Design and Analysis of Experiments

Larry Winner

2025-06-23

Contents

Preface

These notes are for STA 4211 at the University of Florida. All examples are based on published articles, though most data has been generated to match summary statistics.

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.1.3
```

```
## -- Attaching packages ------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.8     v dplyr   1.0.10
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

```
## Warning: package 'tibble' was built under R version 4.1.3
```

```
## Warning: package 'tidyr' was built under R version 4.1.2
```

```
## Warning: package 'readr' was built under R version 4.1.2
```

```
## Warning: package 'purrr' was built under R version 4.1.2
```

```
## Warning: package 'dplyr' was built under R version 4.1.3
```

```
## Warning: package 'forcats' was built under R version 4.1.2
```

```
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Chapter 1

Introduction

In this chapter, we will introduce terminology and briefly describe experimental designs used in a wide variety of research fields.

1.1 Terminology

Studies involve making observations on individual units under various conditions. When units have been randomly assigned to a treatment condition, they may be referred to as **Experimental Units**. When units have been sampled and observed from an existing population they may be referred to as **Observational Units**. In some studies, larger blocks of units may be randomized to a treatment, with subunits being observed (as when classrooms are randomized to various conditions), in this setting the measured units (e.g. students) may be referred to as **Measurement Units**.

- **Experimental Studies** - Observational units are assigned at random to treatments/conditions
  - **Experimental Factors** - Conditions with two or more levels that are assigned (at random) to units. Many experiments include multiple factors and treatments are the combinations of factor levels.
- **Observational Studies** - Observational units are sampled from various populations/subpopulations
  - **Observational Factors** - Set of populations/subpopulations observed in a study
- **Mixed Studies** - Studies that have both experimental and observational factors (not to be confused with Mixed Effects Designs)

Example 1.1 - Effect of Container Size on Food Intake

A study was conducted to compare three conditions on food intake in students [Marchiori et al., 2012]. A sample of 88 subjects was obtained, and randomly assigned to one of 3 conditions involving bowl size and portion of M&Ms while watching a television program: 1) Medium portion size/Small container ($n_1 = 30$), 2) Medium portion/Large container ($n_2 = 29$), and 3) Large portion/Large container ($n_3 = 29$). Researchers measured the food intake among the students. Note that this is an **Experimental Study**, as Subjects were assigned at random to treatments.

∇

Example 1.2 - Waste in the Mediterranean Sea

A study measured the amounts of Natural and Artificial floating debris at samples of transects at 14 locations in the Mediterranean Sea [Suaria and Aliani, 2014]. The researchers wished to compare the amounts of debris of each type among the locations. Note this is an **Observational Study**, as transects were sampled within the selected locations.

∇

Example 1.3 - Quilting Layers in Body Armour

A study was conducted to determine whether the number of quilting layers improved the fragment protective performance of body armour [Carr et al., 2012]. The researchers sampled 36 specimens of each number of layers (1,2,3, and 5), assigning 12 at random to each of 3 bullet impacts (slow, fast, and edge). The energy absorbed by each specimen was measured. Note this is a **Mixed Study**, as Layers is Observational, and Impact is Experimental.

∇

1.2 Basics of Controlled Experiments

In this section we describe some aspects and terminology of controlled experiments and give brief examples of them.

- **Explanatory Factors** – Conditions (with 2 or more levels) that are assigned to units.
- **Crossed Factors** – Factors with levels that are the same within levels of the other factor(s)
- **Nested Factors** – Factors with levels that are different within levels of the other factor(s)
- **Treatments** – Combinations of factor levels given to units
- **Experimental Units** – Units used in the study, which are subject to randomization to treatments.
- **Randomization Process** - Use of random number generator to assign units to treatments
- **Response(s)** - Outcome measurement(s) obtained from treated units

Example 1.4 - Reading Times on 3 Electronic Readers at 4 Illumination Levels

An experiment was conducted to measure reading times on 3 e-reader devices at 4 illumination levels [Chang et al., 2013]. A sample of 60 subjects were randomly assigned so that 5 received each of 12 treatments (combinations of 3 e-reader models and 4 illumination levels). The experiment was **Crossed** in the sense that each e-reader model was set at the same 4 illumination levels (200, 500, 1000, 1500Lx). The time to complete a reading task was the measured response.

∇

Example 1.5 - Combability of Hair for Two Shampoo Formulations

An experiment was conducted to compare two shampoo formulations with respect to combability of hair [Garcia and Diaz, 1976]. A sample of 16 hair swatches were created and randomly assigned, such that 8 received shampoo A, and 8 received shampoo B. Each swatch was washed 5 times, and the combability was measured. The experiment was **Nested**, as the swatches receiving shampoo A were different from the swatches receiving shampoo B. Note that the swatches are the experimental units, as they are randomly assigned to treatments (shampoos). The replicates measured are the measurement units.

∇

1.3 Completely Randomized Design (CRD)

In the **Completely Randomized Design**, experimental units are randomly assigned to treatments, and responses are recorded on the units after treatments are applied. We will refer to the number of treatments as r, with the number of replicates for the i^{th} treatment being n_i. When all treatments have the same number of replicates ($n_1 = ... = n_r = n$), the design is said to be **balanced**. The total sample size across all treatments will be labelled $n_T = n_1 + \cdots + n_r$. When the experiment is balanced, $n_T = rn$.

The statistical model for the One-Way Analysis of Variance based on the Completely Randomized Design is as follows, where the subscript i represents the treatment and j represents the replicate within the treatment.

$$Y_{ij} = \mu_i + \epsilon_{ij} = \mu + \tau_i + \epsilon_{ij} \qquad i = 1, \dots, r; \quad j = 1, \dots, n_i \qquad \epsilon_{ij} \sim NID\left(0, \sigma^2\right)$$

Example 1.6 - Anthocyanin Extractability in Cabernet Franc Grapes

In a study conducted by researchers in France and Italy, Cabernet Franc grapes were harvested at $r = 6$ different classes of sugar content (176.5, 192.6, 209.3, 225.0, 242.1, and 258.5 grams/litre). While these are numeric levels, the authors treated sugar content as a factor variable. There were $n = 15$ berries within each treatment for a total of $n_T = 6(15) = 90$ berries included in the study. Various physical, textural, and anthocyanin extractability measurements were made. We will focus on extraction yield of anthocyanin, which was labelled in the paper as EA%, [Zouid et al., 2013]. A plot of the data is given in Figure 1.1.

```
## [1] "sugar"   "anthExt"
```

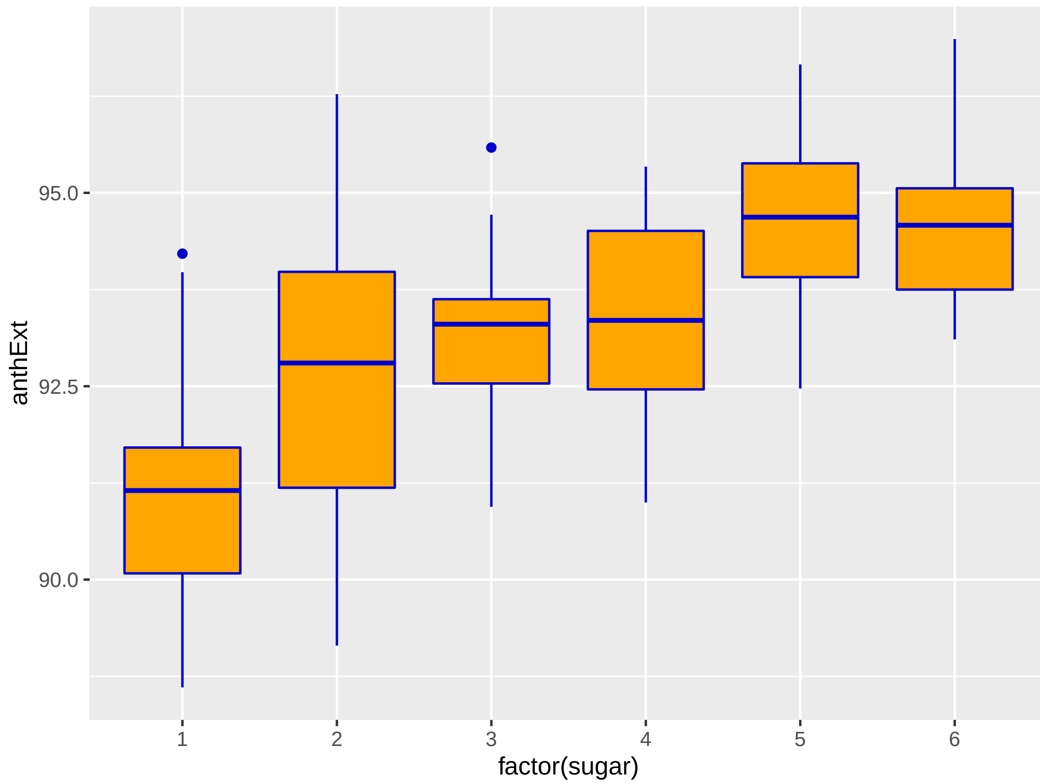


Figure 1.1: Berry texture scores by sugar content

∇

1.4 Randomized Complete Block Design (RCBD or RBD)

In the **Randomized Complete Block Design** (which is often simply referred to as the **Randomized Block Design**), experimental units are blocked into "groups" of homogeneous units. Within each block, units are randomly assigned to the treatments, with each treatment being applied to one unit within each block. We will continue using r as the number of treatments and will use b for the number of blocks. The goal is to remove the heterogeneity across blocks to obtain more precise comparisons among treatments, when possible. In many instances, blocks will be the same individual that will receive each treatment when this is feasible. Blocks are typically treated as a random factor, in the sense that results are to be generalized across a population of such blocks or individuals. When the blocks are individuals who receive each treatment, this design is often referred to as a **Repeated Measures Design** or a **Crossover Design**.

The statistical model can be written as follows where the subscript i represents the treatment and j represents the block.

$$Y_{ij} = \mu_i + \beta_j + \epsilon_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \qquad i = 1, \ldots, r; \quad j = 1, \ldots, b \qquad \epsilon_{ij} \sim NID\left(0, \sigma^2\right)$$

Example 1.7 - Comparison of 4 Treadmill Models for User Satisfaction

Researchers in Italy and New Zealand conducted an experiment to compare $r = 4$ treadmill models (Life Fitness, Precor, Matrix, Technogym) among $b = 57$ trained runners [Carraro et al., 2019]. Each runner rated each treadmill model in terms of seven characteristics: Running Surface, Controls, Stability and Safety, Physical Interaction, Console Readability, Aesthetic Appeal, and Enjoyment of Use. The responses were measured on visual Analogue Scales (VAS) from very unpleasant to very pleasant. We will consider Enjoyment of Use (enjoyUse) in Figure 1.2 and Figure 1.3.

