WEEK 12: CHAPTER 12 THE ANALYSIS OF VARIANCE , 13 RANDOMIZED BLOCK DESIGNS & 14 KRUSKAL-WALLIS TEST & FRIEDMAN ANOVA

TABLE OF CONTENTS

Page:

List of Figures . . . 2

List of Tables . . . 3

List of m.files . . . 3

Assignment . . . 4
Required reading . . . 4

Understanding by Design Templates . . . 5

Understanding By Design Stage 1 — Desired Results Week 12 . . . 5

Understanding by Design Stage II — Assessment Evidence Week 12 8/16-8/22 . . . 5

Introduction . . . 5

Principles of ANOVA design . . . 7

Fixed Effects ANOVA . . . 9

A priori & a posteriori tests . . . 10

Case Studies . . . 11
Case Study 12.2.1: Does smoking affect exercise heart rate? . . . 11
Tests of assumptions . . . 11
A priori hypotheses . . . 12
Results & Discussion . . . 13
A posteriori tests . . . 16
Statistical Inference allowed on Case Study 12.2.1 . . . 17
Case Study 12.3.1: Binding of antibiotics to serum proteins . . . 17
Case Study 12.4.1: Infants walking . . . 19
Introduction to the case study . . . 19
Results & Discussion . . . 19
Case Study 13.2.1: Fear of heights . . . 20
Introduction . . . 20

Experimental design issues: lack of replication . . . 21
Results . . . 21
A posteriori tests . . . 22
Case Study 13.2.2: Rat poison . . . 23
Introduction . . . 23
Results and Discussion . . . 23
Case Study 13.2.3: Transylvannia Effect . . . 25
Introduction . . . 25
Results & Discussion . . . 25
Testing the factorial model with interaction term . . . 26
Case Study 14.4.1: 1969 Lottery . . . 27
Case Study 14.5.1: Base Running . . . 28

References . . . 29

Annotated outline (with Matlab scripts) for Larsen & Marx Chapter 12-13 . . . 29

Index . . . 56

List of Figures

Figure 1. Cobb's 1997 book begins with a split-plot ANOVA design and ends with t tests. . . . 6
Table 12.2.1 . . . 11
Figure 3. Notched boxplots for the four groups. 1 is non-smoking, 2, light, 3, moderate and 4 Heavy. . . . 13
Figure 4. Plot of maximum likelihood of lambda (up to a constant) vs. lambda. The plot indicates the appropriate lambda is zero (a ln transform), but the 95% confidence interval includes 1, incdicating no transform required. . . . 13
Figure 5. Plot of the $F_{3,20}$ distribution showing the 3.1 critical value. . . . 14
Figure 6. Results of the Scheffé multiple comparison procedure indicating that the Heavy smoker group differs from the Non-smoker and Light smoker groups. No other differences have a p value less than 0.05. . . . 16
Figure 7. Results of the Tukey HSD multiple comparison procedure indicating that the Heavy smoker group differs from the Non-smoker and Light smoker groups. No other differences have a p value less than 0.05. . . . 16
Figure 8. Notched boxplots for the five types of antibiotics: 1) Penicillin G, 2) Tetracyline, 3) Streptomycin, 4) Erythromycin and 5) Chloramphenicol. . . . 17
Figure 9. Means and Tukey HSD 95% confidence intervals are displayed with with antibiotics differing in mean serum binding indicated by different colors and line styles . . . 18
Figure 10. Notched boxplots for the four groups. A) 12-min walking & placing, B) 12-min exercise, C) No exercise weekly monitoring, and D) No exercise without weekly monitoring . . . 19
Figure 11. Notched boxplots for the three treatments: Contact Desensitization, Demonstration Participation, and Live Modeling. . . . 21
Figure 12. Interaction between therapies and fear of height blocks. Notched boxplots for the four groups. Group A had the greatest fear of heights and Group C the least. . . . 22
Figure 13. Treatment means and Tukey HSD 95% confidence limits. . . . 22
Figure 14. Notched boxplots for the three treatments: Contact Desensitization, Demonstration Participation, and Live Modeling. . . . 23
Figure 15. Treatment by block plot. The set of nearly parallel lines indicates little problem with interaction of flavors and surveys. . . . 24
Figure 16. Treatment means and Tukey HSD 95% confidence limits. . . . 24
Figure 17. Boxplots for the three moon phases: before, during and after full-moon phases. . . . 25
Figure 18. Interaction plot showing strong interaction. The three lowest admission months during the full moon, August, December and January all have a pattern of decreased admissions during the full moon compared to the two non-full moon treatments. May also declines between non-full and full moon periods. These four months produce a strong interaction effect in the Tukey additivity test. . . . 27
Figure 19. Notched boxplots for the lottery numbers by month (1 is January). . . . 27
Figure 20. Notched boxplots for the lottery numbers by month (1 is January). . . . 27

Figure 21. Table 12.1.1. 29
Figure 22 Figure 12.2.1 30
Case Study 12.2.1 31

List of Tables

Table 1. ANOVA Table for Case Study 12.2.1 14
Table 2. ANOVA Table for Case Study 12.2.1. The linear contrast the heart rate of the non-smokers vs. the weighted average of the 3 smoking categories. 14
Table 3. ANOVA Table for Case Study 12.2.1 The linear contrast tests the non-smokers vs. the heavy smokers. 15
Table 4. ANOVA Table for Case Study 12.2.1 This linear contrast tests for a linear trend among the four smoking categories. 15
Table 5. ANOVA Table for Case Study 12.2.1 This contrast tests for a quadratic trend, or hump shaped pattern, among the four smoking categories. 15
Table 6. ANOVA Table for Case Study 12.2.1 This contrast tests for a cubic or S-shaped trend among the four smoking categories 16
Table 7. ANOVA Table for Case Study 12.3.1 18
Table 8. Results of HSD tests using Case Study 12.2 antibiotic data 18
Table 9. ANOVA Table for Case Study 12.4.1 20
Table 9. ANOVA Table for Case Study 12.4.1 linear contrast comparing group A vs. Groups B+C+D 20
Table 9. ANOVA Table for Case Study 13.2.1 22

List of m.files

LMcs120201_4th 31
LMcs120301_4th 33
LMcs120401_4th 34
LMcs130201_4th 36
LMcs130202_4th 36
LMcs130203_4th 37
anovalc 40
LMcs130301_4th 42
LMcs130302_4th 43
LMcs140201_4th 44
LMcs140202_4th 44
LMcs140203_4th 45
LMcs140301_4th 45
LMcs140302_4th 46
LMcs140303_4th 46
LMcs140304_4th 47
function [pvalue,W,U,Wstar]=**Wilcoxranksum**(X,Y,Ex) 48
function pexuptail=**Wilcoxrsexact**(n,m,W,ranks); 49
function [T,R,ind]=**ties**(A) 52
LMcs140401_4th 53
LMcs140501_4th 54

Assignment

REQUIRED READING

- Larsen, R. J. and M. L. Marx. 2006. An introduction to mathematical statistics and its applications, 4th edition. Prentice Hall, Upper Saddle River, NJ. 920 pp.
  - Read All of Chapter 12 & 13,
  - Chapter 14.4 Kruskal Wallis ANOVA & 14.5 Friedman ANOVA

Understanding by Design Templates

Understanding By Design Stage 1 — Desired Results Week 12

LM Chapter 12 & 13 The Analysis of Variance

G Established Goals

- Become familiar with ANOVA, the foundation of experimental and survey design

U Understand

- Model I ANOVA uses the ratio of variances to test for difference in means.

Q Essential Questions

- What is the difference between a Model I and Model II ANOVA?
- Why can't all possible pairs of groups be tested at α=0.05

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

- ANOVA types (**randomized block**, factorial, Friedman, Kruskal-Wallis, Model I and Model II, One-way), **Bonferroni**, **Box-Cox transformation**, **linear contrast**, **multiple comparisons problem**, **orthogonal contrasts**, **pseudoreplication**, **Scheffé multiple comparisons procedure**, **Treatment Mean Square**,**Treatment Sum of Squares**, **Tukey-Kramer test (Tukey's HSD)**, **LSD**

S *Students will be able to*

- Perform parametric and non-parametric ANOVAs, including
- Graphically and statistically analyzing the equal spread assumption
- Setting up and performing linear contrasts among ANOVA groups
- Pereform the appropriate a posteriori ANOVA test

Understanding by Design Stage II — Assessment Evidence Week 12 8/16-8/22

Chapter 12, 13 & 14 (Kruskal-Wallis) The Analysis of Variance

- **Post in the discussion section** by 8/22/11 T
  - Any questions on the final exam.
  - The final exam will take place between 8/23 & 8/25

Introduction

When I planned the revision of EEOS601 several years ago, I considered two drastically different types of course. One was this course, based on a strong foundation in probability, moving to hypothesis testing and finishing with ANOVA. The alternate approach that I almost followed was to use George Cobb's 1997 textbook, The Design and Analysis of Experiments

(Figure 1). In that remarkable text, Cobb, professor emeritus at Mt. Holyoke, begins chapter 1 with a factorial ANOVA to teach the fundamental principles in analyzing real data. In a way, Cobb works backwards through the traditional introductory statistics curriculum, finishing with *t* tests near the end of the book (p 733).

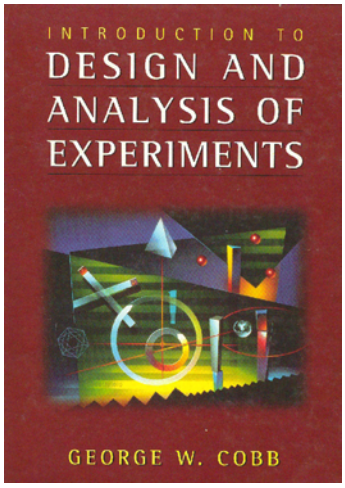


Figure 1. Cobb's 1997 book begins with a split-plot ANOVA design and ends with t tests.

http://c.chegg.com/covers2/9520000/9523597_1295722792.jpg

ANOVA stands for analysis of variance. A one-factor or **one-way ANOVA** involves one factor with two or more levels. As shown in Larsen & Marx (2006), an independent samples *t* test is equivalent to a 1-factor ANOVA with two levels, only the test statistic is a Student's *t* statistic whereas ANOVA uses an *F* statistic. The test statistics and *p* values are closely related in that the *F* statistic with 1 numerator *df* is the square of the t test's *t* statistic and the *p* values for the *F* and *t* statistics are identical. A **factorial ANOVA**, covered in Larsen & Marx Chapter 13 involves two or more factors with two or more levels of each factor. Factorial ANOVA is the key design for experiments because it can assess the interactions among variables, say for example the interacting effects of salinity and temperature on animal growth.

A **randomized block ANOVA** is a subset of factorial ANOVA in which the block is designed to reduce the contribution of the blocking factor in estimating the error variance, producing a much more powerful design. In agricultural experiments, fields are often the blocking factor. Fertilizer levels might be set on different fields, and field-to-field differences are not the primary objects of study. It is always important to assess the potential interaction of blocks and other treatment levels, which usually requires that the treatments be replicated within blocks. Tukey devised an additivity test that allows for block x treatment interactions to be assessed with unreplicated designs.

In a **split-plot ANOVA**, introduced by **Cobb (1997)** in Chapter 1, full replication of treatment levels isn't possible so treatments are divided into whole plots and subplots. For example, experiments involving temperature or humidity might be conducted in greenhouses, which would constitute the whole plot, while different strains of plant might be tested within each greenhouse (the subplots).

In a **nested or hierachic ANOVA**, the experimental units are nested within treatment levels. For example, a study of predation among the animals that live within mud and sand (i.e., the soft-bottom benthos) might involve replicated 1-m^2 quadrats in which a predator has been added. The benthic animals are sampled using 1-cm^2 cores. The quadrats are the experimental units, nested within treatments. The 1-cm^2 cores are sampling units nested within the experimental units, the quadrats.

In a **repeated measures ANOVA**, the same individual or plot is sampled through time. Drug trials often involve the same individuals receiving different treatments. The repeated measures

design, by controlling for individual-to-individual variability produces much more powerful designs than if separate individuals were sampled.

