3, 5-10, 12-26, 28-38, 42-45
Student's t distribution.	12
Student's t.	3, 7, 8, 10, 12
Test statistic.	8, 9, 17, 22, 28-31, 33, 35, 44
Type I error.	12
t-test.	3, 10
variance.	3, 5, 7-9, 12, 16, 18, 20, 21, 24, 37, 38, 41
Welch's test.	4
Wilcoxon rank sum test.	3, 4, 7-10, 12, 36
Wilcoxon signed rank test.	5, 31-35