```
##   trdMill subject runSurf controls stblSfty physIntrct consRead aesthApp
## 1       1       1  8.5473   6.2358   5.0478     6.8996   6.2044   4.5798
## 2       1       2  7.4018   7.2364   6.0081     2.5457   7.1048   6.4683
## 3       1       3  5.1259   8.6687   8.8688     6.4669   4.6827   9.3432
## 4       1       4  4.8110   4.8361   9.5405     5.1269   6.7633   6.4748
## 5       1       5 10.1249   4.2111   9.5724     6.4366   6.6144   7.0896
## 6       1       6  5.1529   6.5803   8.0583     5.8038   3.7302   6.3881
##   enjoyUse
## 1   8.9383
## 2   4.8944
## 3   7.9970
## 4   7.4555
## 5   3.7710
## 6   5.2169
```

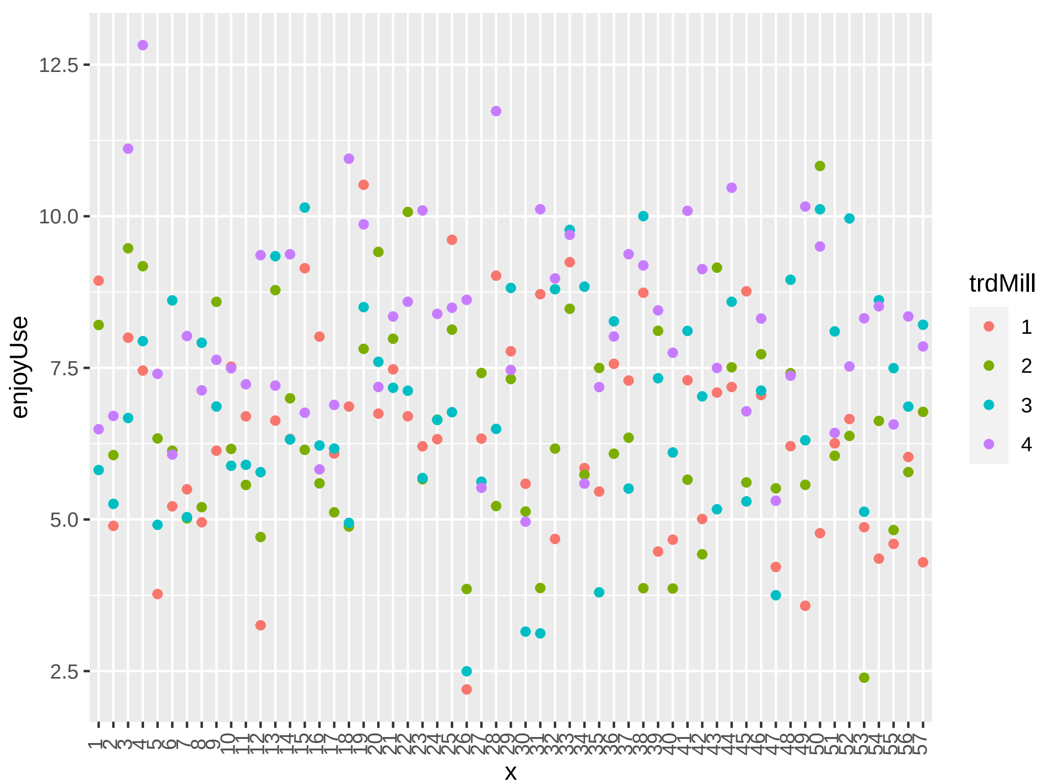


Figure 1.2: Enjoyment of use of 4 treadmills for 57 runners

∇

1.5 Overview of Some Standard Experimental Designs

In this section we list some commonly used experimental designs as well as some examples of them.

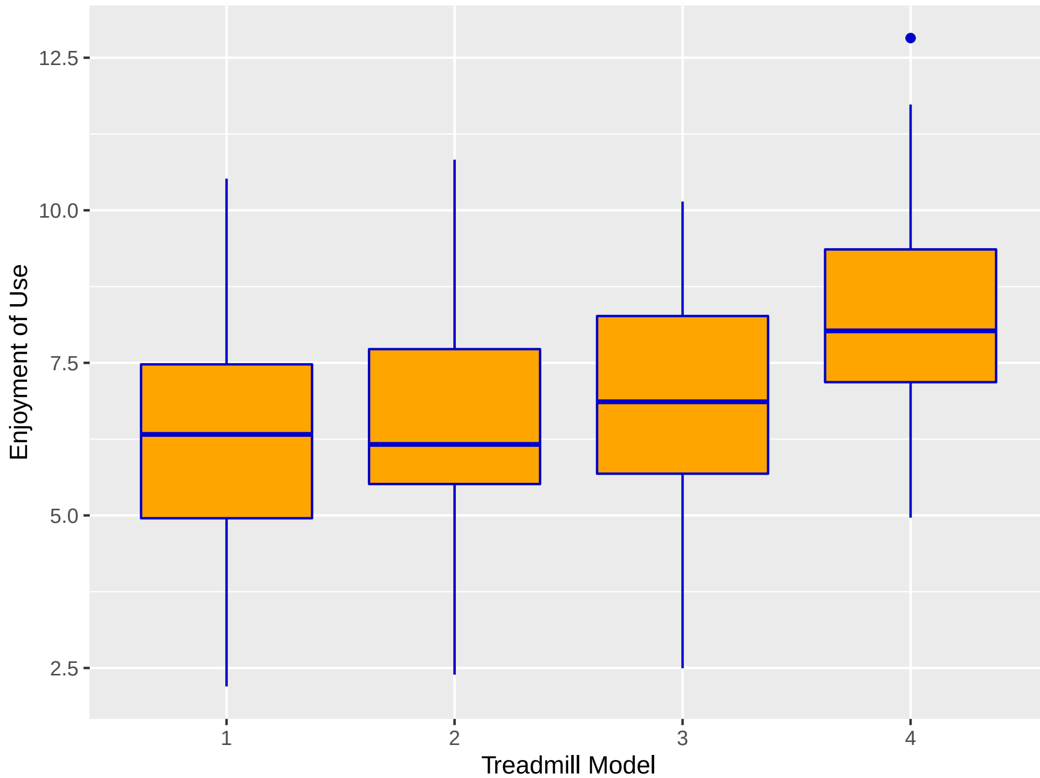


Figure 1.3: Boxplots of Enjoyment of use of 4 treadmills for 57 runners

- **Completely Randomized Design (CRD)** – Units randomized to treatments with no restrictions on randomization process
- **Factorial Experiments** – CRD with two or more crossed factors. Treatment effects are made up of main factor effects and interaction effects
- **Randomized Complete Block Design (RCBD)** – Units are grouped into blocks. Treatments are randomly assigned to units within blocks
- **Nested Designs** – Levels of Factor B differ across levels of Factor A
- **Crossed/Nested Designs** – Designs with both crossed and nested factors
- **Repeated Measures Designs** – Each unit is measured multiple times
  - **RBD** - Each subject receives each treatment once
  - **CRD** Each subject receives only one treatment, but is measured at multiple time points
- **Split-Plot Designs** – Two (or more) sizes of experimental units due to randomization restrictions for factors
- **Incomplete Block Designs** – Block Designs with block sizes smaller than the number of treatments
- **2-Level Factorial Experiments** – Several (possibly many) factors, each at 2 levels (low/high). With k factors, there will be 2^k treatments
- **2-Level Fractional Factorial Designs** – Experiments with only a subset of all 2^k treatments to reduce cost, but still obtain estimates of main effects and lower-order interactions
- **Response Surface Designs** – Designs used to fit polynomial regression models to optimize responses for numeric factors
- **Mixture Designs** – Designs used to fit models to optimize responses among mixtures (components sum to 1) of numeric factors.

Example 1.8 - Advertising Messaging Strategy and Attitude to the Firm

An experiment was conducted to compare 4 advertisement conditions [Hyllegard et al., 2009]. A sample of 425 students were selected and randomly assigned to one of 4 conditions. The ads were:

- Ad1: Firm as "pioneer of industry standards in social responsibility" and US location
- Ad2: Young woman partially clothed in shower, winner of wet t-shirt contest

- Ad3: Female co-founder of porn mag for women, in jogging shorts/hoodie
- Ad4: Female and male partially clothed couple in bed, faces cropped out of image. Female on top of male.

The response was an overall attitude toward the firm based on a series of rating items. Note that each student was exposed to only one condition. Condition 1: Ad1 Only, Condition 2: Ad1&Ad2, Condition 3: Ad1&Ad3, Condition 4: Ad1&Ad4.

∇

Example 1.9 - Energy Efficiency of 4 Dryer Types and 3 Clothing Categories

A study compared combinations of 4 dryer types and 3 clothing categories on energy efficiency [To et al., 2007]. The dryer types were (1=Electric Dryer, 2=Bi-directional Electric dryer, 3=Town Gas-Fired Dryer, 4=LPG-Fired dryer) and the clothing categories were (1=Towels, 2=Jeans, 3=Thermal Clothing). The response was Energy Efficiency (kWh/kg), and there were 3 replications per treatment.

∇

Example 1.10 - Comparison of 6 Chopstick Lengths on Feeding Efficiency

A study compared 6 chopstick lengths (180, 210, 240, 270, 300, 330mm) in terms of the numbers of peanuts picked up and placed in a cup [Hsu and Wu, 1991]. There were 31 subjects, and each subject used each chopstick. The subjects act as blocks. This can also be treated as a Repeated Measures Design, where each subject receives each treatment.

∇

Example 1.11 - Caffeine Content of Coke and Pepsi Products at Various Restaurants

A study compared Coca-Cola and Pepsi-Cola at various restaurants with respect to caffeine content [Grand and Bell, 1997]. There were 5 restaurants that sold Coca-Cola brand (1=Red Lobster, 2=Applebees, 3=McDs, 4=BK, 5=Hardees) and 7 that sold Pepsi products (6=Arbys, 7=Subway2, 8=Subway1, 9=KFC, 10=PizzaHut, 11=TacoBell, 12=Wendys). Each restaurant sold both sugar and diet formulations. There were 10 measurements per restaurant per formulation. Note that restaurant is nested within brand, but crossed with formulation.

∇

Example 1.12 - Zylkene v Placebo for Cats with Anxiety Over Time

A study compared the effects of Zylkene v Placebo in cats with anxiety [Beata et al., 2007]. A sample of 34 cats with anxiety was obtained, and randomized to receive either Zylkene or Placebo (17 cats per treatment). Each cat was observed on a global anxiety scale at each of 5 time points. The goal is to compare the treatments and determine whether time effects occur, and whether the treatment effects differ over time.

∇

Example 1.13 - Effects of Seeding Rates of Hardinggrass and Ryegrass on Growth

An experiment was conducted to measure the effects of 4 rates of seeding of the perennial hardinggrass and 6 rates for ryegrass [Schultz and Biswell, 1952]. Hardinggrass was measured on whole (larger) plots, with levels of 1,2,3,4 pounds per acre) while ryegrass which was measured on subplots within the whole plots with levels of 0,3,6,9,12,15 pounds per acre). The experiment was conducted in 3 blocks (replicates). The response measured was the density of hardinggrass. Note that the experimental units for levels of hardinggrass are larger than those for levels of ryegrass.

∇

Example 1.14 - Consumer Liking of 16 Dark Chocolate Formulations

An experiment was conducted to compare computer liking among 16 dark chocolate formulations [Hinneh et al., 2020]. As consumers have quite different tastes and preferences, they are treated as blocks, and there were 16 raters. However, due to fatigue, the researchers had each consumer rate only 6 dark chocolate formulations. In a balanced experiment, each formulation would be rated by the same numbers of consumers, and each pair of formulations would be tasted by equal numbers of consumers. In this case, each chocolate was rated by 6 raters, and each pair of chocolates were rated by 2 raters. This represents a Balanced Incomplete Block Design.

∇

Example 1.15 - Using Seaweed to Extract Phenol from Aqueous Solution

An experiment was conducted to study the effects of 3 factors (pH (3, 9), adsorbent dosage (1, 10 g/L), and temperature (30, 60C)) on phenol extraction efficiency (%) from an aqueous solution [Ranthinam et al., 2011]. Dried seaweed was treated with zinc chloride, then applied to one of the 8 combinations of the 3 factors. There were two replicates at each factor level. This is an example of a 2^3 full-factorial design.

∇

Example 1.16 - Factors Affecting Damage to Motorcycle Wheels

An experiment was conducted to determine the effect of 5 factors on the Crush Radius on the front wheel of a motorcycle [Tan et al., 2009]. The factors were: Impact speed (3, 6), Impact Mass (51.18, 101.33), Tire Pressure (148, 252), Striker Contact Geometry (0.03, 0.10), and Impact Offset Distance (0, 0.108). Although there are $2^5 = 32$ combinations of the factors, the experimenters ran it in $2^{5-1} = 16$ combinations to reduce costs. There were four replicates at each combination of factor settings. This is an example of a 2^{5-1} fractional factorial design.

∇

Example 1.17 - Optimizing Qualities of Potato Chips

An experiment varied 3 factors to optimize responses regarding potato chips [Song et al., 2007]. The factors (and levels) were: Vacuum microwave pre-drying time (0.95, 3, 6, 9, 11.05 minutes), Vacuum temperature (83.18, 90, 100, 110, 116.82C), and Frying Time (11.59, 15, 20, 25, 28.41 minutes). Three responses were measured (analyzed one at a time): Moisture Content, Fat Content, and Breaking Force. The goal was to choose factor levels that optimize the response. This is an example of a response surface design.

∇

Example 1.18 - Optimizing Antibacterial Effect of Mixtures of 3 Essential Oils

An experiment varied mixtures of 3 types of essential oils in terms of optimizing a response [Ouedrhiria et al., 2016]. The three oils were: (O. compactum, O. majorana, and T. serypyllum). Three responses were measured (analyzed one-at-a-time): minimum inhibitory concentration (MIC %) of 3 types of bacterium: B. subtilis. S. aureus, and E. coli. The goal was to choose the mixture of the 3 types of essential oils that minimizes the MIC of the bacterium.

∇

1.6 Overview of Observational Study Designs

- **Cross-Sectional Studies** – Observations made from populations/subpopulations at a single time point or interval.
- **Prospective Studies** – Groups are formed by levels of a potential causal factor, then observed over time for some measurable outcome.
- **Retrospective Studies** – Studies where subjects are identified based on the outcome of interest and potential risk factors are identified that previously occurred.

- **Matching** – Subjects from different populations are matched, based on external factors, similar to blocking in experimental studies.

Example 1.19 - Medical Profession Students Attitudes Toward Interdisciplinary Studies

A survey was conducted to measure students' in medical professions Readiness for Inter-Professional Learning [Keshtkaran et al., 2014]. Students were sampled from 3 groups: Nursing, Science in Surgical Technology, and Medicine, and given the scale measuring their readiness. This was cross-sectional in the sense that it was taken at one point in time.

∇

Example 1.20 - Blood Transfusions and Caesarean Deliveries in 3 Pakistan Hospitals

A prospective study was conducted to compare birth deliveries in 3 Pakistan hospitals over the period of January-June 2010 [Ismail et al., 2014]. In particular, the authors were interested in whether the mother had acCaesarean Section and whether the patient had a subsequent blood transfusion.

∇

Example 1.21 - Fertilization Time-Lapse Variables and Embryo Sex

A study at a University-affiliated private fertility center considered the gender of an embryo, as well as various cleavage timing variables, from the time of the fertilization, retrospectively [Bronet et al., 2015]. The researchers were interested in determining whether any of the timing variables could predict the eventual sex of the embryo.

∇

Example 1.22 - Recidivism Rates for Juvenile Offenders

A study compared recidivism rates among juvenile offenders in 2 conditions: transferred to adult court and not transferred to adult court, that is, tried in juvenile court [Bishop et al., 1996]. A database of past criminal record including number and severity of prior and current charges, gender, and age was created, and matches were created where within each pair, one had been transferred, the other had not. Subsequent recidivism was observed within each pair.

∇