Principles of ANOVA design

- Design the experiment or survey after identifying the hypotheses and specify the statistical tests BEFORE collecting data (*a priori*)
- Do a pre-test or preliminary survey
  - If the variance is unknown, consider doing a preliminary experiment to calculate power for the full analysis
  - Also, a pre-test will allow covariates and problems to be identified
- Do a power analysis based on literature data or a preliminary survey.
  - Ensure that the number of experimental units assures sufficient power so that any results obtained are unlikely to be due to chance (the positive predictive value is different from the Probability of Type I error, **Ioannidis (2005)**)
  - If constraints don't permit a sufficient number of samples
    - Use a different measurement procedure, less prone to error
    - Reduce the mean squared error through blocking or stratification
    - Use a repeated measures design
    - If sufficient power still can't be attained, stop & pick a new research problem.
- Endeavor to create balanced designs with equal number of replicates at each combination of treatment and block levels
  - ANOVA is robust to unequal spread (*i.e.,* heteroscedasticity) if the design is balanced (**Winer et al. 1991**)
- ANOVA tests for difference in means (fixed effect) or whether $(\sigma_i^2+\sigma^2)/\sigma^2=1$ (random effect) or both (mixed model)
- Fixed vs. random effects
  - The choice of fixed vs. random effects is often crucial and depends on whether the factor levels represent a random or representative sample from some larger statistical population
  - The *F* statistics, the interpretation of the results, and the extent of statistical inference often change depending on whether factors are fixed or random.
- Avoid pseudoreplication (**Hurlbert 1984**)
  - Pseudoreplication, a term coined by Hurlbert is also called model misspecification, has two causes: inadequate replication at the design stage, or
  - Using an inappropriate model especially the wrong ANOVA model with an inappropriate error mean square and error d.f.
  - Examples of model misspecification
    - Failing to use a repeated measures design for longitudinal data
    - Confusing factorial and nested ANOVA
    - Inappropriately pooling terms in a nested, randomized block, or factorial ANOVA

- The alpha level for hypothesis tests (i.e., the critical values) must be set in advance. Tests and hypothesis, as far as possible, should be specified in advance. *A priori* hypotheses, if a small subset of the possible tests that might be performed, can be tested at the experiment- wise alpha level, usually α=0.05.
- Patterns which reveal themselves after the data have been analyzed, or even graphed, must be assessed using an appropriate multiple comparison procedure that reduces the test α to maintain the experiment-wise or family-wise α level (usually α=0.05)
- After writing down the hypotheses to be tested and the tests to be performed, graph the data and critically look for violations of the assumptions, especially unequal spread
  - Use boxplots & Levene's tests to assess unequal variance & detect outliers
  - =unequal variance = heteroscedasticity = heteroskedacity = lack of homoscedasticity
  - Unequal variance is best revealed by box plots
  - Unequal spread can be tested with Levene's test
- Transform the data to correct unequal spread
  - $\sqrt{}$ transform for Poisson-distributed counts, log (X+1) for logarithmically or log-normally distributed data
  - Logit (log (p/(1-p)) transform or arcsin $\sqrt{P}$ for binomial data
- Perform the ANOVA
- Assess higher order interaction effects and analyze the influence of outliers
    - Graphically display residuals vs. expected values & assess heteroscedasticity (again) and effects of outliers
    - Note that an outlier is only an outlier when compared to an underlying probability model
- Use appropriate rules for pooling sums of squares to produce more powerful tests of lower order interactions & main effects
- Examine the data for outliers, but
  - Never remove outliers without strong justification
  - Examine data notebooks to find out if there were conditions that justify treating outliers as a different statistical population (e.g., different analyst or different analytical instrument)
  - If the outlier's removal might be justified
      - Do the analysis with and without the outlier
        - If the conclusion remains the same, leave the outlier in, unless it has caused a major violation in assumptions
        - If the conclusion differs, drop the outlier and all similar data
    - If there is no reason for removing the outlier
      - Use rank-based methods, like Kruskal-Wallis or Friedman's ANOVA which are resistant to outlier effects
      - Report the results with and without the oultier
- Evaluate null hypotheses, report p values & effect sizes
- Multiple comparisons procedures, from most to least conservative
  - Scheffé: must be used whenever more than one group is combined in a linear contrast, more conservative than Bonferroni

        - Scheffé multiplier:
            - $\sqrt{[(I-1)F_{(I-1),d.f.,,(1-\alpha)}]}$
            - Where, I is number of groups, d.f. =error df, $F_{(I-1),d.f.,(1-\alpha)}$ is 95th percentile of F distribution
    - Bonferroni: insufficiently conservative for all possible linear contrasts, but the most conservative for pair-wise contrasts
    - Tukey's Honestly Significant Difference (HSD), also called Tukey-Kramer if sample sizes are unequal
    - Student-Newman-Keuls More powerful than HSD
    - Dunnet's, appropriate if there is a control group
    - Tukey's LSD with F-protection: Use LSD if the overall F statistic is significant; not sufficiently conservative
- Report all relevant p values and *df* needed to reconstruct the ANOVA table
    - **Hurlbert (1984)**: it wasn't clear in the majority of ecological studies what test was performed
    - Avoid the significant/non-significant dichotomy (see **Sterne & Smith (2001)**)
- Summarize the results of the ANOVA in the text, table or figure. It is unlikely that a journal will allow both a table and figure, but summary in the text is essential
- Report the effect size (*i.e.*, difference in means with 95% confidence intervals)
- Report negative results, (e.g., failure to reject the null hypothesis)

Fixed Effects ANOVA

A fixed effects ANOVA tests for differences in means by testing the treatment mean square over the error mean square. Larsen & Marx Theorem 12.2.1 provides the expected value for the treatment sum of squares:

Theorem 12.2.1. *Let SSTR be the treatment sum of squares defined for k independent random samples of sizes* $n_1, n_2, \ldots,$ and n_k. *Then*

$$E(SSTR) = (k - 1)\sigma^2 + \sum_{j=1}^{k} n_j(\mu_j - \mu)^2$$

Theorem 12.2.5 describes the standard F test for testing whether means among treatment levels in a one-factor ANOVA are different.

Theorem 12.2.5. *Suppose that each observation in a set of k independent random samples is normally distributed with the same variance σ^2. Let $\mu_1, \mu_2, \ldots,$ and μ_k be the true means associated with the k samples. Then*

a. *If $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ is true,*

$$F = \frac{SSTR/(k-1)}{SSE/(n-k)}$$

has an F distribution with $k-1$ and $n-k$ degrees of freedom.

b. *At the α level of significance, $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ should be rejected if $F \geq F_{1-\alpha,k-1,n-k}$.*

The results of an ANOVA are usually presented in the form of an ANOVA table, showing the key F test, the treatment mean square divided by the error mean square.

Source	df	*SS*	*MS*	*F*	*P*
Treatment	$k-1$	*SSTR*	*MSTR*	$\frac{MSTR}{MSE}$	$P(F_{k-1,n-k} \geq$ observed $F)$
Error	$n-k$	*SSE*	*MSE*		
Total	$n-1$	*SSTOT*			

FIGURE 12.2.1

A priori & a posteriori tests

For all but the simplest ANOVA results, if the null hypothesis is rejected, say $\mu_1 = \mu_2 = \mu_3$, there is still interest in finding out which differences led to the rejection of the null. If the experimenter has followed proper procedures, the key hypotheses should have been set in advance. They can now be tested using powerful tests such as the F-protected least significant difference or linear contrasts. These tests can be performed at an α level of 0.05, or whatever the accepted α level is for the study.

If the tests haven't been specified in advance, then the α level must be adjusted for the number of possible comparisons that could have been performed. The most common multiple comparison procedure is also just about the most conservative, the Bonferroni procedure. If there were 5 groups, there are $\binom{5}{2}$ or 10 different ways the groups can be compared 2 at a time. In the simplest version of the Bonferroni correction, the α level would be divided by the number of possible tests or 10. To maintain an experiment-wise of family-wise alpha level of 0.05, each test must be performed at the α/Number of tests = 0.05/10 =0.005 level. Without this correction, the probability of rejecting the null hypothesis after performing 10 independent tests is not 0.05 but is instead $\alpha_{experimental} = 1 - (1\text{-alpha}_{test})^{10} = 1\text{-}0.95^{10} = 0.4013$. If the alpha level is divided by the number of tests, the experiment-wise alpha level is maintained: $1\text{-}(1\text{-}0.05/10)^{10} = 0.0489$.

If the investigator compares averages of more than one treatment group, even the Bonferroni correction is inadequate to properly protect against multiple hypothesis testing. The Scheffé procedure is the only standard procedure for adjusting α levels for unplanned linear contrasts.

There are dozens of other multiple comparison procedures that have been proposed. **Quinn & Keough (2002)** provide a nice summary. The Tukey-Kramer honestly significant difference (HSD) is one of the most widely used. The Student-Newmann-Keuls test, based on the studentized range is more powerful than the HSD.

If there is an unplanned comparison with a control group, then Dunnet's procedure is appropriate. Since there are fewer pairwise comparisons when one must be the control group, Dunnet's procedure is more powerful in detecting an effect.

Case Studies

Case Study 12.2.1: Does smoking affect exercise heart rate?

CASE STUDY 12.2.1

Generations of athletes have been cautioned that cigarette smoking retards performance. One measure of the truth of that warning is the effect of smoking on heart rate. In one study (72) examining that impact, six each of non-smokers, light smokers, moderate smokers, and heavy smokers undertook sustained physical exercise. Their heart rates were measured after resting for three minutes. The results appear in Table 12.2.1. Are the differences among the $\bar{Y}_{.j}$s statistically significant? That is, if

TABLE 12.2.1

	Non-Smokers	Light Smokers	Moderate Smokers	Heavy Smokers
	69	55	66	91
	52	60	81	72
	71	78	70	81
	58	58	77	67
	59	62	57	95
	65	66	79	84
$T_{.j}$	374	379	430	490
$\bar{Y}_{.j}$	62.3	63.2	71.7	81.7

Twenty four individuals undertook sustained physical exercise and their pulse was measured after resting for 3 minutes. The results are shown in Table 12.2.1. With this experiment, we should set our null hypotheses in advance, and they are:

H_o: $\mu_{HR\ Non\text{-}Smokers} = \mu_{HR\ Light\ Smokers} = \mu_{HR\ Moderate\ Smokers} = \mu_{HR\ Heavy\ Smokers}$

H_a: $\mu_{HR\ Non\text{-}Smokers} < \mu_{HR\ Light\ Smokers} < \mu_{HR\ Moderate\ Smokers} < \mu_{HR\ Heavy\ Smokers}$

Tests of assumptions

The main assumption to be tested with these data is homoscedasticity, or equality of variances among groups. This will be tested with a box plot, and if there is an appearance of heteroscedasticity, or unequal variance, then by the formal Levene's test. A Levene's test

performs another ANOVA on the absolute value of the deviations from the means among groups. There is an m.file submitted by A. Trujillo-Ortiz and R. Hernandez-Walls to the Matlab user's group that performs the Levene's test using the absolute value of the difference between cases and group means. There are two other common ways for performing the Levene's test: squared deviation from group means and absolute value of deviations between group medians. I suppose one could calculate squared deviation from group medians as well, but I've never seen that proposed.

I used the Box-Cox transformation procedure on these data to find out what transform was appropriate. As described in my handout 2 (statistical definitions), the Box-Cox procedure will test a family of transformations including the inverse, power, log, and square root transformations. Using a user-contribued m.file by Hovav Dror on the Matlab file exchange, the lambda parameter can be found.

A priori hypotheses

There are many tests that could be peformed with these 4 categories of smoking. There are $\binom{4}{2}$, or 6, two-at-a-time contrasts (e.g., Non-smokers vs. Heavy Smokers). There are 7 simple linear contrasts, 4 1-group vs. three-group contrasts (e.g., A vs. B+ C+D) and 3 two-group vs. two-group contrasts (e.g., A+B vs. C+D). There are 8 ways that groups can be compared 3 at a time (e.g., B vs. CD or C vs. BD). Thus, there are 6+4+3+8 or 21 ways the four groups of data can be compared using simple contrasts and more possible ways to analyze the data with more complicated linear contrasts. For example one could hypothesize a linear trend among categories with heart rate increasing with contrast coefficients -3/2, -1/2, 1/2 , 3/2 or one could test for a hump-shaped pattern in heart rate with a quadratic orthogonal polynomial -3/2 -11/6 -1/6 7/2. The linear and quadratic contrasts are set to be uncorrelated or othogonal. There is a wonderful program, called orthpoly.m in the stixbox.m free toolbox for Matlab that allows the calculation of othogonal polynomials of any degree for an input vector like 1, 2, 3, 4. If we had actual data on packs smoked per day, we could have used that data to set up an **orthogonal contrast**.

To keep the analysis simple, I'll just request five linear contrasts. The first tests whether the non-smokers differ from the weighted average of the non-smoking group. We will set a one-sided alternative hypothesis in that our strong expectation is that the smokers will have a higher heart rate.

$H_{o1}: \mu_{HR\ Non\text{-}Smokers} = \mu_{HR\ Light\ Smokers} + \mu_{HR\ Moderate\ Smokers} + \mu_{HR\ Heavy\ Smokers}$
$H_a: \mu_{HR\ Non\text{-}Smokers} < \mu_{HR\ Light\ Smokers} + \mu_{HR\ Moderate\ Smokers} + \mu_{HR\ Heavy\ Smokers}.$

The second *a priori* hypothesis simply tests whether there is a difference between the Non-smokers and Heavy Smokers. Again, we will use a one-sided alternative hypothesis.