```
library(tidyverse)
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
library(effectsize)
```

Chapter 2

Single Factor Studies

In this chapter, we consider the case where there is a single factor. The factor can be **qualitative** or **quantitative**. Qualitative factors can be **nominal** or **ordinal**. Nominal factors have no inherent ordering, while ordinal factors have an underlying ordering that is non-numeric. Quantitative factors can be used in linear or nonlinear regression models, however in some cases researchers don't wish to place a structure to the model and treat levels as categories.

The factor levels, unless specified otherwise, will be treated as all levels of interest to researchers. These are referred to as **fixed factors**. Later in the course, we will consider factors with levels that are a sample from a population of levels, which will be referred to as **random factors**.

We will let the number of levels of the factor be represented as r where $r \geq 2$. When $r = 2$, the analysis can be analyzed as an independent sample t-test, which we will see gives identical conclusions as the 1-Way Analysis of Variance described below.

The analysis will be the same whether the data are obtained from a Controlled Experiment or Observational Study, however interpretations of cause and effect are stronger for Controlled Experiments.

In this chapter, we will focus on models with independent, normally distributed errors with constant variance. We will consider other tests in subsequent chapters.

2.1 Models with $r = 2$ Factor Levels

In this section, we consider a Controlled Experiment and an Observational Study, each with two groups. We use the independent sample t-test in each case to compare the two means.

Example 2.1 - Tai Chi and Strength Training in Hip Replacement Patients

A controlled experiment was conducted to compare post-operative effects of an in-home training treatment for patients who have received hip replacement surgery [Zeng et al., 2015]. The final analysis was based on $n_T = 59$ subjects who completed the protocol, 32 in the treatment group (Tai Chi and strength training) and 27 in the control group. Patients were randomly assigned to the $r = 2$ conditions. One response that was reported was Timed Up and Go Test (TUG) to measure mobility (lower times are better). The means, standard deviations, and sample sizes are given below for the treatment and control groups (the units are seconds).

$$\text{Treatment: } \overline{y}_T = 14.61 \quad s_T = 2.60 \quad n_T = 32$$

$$\text{Control: } \overline{y}_C = 19.06 \quad s_C = 3.37 \quad n_C = 27$$

For the 2-sample t-test (aka independent sample t-test), we will first compute the pooled variance then the t-statistic, critical value ($\alpha = 0.05$) and P-value for a 2-sided test.

$$\text{Pooled Variance:} \quad s_p^2 = \frac{(32-1)(2.60)^2 + (27-1)(3.37)^2}{32+27-2} = 8.86 \qquad s_p = \sqrt{8.86} = 2.98$$

$$\text{Test Statistic:} \quad t^* = \frac{14.61 - 19.06}{\sqrt{8.86\left(\frac{1}{32} + \frac{1}{27}\right)}} = \frac{-4.45}{0.78} = -5.71$$

$$\text{Rejection Region:} \quad t_{.975,57} = 2.002 \qquad \text{P-value:} \quad 2P\left(t_{57} \geq |-5.71|\right) < .0001$$

$$\text{95\% Confidence Interval:} \quad -5.71 \pm 2.002(0.78) \quad \equiv \quad -5.71 \pm 1.56 \quad \equiv \quad (-7.27, -4.15)$$

The **Effect Size** is the absolute difference in the sample means in units of the pooled standard deviation. In this example, $e = 4.45/2.98 = 1.49$ which is considered a large effect size.

∇

Example 2.2 - Firefighter Air Consumption

An observational study was conducted to compare firefighters that were classified into two categories of air consumption, slow and fast [Wohlgemuth et al., 2024]. The firefighters were classified based on scores for 10 tasks pertinent to firefighting. Once the $n_T = 160$ firefighters were classified into the two groups (94 fast and 66 slow), the groups were compared with respect to several physical characteristics including Age, Body Mass Index, and Peak Heart Rate (none of these were used in classifying as Fast/Slow Air Consumption). Summary statistics for the response Body Fat Percentage and the corresponding t-test are given below.

$$\text{Fast: } \overline{y}_F = 20.71 \quad s_F = 6.65 \quad n_F = 94 \qquad \text{Slow: } \overline{y}_S = 25.11 \quad s_S = 8.93 \quad n_S = 66$$

$$s_p^2 = \frac{(94-1)(6.65)^2 + (66-1)(8.93)^2}{94 + 66 - 2} = 58.84 \qquad s_p = 7.67$$

$$t^* = \frac{20.71 - 25.11}{\sqrt{58.85\left(\frac{1}{94} + \frac{1}{66}\right)}} = \frac{-4.40}{1.23} = -3.57 \qquad t_{.975,158} = 1.975 \qquad P = .0005$$

$$\text{95\% Confidence Interval:} \quad -4.40 \pm 1.975(1.23) \quad \equiv \quad -4.40 \pm 2.43 \quad \equiv \quad (-6.83, -1.97)$$

For this example, the effect size is $e = 4.40/7.67 = 0.57$.

Idealized normal distributions for the Tai Chi and Firefighter Air Consumption studies are given in Figure 2.1 and Figure 2.2.

∇

There are various ways to parameterize single factor models. We will consider the following three formulations in this chapter.

- **Cell Means Model** - Each treatment mean is its own parameter (no intercept term)
- **Treatment Effect Model** - Intercept is the mean of the group means, each treatment has an effect relative to the mean, with effects summing to 0
- **Reference Group Model** - Intercept corresponds to the mean for Group 1, remaining parameters are differences between remaining group means and group 1 mean

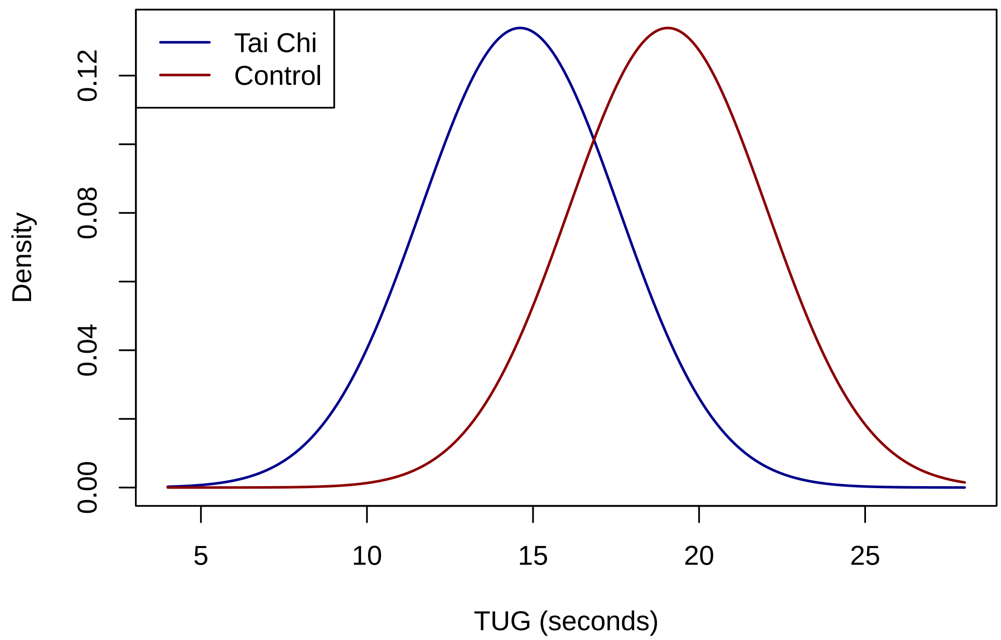


Figure 2.1: Idealized normal distibutions for the Tai Chi study

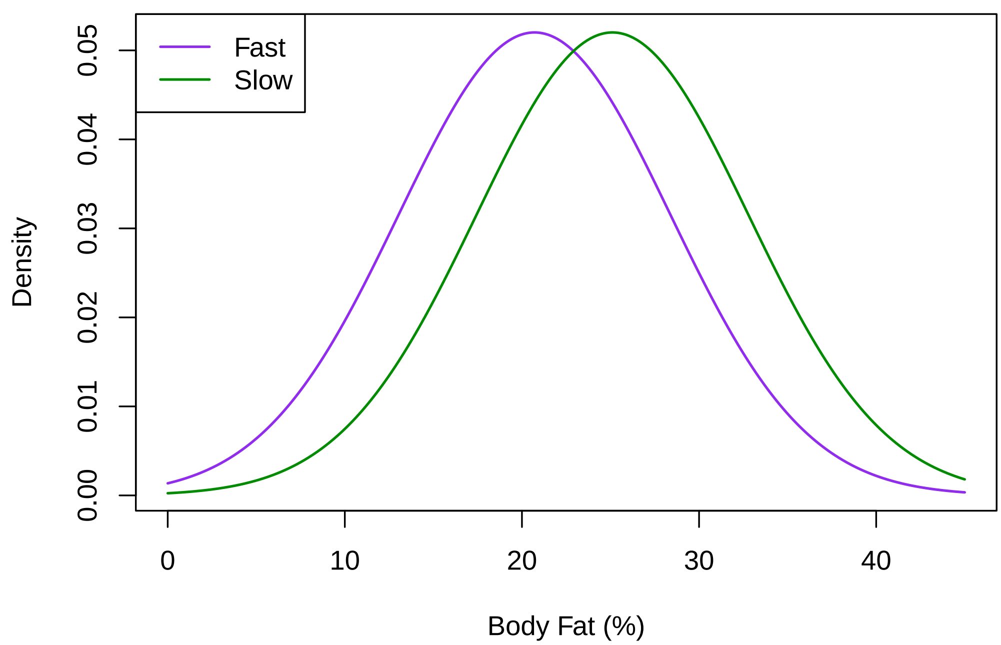


Figure 2.2: Idealized normal distibutions for the Firefighter air consumtion study

2.2 Three Model Formulations

2.2.1 Cell Means Model

The model and assumptions are given below for the **Cell Means Model**. Note that the total sample size is $n_T = n_1 + \cdots + n_r = \sum_{i=1}^{r} n_i$. When all sample sizes are equal ($n_i = n$), the design is said to be **balanced**.

$$Y_{ij} = \mu_i + \epsilon_{ij} \qquad i = 1, \dots, r; \quad j = 1, \dots, n_i \qquad \epsilon_{ij} \sim NID\left(0, \sigma^2\right)$$

$$E\{Y_{ij}\} = E\{\mu_i + \epsilon_{ij}\} = \mu_i + 0 = \mu_i \qquad \sigma^2\{Y_{ij}\} = \sigma^2\{\mu_i + \epsilon_{ij}\} = 0 + \sigma^2 = \sigma^2$$

For this model, the treatment/population mean for the i^{th} condition is μ_i and the variance of the individual measurements within each condition is assumed to be σ^2 (standard deviation $= \sigma$) . Further, the individual measurements within each condition are assumed to be independent and normally distributed.

The point estimators for the population means, which are derived below, are the sample means for the various groups. The estimator for the population variance is the pooled variance among the individual groups, which is the same for all three model forms.

2.2.2 Treatment Effects Model

Let μ_i be the mean for treatment i, and let $\mu_\bullet$ be the unweighted mean of the μ_i^s.

$$\mu_\bullet = \frac{\sum_{i=1}^{r} \mu_i}{r} \qquad \mu_i = \mu_\bullet + (\mu_i - \mu_\bullet) = \mu_\bullet + \tau_i \quad \Rightarrow \quad \tau_i = \mu_i - \mu_\bullet$$
$$\mu_1 = \cdots = \mu_r = \mu_\bullet \quad \Rightarrow \quad \tau_1 = \cdots = \tau_r = 0$$

$$Y_{ij} = \mu_\bullet + \tau_i + \epsilon_{ij} \qquad i = 1, \dots, r; \quad j = 1, \dots, n_i \qquad \epsilon_{ij} \sim NID\left(0, \sigma^2\right)$$
$$E\{Y_{ij}\} = E\{\mu_\bullet + \tau_i + \epsilon_{ij}\} = \mu_\bullet + \tau_i + 0 = \mu_\bullet + \tau_i \qquad \sigma^2\{Y_{ij}\} = \sigma^2\{\mu_\bullet + \tau_i + \epsilon_{ij}\} = 0 + \sigma^2 = \sigma^2$$

$$\sum_{i=1}^{r} \tau_i = 0 \quad \Rightarrow \quad \tau_r = -\sum_{i=1}^{r-1} \tau_i$$

The point estimator for the overall population mean is the unweighted mean of group sample means. The estimator for the effect of group i is the difference between the sample mean for group i and the overall unweighted mean.

2.2.3 Reference Group Model

This model, which is the default model for the **lm** and **aov** functions in R, uses group 1 (based on alpha-numeric ordering of the group levels) as the intercept. The remaining parameters are the differences between the remaining $r - 1$ group means and the mean for group 1. The point estimators are based on the corresponding sample means.

An example where $r = 3$ conditions have means of $\mu_1 = 40$, $\mu_2 = 50$, and $\mu_3 = 60$, respectively and standard deviation $\sigma = 10$ is given in Figure 2.3.

Example 2.3 - Virtual Training for a Lifeboat Launching Task A study in South Korea compared $r = 4$ methods of training to conduct a lifeboat launching task [Jung and Ahn, 2018]. The treatments and their labels from the paper are given below and the response Y was a procedural knowledge score for subjects post training.

- **Lecture/Materials** - Traditional Lecture with no computer component (LEC/MAT)
- **Monitor/Keyboard** - Trained virtually with a monitor, keyboard, and mouse (MON/KEY)
- **Head-Mounted Display/Joypad** - Trained virtually with HMD and joypad (HMD/JOY)
- **Head-Mounted Display/Wearable Sensors** - (HMD/WEA)

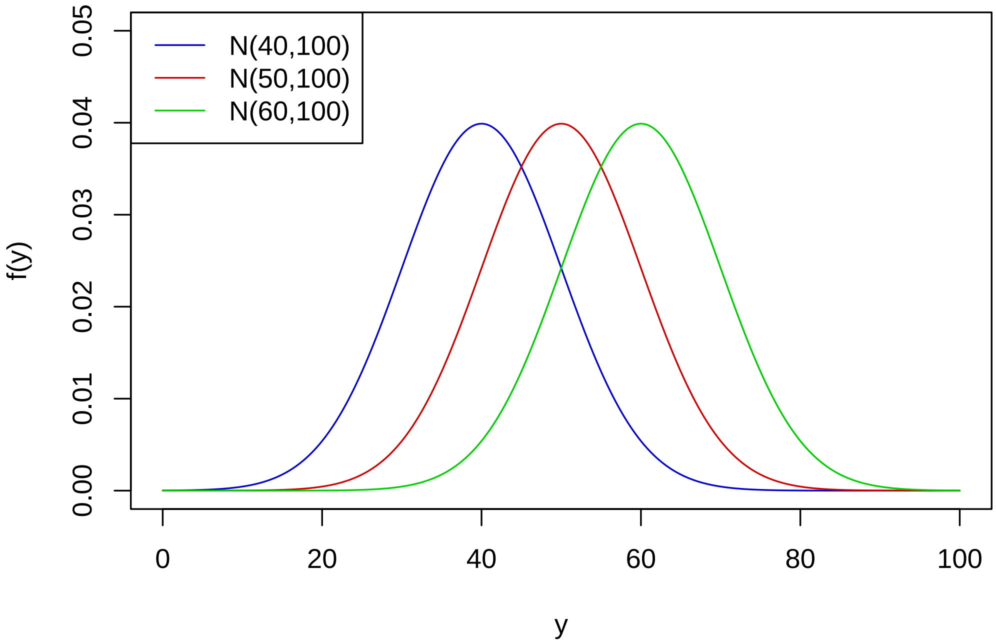


Figure 2.3: Normal distibutions with means=40,50,60 and standard deviation=10

There were a total of $n_T = 64$ subjects and they were randomized so that 16 subjects received each treatment ($n_1 = n_2 = n_3 = n_4 = 16$). Data that have been generated to match the authors' reported means and standard deviations are given in Table 2.1. Note that when we later run this analysis in R, there will be a variable for treatment and a variable for procedural score. Figure 2.4 gives the generated observed values (dots), treatment means (green lines), and the overall mean (red line).

```
##   grp.trt procKnow
## 1       1   4.5614
## 2       1   6.6593
## 3       1   5.6427
## 4       1   6.4394
## 5       1   4.8635
## 6       1   0.3268
```

Using the summary data from the table, we will obtain point estimates and standard errors directly for the three model formulations directly, then using R. First, we compute the pooled standard deviation (aka residual standard error).

$$s = \sqrt{\frac{\sum_{i=1}^{r} (n_i - 1)\, s_i^2}{n_T - r}} = \sqrt{\frac{(16-1)\left(1.94^2 + 1.43^2 + 2.82^2 + 1.99^2\right)}{64-4}} = 2.105$$

Cell Means Model

$$\hat{\mu}_1 = \overline{y}_1 = 4.931 \quad \hat{\mu}_2 = \overline{y}_2 = 7.708 \quad \hat{\mu}_3 = \overline{y}_3 = 6.736 \quad \hat{\mu}_4 = \overline{y}_4 = 6.875$$

$$s\left\{\hat{\mu}_i\right\} = \frac{s}{\sqrt{n_i}} = \frac{2.105}{\sqrt{16}} = 0.526 \qquad i = 1, 2, 3, 4$$

Table 2.1: Lifeboat Virtual Training Data

	LEC/MAT (i=1)	MON/KEY (i=2)	HMD/JOY (i=3)	HMD/WEA (i=4)
j=1	4.5614	5.4532	1.8773	9.8257
j=2	6.6593	8.9208	6.1884	8.9988
j=3	5.6427	8.3489	6.8029	5.4112
j=4	6.4394	8.3175	9.4535	4.9396
j=5	4.8635	9.7848	8.7312	8.6486
j=6	0.3268	6.2697	6.4198	5.1459
j=7	5.6990	7.1327	8.4955	7.0656
j=8	6.3545	7.0886	5.9008	7.9552
j=9	6.7509	5.0733	3.3336	5.7284
j=10	6.6019	8.3865	1.6530	6.6314
j=11	3.2365	9.4227	9.4710	7.5014
j=12	6.1655	8.9142	7.2818	8.2456
j=13	2.4060	8.5150	9.7649	4.2465
j=14	1.9851	6.8198	3.4931	3.2286
j=15	5.2198	8.6390	8.9848	6.5520
j=16	5.9765	6.2467	9.9261	9.8754
n	16.0000	16.0000	16.0000	16.0000
Mean	4.9306	7.7083	6.7361	6.8750
SD	1.9400	1.4300	2.8200	1.9900

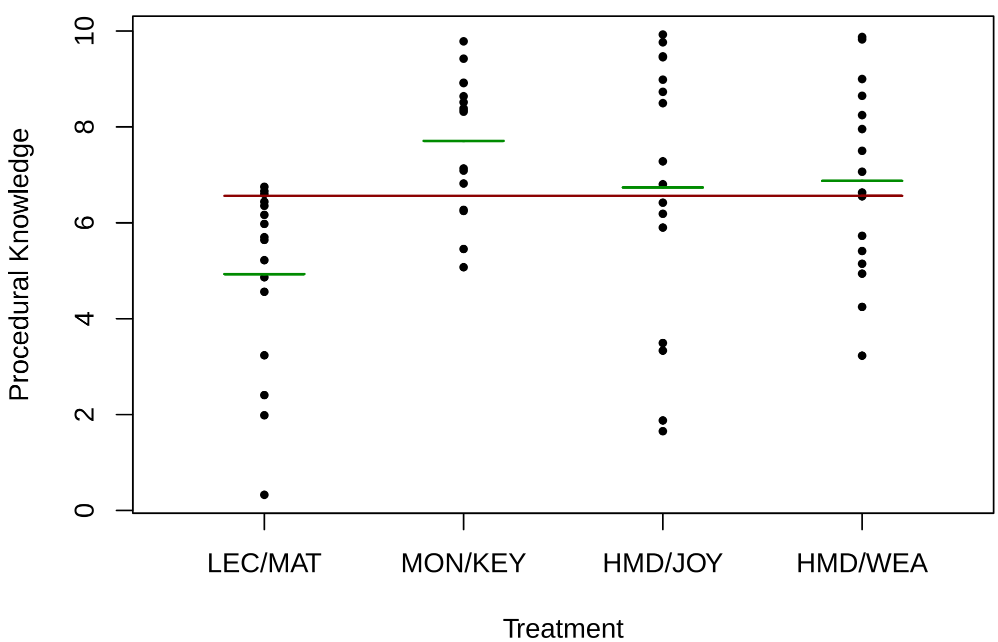


Figure 2.4: Procedural knowledge scores for lifeboat training study

Treatment Effects Model

$$\hat{\mu} = \frac{4.931 + 7.708 + 6.736 + 6.875}{4} = 6.563$$

$$s\{\hat{\mu}\} = s\sqrt{\frac{1}{r^2}\sum_{i=1}^{r}\frac{1}{n_i}} = 2.105\sqrt{\frac{1}{4^2}4\left(\frac{1}{16}\right)} = 0.263$$

$$\hat{\tau}_1 = 4.931 - 6.563 = -1.632 \quad \hat{\tau}_2 = 1.145 \quad \hat{\tau}_3 = 0.173 \quad \hat{\tau}_4 = 0.312$$

To obtain the standard error for $\hat{\tau}_k$, we can re-write it as follows and make use of independence among group means.

$$\hat{\tau}_k = \overline{y}_k - \hat{\mu} = \overline{y}_k - \frac{1}{r}\sum_{i=1}^{r}\overline{y}_i = \frac{r-1}{r}\overline{y}_k - \frac{1}{r}\sum_{\substack{i=1\\ i\neq k}}^{r}\overline{y}_i$$

$$s\{\hat{\tau}_k\} = s\sqrt{\frac{1}{r^2}\left(\frac{(r-1)^2}{n_k} + \sum_{\substack{i=1\\ i\neq k}}^{r}\frac{1}{n_i}\right)} = 2.105\sqrt{\frac{1}{4^2}\left(\frac{(4-1)^2}{16} + 3\frac{1}{16}\right)} = 0.456$$

Reference Group Model

$$\hat{\mu}_1 = \overline{y}_1 = 4.931 \qquad s\{\hat{\mu}_1\} = \frac{s}{\sqrt{n_1}} = \frac{2.105}{\sqrt{16}} = 0.526$$

$$i = 2,\ldots,r \quad \hat{\tau}_i = \overline{y}_i - \overline{y}_1 \qquad s\{\hat{\tau}_i\} = s\sqrt{\frac{1}{n_i} + \frac{1}{n_1}}$$

$$\hat{\tau}_2 = 7.708 - 4.931 = 2.777 \quad \hat{\tau}_3 = 6.736 - 4.931 = 1.805 \quad \hat{\tau}_4 = 6.875 - 4.931 = 1.944$$

$$s\{\hat{\tau}_i\} = 2.105\sqrt{\frac{1}{16} + \frac{1}{16}} = 0.744$$

```
##   grp.trt procKnow
## 1       1   4.5614
## 2       1   6.6593
## 3       1   5.6427
## 4       1   6.4394
## 5       1   4.8635
## 6       1   0.3268
```

```
##
## Call:
## lm(formula = procKnow ~ factor(grp.trt) - 1, data = vt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0831 -1.4444  0.5774  1.5495  3.1900
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## factor(grp.trt)1   4.9306     0.5262    9.37 2.37e-13 ***
## factor(grp.trt)2   7.7083     0.5262   14.65  < 2e-16 ***
## factor(grp.trt)3   6.7361     0.5262   12.80  < 2e-16 ***
## factor(grp.trt)4   6.8750     0.5262   13.06  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 2.105 on 60 degrees of freedom
## Multiple R-squared:  0.9139, Adjusted R-squared:  0.9082
## F-statistic: 159.2 on 4 and 60 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = procKnow ~ factor(grp.trt), data = vt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0831 -1.4444  0.5774  1.5495  3.1900
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        6.5625     0.2631  24.943  < 2e-16 ***
## factor(grp.trt)1  -1.6319     0.4557  -3.581 0.000686 ***
## factor(grp.trt)2   1.1458     0.4557   2.514 0.014622 *
## factor(grp.trt)3   0.1736     0.4557   0.381 0.704573
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.105 on 60 degrees of freedom
## Multiple R-squared:  0.1981, Adjusted R-squared:  0.158
## F-statistic: 4.941 on 3 and 60 DF,  p-value: 0.003931
```