$H_{o1}: \mu_{HR\ Non\text{-}Smokers} = \mu_{HR\ Heavy\ Smokers}$
$H_a: \mu_{HR\ Non\text{-}Smokers} < \mu_{HR\ Heavy\ Smokers}.$

Just for interest's sake, I'll also test the contrasts for a linear trend and quadratic trend in the smoker data. The first of these will be tested against a one-sided alternative. I have no prior expectation about the 2nd pattern, but I expect that it will be there. It could be a concave-up hump-shaped pattern or a concave-down hump-shaped pattern. For a publication, I wouldn't do this because we haven't established that smoking effects should be analyzed as an interval-scaled variable and we have little justification for proposing the 1-to-4 scale. All of these contrasts will use Gallagher's lcanova program, which tests linear contrasts with Matlab's ANOVA programs. The contrast matrix will be:

Lmatrix= [	-1	1/3	1/3	1/3
	-1	0	0	1
	-3/2	-½	½	3/2
	-3/2	-11/6	-1/6	7/2
	747/310	-141/310	-909/310	303/310]

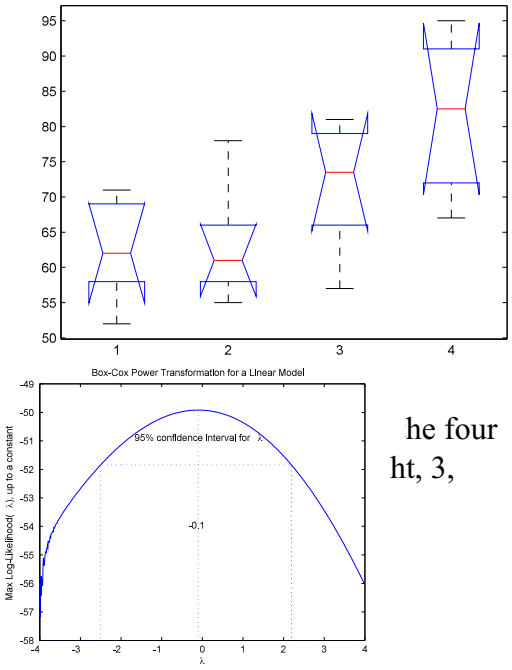
The final three orthogonal contrasts — the linear, quadratic, and cubic — comprise a set of mutually orthogonal contrasts partitioning the treatment sum of squares (See **Larsen & Marx 2006 Theorem 12.4.1, p 754**).

Results & Discussion

Overall ANOVA

As shown in Figure 3, there is some evidence for unequal variance among the four groups. Levene's test indicated that little evidence ($P(F_{3,20} \geq 0.378)=0.77$) that these data violated the homoscedasticity hypothesis. **Larsen & Marx (2006)**, following a tradition in introductory statistics texts, don't discuss the use of transformations, but the data in Figure 3 should be transformed with a log or square-root transform to equalize the variance among groups. Since the sample sizes are equal in this study, **Winer et al. (1991)** present analyses indicating the the conclusions of the ANOVA will be robust to minor violations of the homoscedasticity assumption.

The Box-Cox Analysis with results shown in Figure 5 indicated that a transformation with lamba=0, indicating a log transform, would equalize the variance among groups. But, the 95% CI for lambda was -2.5 to 2.2, which includes 1 which is the lambda indicating no transformation. For the remainder of the analysis and to conform with the solutions in **Larsen & Marx (2006)**, the untransformed data will be used.



he four
ht, 3,

Figure 4. Plot of maximum likelihood of lambda (up to a constant) vs. lambda. The plot indicates the appropriate lambda is zero (a ln transform), but the 95% confidence interval includes 1, incdicating no transform required.

Table 1 shows the results of the ANOVA. There is strong evidence to reject the null hypothesis of equal heart rates among the four groups. The F statistic, 6.12, far exceeds the 3.1 critical value (in Matlab: finv(0.95,3,20)) as shown in Figure 5.

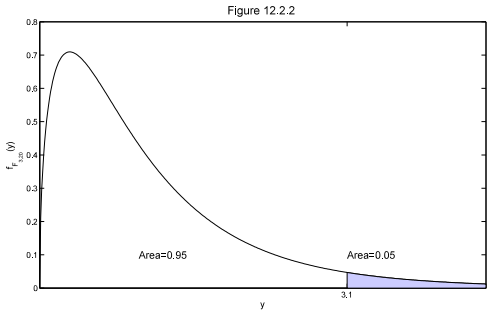


Figure 5. Plot of the $F_{3,20}$ distribution showing the 3.1 critical value.

Table 1. ANOVA Table for Case Study 12.2.1

Source	Sum Sq.	d.f.	Mean Square	F	Prob>F
Smoking Level	1464.13	3	488.04	6.12	0.004
Error	1594.83	20	79.74		
Total	3058.96	23			

Linear Contrasts

The first linear contrast tested the non-smokers vs. the three smoking categories, producing the ANOVA Table shown in Table 2. With a *p* value of 0.03, this ANOVA provides modest evidence against the null hypothesis of equal heart rates between these two groups. The average smoker has a heart rate 3 minutes after exercise 10 ± 9 beats per minute higher than the non-smokers. Note that the 95% CI doesn't include 0, consistent with the p value of 0.03.

Table 2. ANOVA Table for Case Study 12.2.1. The linear contrast the heart rate of the non-smokers vs. the weighted average of the 3 smoking categories.

Source	Sum Sq.	d.f.	Mean Square	F	Prob>F
Non-smoking *vs.* 3 smoking categories	435.1	1	435.1	5.5	0.03
Error	1594.83	20	79.74		
Total	3058.96	23			

The second linear contrast tested the Non-smokers vs. the Heavy smokers, producing the ANOVA Table shown in Table 3. With a *p* value of 0.00126, this ANOVA provides very strong evidence against the null hypothesis of equal heart rates between these two groups. Heavy

smokers hav a heart rate 3 minutes after exercise 19 ± 11 beats per minute higher than the non-smokers.

Table 3. ANOVA Table for Case Study 12.2.1 The linear contrast tests the non-smokers vs. the heavy smokers.

Source	Sum Sq.	d.f.	Mean Square	F	Prob>F
Non-smokers *vs.* Heaviest Smokers	1121.3	1	1121.3	14.1	0.00126
Error	1594.83	20	79.74		
Total	3058.96	23			

The third linear contrast tested for a linear trend among the four smoking categories, producing the ANOVA Table shown in Table 4. With a p value of 0.00059, this ANOVA provides very strong evidence against the null hypothesis of no linear trend. The value for the contrast indicates that in moving from one category to the next, heart rate increases by 8.3 ± 4.3 beats per minute. A least squares regression, which requires more assumptions be met produces a slope of 6.65 ± 3.36 between categories.

Table 4. ANOVA Table for Case Study 12.2.1 This linear contrast tests for a linear trend among the four smoking categories.

Source	Sum Sq.	d.f.	Mean Square	F	Prob>F
Linear trend among smoker categories	1326.7	1	1326.7	16.6	0.00059
Error	1594.83	20	79.74		
Total	3058.96	23			

The fourth linear contrast tests for a quadratic or hump-shaped trend among the four smoking categories, producing the ANOVA Table shown in Table 5. With a p value of 0.00045, this ANOVA provides very strong evidence against the null hypothesis of no quadratic trend. The presence of a concave-up pattern in the heart-rate data is consistent with the finding of a strong hump shaped pattern in addition to the linear pattern.

Table 5. ANOVA Table for Case Study 12.2.1 This contrast tests for a quadratic trend, or hump shaped pattern, among the four smoking categories.

Source	Sum Sq.	d.f.	Mean Square	F	Prob>F
Linear trend among smoker categories	1399.0	1	1399.0	17.5	0.00045
Error	1594.83	20	79.74		
Total	3058.96	23			

The fifth and final contrast tests for a cubic or S-shaped trend among the four smoking categories, producing the ANOVA Table shown in Table 6. With a *p* value of 0.55, this ANOVA provides very little evidence against the null hypothesis of no S-shaped pattern.

Table 6. ANOVA Table for Case Study 12.2.1 This contrast tests for a cubic or S-shaped trend among the four smoking categories

Source	Sum Sq.	d.f.	Mean Square	F	Prob>F
Linear trend among smoker categories	30.2	1	30.2	0.4	0.55
Error	1594.83	20	79.74		
Total	3058.96	23			

A posteriori tests

The five a priori tests (out of 24 possible tests) provide an excellent summary of the data. But there are several questions left unanswered. For example, do moderate smokers have heart rates after exercise different from the other three groups. These can be answered using appropriate a posteriori tests. **Larsen & Marx (2006)** discuss the Tukey HSD, which is available in Matlab and included in the program for this case study (Matlab: multcompare(stats,'ctype','hsd','alpha',0.05)). In this exegesis Larsen & Marx case study 12.2.1, we've used linear contrasts, so the only appropriate a posteriori adjustment procedure is the conservative Scheffé procedure (Matlab: multcompare(stats,'ctype','scheffe','alpha',0.05)). The results of that analysis are shown in Figure 6. The Heavy Smoker heart rates differed from the non-smoker and light smoker groups, but no other differences had a p value less than 0.05. There was insufficient evidence tha the heart rate of moderate smokers differed from any of the three other groups.

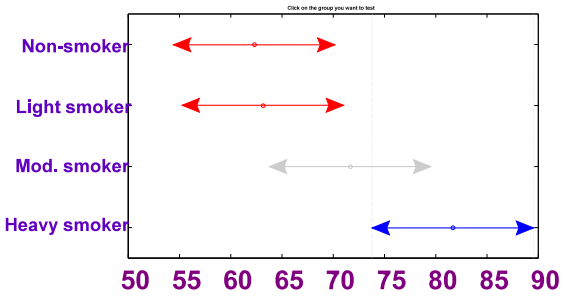


Figure 6. Results of the Scheffé multiple comparison procedure indicating that the Heavy smoker group differs from the Non-smoker and Light smoker groups. No other differences have a p value less than 0.05.

These are the same conclusion that would be reached with the more liberal Tukey Honestly Sigificant Difference (HSD) multiple comparisons procedure, as shown in Figure 7.

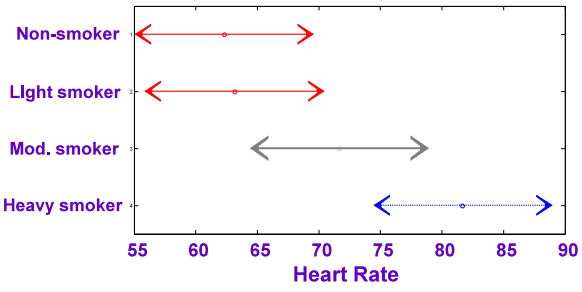


Figure 7. Results of the Tukey HSD multiple comparison procedure indicating that the Heavy smoker group differs from the Non-smoker and Light smoker groups. No other differences have a p value less than 0.05.

Statistical Inference allowed on Case Study 12.2.1

This was an example drawn from a biostatistics text, so inferences based on the data would be speculative. There is no indication that the subjects were randomly assigned to treatment groups, as would be required for a proper experiment. It seems unlikely that randomly selected individuals could be ethically required to smoke large numbers of cigarettes in the days, weeks or months before exercising for this experiment. R. A Fisher, a pipe smoker, denied that there was a causal link between smoking and cancer and presumably denied the link between smoking and heart disease. **Fisher (1958)** argued that it wasn't the fault of the early cancer investigators that a proper experiment to demonstrate a causal link between smoking and cancer couldn't be performed:

> "Now, randomization is totally impossible, so far as I can judge, in an inquiry of this kind. It is not the fault of the medical investigators. It is not the fault of Hill or Doll or Hammond that they cannot produce evidence in which a thousand children of teen age have been laid under a ban that they shall never smoke, and a thousand or more chosen at random from the same age group have been under compulsion to smoke at least thirty cigarettes a day. If that type of experiment could be done, there would be no difficulty."

Since it was unlikely that individuals weren't assigned to different categories through randomization, then no causal link can be claimed between smoking and exercise heart rate. Indeed, one can guess that heavy smokers are not prone to exercise much, so the results could have been due to the overall fitness of the four groups and not necessarily to their smoking habits.

CASE STUDY 12.3.1: BINDING OF ANTIBIOTICS TO SERUM PROTEINS

A boxplot of the data, Figure 8, indicates no problems with unequal variance.

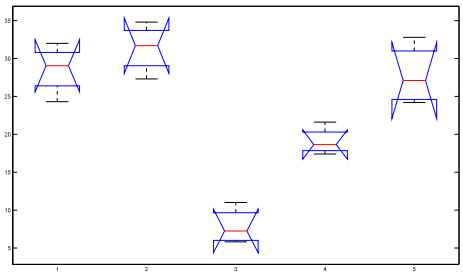


Figure 8. Notched boxplots for the five types of antibiotics: 1) Penicillin G, 2) Tetracyline, 3) Streptomycin, 4) Erythromycin and 5) Chloramphenicol.