```
##
## Call:
## lm(formula = procKnow ~ factor(grp.trt), data = vt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0831 -1.4444  0.5774  1.5495  3.1900
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        4.9305     0.5262   9.370 2.37e-13 ***
## factor(grp.trt)2   2.7778     0.7442   3.733 0.000423 ***
## factor(grp.trt)3   1.8056     0.7442   2.426 0.018277 *
## factor(grp.trt)4   1.9444     0.7442   2.613 0.011329 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.105 on 60 degrees of freedom
## Multiple R-squared:  0.1981, Adjusted R-squared:  0.158
## F-statistic: 4.941 on 3 and 60 DF,  p-value: 0.003931
```

∇

2.3 The Analysis of Variance and F-test

In this section, we obtain a decomposition of the total variation of the individual measurements around the overall mean, the **Total Sum of Squares**, which we will denote $SSTO$. The **Degrees of Freedom** associated with the total sum of squares is $df_{TO} = n_T - 1$.

$$\overline{Y}_{\bullet\bullet} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i} Y_{ij}}{n_T} \qquad SSTO = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_{\bullet\bullet}\right)^2 \qquad df_{TO} = n_T - 1$$

The total sum of squares can be partitioned into the **Error Sum of Squares** and the **Treatment Sum of Squares**. The error sum of squares can be thought of as how variable are measurements within the same treatment or group. The treatment sum of squares represents how variable the treatment or group means are around the overall mean (weighted by their sample sizes). The sums of squares are given below (note that in R, for linear models, the error sum of squares is labelled as RSS, for the residual sum of squares). As with linear regression models, the error sum of squares is the sum of the squared distances from the observed values to their predicted values.

$$SSE = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(Y_{ij} - \hat{Y}_{ij}\right)^2$$

Noting that for this model, $\hat{Y}_{ij} = \overline{Y}_{i\bullet}$, we obtain the following formula for SSE. The degrees of freedom for the error sum of squares is $df_E = n_T - r$.

$$SSE = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_{i\bullet}\right)^2 \qquad df_E = n_T - r$$

Another useful way of computing SSE based on published summary statistics is given below. Suppose the authors have published the sample sizes $\{n_i\}$, the sample means $\left\{\overline{y}_{i\bullet}\right\}$ using lower case y as these are observed means from a realization of this experiment, and the sample standard deviations $\{s_i\}$ for the r treatments. Then SSE can be computed as below.

$$s_i = \sqrt{\frac{\sum_{j=1}^{n_i}\left(y_{ij} - \overline{y}_{i\bullet}\right)^2}{n_i - 1}} \qquad \Rightarrow \qquad SSE = \sum_{i=1}^{r}\left(n_i - 1\right)s_i^2$$

The **Treatment Sum of Squares**, denoted as $SSTR$, acts like the regression sum of squares for linear regression models. That is, $SSTR$ represents the sum of squared differences from the predicted values to the overall mean. The degrees of freedom for the treatment sum of squares is $df_{TR} = r - 1$. Keeping in mind that the fitted value for Y_{ij} is $\hat{Y}_{ij} = \overline{Y}_{i\bullet}$, we obtain SSTR as follows.

$$SSTR = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet}\right)^2 = \sum_{i=1}^{r} n_i\left(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet}\right)^2 \qquad df_{TR} = r - 1$$

For each sum of squares, the **Mean Square** is the sum of squares divided by its corresponding degrees of freedom, and are used for inferences among the means.

In single factor models, the goal is typically to test for differences among the treatment or group means. The null hypothesis is that the means are all equal, while the alternative hypothesis is that there are differences among the means. In the next chapter, we consider specific comparisons among treatments.

$$H_0 : \mu_1 = \cdots = \mu_r \quad \Rightarrow \quad \tau_1 = \cdots = \tau_r = 0$$

The test statistic is the F-statistic from the Analysis of Variance and will be used in many forms throughout this course. Under the null hypothesis, the F-statistic follows the F-distribution, based on its numerator and denominator degrees of freedom. Large values of the F-statistic are evidence against the null hypothesis. Technical details of the test are given in a subsequent section.

$$\text{Test Statistic:} \quad F^* = \frac{\left[\frac{SSTR}{df_{TR}}\right]}{\left[\frac{SSE}{df_E}\right]} \quad = \quad \frac{MSTR}{MSE}$$

$$\text{Rejection Region:} \quad F^* \geq F_{1-\alpha;r-1,n_T-r} \qquad P = P\left(F_{r-1,n_T-r} \geq F^*\right)$$

The **effect size** for the One-Way ANOVA model is $\eta^2 = SSTR/SSTO$, that is, the fraction of the total variation that is "attributable" to differences in the group means.

Example 2.4 - Virtual Training for a Lifeboat Launching Task

$$n_1 = n_2 = n_3 = n_4 = 16 \qquad \overline{y}_{1\bullet} = 4.931 \quad \overline{y}_{2\bullet} = 7.708 \quad \overline{y}_{3\bullet} = 6.736 \quad \overline{y}_{4\bullet} = 6.875 \quad \overline{y}_{\bullet\bullet} = 6.563$$

$$SSTR = 16\left[(4.931-6.563)^2 + (7.708-6.563)^2 + (6.736-6.563)^2 + (6.875-6.563)^2\right] = 65.628$$

$$s_1 = 1.94 \quad s_2 = 1.43 \quad s_3 = 2.82 \quad s_4 = 1.99$$
$$\Rightarrow \quad SSE = (16-1)\left(1.94^2 + 1.43^2 + 2.82^2 + 1.99^2\right) = 265.815$$

Now, we test for differences among the effects of the $r = 4$ teaching methods. The degrees of freedom for treatments is $df_{TR} = r - 1 = 3$ and for error is $df_E = n_T - r = 60$.

$$H_0 : \mu_1 = \cdots = \mu_4 \quad \Rightarrow \quad \tau_1 = \cdots = \tau_4 = 0$$
$$\text{Test Statistic:} \quad F^* = \frac{\left[\frac{SSTR}{df_{TR}}\right]}{\left[\frac{SSE}{df_E}\right]} = \frac{\left[\frac{65.789}{3}\right]}{\left[\frac{265.815}{60}\right]} = \frac{21.930}{4.430} = 4.950$$
$$\text{Rejection Region:} \quad F^* \geq F_{.95;3,60} = 2.578 \qquad P = P\left(F_{3,60} \geq 4.950\right) = .0039$$
$$\eta^2 = \frac{SSTR}{SSTO} = \frac{SSTR}{SSTR+SSE} = \frac{65.789}{65.789+265.815} = \frac{65.789}{331.604} = 0.1984$$

Approximately 20% of the total variation is due to differences in the training methods.

We obtain the Sums of Squares for Treatments and Error below using the **anova** of the model. A warning involves the fact that when using the cell means model (with no intercept) the Treatment Sum of Squares and Degrees of Freedom are incorrect here. Under this formulation, we are testing $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$, which is a much stronger hypothesis than testing $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$. We will use the Treatment Effects model or the Reference Group model function to obtain the correct degrees of freedom and F-statistic below.

```
vt <- read.csv("http://www.stat.ufl.edu/~winner/data/virtual_training.csv")
head(vt)
```

```
##   grp.trt procKnow
## 1       1   4.5614
## 2       1   6.6593
## 3       1   5.6427
## 4       1   6.4394
## 5       1   4.8635
## 6       1   0.3268
```

```
vt.mod1 <- lm(procKnow ~ factor(grp.trt) - 1, data=vt)
summary(vt.mod1)
```

```
##
## Call:
## lm(formula = procKnow ~ factor(grp.trt) - 1, data = vt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0831 -1.4444  0.5774  1.5495  3.1900
```

```
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## factor(grp.trt)1   4.9306     0.5262    9.37 2.37e-13 ***
## factor(grp.trt)2   7.7083     0.5262   14.65  < 2e-16 ***
## factor(grp.trt)3   6.7361     0.5262   12.80  < 2e-16 ***
## factor(grp.trt)4   6.8750     0.5262   13.06  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.105 on 60 degrees of freedom
## Multiple R-squared:  0.9139, Adjusted R-squared:  0.9082
## F-statistic: 159.2 on 4 and 60 DF,  p-value: < 2.2e-16
```

```
anova(vt.mod1)    ## Incorrect df and MS for Treatments and F-stat
```

```
## Analysis of Variance Table
##
## Response: procKnow
##                 Df  Sum Sq Mean Sq F value    Pr(>F)
## factor(grp.trt)  4 2821.91  705.48  159.24 < 2.2e-16 ***
## Residuals       60  265.81    4.43
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
vt.mod2 <- aov(procKnow ~ factor(grp.trt), data=vt)
anova(vt.mod2)
```

```
## Analysis of Variance Table
##
## Response: procKnow
##                 Df  Sum Sq Mean Sq F value   Pr(>F)
## factor(grp.trt)  3  65.664 21.8880  4.9406 0.003931 **
## Residuals       60 265.815  4.4302
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
options(es.use_symbols=TRUE)

eta_squared(vt.mod1)
```

```
## For one-way between subjects designs, partial eta squared is
##   equivalent to eta squared. Returning eta squared.
```

```
## # Effect Size for ANOVA
##
## Parameter       | Eta2 |       95% CI
## -------------------------------------
## factor(grp.trt) | 0.91 | [0.88, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

```
eta_squared(vt.mod2)
```

```
## For one-way between subjects designs, partial eta squared is
##   equivalent to eta squared. Returning eta squared.
```

```
## # Effect Size for ANOVA
##
## Parameter       | Eta2 |       95% CI
## -------------------------------------
```

```
## factor(grp.trt) | 0.20 | [0.05, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

Note that for the cell means model (vt.mod1) the sum of squares for treatments includes the "correction for the mean," thus it is larger than $SSTR$ by $n_T\overline{y}^2_{\bullet\bullet}$. The effect size for that model is also too large $\eta^2 = 2821.91/(2821.91 + 265.81) = 0.91$.

Here we will conduct the analysis directly from the raw data, computing $\{n_i\}$, $\{\overline{y}_{i\bullet}\}$, and $\{s_i\}$ as well as sums of squares and degrees of freedom.

```
##   grp.trt procKnow
## 1       1   4.5614
## 2       1   6.6593
## 3       1   5.6427
## 4       1   6.4394
## 5       1   4.8635
## 6       1   0.3268
```

```
##           df       SS      MS     F* F(.95) P(>F*)   eta2
## Treatment  3  65.6639 21.8880 4.9406 2.7581 0.0039 0.1981
## Error     60 265.8149  4.4302     NA     NA     NA     NA
## Total     63 331.4788      NA     NA     NA     NA     NA
```

∇

2.3.1 General Linear Test Approach

We begin with r treatments with no restrictions on their population means $\{\mu_i\}$. We then place a restriction on the population means. Both models are fit, and the Error Sums of Squares and Degrees of Freedom are obtained for each model.

Example 2.5 - Virtual Training for a Lifeboat Launching Task

Consider the following restrictions in light of the virtual training experiment.

- No treatment differences among the 4 treatments: $\mu_1 = \mu_2 = \mu_3 = \mu_4$
- No treatment differences among the 3 virtual reality treatments: $\mu_2 = \mu_3 = \mu_4$
- No treatment differences among the 2 Head-Mounted Display treatments: $\mu_3 = \mu_4$

All of these can be tested using a general linear test. Note that the first test is the same as the F-test conducted from the Analysis of Variance.

The form of the test is as follows. Let the Complete model be the model with all r means allowed to be different values. In the virtual reality case we have the following numbers of restrictions.

- $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ - There are 3 restrictions
- $H_0 : \mu_1, \mu_2 = \mu_3 = \mu_4$ - There are 2 restrictions
- $H_0 : \mu_1, \mu_2, \mu_3 = \mu_4$ - There is 1 restriction

Once we fit the Complete (No restriction) and Reduced (Restricted) models, we can obtain the Test Statistic as follows. Start with the first restriction: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_c$ where μ_c is the common mean.

$$\text{Reduced Model: (1 Mean)} \qquad \hat{\mu}_c = \overline{Y}_{\bullet\bullet} = \hat{Y}_{ij}(R)$$

$$\Rightarrow \quad SSE(R) = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(Y_{ij} - \hat{Y}_{ij}(R)\right)^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_{\bullet\bullet}\right)^2 = SSTO \qquad df_R = n_T - 1$$

$$\text{Complete Model: (4 Means)} \qquad \hat{\mu}_i = \overline{Y}_{i\bullet} = \hat{Y}_{ij}(C)$$

$$\Rightarrow \quad SSE(C) = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(Y_{ij} - \hat{Y}_{ij}(C)\right)^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_{i\bullet}\right)^2 = SSE \qquad df_C = n_T - r$$

The general linear F-test is conducted as follows once the previous models have been fit.

$$F^* = \frac{\left[\frac{SSE(R)-SSE(C)}{df_R - df_C}\right]}{\left[\frac{SSE(C)}{df_C}\right]} \qquad \text{Reject } H_0 \text{ if: } \quad F^* \geq F_{1-\alpha, df_R - df_C, df_C}$$

In the case of the test of $H_0 : \mu_1 = \mu_2 = \cdots = \mu_r$, we obtain the following result where we reproduce the ANOVA F-test.

$$SSE(R) = SSTO \quad SSE(C) = SSE \quad \Rightarrow \quad SSE(R) - SSE(C) = SSTR$$

$$df_R = n_T - 1 \quad df_C = n_T - r \quad \Rightarrow \quad df_R - df_C = (n_T - 1) - (n_T - r) = r - 1$$

$$F^* = \frac{\left[\frac{SSE(R)-SSE(C)}{df_R - df_C}\right]}{\left[\frac{SSE(C)}{df_C}\right]} = \frac{\left[\frac{SSTR}{r-1}\right]}{\left[\frac{SSE}{n_T - r}\right]} = \frac{MSTR}{MSE}$$

For the second model described above (equality of virtual reality means), the reduced model would have a single mean for the LEC/MAT control group and common means for the three virtual reality treatments.

$$\hat{Y}_{1j} = \overline{Y}_{1\bullet} \qquad \hat{Y}_{2j} = \hat{Y}_{3j} = \hat{Y}_{4j} = \frac{\sum_{i=2}^{4}\sum_{j=1}^{n_i} Y_{ij}}{n_2 + n_3 + n_4} = \frac{\sum_{i=2}^{4} n_i \overline{Y}_{i\bullet}}{\sum_{i=2}^{4} n_i}$$

The third model involves leaving treatments 1 and 2 as individual groups and combines treatments 3 and 4 into a single group.

$$\hat{Y}_{1j} = \overline{Y}_{1\bullet} \qquad \hat{Y}_{2j} = \overline{Y}_{2\bullet} \qquad \hat{Y}_{3j} = \hat{Y}_{4j} = \frac{\sum_{i=3}^{4} n_i \overline{Y}_{i\bullet}}{\sum_{i=3}^{4} n_i}$$

We will fit these models in R and use the **anova** command to conduct the F-tests from the various model fits. The trick is to create new treatment factors for the various restrictions, and fit an **lm** or **aov** object.

```
##   grp.trt procKnow
## 1       1   4.5614
## 2       1   6.6593
## 3       1   5.6427
## 4       1   6.4394
## 5       1   4.8635
## 6       1   0.3268
```

```
##
##   1 234
##  16  48
```

```
##
##  1  2 34
## 16 16 32
```

```
## Analysis of Variance Table
##
## Response: procKnow
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 63 331.48  5.2616
```

```
## Analysis of Variance Table
##
## Response: procKnow
##                 Df  Sum Sq Mean Sq F value   Pr(>F)
## factor(grp.trt)  3  65.664 21.8880  4.9406 0.003931 **
## Residuals       60 265.815  4.4302
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table
##
## Model 1: procKnow ~ 1
## Model 2: procKnow ~ factor(grp.trt)
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     63 331.48
## 2     60 265.81  3    65.664 4.9406 0.003931 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table
##
## Response: procKnow
##                       Df  Sum Sq Mean Sq F value   Pr(>F)
## factor(grp.trt_1_234)  1  56.816  56.816  12.825 0.000672 ***
## Residuals             62 274.663   4.430
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table
##
## Model 1: procKnow ~ factor(grp.trt_1_234)
## Model 2: procKnow ~ factor(grp.trt)
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     62 274.66
## 2     60 265.81  2    8.8479 0.9986 0.3744

## Analysis of Variance Table
##
## Response: procKnow
##                        Df Sum Sq Mean Sq F value   Pr(>F)
## factor(grp.trt_1_2_34)  2  65.51  32.755  7.5123 0.001212 **
## Residuals              61 265.97   4.360
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table
##
## Model 1: procKnow ~ factor(grp.trt_1_2_34)
## Model 2: procKnow ~ factor(grp.trt)
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     61 265.97
## 2     60 265.81  1   0.15432 0.0348 0.8526
```

These are the interpretations of the results of the three tests from the models fit above.

- $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ - Conclude the 4 means are NOT all equal
- $H_0 : \mu_1, \mu_2 = \mu_3 = \mu_4$ - Fail to reject that 3 virtual means are equal
- $H_0 : \mu_1, \mu_2, \mu_3 = \mu_4$ - Fail to reject that 2 HMD means are equal

∇

Later we will see formally how to make contrasts among treatments to answer specific research questions.

2.3.2 Power Calculations

When the null hypothesis $H_0 : \mu_1 = \cdots = \mu_r$ is false, the F-statistic is distributed non-central F with degrees of freedom $\nu_1 = r - 1$ and $\nu_2 = n_T - r$, and non-centrality parameter λ defined below. We will use the notation $F^* \sim F_{\nu_1,\nu_2}(\lambda)$.

$$\lambda = \frac{\sum_{i=1}^{r} n_i \left(\mu_i - \mu_\bullet\right)^2}{\sigma^2} \qquad \mu_\bullet = \frac{\sum_{i=1}^{r} n_i \mu_i}{n_T}$$

For a given configuration $\{n_i\}$, $\{\mu_i\}$, and σ^2, the **power** of the test, the probability that it will reject H_0 can be computed as follows.

- Compute the non-centrality parameter λ
- Obtain the critical value for the F-statistic: $F_{1-\alpha;r-1,n_T-r}$
- Obtain the area in the non-central F-distribution that exceeds the critical value in the previous step: $P\left(F_{\nu_1,\nu_2}(\lambda) \geq F_{1-\alpha;r-1,n_T-r}\right)$

In many practical applications, researchers are interested in detecting a certain effect size among treatments and must iteratively solve for sample sizes that assure a certain power, such as 0.80.

Example 2.6 - Virtual Training for a Lifeboat Launching Task

Consider the example described previously, and suppose we have samples of size $n_i = 5$ from each treatment.

$$n_1 = n_2 = n_3 = 5 \qquad \mu_1 = 40 \quad \mu_2 = 50 \quad \mu_3 = 60 \quad \mu_\bullet = 50 \qquad \sigma^2 = 10^2 = 100$$

$$\lambda = \frac{5(40-50)^2 + 5(50-50)^2 + 5(60-50)^2}{100} = \frac{1000}{100} = 10$$

If we were conducting a test of equal means, we would reject the null hypothesis if the F-statistic $MSTR/MSE$ exceeds the $1 - \alpha$ quantile of the F_{r-1,n_T-r} distribution. The following plot gives the central and non-central F-distributions for this scenario with with critical value for $\alpha = 0.05$ significance level $F_{.95;2,12} = 3.885$. Note that the area above 3.885 in the non-central F-distribution represents the **power**, the probability that we reject H_0 under this scenario for r, $\{\mu_i\}$, σ, and $\{n_i\}$. Using the **pf** function in R, we find the power to be .7015. In many experimental settings, the goal is to reach a power of .80 to detect a particular effect. In that case, given the settings for $\{\mu_i\}$, and σ, we would need to increase the sample sizes within treatments, $\{n_i\}$.

```
## [1] 3.885294
```

```
## [1] 50
```

```
## [1] 10
```

```
## [1] 0.7015083
```

∇

2.4 Technical Details

2.4.1 Least Squares Estimation for the Cell Means Model (Regression Approach)

The cell means model can be written as a regression model in matrix form as follows. Note that there will not be an intercept in this formulation. We assume the data are stacked by treatment, then replicate within treatment in the $n_T \times 1$ vector Y. The vector β representing the regression coefficients is the $r \times 1$ vector containing $\mu_1, \ldots, \mu_r$

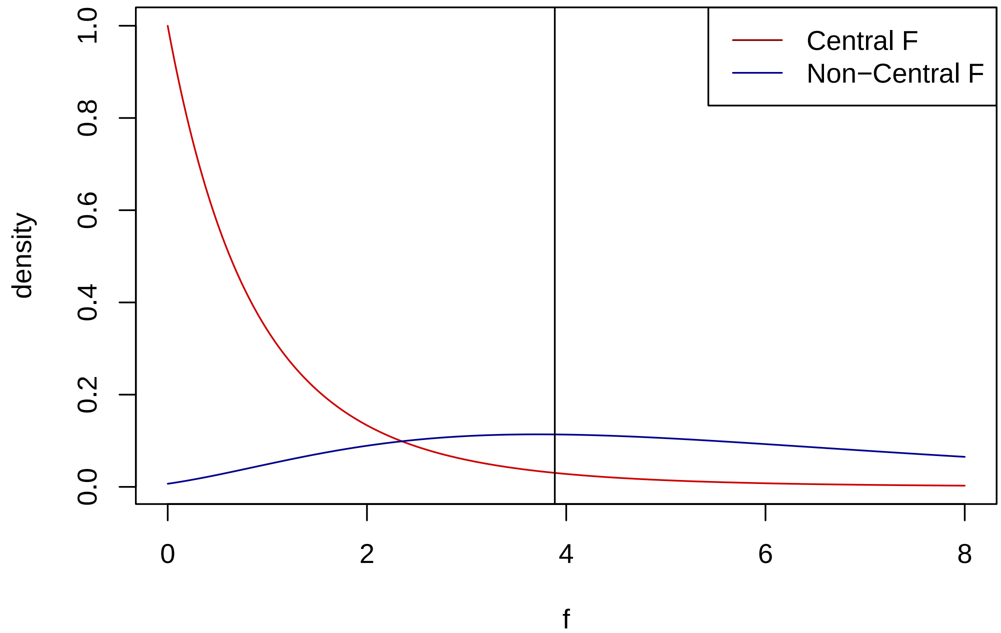


Figure 2.5: Central and non-central F-distributions used for testing for treatment effects and power calculations

and the X matrix is the $n_T \times r$ matrix containing a 1 if the observation in row i is in the treatment (column) j and 0 otherwise. We will formalize this for the simple case of $r = 3$ treatments and $n_1 = n_2 = n_3 = 2$ replicates per treatment, then we will generalize it. As with the linear regression model, we will write the model in matrix form as follows.

$$Y = X\beta + \epsilon$$

where:

$$Y = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad \beta = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \qquad \epsilon = \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{31} \\ \epsilon_{32} \end{bmatrix}$$

Recall that in linear regression, we obtain the ordinary least squares estimator of β, which is labelled b as follows. Further the variance-covariance matrix of b is given as well.

$$b = \left(X'X\right)^{-1} X'Y \qquad \sigma^2\{b\} = \sigma^2 \left(X'X\right)^{-1}$$

For this example, we obtain the following results.

$$X'X = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$\Rightarrow \quad (X'X)^{-1} = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}$$

Continuing on the computations, we obtain the following results.

$$X'Y = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix} = \begin{bmatrix} Y_{11} + Y_{12} \\ Y_{21} + Y_{22} \\ Y_{31} + Y_{32} \end{bmatrix}$$

$$\Rightarrow \quad b = (X'X)^{-1} X'Y = \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} Y_{11} + Y_{12} \\ Y_{21} + Y_{22} \\ Y_{31} + Y_{32} \end{bmatrix} = \begin{bmatrix} \frac{Y_{11}+Y_{12}}{2} \\ \frac{Y_{21}+Y_{22}}{2} \\ \frac{Y_{31}+Y_{32}}{2} \end{bmatrix}$$

$$\sigma^2\{b\} = \sigma^2 (X'X)^{-1} = \sigma^2 \begin{bmatrix} \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}$$

So, this leads to the (hopefully) not surprising general results.

$$\overline{Y}_{i\bullet} = \frac{\sum_{j=1}^{n_i} Y_{ij}}{n_i} \qquad b = \begin{bmatrix} \overline{Y}_{1\bullet} \\ \overline{Y}_{2\bullet} \\ \vdots \\ \overline{Y}_{r-1,\bullet} \\ \overline{Y}_{r\bullet} \end{bmatrix}$$

$$\sigma^2\{b\} = \sigma^2 \begin{bmatrix} \frac{1}{n_1} & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{n_2} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{n_{r-1}} & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{n_r} \end{bmatrix}$$

Example 2.7 - Virtual Training for a Lifeboat Launching Task

For this example, we will set up the X matrix and Y vector as follow. Let 1_{16} represent a 16×1 vector of 1^s and 0_{16} represent a 16×1 vector of 0^s and Y_i be the 16×1 vector of $Y_{i1}, \ldots, Y_{i,16}$ for $i = 1, 2, 3, 4$. Further, let $Y_{i\bullet} = \sum_{j=1}^{n_i} Y_{ij}$ and $\overline{Y}_{i\bullet} = Y_{i\bullet}/n_i$.

$$X = \begin{bmatrix} 1_{16} & 0_{16} & 0_{16} & 0_{16} \\ 0_{16} & 1_{16} & 0_{16} & 0_{16} \\ 0_{16} & 0_{16} & 1_{16} & 0_{16} \\ 0_{16} & 0_{16} & 0_{16} & 1_{16} \end{bmatrix} \qquad X'X = \begin{bmatrix} 16 & 0 & 0 & 0 \\ 0 & 16 & 0 & 0 \\ 0 & 0 & 16 & 0 \\ 0 & 0 & 0 & 16 \end{bmatrix}$$

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} \qquad X'Y = \begin{bmatrix} Y_{1\bullet} \\ Y_{2\bullet} \\ Y_{3\bullet} \\ Y_{4\bullet} \end{bmatrix} = \begin{bmatrix} 78.8888 \\ 123.3334 \\ 107.7777 \\ 109.9999 \end{bmatrix}$$

$$b = (X'X)^{-1}X'Y = \begin{bmatrix} \frac{1}{16} & 0 & 0 & 0 \\ 0 & \frac{1}{16} & 0 & 0 \\ 0 & 0 & \frac{1}{16} & 0 \\ 0 & 0 & 0 & \frac{1}{16} \end{bmatrix} \begin{bmatrix} 78.8888 \\ 123.3334 \\ 107.7777 \\ 109.9999 \end{bmatrix} = \begin{bmatrix} 4.9306 \\ 7.7083 \\ 6.7361 \\ 6.8750 \end{bmatrix}$$

The variance of each element of b is $\sigma^2/16$, and the standard error is $\sigma/\sqrt{16} = \sigma/4$. Here, we make use of the **lm** function in R to obtain point estimates and estimated standard errors for the cell means model. Note that the intercept is removed (-1) in the command and we have to define the treatment as a factor variable (so as not to fit a straight line).