The results of the ANOVA are shown in Table 7. Differences among antibiotics were tested using Tukey's HSD and the results are shown graphically in Figure 9 and in Table 8.

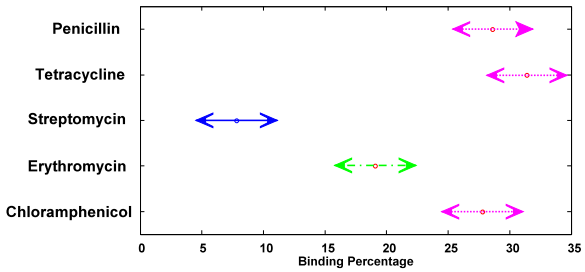


Figure 9. Means and Tukey HSD 95% confidence intervals are displayed with with antibiotics differing in mean serum binding indicated by different colors and line styles

There is exceptionally strong evidence for differences in binding percentage among antibiotics (ANOVA: $P(F_{4,15} \geq 40.9) < 10^{-7}$). At α=0.05, Tukey's HSD revealed that streptomycin binding percentage was lower than the other four antibiotics, and erythromycin binding percentage was less than penicillin, tetracycline and chloramphenicol. Tukey's HSD provided little evidence at α=0.05 for differences in binding among penicillin, tetracycline and chloramphenicol.

Table 7. ANOVA Table for Case Study 12.3.1

Source	**Sum Sq.**	**d.f.**	**Mean Square**	**F**	**Prob>F**
Smoking Level	1480.8	4	370.2	40.9	$< 10^{-7}$
Error	135.82	15	9.1		
Total	1616.7	19			

Table 8. Results of HSD tests using Case Study 12.2 antibiotic data, produced by Matlab's multcompare.m

Level *i*	**Level *j***	**Lower 95% CI**	**Mean difference**	**Upper 95% CI**	**Conclusion**
Pen.	Tetra.	-9.3	-2.8	3.8	NS
Pen.	Strepto.	14.2	20.8	27.3	Reject
Pen.	Erythro.	3.0	9.5	16.1	Reject
Pen.	Chloram.	-5.8	0.8	7.4	NS
Tetra.	Strepto.	17.0	23.6	30.1	Reject
Tetra.	Erythro.	5.7	12.3	18.9	Reject
Tetra.	Chloram.	-3.0	3.6	10.1	NS
Strepto.	Erythro.	-17.8	-11.2	-4.7	Reject
Strepto.	Chloram.	-26.5	-20.0	-13.4	Reject
Erythro.	Chloram.	-15.3	-8.7	-2.2	Reject

CASE STUDY 12.4.1: INFANTS WALKING

Introduction to the case study

Can the age to walking be reduced through walking exercises? Twenty three infants were randomly divided into four groups, A through D. Group A received 12 minutes of walking and placing exercises daily. Group B received 12 minutes of daily exercise without special walking and placing exercise. Group C and D received no special instructions. Group C's progress, like A & B, were checked for weekly progress, but Group D was checked only at the end of the experiment. Table 12.4.1 shows the ages at which the babies walked alone.

In addition to the overall ANOVA, it would be interesting to compare group A with the average of Groups B through D. It would also be interesting to compare groups A vs. B. A check of C vs. D would evaluate whether there were any effects of weekly checks. It would also be interesting to compare the two 12-min exercise groups (A & B) with the two groups that weren't asked to do anything (C & D). The linear contrast coefficients can be expressed in a Linear contrast matrix:

$$\text{Lmatrix} = \begin{bmatrix} 1 & -1/3 & -1/3 & -1/3 \\ 1 & -1 & 0 & 0 \\ ½ & ½ & -½ & -½ \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

Results & Discussion

The notched boxplots, shown in Figure 10, reveal some problems with unequal spread, but there was an extreme outlier in both groups A & B. This could result in an inflated error variance and indicates that the results should be checked with a Wilcoxon rank sum test.

The overall ANOVA (Table 9) indicates little evidence for any group differences.

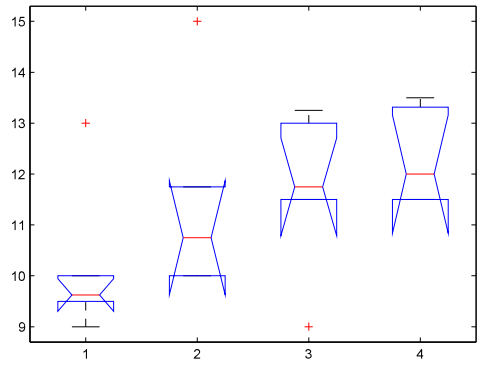


Figure 10. Notched boxplots for the four groups. A) 12-min walking & placing, B) 12-min exercise, C) No exercise weekly monitoring, and D) No exercise without weekly monitoring

Table 9. ANOVA Table for Case Study 12.4.1					
Source	**Sum Sq.**	**d.f.**	**Mean Square**	**F**	**Prob>F**
Baby Group	14.78	3	4.93	2.1	0.13
Error	43.69	19	2.30		
Total	58.47	22			

Linear contrasts

The results of the first contrast indicates moderately strong evidence against the null hypothesis of equal walking times. The baby exercise group walked 1.7 ± 1.5 months earlier than the other groups.

Table 9. ANOVA Table for Case Study 12.4.1 linear contrast comparing group A vs. Groups B+C+D					
Source	**Sum Sq.**	**d.f.**	**Mean Square**	**F**	**Prob>F**
Baby Group	12.6	1	12.6	5.5	0.03
Error	43.7	19	2.3		
Total	58.5	22			

The linear contrast between groups A and B indicated that group A walked 1.2 ± 1.8 months before group B, a difference that could be due to chance ($P(F_{1,19} \geq 2.0)=0.17$ with a 95% CI that includes 0). The linear contrast between the two 12-min exercise groups (A +B) and the two other groups (C+D) indicated that group A & B walked 1.28 ± 1.33 months before groups C+D, a difference that could be due to chance ($P(F_{1,19} \geq 4.1) = 0.058$ with a 95% CI that includes 0). The linear contrast between groups C and D indicated that group C walked 0.6 ± 1.9 months before group D, a difference that could be due to chance ($P(F_{1,19} \geq 0.5) = 0.49$ with a 95% CI that includes 0).

Because of the two extreme outliers, a Kruskal-Wallis ANOVA was performed, but there was only weak evidence against the null hypothesis of equal medians among groups ($P(\chi^2_3>6.88)=0.076$). Thus, the parametric and non-parametric ANOVA's produced similar results.

Case Study 13.2.1: Fear of heights

Introduction

Three different therapies to treat fear of heights was tested on 15 subjects. The subjects were given a HAT test assessing their fear of heights and were divided into 5 groups of 3 based on their initial fear of heights. One individual in each group was assigned randomly to each of the

three treatments. After the treatments, the subjects were given the HAT test again and the response is the difference in scores.

TABLE 13.2.3: HAT Score Changes

	Therapy			
Block	Contact Desensitization	Demonstration Participation	Live Modeling	$T_{i.}$
A	8	2	−2	8
B	11	1	0	12
C	9	12	6	27
D	16	11	2	29
E	24	19	11	54
$T_{.j}$	68	45	17	130

Experimental design issues: lack of replication

Unfortunately, the treatments weren't replicated among blocks, and it is vitally important to assess the block by treatment interaction effect. Does the effect of a treatment differ based on the initial classification of fear of heights (i.e., the assignment to groups A through D). Fortunately, this can be tested with the Tukey additivity test, discussed in **Quinn & Keough (2002)** and available as a user-contributed m.file for Matlab.

Results

The boxplots (Figure 11) indicate little evidence for unequal spread.

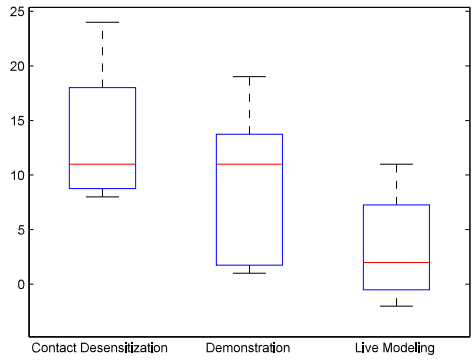


Figure 11. Notched boxplots for the three treatments: Contact Desensitization, Demonstration Participation, and Live Modeling.

The interaction between blocks and treatments can usually be qualitatively evaluated by a plot of treatment level means by block as shown in Figure 12. The Tukey additivity test with a *p* value of 0.6 provided little evidence to reject the assumption of additivity of block and treatment effects.

There is exceptionally strong evidence for a Therapy effect on change in HAT scores (Randomized blocked ANOVA $P\{F_{2,8} \geq 15.3 \mid H_o\}=0.002$). There was also a pronounced block effect with the groups with the strongest acrophobia showing the least improvement in HAT scores (ANOVA $P\{F_{4,8} \geq 12.8 \mid H_o\}=0.002$).

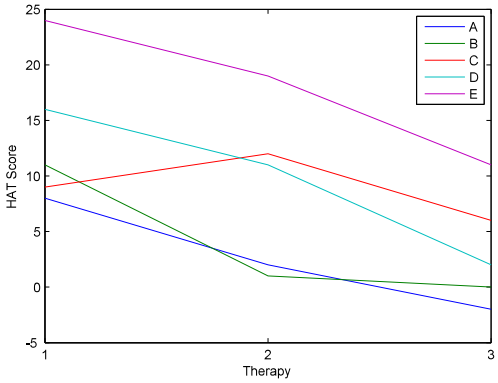


Figure 12. Interaction between therapies and fear of height blocks. Notched boxplots for the four groups. Group A had the greatest fear of heights and Group C the least.

Table 9. ANOVA Table for Case Study 13.2.1

Source	Sum Sq.	d.f.	Mean Square	F	Prob>F
Treatment	260.9	2	130.5	15.3	0.002
Fear Block	438	4	109.5	12.8	0.002
Error	68.4	8	8.6		
Total	767.3	14			

A posteriori tests

A posteriori tests using Tukey's HSD are shown in Figure 13. There is very strong evidence that the contact desensitization (CD) therapy increased mean HAT scores relative to Live Modeling (LM) (Difference ± half 95% CI I = 10.2 ± 5.3 using Tukey's HSD). There is little evidence for a difference between CD and Demonstration Participation (DP): difference=4.6±5.3. There was modest evidence that the mean DP HAT score exceeded LM: difference ± half 95% CI = 5.6±5.3.

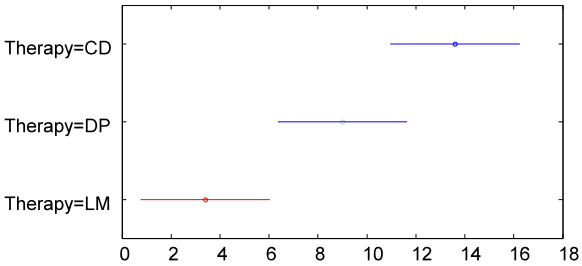


Figure 13. Treatment means and Tukey HSD 95% confidence limits.

CASE STUDY 13.2.2: RAT POISON

Introduction

Rats are treated by poisoning cornmeal, but in some areas, rats won't eat the cornmeal unless it is flavored. Mixing it with real food leads to spoilage so in 5 different surveys, corn meal was mixed with artificial flavoring and the response measured relative to a cornmeal control. The response variable is the fraction of cornmeal eaten by rats.

TABLE 13.2.5

Survey Number	Plain	Butter Vanilla	Roast Beef	Bread
1	13.8	11.7	14.0	12.6
2	12.9	16.7	15.5	13.8
3	25.9	29.8	27.8	25.0
4	18.0	23.1	23.0	16.9
5	15.2	20.2	19.0	13.7

Results and Discussion

The boxplots shown in Figure 14 revealed no problems with the homoscedasticity assumption.

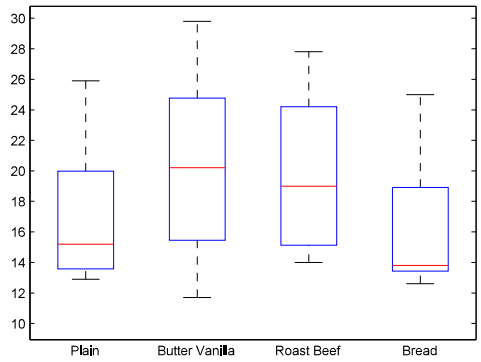


Figure 14. Notched boxplots for the three treatments: Contact Desensitization, Demonstration Participation, and Live Modeling.

The plot of treatments x block means shown in Figure 15 reveals sets of nearly parallel lines indicating no evident block by treatment interactions. There is little evidence for an interaction between Survey and Flavor (Tukey additivity test $P\{F_{1,11} \geq 2.4 | H_o\}=0.16$).

The ANOVA Table, shown in Table 10 provides strong evidence for differences in the percentage of bait eaten among flavors (randomized block ANOVA $P\{F_{3,12} \geq 7.6 | H_o\}=0.0042$). There were also substantial differences in bait consumed among the 5 surveys ($P\{F_{4,12} \geq 50 | H_o\}<10^{-6}$).

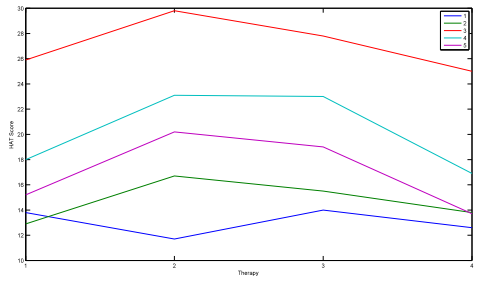


Figure 15. Treatment by block plot. The set of nearly parallel lines indicates little problem with interaction of flavors and surveys.

Table 10. ANOVA Table for Case Study 13.2.2

Source	Sum Sq.	d.f.	Mean Square	F	Prob>F
Survey	495.3	4	123.8	50	$<10^{-6}$
Flavor	56.4	3	18.8	7.6	0.0042
Error	29.8	12	2.5		
Total	581.5	19			

Results of Tukey's HSD tests are presented in Figure 16. Groups for which there is insufficient evidence to reject the equal means hypothesis (at $\alpha=0.05$) are indicated by the same letters. For example, bread flavor isn't different from plain but is less than roast beef and butter flavor.

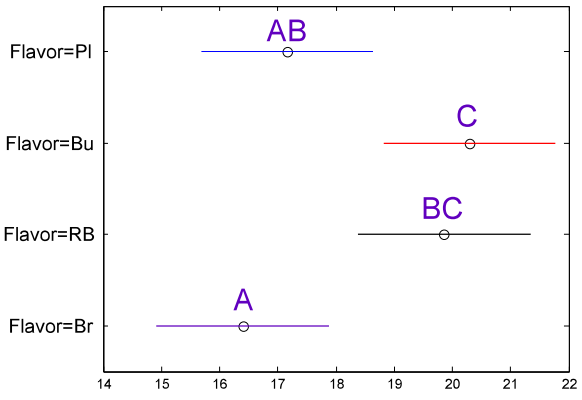


Figure 16. Treatment means and Tukey HSD 95% confidence limits.

CASE STUDY 13.2.3: TRANSYLVANNIA EFFECT

Introduction

Are hospital admission rates higher during the full moon? This case study will apply linear contrasts to a factorial model.

TABLE 13.2.8

Admission Rates (patients/day)

Month	(1) Before Full Moon	(2) During Full Moon	(3) After Full Moon	$\overline{Y}_{i\cdot}$
Aug.	6.4	5.0	5.8	5.73
Sept.	7.1	13.0	9.2	9.77
Oct.	6.5	14.0	7.9	9.47
Nov.	8.6	12.0	7.7	9.43
Dec.	8.1	6.0	11.0	8.37
Jan.	10.4	9.0	12.9	10.77
Feb.	11.5	13.0	13.5	12.67
Mar.	13.8	16.0	13.1	14.30
Apr.	15.4	25.0	15.8	18.73
May	15.7	13.0	13.3	14.00
June	11.7	14.0	12.8	12.83
July	15.8	20.0	14.5	16.77
$\overline{Y}_{\cdot j}$	10.92	13.33	11.46	

Results & Discussion

The boxplot shown in Figure 17 reveals no evident problems with heteroscedasticity.

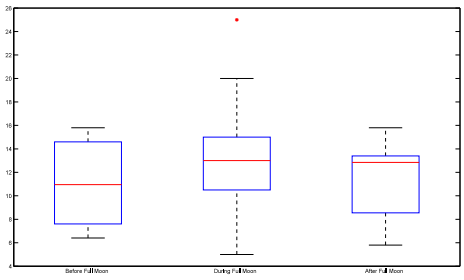


Figure 17. Boxplots for the three moon phases: before, during and after full-moon phases.

The month by phase plot shown in Figure 18 reveals some major problems with interactions between blocks (months) and treatments.There is moderate evidence to reject the null hypothesis of no interaction among phases and months (Tukey additivity test $P\{F_{1,21} \geq 4.5 \mid H_o\} = 0.046$)

With the significant interaction effect, it isn't appropriate to rely on the overall ANOVA test. All one should do is display Figure 18 and state that the effect depends on month. There is not evident reason for the interaction with August, June and February all showing the three lowest full moon admissions and all three of these months showing a decline in admissions relative to other months.

Ignoring the evident interaction effect, Table 11 shows the ANOVA results. There is a strong monthly effect on admissions ($p < 10^{-4}$) but only very modest evidence (p=0.06) for a Transylvania effect.

Table 11. ANOVA Table for Case Study 13.2.3

Source	Sum Sq.	d.f.	Mean Square	F	Prob>F
Phases	38.6	2	19.3	3.21	0.06
Months	451.1	11	41.0	6.8	0.0001
Error	132.1	22	6.0		
Total	621.8	35			

Testing the factorial model with interaction term

Instead of considering all 3 phases of the moon, if we consider just two groupings, full moon and not full moon, we can free up degrees of freedom to formally test for the interaction effect. Table 12 shows the ANOVA results. There is very strong evidence (p=0.001) for rejecting the no interaction null hypothesis.

Table 12. ANOVA Table for Case Study 13.2.3

Source	Sum Sq.	d.f.	Mean Square	F	Prob>F
Lunar Phases	36.8	1	36.8	24.9	0.0003
Months	483.8	11	43.9	29.7	$<10^{-6}$
Lunar Phases x Months	116.1	11	10.6	7.1	0.001
Error	17.8	12	1.5		
Total	621.8	35			

and July to December (Figure 20). There is an obvious pattern of having lower lottery numbers with later months.

This lack of independence among months can be formally tested with a Kruskal-Wallis ANOVA. Two separate analyses were done, one with all 12 months and another with just the two 6-month periods. There is strong evidence that the 1969 draft lottery was not random (Kruskal Wallis ANOVA of median ranks among months, $P\{\chi_{11}^2 \geq 26 \mid H_o\}<0.007$). When months are pooled {Jan-Jun vs. July-Dec}, there was striking evidence against the null hypothesis that lottery number is independent of time of year (Kruskal Wallis ANOVA of median ranks among months, $P\{\chi_1^2 \geq 16.8 \mid H_o\} < 10^{-4}$).

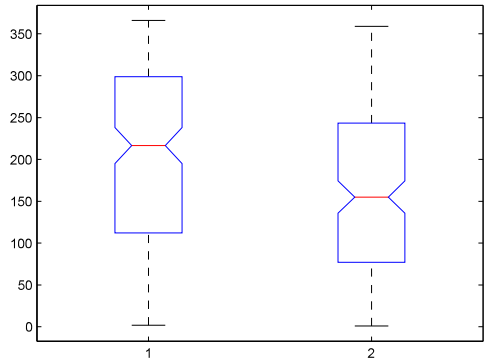


Figure 20. Notched boxplots for the lottery numbers by month (1 is January).

It is obvious that the 1969 draft lottery was not fair. Despite this, **Larsen & Marx (2006, p 830)** note that these draft lottery numbers were those that were used.

CASE STUDY 14.5.1: BASE RUNNING

There are several different strategies for base runners in baseball to go from home to 2nd base. One is narrow angle and the other is wide angle, shown at right.

22 base runners were asked to run from from home the 2nd base and their times recorded from a position 35 feet from home plate to a point 15 feet from second base. Those times are shown below. The data were analyzed with a Friedman's ANOVA.

There is very strong evidence that the median base running speed from home to 2nd base with the wide-angle approach is faster than the narrow-angle approach (Friedman's ANOVA: $\{P\{\chi_1^2\} \geq 6.55 = 0.011\}$)

CASE STUDY 14.5.1

Baseball rules allow a batter considerable leeway in how he is permitted to run from home plate to second base. Two of the possibilities are the narrow-angle and the wide-angle paths diagrammed in Figure 14.5.1. As a means of comparing the two, time trials were held involving 22 players (217). Each player ran both paths. Recorded for each runner was the time it took to go from a point 35 feet from home plate to a point 15 feet from second base. Based on those times, ranks (1 and 2) were assigned to each path for each player (see Table 14.5.1).

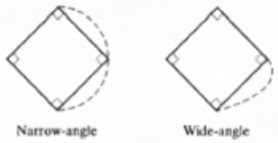


FIGURE 14.5.1: Batter's path from home plate to second base.

(Continued on next page)

Table 14.5.1. Times (sec) Required to Round First Base

Player	Narrow-Angle	Rank	Wide-Angle	Rank
1	5.50	1	5.55	2
2	5.70	1	5.75	2
3	5.60	2	5.50	1
4	5.50	2	5.40	1
5	5.85	2	5.70	1
6	5.55	1	5.60	2
7	5.40	2	5.35	1
8	5.50	2	5.35	1
9	5.15	2	5.00	1
10	5.80	2	5.70	1
11	5.20	2	5.10	1
12	5.55	2	5.45	1
13	5.35	1	5.45	2
14	5.00	2	4.95	1
15	5.50	2	5.40	1
16	5.55	2	5.50	1
17	5.55	2	5.35	1
18	5.50	1	5.55	2
19	5.45	2	5.25	1
20	5.60	2	5.40	1
21	5.65	2	5.55	1
22	6.30	2	6.25	1
		39		27

References

Cobb, G. W. 1997. Introduction to design and analysis of experiments. Springer, New York. 795 pp. [6]

Fisher, R. A. 1958. Cigarettes, cancer, and statistics. Centennial Review *2*: 151-166. [17]

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs *54*: 187-211. [7, 9]

Ioannidis, J. P. A. 2005. Why most published research findings are false. PLos Med *2*: 696-701. [7]

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. [13, 16, 28]

Quinn, G. P. and M. J. Keough. 2002. Experimental Design and Data Analysis for Biologists. Cambridge University Press. 520 p [11, 21]

Sterne, J. A. C. and G. D. Smith. 2001. Sifting the evidence — What's wrong with significance tests? British Medical Journal *322*: 226-231. [9]

Winer, B. J., D. R. Brown, and K. M. Michels. 1991. Statistical principles in experimental design, Third Edition. McGraw-Hill, New York. 1057 pp. [7, 13]

Annotated outline (with Matlab scripts) for Larsen & Marx Chapter 12-13

12 **The analysis of variance** (Week 12)
Ronald A. Fisher

12.1 **INTRODUCTION**

12.1.1 ANOVA, short for analysis of variance

12.1.2 **Comment**. Fisher was the major early developer of ANOVA

Table 12.1.1

TABLE 12.1.1

	Treatment Level			
	1	2	...	k
	Y_{11}	Y_{12}		Y_{1k}
	Y_{21}	Y_{22}		
	⋮	⋮	...	⋮
	$Y_{n_1 1}$	$Y_{n_2 2}$		$Y_{n_k k}$
Sample sizes:	n_1	n_2	...	n_k
Sample totals:	$T_{.1}$	$T_{.2}$		$T_{.k}$
Sample means:	$\overline{Y}_{.1}$	$\overline{Y}_{.2}$		$\overline{Y}_{.k}$
True means:	μ_1	μ_2		μ_k

Figure 21. Table 12.1.1.

12.2 THE F TEST

12.2.1 Distribution assumption

the Yij's will be presumed to be independent and normally distributed with μj, j=1, 2, ..., k and variance σ² (constant for all j)

12.2.2 Sum of squares

12.2.2.1 Treatment sum of squares

Theorem 12.2.1 Let SSTR be the treatment sum of squares defined for k independent random samples of sizes n1, n2, ..., and nk. Then

$$E(SSTR) = (k-1)\sigma^2 + \sum_{j=1}^{k} n_j(\mu_j - \mu)^2$$

12.2.3 Testing $\mu_1=\mu_2=\ldots=\mu_k$ when σ^2 is known

Theorem 12.2.2 When H_o: $\mu_1=\mu_2=\ldots=\mu_k$ is true, $SSTR/\sigma^2$ has a chi square distribution with k-1 degrees of freedom.

12.2.4 Testing $H_o: \mu_1=\mu_2=\ldots=\mu_k$ when σ^2 is unknown

Theorem 12.2.3 Whether or not H_o: $\mu_1=\mu_2=\ldots=\mu_k$ is true.

1. SSE/σ^2 has a chi square distribution with n-k degrees of freedom
2. SSE and SSTR are independent.

Theorem 12.2.4 If n observations are divided into k samples of sizes n_1, n_2, ..., and n_k

SSTOT=SSTR+SSE

Theorem 12.2.5 Suppose that each observation in a set of k independent random samples is normally distributed with the same variance σ^2. The μ_1, μ_2, ... and μ_k be the true means associated with the k samples. Then

a. If H_o: μ_1, μ_2, ... = μ_k is true

$$F = \frac{SSTR/(k-1)}{SSE/(n-k)}$$

b. At the α level of significance, H_o: μ_1, μ_2, ... = μ_k should be rejected if $F \geq F_{1-\alpha,\, k-1,\, n-k}$.