```
vt <- read.csv("http://www.stat.ufl.edu/~winner/data/virtual_training.csv")
head(vt, 2)
```

```
##   grp.trt procKnow
## 1       1   4.5614
## 2       1   6.6593
```

```
vt.mod1 <- lm(procKnow ~ factor(grp.trt) - 1, data=vt)
summary(vt.mod1)
```

```
##
## Call:
## lm(formula = procKnow ~ factor(grp.trt) - 1, data = vt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0831 -1.4444  0.5774  1.5495  3.1900
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## factor(grp.trt)1   4.9306     0.5262    9.37 2.37e-13 ***
## factor(grp.trt)2   7.7083     0.5262   14.65  < 2e-16 ***
## factor(grp.trt)3   6.7361     0.5262   12.80  < 2e-16 ***
## factor(grp.trt)4   6.8750     0.5262   13.06  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.105 on 60 degrees of freedom
## Multiple R-squared:  0.9139, Adjusted R-squared:  0.9082
## F-statistic: 159.2 on 4 and 60 DF,  p-value: < 2.2e-16
```

The summary output includes the residual standard error, which is our estimate of σ, the standard deviation of the model errors. Derivation of the unbiased estimator of σ^2 is described below.

∇

2.4.2 Scalar Estimation for the Cell Means Model

We re-write the model here in scalar form.

$$Y_{ij} = \mu_i + \epsilon_{ij} \qquad i = 1, \dots, r; \quad j = 1, \dots, n_i \qquad \epsilon_{ij} \sim NID\left(0, \sigma^2\right)$$

Let $Q = \sum_{i=1}^{r} \sum_{j=1}^{n_i} \epsilon_{ij}^2$. We choose $\hat{\mu}_k$ that minimizes Q for $k = 1, \dots, r$. This is done by taking the derivative of Q with respect to μ_k, setting it equal to zero and solving for the least squares estimator.

$$Q = \sum_{i=1}^{r} \sum_{j=1}^{n_i} \epsilon_{ij}^2 = \sum_{i=1}^{r} \sum_{j=1}^{n_i} \left(Y_{ij} - \mu_i\right)^2$$

$$\Rightarrow \quad \frac{\partial Q}{\partial \mu_k} = \frac{\partial}{\partial \mu_k} \sum_{i=1}^{r} \sum_{j=1}^{n_i} \left(Y_{ij} - \mu_i\right)^2 = 2 \sum_{j=1}^{n_k} \left(Y_{kj} - \mu_k\right)(-1)$$

Note that this only involves the n_k observations for treatment k. Now we set the partial derivative to zero for the least squares (and maximum likelihood) estimator of μ_k.

$$-2 \sum_{j=1}^{n_k} \left(Y_{kj} - \hat{\mu_k}\right) = 0 \quad \Rightarrow \quad \sum_{j=1}^{n_k} Y_{kj} = n_k \hat{\mu_k} \quad \Rightarrow \quad \hat{\mu_k} = \frac{\sum_{j=1}^{n_k} Y_{kj}}{n_k} = \overline{Y}_{k\bullet}$$

Thus the fitted values for Y_{ij} are $\hat{Y}_{ij} = \overline{Y}_{i\bullet}$.

2.4.3 Estimation for the Treatment Effects Model

Let μ_i be the population mean for treatment i, and let $\mu_\bullet$ be the unweighted mean of the μ_i^s. Note that this will differ from $\mu_\bullet$ used in the cell means model when the sample sizes are not all equal.

$$\mu_\bullet = \frac{\sum_{i=1}^{r} \mu_i}{r} \qquad \mu_i = \mu_\bullet + (\mu_i - \mu_\bullet) = \mu_\bullet + \tau_i \quad \Rightarrow \quad \tau_i = \mu_i - \mu_\bullet$$

We will refer to τ_i as the effect of treatment i relative to the mean among all treatments. The sum of τ_i is zero, when we use the unweighted mean $\mu_\bullet$.

$$\mu_1 = \cdots = \mu_r = \mu_\bullet \quad \Rightarrow \quad \tau_1 = \cdots = \tau_r = 0$$

$$Y_{ij} = \mu_\bullet + \tau_i + \epsilon_{ij} \qquad i = 1, \dots, r; \quad j = 1, \dots, n_i \qquad \epsilon_{ij} \sim NID\left(0, \sigma^2\right)$$

$$E\{Y_{ij}\} = E\{\mu_\bullet + \tau_i + \epsilon_{ij}\} = \mu_\bullet + \tau_i + 0 = \mu_\bullet + \tau_i \qquad \sigma^2\{Y_{ij}\} = \sigma^2\{\mu_\bullet + \tau_i + \epsilon_{ij}\} = 0 + \sigma^2 = \sigma^2$$

$$\sum_{i=1}^{r} \tau_i = 0 \quad \Rightarrow \quad \tau_r = -\sum_{i=1}^{r-1} \tau_i$$

2.4.4 Least Squares Estimation (Regression Approach)

Let Y be defined as before. We will parameterize β in terms of $\mu_\bullet$ and $\tau_1, \dots, \tau_{r-1}$, keeping in mind the result above regarding τ_r. First consider the case as we did previously with $r = 3$ and $n_1 = n_2 = n_3 = 2$.

$$Y = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \qquad \beta = \begin{bmatrix} \mu_\bullet \\ \tau_1 \\ \tau_2 \end{bmatrix} \qquad \epsilon = \begin{bmatrix} \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{21} \\ \epsilon_{22} \\ \epsilon_{31} \\ \epsilon_{32} \end{bmatrix}$$

$$b = \left(X'X\right)^{-1} X'Y \qquad \sigma^2\{b\} = \sigma^2 \left(X'X\right)^{-1}$$

For this example, we obtain the following results.

$$X'X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 4 & 2 \\ 0 & 2 & 4 \end{bmatrix}$$

Making use of cofactors, we obtain $(X'X)^{-1}$ below.

$$\Rightarrow \quad (X'X)^{-1} = \begin{bmatrix} \frac{1}{6} & 0 & 0 \\ 0 & \frac{1}{3} & -\frac{1}{6} \\ 0 & -\frac{1}{6} & \frac{1}{3} \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

Continuing on the computations, we obtain the following results.

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \\ Y_{31} \\ Y_{32} \end{bmatrix} = \begin{bmatrix} Y_{11} + Y_{12} + Y_{21} + Y_{22} + Y_{31} + Y_{32} \\ Y_{11} + Y_{12} - Y_{31} - Y_{32} \\ Y_{21} + Y_{22} - Y_{31} - Y_{32} \end{bmatrix}$$

When the dust settles, the estimator b for β is obtained below.

$$b = (X'X)^{-1} X'Y = \frac{1}{6} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} Y_{11} + Y_{12} + Y_{21} + Y_{22} + Y_{31} + Y_{32} \\ Y_{11} + Y_{12} - Y_{31} - Y_{32} \\ Y_{21} + Y_{22} - Y_{31} - Y_{32} \end{bmatrix} = \begin{bmatrix} \overline{Y}_{\bullet\bullet} \\ \overline{Y}_{1\bullet} - \overline{Y}_{\bullet\bullet} \\ \overline{Y}_{2\bullet} - \overline{Y}_{\bullet\bullet} \end{bmatrix}$$

The general result is as follows, allowing for unequal sample sizes.

$$b = \begin{bmatrix} \frac{1}{r}\sum_{i=1}^{r} \overline{Y}_{i\bullet} \\ \frac{r-1}{r}\overline{Y}_{1\bullet} - \frac{1}{r}\sum_{i=2}^{r} \overline{Y}_{i\bullet} \\ \vdots \\ \frac{r-1}{r}\overline{Y}_{r-1,\bullet} - \frac{1}{r}\sum_{\substack{i=1 \\ i\neq r-1}}^{r} \overline{Y}_{i\bullet} \end{bmatrix} \qquad \hat{\tau}_r = -\sum_{i=1}^{r-1} \hat{\tau}_i = \frac{r-1}{r}\overline{Y}_{r\bullet} - \frac{1}{r}\sum_{i=1}^{r-1} \overline{Y}_{i\bullet}$$

From these expressions, we obtain the following variances for $\hat{\mu}$ and $\hat{\tau}_k$.

$$\sigma^2\{\hat{\mu}\} = \frac{\sigma^2}{r^2}\sum_{i=1}^{r}\frac{1}{n_i} \qquad \sigma^2\{\hat{\tau}_k\} = \frac{\sigma^2}{r^2}\left[\frac{(r-1)^2}{n_k} + \sum_{\substack{i=1 \\ i\neq k}}^{r}\frac{1}{n_i}\right]$$

2.4.5 The Analysis of Variance

In this section, we obtain a decomposition of the total variation of the individual measurements around the overall mean, the **Total Sum of Squares**, which we will denote $SSTO$. Further, we will derive the expectations of the error and treatment sums of squares (and mean squares).

$$\overline{Y}_{\bullet\bullet} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i} Y_{ij}}{n_T} \qquad SSTO = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_{\bullet\bullet}\right)^2$$

A useful expansion of this is helpful for derivations below.

$$SSTO = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(Y_{ij}^2 + \overline{Y}_{\bullet\bullet}^2 - 2Y_{ij}\overline{Y}_{\bullet\bullet}\right)$$

$$\Rightarrow \quad SSTO = \sum_{i=1}^{r}\sum_{j=1}^{n_i}Y_{ij}^2 + n_T\overline{Y}_{\bullet\bullet}^2 - 2\overline{Y}_{\bullet\bullet}\sum_{i=1}^{r}\sum_{j=1}^{n_i}Y_{ij}$$

$$\Rightarrow \quad SSTO = \sum_{i=1}^{r}\sum_{j=1}^{n_i}Y_{ij}^2 - n_T\overline{Y}_{\bullet\bullet}^2$$

As was the case for the linear regression model, we obtain fitted values for the observations in the form $\hat{Y} = Xb$. This leads to the error sum of squares, labelled as $SSE = \left(Y - \hat{Y}\right)'\left(Y - \hat{Y}\right)$ here (note that in R, for linear models, it is labelled as RSS, for the residual sum of squares).

$$SSE = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(Y_{ij} - \hat{Y}_{ij}\right)^2$$

Noting that for this model, $\hat{Y}_{ij} = \overline{Y}_{i\bullet}$, we obtain the following formula for SSE.

$$SSE = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y}_{i\bullet}\right)^2$$

We can expand this form of SSE which will be used to derive the unbiased estimator for σ^2, the error variance.

$$SSE = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(Y_{ij}^2 + \overline{Y}_{i\bullet}^2 - 2Y_{ij}\overline{Y}_{i\bullet}\right)$$

$$\Rightarrow \quad SSE = \sum_{i=1}^{r}\sum_{j=1}^{n_i}Y_{ij}^2 + \sum_{i=1}^{r}n_i\overline{Y}_{i\bullet}^2 - 2\sum_{i=1}^{r}\overline{Y}_{i\bullet}n_i\overline{Y}_{i\bullet}$$

$$\Rightarrow \quad SSE = \sum_{i=1}^{r}\sum_{j=1}^{n_i}Y_{ij}^2 - \sum_{i=1}^{r}n_i\overline{Y}_{i\bullet}^2$$

To obtain an unbiased estimator for σ^2, consider the following results from the Cell Means Model.

$$E\left\{Y_{ij}\right\} = \mu_i \qquad \sigma^2\left\{Y_{ij}\right\} = \sigma^2$$

$$\Rightarrow \quad E\left\{Y_{ij}^2\right\} = \sigma^2\left\{Y_{ij}\right\} + \left[E\left\{Y_{ij}\right\}\right]^2 = \sigma^2 + \mu_i^2$$

Similarly, we obtain the expected value of the treatment sample mean squared below.

$$E\left\{\overline{Y}_{i\bullet}\right\} = \frac{1}{n_i}\sum_{j=1}^{n_i}E\left\{Y_{ij}\right\} = \mu_i \qquad \sigma^2\left\{\overline{Y}_{i\bullet}\right\} = \frac{1}{n_i^2}\sum_{j=1}^{n_i}\sigma^2\left\{Y_{ij}\right\} = \frac{\sigma^2}{n_i}$$

$$\Rightarrow \quad E\left\{\overline{Y}_{i\bullet}^2\right\} = \frac{\sigma^2}{n_i} + \mu_i^2$$

Next, we obtain $E\left\{SSE\right\}$ making use of these results.

$$E\left\{SSE\right\} = \sum_{i=1}^{r}\sum_{j=1}^{n_i}E\left\{Y_{ij}^2\right\} - \sum_{i=1}^{r}n_iE\left\{\overline{Y}_{i\bullet}^2\right\}$$

$$\Rightarrow \quad E\{SSE\} = \sum_{i=1}^{r} n_i \left(\sigma^2 + \mu_i^2\right) - \sum_{i=1}^{r} n_i \left(\frac{\sigma^2}{n_i} + \mu_i^2\right)$$

$$\Rightarrow \quad E\{SSE\} = n_T\sigma^2 - r\sigma^2 = (n_T - r)\,\sigma^2$$

Thus, if we divide SSE by $n_T - r$, we obtain an unbiased estimator of σ^2. This is the **Mean Square Error**, denoted as MSE, and often labelled as S^2 (as an estimator) and s^2 when computed on an observed set of data. The denominator $n_T - r$ is the **Error Degrees of Freedom**.

$$MSE = \frac{SSE}{n_T - r} \qquad E\{MSE\} = \sigma^2$$

The **Treatment Sum of Squares**, denoted as $SSTR$, acts like the regression sum of squares for linear regression models. That is, $SSTR = \left(\hat{Y} - \overline{Y}1\right)' \left(\hat{Y} - \overline{Y}1\right)$. Keeping in mind that the fitted value for Y_{ij} is $\hat{Y}_{ij} = \overline{Y}_{i\bullet}$, we obtain SSTR as follows.

$$SSTR = \sum_{i=1}^{r}\sum_{j=1}^{n_i} \left(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet}\right)^2 = \sum_{i=1}^{r} n_i \left(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet}\right)^2$$

Again, it is helpful to expand this as before.

$$SSTR = \sum_{i=1}^{r} n_i \left(\overline{Y}_{i\bullet}\right)^2 - n_T\overline{Y}_{\bullet\bullet}^2$$

We know the expectation of the first term from the derivation of $E\{SSE\}$, so we just need the second term.

$$E\left\{\overline{Y}_{\bullet\bullet}\right\} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i} E\left\{Y_{ij}\right\}}{n_T} = \frac{\sum_{i=1}^{r} n_i\mu_i}{n_T} = \mu_\bullet$$

$$\sigma^2\left\{\overline{Y}_{\bullet\bullet}\right\} = \frac{1}{n_T^2}\sum_{i=1}^{r}\sum_{j=1}^{n_i}\sigma^2\left\{Y_{ij}\right\} = \frac{\sigma^2}{n_T}$$

$$\Rightarrow \quad E\{SSTR\} = \sum_{i=1}^{r} n_i\left(\frac{\sigma^2}{n_i} + \mu_i^2\right) - \left(\sigma^2 + n_T\mu_\bullet^2\right)$$

$$\Rightarrow \quad E\{SSTR\} = \sigma^2(r-1) + \sum_{i=1}^{r} n_i\mu_i^2 - n_T\mu_\bullet^2$$

$$\Rightarrow \quad E\{SSTR\} = \sigma^2(r-1) + \sum_{i=1}^{r} n_i\left(\mu_i - \mu_\bullet\right)^2$$

The treatment degrees of freedom are $r - 1$, the number of treatments minus one. We compute the **Treatment Mean Square** by dividing the Treatment Sum of Squares by its degrees of freedom and obtain its expectation below.

$$MSTR = \frac{SSTR}{r-1} \qquad E\{MSTR\} = \sigma^2 + \frac{\sum_{i=1}^{r} n_i\left(\mu_i - \mu_\bullet\right)^2}{r-1}$$

When all treatment means are equal, that is, $\mu_1 = \cdots = \mu_r = \mu_\bullet$, then the expected mean squares for treatment and error are equal ($E\{MSTR\} = E\{MSE\} = \sigma^2$), otherwise it is higher for treatments than for error. Further, under the assumptions of the model thus far (normal, independent errors with constant variance) we obtain the following results.

$$\frac{SSE}{\sigma^2} = \frac{(n_T - r)\, MSE}{\sigma^2} \sim \chi^2_{n_T - r}$$

$$\frac{SSTR}{\sigma^2} = \frac{(r-1) MSTR}{\sigma^2} \sim \chi^2_{r-1}(\lambda)$$

where λ is the non-centrality parameter of the chi-square distribution. In the One-Way Analysis of Variance, λ is computed as follow.

$$\lambda = \frac{\sum_{i=1}^{r} n_i \left(\mu_i - \mu_\bullet\right)^2}{\sigma^2}$$

Further $MSTR$ and MSE are independent. When all μ_i are equal, then the chi-square distribution for $SSTR/\sigma^2$ is "central" chi-square, and we can use the (central) F-distribution to test the hypothesis $H_0 : \mu_1 = \cdots = \mu_r$.

We will make use of the following results when testing for differences among means. Suppose W_1 is chi-square with ν_1 degrees of freedom, W_2 is chi-square with ν_2 degrees of freedom and W_1 and W_2 are independent, then we get the following distributional result.

$$W_1 \sim \chi^2_{\nu_1} \quad W_2 \sim \chi^2_{\nu_2} \quad W_1 \perp W_2 \quad \Rightarrow \quad F = \frac{W_1/\nu_1}{W_2/\nu_2} \sim F_{\nu_1, \nu_2}$$

When W_1 is chi-square with ν_1 degrees of freedom and non-centrality parameter λ and W_2 is chi-square with ν_2 degrees of freedom and W_1 and W_2 are independent, then we get the following distributional result.

$$W_1 \sim \chi^2_{\nu_1}(\lambda) \quad W_2 \sim \chi^2_{\nu_2} \quad W_1 \perp W_2 \quad \Rightarrow \quad F = \frac{W_1/\nu_1}{W_2/\nu_2} \sim F_{\nu_1, \nu_2}(\lambda)$$

This is the non-central F-distribution. We can use R to obtain probabilities, densities, quantiles, and random samples for non-central F-distributions.