12.2.5 ANOVA tables

Source	df	SS	MS	F	P
Treatment	k - 1	SSTR	MSTR	$\frac{MSTR}{MSE}$	$P(F_{k-1,n-k} \geq$ observed F)
Error	n - k	SSE	MSE		
Total	n - 1	SSTOT			

Figure 22 Figure 12.2.1

Case Study 12.2.1

CASE STUDY 12.2.1

Generations of athletes have been cautioned that cigarette smoking retards performance. One measure of the truth of that warning is the effect of smoking on heart rate. In one study (72) examining that impact, six each of non-smokers, light smokers, moderate smokers, and heavy smokers undertook sustained physical exercise. Their heart rates were measured after resting for three minutes. The results appear in Table 12.2.1. Are the differences among the $\overline{Y}_{.j}$s statistically significant? That is, if

TABLE 12.2.1

	Non-Smokers	Light Smokers	Moderate Smokers	Heavy Smokers
	69	55	66	91
	52	60	81	72
	71	78	70	81
	58	58	77	67
	59	62	57	95
	65	66	79	84
$T_{.j}$	374	379	430	490
$\overline{Y}_{.j}$	62.3	63.2	71.7	81.7

Case Study 12.2.1

Hypotheses:

H_o: $\mu_{Non\text{-}smoking} = \mu_{Light} = \mu_{Moderate} = \mu_{High}$

H_a: $\mu_{Non\text{-}smoking} < \mu_{Light} < \mu_{Moderate} < \mu_{High}$

Heart rate increases with increasing smoking

Statistical test: One-way ANOVA

Alpha level = 0.05 for assessing effects and reporting confidence limits

Multiple comparison & Linear contrast

Compare non-smoking with smoking using Dunnet's procedure, or this a priori linear contrast

Test C = 1Non-1/3Light-1/3Mod-1/3 High

```
% LMcs120201_4th.m
% LMcs120201_4th.m
% Case Study 12.2.1, Smoking & Exercise Study
% Larsen & Marx (2006) Introduction to Mathematical Statistics 4th edition
% Page 740
% Written by Eugene.Gallagher@umb.edu 12/3/2010; revised 2/15/11
% There are at least 2 ways to analyze these data with Matlab: anova1 &
% anovan
DATA=[69 55 66 91
   52 60 81 72
   71 78 70 81
   58 58 77 67
   59 62 57 95
   65 66 79 84];
DATA=DATA;
Tdotj=sum(DATA);
Ymeandotj=mean(DATA);
```

```
[p,table,stats] = anova1(DATA);
pause
% ANOVA1 can only be used for balanced data. ANOVAN is the more
% general approach, but the data have to be restructured
multcompare(stats,'ctype','hsd','alpha',0.05)
pause
% Since a linear contrast was used below, the Scheffe procedure should be
% used for the pair-wise contrasts.
multcompare(stats,'ctype','scheffe','alpha',0.05)
pause
y=DATA(:); % convert the data into columns
group=repmat(1:4,6,1);group=group(:);
% Levene test downloaded from Matlab Central
Levenetest([y group],0.05);
% Levene's test indicates no evident problems
% Try a box-cox transformation, but 1st set up dummy variables
X=[ones(length(y),1) [ones(6,1);zeros(18,1)] ...
   [zeros(6,1);ones(6,1);zeros(12,1)] [zeros(12,1);ones(6,1);zeros(6,1)]];
PlotLogLike=1;LambdaValues=1;alpha=0.05;
[LambdaHat,LambdaInterval]=boxcoxlm(y,X,1,[-4:0.01:4]);
pause
[p,table,stats,terms] = anovan(y,group,'varnames','Smoking Level');
% Generate Figure 12.2.2 using LMex040307_4th.m as a model.
X=0:.01:4.5;
Y = fpdf(X,3,20);
plot(X,Y,'-k');
axis([0 4.5 0 0.8]);title('Figure 12.2.2','FontSize',20);
ax1=gca;
xlabel('y','FontSize',16),
ylabel('f_{F_{3,20}}(y)','FontSize',16);
ax1=gca;
set(ax1,'xtick',3.1,'FontSize',14);
hold on;
xf=3.1:.01:4.5;yf=fpdf(xf,3,20);
fill([3.1 xf 4.5]',[0 yf 0]',[.8 .8 1])
text(1,0.1,'Area=0.95','FontSize',18);
text(3.1,0.1,'Area=0.05','FontSize',18)
figure(gcf);pause
hold off;
% The following analysis uses the concept of linear contrast presented on
% page 751-758 in Larsen & Marx. The linear contrast between smokers and
% the one non-smoking group was set a priori, so it can be tested and
% reported with an alpha level of 0.05.
format rat; LM=center(orthpoly(1:4,3,1))
LMatrix=[   -1   1/3   1/3  1/3
```

```
    -1     0    0      1
  -3/2   -1/2   1/2   3/2
  -3/2 -11/6 -1/6 7/2
 747/310      -141/310      -909/310      303/310];
% Call's Gallagher's anova linear contrast function
anovalc(LMatrix, y, group, stats);
% To compare the slope of the 3rd linear contrast with a regression, slope
% do a linear regression.
x=[ones(length(y),1) [repmat(1,6,1);repmat(2,6,1);
  repmat(3,6,1);repmat(4,6,1)]];
[b,bint,R,RINT,STATS] = regress(y,x)
```

Source	SS	df	MS	F	Prob>F
Columns	1464.13	3	488.042	6.12	0.004
Error	1594.83	20	79.742		
Total	3058.96	23			

Figure 12.2.3

12.2.6 **Computing formulas**

Questions

12.2.7 **Comparing the Two-Sample t Test with the Analysis of Variance**

Example 12.2.2 Demonstrating the equivalence of the Students' t and ANOVA F tests

Questions

12.3 **Multiple comparisons: Tukey's method**

12.3.1 **Multiple comparisons problem** Keeping the probability of Type I error small even when many tests are performed.

12.3.2 **A Background result: the studentized range distribution**

Definition 12.3.1 The studentized range

Theorem 12.3.1

Case Study 12.3.1 Serum protein-bound antibiotics.

```
% LMcs120301_4th.m
% Larsen & Marx (2006) Introduction to Mathematical Statistics 4th edition
% Case Study 12.3.1 Page 749
% Written by Eugene.Gallagher@umb.edu 2/15/11
% There are at least 2 ways to analyze these data with Matlab & ANOVA
DATA=[29.6 27.3 5.8 21.6 29.2
   24.3 32.6  6.2 17.4 32.8
   28.5 30.8 11.0 18.3 25.0
   32.0 34.8  8.3 19.0 24.2]
Tdotj=sum(DATA)
Ymeandotj=mean(DATA)
[p,table,stats] = anova1(DATA);
```

```
pause % The pause is so that the boxplot can be examined before it is
    % overwritten by multcompare's graph
multcompare(stats,'ctype','hsd', 'alpha',0.05)
% ANOVA1 can only be used for balanced data,i.e., data with an equal number
% of cases per treatment level
% ANOVAN is the more
% general approach, but the data have to be restructured so that they are
% all in one column.
% This will be introduced later in the chapter
% multcompare(stats,'ctype','hsd')
y=DATA(:); % convert the data into columns
group=repmat(1:5,4,1);group=group(:);
[p,table,stats,terms] = anovan(y,group,'varnames','Among Antibiotics');
multcompare(stats,'ctype','hsd');
```

12.4 TESTING HYPOTHESES WITH CONTRASTS

Definition 12.4.1

Orthogonal contrasts: Two contrasts are said to be orthogonal if $\sum_{j=1}^{k} \frac{c_{1j}c_{2j}}{n_j} = 0$.

Definition 12.4.2

Theorem 12.4.1 **mutually orthogonal contrasts**

Theorem 12.4.2

12.4.1

12.4.2

12.4.3

12.4.4

12.4.5

12.4.6

12.4.7 Testing subhypotheses with orthogonal contrasts

Definiton 12.4.1

Comment

Definition 12.4.2

Theorem 12.4.1

Theorem 12.4.2

Comment

Case Study 12.4.1 Infant walking

```
% LMcs120401_4th.m
% Larsen & Marx (2006) Introduction to Mathematical Statistics 4th edition
% Case Study 12.4.1 Page 755-756
% Are there differences in infant walking times based on exercise?
% Written by Eugene.Gallagher@umb.edu 12/3/2010, revised 2/15/11
% There are at least 2 ways to analyze these data with Matlab & ANOVA
% Note that I'm using 'not a number' to balance the matrix.
DATA=[9 11 11.5 13.25
```

```
   9.5 10 12 11.5
   9.75 10 9 12
   10 11.75 11.5 13.5
   13 10.5 13.25 11.5
   9.5 15 13 NaN]
sumDATA=sum(DATA);
meanDATA=mean(DATA);
[p,table,stats] = anova1(DATA);
pause
multcompare(stats,'ctype','hsd');
% ANOVA1 can only be used for balanced data,i.e., data with an equal number
% of cases per treatment level
% ANOVAN is the more
% general approach, but the data have to be restructured so that they are
% all in one column.
pause
% ANOVA1 can only be used for balanced data,i.e., data with an equal number
% of cases per treatment level
% ANOVAN is the more
% general approach, but the data have to be restructured so that they are
% all in one column.
pause
y=DATA(:); % convert the data into columns; drop the NaN elements
group=repmat(1:4,6,1);group=group(:);i=~isnan(y);y=y(i);group=group(i);
[p,table,stats] = anovan(y,group,'varnames','Exercise');
multcompare(stats,'ctype','hsd');
% Levene test downloaded from Matlab Central
Levenetest([y group],0.05);
% Program the linear contrast from Definition 12.4.2 (confirmed with
% PASW oneway and UNIANOVA)
% By using fractional coefficients, the difference in means for the
% contrasts will be of the right size.
LMatrix = [1   -1/3     -1/3     -1/3
             1       -1         0        0
            1/2      1/2 -1/2         -1/2
            0         0         1        -1]
% Call's Gallagher's anova linear contrast function
anovalc(LMatrix, y, group, stats);
% Because of the two outliers, use a Kruskal-Wallis ANOVA to check results
[P,ANOVATAB,STATS] = kruskalwallis(y,group)
multcompare(STATS)
```

12.5 DATA TRANSFORMATIONS

Example 12.5.1

Example 12.5.2

Questions

12.6 Taking a second look at statistics (putting the subject of statistics together — the contributions of Ronald A. Fisher)
12.6.1
12.6.2
12.6.3
Appendix 12.A.1. Minitab applications
Appendix 12.A.2. A proof of theorem 12.2.2
Appendix 12.A.3. The distribution of SSTR/(k-1)/SSE(n-k) when $\mathbf{H_1}$ is true
Definition 12.A.3.2
Theorem 12.A.3.2

13 Randomized Block Designs

13.1 **INTRODUCTION**
13.2 **THE F TEST FOR A RANDOMIZED BLOCK DESIGN**

Theorem 13.2.1 & 13.2.2

Theorem 13.2.1. *Suppose that k treatment levels are measured over a set of b blocks. Then*

a. $SSTOT = SSTR + SSB + SSE$

b. *SSTR, SSB, and SSE are independent random variables.*

Proof. The independence of the three terms that combine to give *SSTOT* can be established using the same approach that was taken in Chapter 12. The details will be omitted. □

Theorem 13.2.2. *Suppose that k treatment levels, with means $\mu_1, \mu_2, \ldots, \mu_k$, are measured over a set of b blocks, where the block effects are $\beta_1, \beta_2, \ldots, \beta_b$. Then*

a. *When $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ is true, $SSTR/\sigma^2$ has a chi square distribution with $k - 1$ degrees of freedom.*

b. *When $H_0: \beta_1 = \beta_2 = \ldots = \beta_b$ is true, SSB/σ^2 has a chi square distribution with $b - 1$ degrees of freedom.*

c. *Regardless of whether the μ_j's and/or the β_i's are equal, SSE/σ^2 has a chi square distribution with $(b - 1)(k - 1)$ degrees of freedom.*

Theorem 13.2.3. *Suppose that k treatment levels with means $\mu_1, \mu_2, \ldots, \mu_k$ are measured over a set of b blocks. Then*

a. *If $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ is true,*

$$F = \frac{SSTR/(k-1)}{SSE/(b-1)(k-1)}$$

has an F distribution with $k - 1$ and $(b - 1)(k - 1)$ degrees of freedom.

b. *At the α level of significance, $H_0: \mu_1 = \mu_2 = \ldots = \mu_k$ should be rejected if $F \geq F_{1-\alpha,k-1,(b-1)(k-1)}$.*

Theorem 13.2.4

Case Study 13.2.1 Acrophobia
%LMcs130201_4th.m

Case Study 13.2.2 Rat poison
%LMcs130202_4th.m
% Larsen & Marx (2006) Introduction to Mathematical Statistics 4th edition
% Case Study 13.2.1 Page 780-781
% Written by Eugene.Gallagher@umb.edu 12/3/2010; revised 12/7/2010
% Calls other files
DATA=[13.8 11.7 14.0 12.6
12.9 16.7 15.5 13.8
25.9 29.8 27.8 25.0

```
18.0 23.1 23.0 16.9
15.2 20.2 19.0 13.7];
boxplot(DATA,'sym','r*','labels',...
{'Plain','Butter Vanilla','Roast Beef','Bread'});
figure(gcf);pause
plot(DATA');
ax1=gca;
set(ax1,'xtick',[1 2 3 4]);legend('1','2','3','4','5');
xlabel('Therapy');ylabel('HAT Score');
figure(gcf);pause
pause % Needed so can see the box plots
sumDATA=sum(DATA);
meanDATA=mean(DATA);
y=DATA(:);
% convert the data into columns;
g1=repmat(['S1';'S2';'S3';'S4';'S5'],4,1);
g2=[repmat('Pl',5,1);repmat('Bu',5,1);repmat('RB',5,1);repmat('Br',5,1)];
% find and delete any NaN elements, if any
i=~isnan(y);y=y(i);g1=g1(i,:);g2=g2(i,:);
[p,table,stats] = anovan(y,{g1 g2},'model','linear',...
   'varnames',{'Survey';'Flavor'});
[c,m,h] = multcompare(stats,'display','on','dimension',2);
figure(h)
title(' '),xlabel(' ');figure(gcf)
pause;
% Since the treatments were not replicated within blocks, Tukey's test
% for additivity should be run:
r1=repmat([1:5]',4,1);
r2=[repmat(1,5,1);repmat(2,5,1);repmat(3,5,1);repmat(4,5,1)];
X=[y r1 r2];
adTukeyAOV2(X,2,0.05)
```

13.2.1 **Tukey Comparisons for Randomized Block Data**

Theorem 13.2.5

Example 13.2.1

Tukey tests already incorporated in previous m.files for the case studies

13.2.2 **Contrasts for randomized block designs**

Case Study 13.2.3

```
%LMcs130203_4th.m
%LMcs130203_4th.m
% Larsen & Marx (2006) Introduction to Mathematical Statistics 4th edition
% Case Study 13.2.3 Page 778-779, The Transylvannia5 effect
% An example of linear contrasts for Randomized Block data.
% Written by Eugene.Gallagher@umb.edu 2/15/11
% Calls Trujillo-Ortiz et al. adTukeyAOV2.m from Matlab file central
```

```
% Tukey's test for additivity.
fprintf('\nAnalysis of Case Study 13.2.3: The Transylvannia Effect\n')
DATA=[6.4 5.0 5.8
   7.1 13.0 9.2
   6.5 14.0 7.9
   8.6 12 7.7
   8.1 6 11
   10.4 9 12.9
   11.5 13.0 13.5
   13.8 16.0 13.1
   15.4 25.0 15.8
   15.7 13.0 13.3
   11.7 14.0 12.8
   15.8 20 14.5];
[R,C]=size(DATA);
boxplot(DATA,'sym','r*','labels',...
{'Before Full Moon','During Full Moon','After Full Moon'});
figure(gcf);pause

plot(DATA');
ax1=gca;
set(ax1,'xtick',[1 2 3]);
set(ax1,'XtickLabel',...
  {'Before Full Moon','During Full Moon','After Full Moon'},'FontSize',9);
legend('Au','Se','Oc','Nv','De','Ja','Fe','Mr','Ap','My','Jn','Jl');
xlabel('Moon Phase');ylabel('Hospital Admission Rates');
figure(gcf);pause
pause % Needed so can see the box plots
sumDATA=sum(DATA);
meanDATA=mean(DATA);
% Since the design is balanced, either anova2 or anovan can be used.
[p,table,stats]=anova2(DATA,1)
% This will produce the ANOVA table as a figure. The results are
% printed out in table. stats could be sued for multcompare
pause
% The data can also be analyzed using anovan, producing identical results.
% anovan allows labeling of the ANOVA table.
y=DATA(:);
% convert the data into columns; drop the NaN elements
g1=repmat(['Au';'Se';'Oc';'Nv';'De';'Ja';'Fe';'Mr';'Ap';'My';'Jn';'Jl'],3,1);
g2=[repmat('BFM',12,1);repmat('DFM',12,1);repmat('AFM',12,1)];
group=[repmat(1,12,1) repmat(2,12,1) repmat(3,12,1)];
% find and delete any NaN elements, if any
i=~isnan(y);y=y(i);g1=g1(i,:);g2=g2(i,:);group=group(i);
% Use Trujillo-Ortiz's Levenestest
```

```
levenetest([y group],0.05);
[p,table,stats] = anovan(y,{g1 g2},'model','linear',...
   'varnames',{'Months';'Lunar Cycles'})
% This is Example 13.2.1, comparing treatments
[c,m,h] = multcompare(stats,'ctype','tukey-kramer','display','on',...
   'dimension',2);
fprintf('Pairwise Difference \tLower 95\tEstimate\tUpper 95\n');
fprintf('\t\t%1.0f - %1.0f\t\t\t   %4.1f\t\t%4.1f\t\t%4.1f\n',c')
figure(h);
title(' '),xlabel(' ');xlabel('Hospital Admission Rates');
title('Case Study 13.2.3'),
figure(gcf)
pause;

% Since the treatments were not replicated within blocks, a test
% for additivity should be run. If replicates were available a formal
% block by interaction test could have been run.
r1=repmat([1:R]',C,1);
r2=[repmat(1,R,1);repmat(2,R,1);repmat(3,R,1)];
X=[y r1 r2];
adTukeyAOV2(X,2,0.05)
% Note that there is evidence (p=0.046) to reject the additivity assumption

fprintf('\nCheck the additivity assumption with just 2 groups:\n')
% Reanalyze the data pooling 2 non-full moon periods.
D=mean(DATA(:,[1 3])')';D=[D DATA(:,2)];
plot(D');
ax1=gca;
set(ax1,'xtick',[1 2]);
set(ax1,'XtickLabel',...
   {'Not Full Moon','Full Moon'},'FontSize',9);
legend('Au','Se','Oc','Nv','De','Ja','Fe','Mr','Ap','My','Jn','Jl');
xlabel('Moon Phase');ylabel('Hospital Admission Rates');
figure(gcf);pause
pause % Needed so can see the box plots
[r,c]=size(D);
r1=repmat([1:r]',c,1);
r2=[repmat(1,r,1);repmat(2,r,1)];
X=[D(:) r1 r2];
adTukeyAOV2(X,2,0.05);
% p=0.0367; so still a strong interaction evident.

[p2,table2,stats2] = anovan(D(:),[r1 r2],'model','linear',...
   'varnames',{'Months';'Lunar Cycles'})
```

```
% Not covered in Larsen & Marx, but now it is possible to test formally
% for the interaction term.
Y=DATA(:);
G1=repmat(['Au';'Se';'Oc';'Nv';'De';'Ja';'Fe';'Mr';'Ap';'My';'Jn';'Jl'],3,1);
% set two groups: Not full moon and During Full moon
G2=[repmat('NFM',12,1);repmat('DFM',12,1);repmat('NFM',12,1)];
Group=[repmat(1,12,1) repmat(2,12,1) repmat(1,12,1)];
% find and delete any NaN elements, if any
i=~isnan(Y);Y=Y(i);G1=G1(i,:);G2=G2(i,:);Group=Group(i);
% Use Trujillo-Ortiz's Levenestest
levenetest([Y Group],0.05);
[p,table,stats] = anovan(Y,{G1 G2},'model',2,...
   'varnames',{'Months';'Lunar Cycles'})
% There should be no formal analysis of main effects of the main effects if
% I was taught in my graduate statistics class that if there is a
% significant interaction, show the interactions in an effects plot,discuss
% them and end the analysis.

% If there were no interactions, this would be a valid post hoc analysis:
% The following analysis uses the concept of linear contrast presented on
% page 751-758 in Larsen & Marx. The linear contrast between the full moon
% period and the other two phases was set a priori, so it can be tested and
% reported with an alpha level of 0.05.
LMatrix=[-1/2 1 -1/2];
planned=0;
anovalc(LMatrix, y, group,stats,planned)
```

```
function anovalc(LMatrix, y, group, stats, planned)
% format anovaLC(LMatrix, y, group, stats,planned)
% Input LMatrix
% Each row of the LMatrix should contain a linear contrast
%   LMatrix = [-1 1 0 0;-0.5 0.5 0 0] will return identical contrasts
%   y=data in a column vector
%   group is the column vector indicating group membership
%   stats is output from anova1, anova2 or anovan
%   planned =1 if the contrast was planned a priori
%   planned =0 if the contrast was not planned, in which case Scheffe
%            multipliers will be used.
% Written by Eugene D. Gallagher 12/7/2010
if nargin<5;planned=1;end
[R,C]=size(LMatrix);
% Create placeholder vectors for the output of the data
G=unique(group);   % Contains indices indicating treatment membership
n=zeros(1,C);
meanDATA=zeros(1,C);
sumDATA=zeros(1,C);
```

```
SSC=zeros(R,1);
F=zeros(R,1);
Fprob=zeros(R,1);
g=zeros(R,1);
seg=zeros(R,1);
tdf=tinv(0.975,stats.dfe);
for j=1:C
   i=find(group==G(j));
   n(j)=length(i);
   sumDATA(j)=sum(y(i));
   meanDATA(j)=mean(y(i));
end
for i=1:R % do each linear contrast
      sumLM=sum(LMatrix(i,:));
      sumabsLM=sum(abs(LMatrix(i,:)));
      fprintf('\nContrast Result Number %1.0f:\n',i)
      format rat
       disp(LMatrix(i,:));
       format
      if abs(sumLM)>=3*eps
         error('Linear contrasts must sum to 0');
      elseif abs((sumabsLM-2))>eps
         % This corrects an issue that is found in PASW, in which
         % ANOVA doesn't allow fractional linear contrasts and the
         % effects size and standard error are wrong if a contrast
         % such as [-1 -1 2 0] is used, in which case the sum of
         % the absolute value of the contrasts is 4, not 2 and the
         % estimated effect size and standard are 2x too large.
         LMatrix(i,:)=1/(sumabsLM/2)*LMatrix(i,:);
         fprintf(...
            'Linear Contrast %1.0f converted to equivalent form:\n',i)
         format rat
            disp(LMatrix(i,:))
         format
      end
      SSC(i)=sum(LMatrix(i,:).*sumDATA./n)^2/sum(LMatrix(i,:).^2./n);
      % Calculate the value of the linear contrast g (from Sleuth)
      g(i)=sum(LMatrix(i,:).*meanDATA);
      % The equation for the standard error of the linear contrast
      % can be found in Statistical Sleuth Chapter 6
      seg(i)=sqrt(stats.mse).*sqrt(sum(LMatrix(i,:).^2./n));
      F(i)=SSC(i)/stats.mse;
      Fprob(i)=1-fcdf(F(i),1,stats.dfe);
      if planned==1
         fprintf('The difference in means is %5.2f +/- %5.2f\n',...
```

```
            g(i),seg(i)*tdf)
        else
           Scheffe=sqrt((C-1)*finv(1-0.05,C-1,stats.dfe));
           fprintf(...
      'The difference in means is %5.2f +/- %5.2f (Scheffe Interval)\n',...
            g(i),seg(i)*Scheffe)
        end
        fprintf('\n Source      SS     df    MS    F   Prob\n')
        fprintf(...
          'Contrast    %4.1f     1   %4.1f %4.1f %5.3g\n',SSC(i),SSC(i),...
           F(i),Fprob(i))
  fprintf('   Error   %4.1f    %2.0f %5.2g\n',stats.mse*stats.dfe,...
     stats.dfe,stats.mse)
end
```

Questions 784-788

13.3 THE PAIRED *t* TEST

Theorem 13.3.1

Case Study 13.3.1

```
% LMcs130301_4th.m
% Case Study 13.3.1 p 790-791 in
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% Written by Eugene.Gallagher@umb.edu 11/14/10, revised 1/21/11
% Revised: 1/21/11
% X(:,1) Hemoglobin before 60 km walk and X(:,2) after 60-km walk
X=[14.6 13.8
   17.3 15.4
   10.9 11.3
   12.8 11.6
   16.6 16.4
   12.2 12.6
   11.2 11.8
   15.4 15.0
   14.8 14.4
   16.2 15.0]
D=(X(:,2)-X(:,1));hist(D);figure(gcf)
[H,P,CI,STATS] = TTEST(X(:,1),X(:,2),0.05,'both')
fprintf('The paired t test 2-tailed p=%6.4f\n',P);
[p,h,stats] = signtest(D,0,0.05,'method','exact');
fprintf('The sign test exact p=%6.4f\n',p);
[p,h,stats] = signtest(D,0,'method','approximate');
fprintf('The sign test approximate p=%6.4f\n',p);
[P,H,STATS] = signrank(X(:,1),X(:,2),'alpha',0.05,'method','exact');
fprintf('The sign rank test exact p=%6.4f\n',P);
[P,H,STATS] = signrank(X(:,1),X(:,2),'alpha',0.05,'method','approximate');
fprintf('The sign rank test approximate p=%6.4f\n',P);
```

```
[P,W]=wilcoxsignrank(X(:,1),X(:,2));
fprintf('The sign rank test approximate p=%6.4f\n',P);
```

Case study 13.3.2

```
% LMcs130302_4th.m
% Larsen & Marx (2006) Introduction to Mathematical Statistics, 4th edition
% page 791. A case study solved by the paired t test
% Written by Eugene.Gallagher@umb.edu 11/14/10 Revised 11/16/10
% X(:,1) Alamo rent-a-car, X(:,2)Avis rent-a-car
X=[48.99 51.99
   49.99 55.99
   42.99 47
   34.99 42.99
   42.99 44.95
   33.99 38.99
   59 69
   42.89 50.99
   47.99 49.99
   47.99 53.99
   35.99 42.99
   44.99 44.99]
D=(X(:,2)-X(:,1));hist(D);figure(gcf)
[H,P,CI,STATS] = ttest(X(:,1),X(:,2),0.05,'left')
fprintf('The paired t test 1-tailed p=%6.4f\n',P);
[p,h,stats] = signtest(D,0,0.05,'method','exact');
fprintf('The sign test exact p=%6.4f\n',p);
[p,h,stats] = signtest(D,0,'method','approximate');
fprintf('The sign test approximate p=%6.4f\n',p);
[P,H,STATS] = signrank(X(:,1),X(:,2),'alpha',0.05,'method','exact');
fprintf('The sign rank test exact p=%6.4f\n',P);
[P,H,STATS] = signrank(X(:,1),X(:,2),'alpha',0.05,'method','approximate');
fprintf('The sign rank test approximate p=%6.4f\n',P)
```

13.3.1 Criteria for Pairing
13.3.2 **The equivalence of the paired t test and the randomized block ANOVA when k = 2**

Questions 795-796

13.4 **Taking a second look at statistics (choosing between a two-sample t test and a paired t test)**

Example 13.4.1 Comparing two weight loss plans

Example 13.4.2 Comparing two eye surgery techniques

Appendix 13.A.1 Minitab applications

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
%
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];
hist(D);figure(gcf)
[H,P,CI,STATS] = ttest(D,7.39);
fprintf('\nThe paired 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] = signtest(D,7.39,0.05,'method','exact');
fprintf('The 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);
[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);
```

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
%
D=[4.8 4.0 3.8 4.3 3.9 4.6 3.1 3.7];
hist(D);figure(gcf)
[H,P,CI,STATS] = ttest(D,3.55);
fprintf('\nThe paired t test 2-tailed p=%6.4g\n',P);
fprintf('The mean caffeine = %4.2f with 95%% CI: [%4.2f %4.2f]\n',...
   mean(D), CI(1),CI(2));
[p,h,stats] = signtest(D,3.55,0.05,'method','exact');
fprintf('The sign test exact 2-tailed p=%6.4g\n',p);
[p,h,stats] = signtest(D,3.55,'method','approximate');
fprintf('The sign test approximate 2-tailed p=%6.4g\n',p);
[P,H,STATS] = signrank(D,3.55,'alpha',0.05,'method','exact');
fprintf('The sign rank test exact 2-tailed p=%6.4g\n',P);
[P,H,STATS] = signrank(D,3.55,'alpha',0.05,'method','approximate');
fprintf('The sign rank test approximate 2-tailed p=%6.4g\n',P);
```

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
%
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];
hist(D(:,1)-D(:,2));figure(gcf)
[H,P,CI,STATS] = ttest(D(:,1),D(:,2),0.05,'right');
fprintf('\nThe paired t test 1-tailed p=%6.4g\n',P);
[p,h,stats] = signtest(D(:,1),D(:,2),0.05,'method','exact');
fprintf('The sign test exact 1-tailed p=%6.4g\n',p/2);
[p,h,stats] = signtest(D(:,1),D(:,2),'method','approximate');
fprintf('The sign test approximate 1-tailed p=%6.4g\n',p/2);
[P,H,STATS] = signrank(D(:,1),D(:,2),'alpha',0.05,'method','exact');
fprintf('The sign rank test exact 2-tailed p=%6.4g\n',P/2);
[P,H,STATS] = signrank(D(:,1),D(:,2),'alpha',0.05,'method','approximate');
fprintf('The sign rank test approximate 2-tailed p=%6.4g\n',P/2);
```

Questions p 809-810

14.3 **WILCOXON TESTS**

14.3.1 **Testing H_o: $\mu=\mu_o$**

Theorem 14.3.1

14.3.2 **Calculating $p_w(w)$**

14.3.3 **Tables of the cdf, $F_W(w)$**

Case Study 14.3.1 Swell sharks

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

```
[p,h,stats] = signtest(D,M,'method','approximate');
fprintf('The sign test approximate 2-tailed p=%6.4g\n',p);
[P,H,STATS] = signrank(D,M,'alpha',0.05,'method','exact');
fprintf('The sign rank test exact 2-tailed p=%6.4g\n',P);
[P,H,STATS] = signrank(D,M,'alpha',0.05,'method','approximate');
fprintf('The sign rank test approximate 2-tailed p=%6.4g\n',P);
```

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 Methadone

```
% 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
%
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 11/16/10
%
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];
hist(D(:,1)-D(:,2));figure(gcf)
[H,P,CI,STATS] = ttest(D(:,1),D(:,2),0.05,'both');
fprintf('\nThe paired t test 2-tailed p=%6.4g\n',P);
[p,h,stats] = signtest(D(:,1),D(:,2),0.05,'method','exact');
fprintf('The sign test exact 2-tailed p=%6.4g\n',p);
[p,h,stats] = signtest(D(:,1),D(:,2),'method','approximate');
fprintf('The sign test approximate 2-tailed p=%6.4g\n',p);
[P,H,STATS] = signrank(D(:,1),D(:,2),'alpha',0.05,'method','exact');
fprintf('The sign rank test exact 2-tailed p=%6.4g\n',P);
[P,H,STATS] = signrank(D(:,1),D(:,2),'alpha',0.05,'method','approximate');
fprintf('The sign rank test approximate 2-tailed p=%6.4g\n',P);
```

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 Wilcoxrsexact
% Written by Eugene.Gallagher@umb.edu
% Revised 11/14/10

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

```
  if ~largesample
     pexuptail=wilcoxrsexact(n,m,W,R);
     if pexuptail<=0.5
        pvalue=2*pexuptail;
     else
        pvalue=2*(1-pexuptail);
     end
  end
end
if largesample
   % Large sample approximation;% Hollander & Wolfe p. 108
   EoW=(n*(m+n+1))/2;
   % Calculate the variance of W, without ties and with ties.
   if isempty(T)  % Size of tied groups from ties.m
      VaroW=(m*n*(m+n+1))/12;
   else
      VaroW=(m*n)/12*(m+n+1-(sum((T-1).*T.*(T+1)))/((m+n)*(m+n-1)));
   end
   Wstar=(W-(n*(m+n+1)/2))/sqrt(VaroW);  % Without ties, tends to an asymptotic N(0,1)
distribution.
   % Find the 2-tailedprobability of Wstar from the standard normal distributioin
   pvalue=erfc(abs(Wstar)/sqrt(2));
   % Note that the exact p values are tabulated, and an exact test, even in the presence of ties
   % can be performed, see pp. 113-116 in Hollander & Wolfe.
end
```

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

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

% An exact conditional distribution with ties follows Hollander & Wolfe p. 115
if nargin<4
   ranks=1:n+m;
   notiedr=logical(1);
else
   if length(ranks)<n+m
      error(...
         sprintf(...
         'Number of ranks (%d) doesn''t match n+m (%d)\n',...
         length(ranks),n+m));
   end
   ranks=sort(ranks);
   notiedr=logical(0);  % could do a check to see if there really are ties with ties.m
end
ranks=ranks(:);
fudranks=flipud(ranks);
N=n+m;
r = n;
ncomb = nchoosek(N,r);  % Matlab's built-in combination function.
if W>=n*(n+m+1)-W;
   uppertail=logical(1);
else
   W=n*(n+m+1)-W;
   uppertail=logical(0);
end
if W>sum(fudranks(1:n))
   if uppertail
      error('W impossibly large')
   else
      error('W impossibly small')
   end
elseif W==sum(fudranks(1:n)) & notiedr
   if uppertail
      pexuptail=1/ncomb;
   else
      pexuptail=(ncomb-1)/ncomb;
   end
   return
end
  % Strauss's combval lists combinations in c in lexicographic
  % order, thus the critical values for sum(C) are larger than
  % observed W. We can speed up the process by using
```

```
% Wstar=min(W,n*(m+n+1)-W) and exiting loop when Wstar fails
% to be less than critical value
if ncomb>1e6
   t=sprintf(...
      '%d combinations, T=%d min (1e6 combs take 1 min on p4)\n',...
      ncomb,round(ncomb/1e6));
   toomany=menu(t,'Stop','Continue');
   if toomany==1
      return
   end
end
% c = zeros(ncomb,r);     % Don't need to store values.
Tally=0;
j = zeros(1,r);

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

```
function [T,R,ind]=ties(A)
% format: [T,R,ind]=ties(A)
% a function to return a row vector of tied groups, T,
% Ranks R (including average ranks) and indices of tied elements
% needed to calculate variance of S using Kendall's
% variance formula & Spearman's r.
% input: A is a row or column vector
% T: a row vector containing number of members of tied groups
% T=0 if there are no tied groups
% sum(T) is equal to the number of tied elements.
% each element of T equals the number in each tied group
% tied groups are sorted in ascending order.
% Examples: A=[1 2 3];[T,R,i]=ties(A)=>   T=0,R=[1 2 3],i=[]
%          A=[1 2 3 1];          T=2,R=[1.5 3 4 1.5],i=[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
% i: indices of tied elements, sorted by rank; sorted tied elements=A(i);
% 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)<=eps 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

Index

alpha level 8, 10, 30-32, 40
alternative hypothesis 12
ANOVA 1, 4-10, 12-16, 18-20, 22, 24, 26-31, 33-35, 38, 41, 43
factorial 6, 7, 27
Model I 5
Model II 5
nested 7
Bonferroni multiple comparison procedure 5, 8-11
Boxplot 17, 25, 27, 33, 37, 38, 47
Box-Cox transformation 5, 12
Chi square distribution 30
combinations 48-51
Confidence interval 13
critical value 14, 51
degrees of freedom 6, 7, 9, 14-16, 18, 20, 22, 24, 26, 30, 33, 42
Distributions
F 9
Expected value 9, 48
Experiment 7, 8, 10, 11, 17, 19
experimental design 21, 29
Fisher 17, 29, 36
heteroscedasticity 7, 8, 11, 25
honest significant difference 5, 9, 11, 16, 18, 22, 24
independence 28
Kruskal-Wallis 1, 5, 20, 28, 53
Least significant difference 5, 9, 10
least squares 15
level of significance 30
Levene's test 8, 11-13
likelihood 13
linear contrast 5, 8, 14, 15, 19, 20, 31-33, 35, 40, 41
Linear regression 33
logit 8
Matlab 12, 14, 16, 21, 29, 31-35, 37, 55
Maximum likelihood 13
mean square 5, 7, 9, 10, 14-16, 18, 20, 22, 24, 26
Median 28

Mixed model 7, 27
nonparametric 43, 56
null hypothesis 9, 10, 14-16, 20, 26, 28, 48
P-Value 14-16, 22, 48
paired data 45, 46
Parameter 12
Poisson 8
population 7, 8
Power 7, 12
Probability 1, 5, 7, 8, 10, 33
randomized block design 36
Regression 15, 33
Repeated measures 6, 7
Residuals 8
Runs 55
sample 7, 9, 13, 33, 43, 44, 46-49
Sign test 42-47
Standard error 41
Statistic 6, 9, 14, 48, 49, 55
Studentized range 11, 33
Student's t 6
Sums of squares 8
Test statistic 6, 55
Tukey-Kramer test 5, 11
Type I error 7, 33
variable 13, 23
variance 1, 5-8, 11, 13, 17, 19, 29, 30, 33, 48, 49, 52
Wilcoxon rank sum test 19, 47
Wilcoxon signed rank test 45, 46