```
library(tidyverse)
library(kableExtra)
library(effectsize)
library(agricolae)
```

Chapter 3

Analysis of Treatment Means

In this chapter, we will consider estimation and inference concerning treatment/group means. We will first consider estimating individual treatment means. Then, we will consider **Contrasts** among treatment means. Finally, we consider all pairwise comparisons.

3.1 Individual Treatment Means

For the cell means model, we may be interested in making inference regarding the individual treatment.group means. The model is of the following form.

$$Y_{ij} = \mu_i + \epsilon_{ij} \qquad i = 1, \dots, r; \quad j = 1, \dots, n_i \qquad \epsilon_{ij} \sim NID\left(0, \sigma^2\right)$$

The least squares estimator for μ_i is $\hat{\mu}_i = \overline{Y}_{i\bullet}$.

$$\hat{\mu}_i = \overline{Y}_{i\bullet} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij} \qquad E\left\{\overline{Y}_{i\bullet}\right\} = \frac{1}{n_i} n_i \mu_i = \mu_i \qquad \sigma^2\left\{\overline{Y}_{i\bullet}\right\} = \frac{1}{n_i^2} n_i \sigma^2 = \frac{\sigma^2}{n_i}$$

For this model, the treatment means are independent and we have the following distributional properties.

$$\overline{Y}_{i\bullet} \sim N\left(\mu_i, \frac{\sigma^2}{n_i}\right) \qquad \frac{\left(n_T - r\right) MSE}{\sigma^2} \sim \chi^2_{n_T - r}$$

Further, $\overline{Y}_{i\bullet}$ and MSE are independent. Since σ^2 is unknown in practice and must be estimated with $s^2 = MSE$, we obtain the following results.

$$s^2\left\{\overline{Y}_{i\bullet}\right\} = \frac{MSE}{n_i} \qquad \frac{\overline{Y}_{i\bullet} - \mu_i}{s\left\{\overline{Y}_{i\bullet}\right\}} \sim t_{n_T - r}$$

This leads to a $(1-\alpha)100\%$ Confidence Interval for μ_i of this form.

$$\overline{Y}_{i\bullet} \pm t_{1-\alpha/2; n_T - r}\, s\left\{\overline{Y}_{i\bullet}\right\} \quad \equiv \quad \overline{Y}_{i\bullet} \pm t_{1-\alpha/2; n_T - r}\sqrt{\frac{MSE}{n_i}}$$

In the rare situation that we wish to test $H_0 : \mu_i = c$ for some pre-specified constant c, we can compute the t-statistic as follows.

$$\text{Test Statistic:} \quad t_i^* = \frac{\overline{Y}_{i\bullet} - c}{s\left\{\overline{Y}_{i\bullet}\right\}} \qquad \text{Rejection Region:} \quad |t_i^*| \geq t_{1-\alpha/2; n_T - r}$$

Comparing Two Treatment Means

In some cases, researchers may wish to compare two specific treatment means. In this case, suppose the goal is to compare treatments i and i'. The parameter of interest is $D = \mu_i - \mu_{i'}$ and the estimator is $\widehat{D} = \overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}$. The mean and variance of $\widehat{D}$ are given below.

$$E\left\{\hat{D}\right\} = E\left\{\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}\right\} = \mu_i - \mu_{i'} = D$$

$$\sigma^2\left\{\hat{D}\right\} = \sigma^2\left\{\overline{Y}_{i\bullet}\right\} + \sigma^2\left\{\overline{Y}_{i'\bullet}\right\} = \sigma^2\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)$$

As σ^2 is not known in practice, the estimated variance of $\hat{D}$ replaces σ^2 with MSE.

$$s^2\left\{\hat{D}\right\} = s^2\left\{\overline{Y}_{i\bullet}\right\} + s^2\left\{\overline{Y}_{i'\bullet}\right\} = MSE\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)$$

A $(1-\alpha)100\%$ Confidence Interval for $D = \mu_i - \mu_{i'}$ is obtained as follows, as well as a test of $H_0 : D = 0$ versus $H_A : D \neq 0$.

$$\text{Confidence Interval:} \quad \hat{D} \pm t_{1-\alpha/2;n_T-r} s\left\{\hat{D}\right\} \equiv \hat{D} \pm t_{1-\alpha/2;n_T-r}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)}$$

$$\text{Test Statistic:} \quad t^* = \frac{\hat{D}}{s\left\{\hat{D}\right\}} \qquad \text{Rejection Region:} \quad |t^*| \geq t_{1-\alpha/2;n_T-r}$$

The only difference between this and the 2-sample t-test is that all r treatments/groups are used to estimate σ^2, and increasing the degrees of freedom to $n_T - r$. We will consider adjustments for comparing all pairs of treatments later in the chapter.

3.2 Contrasts Among Treatment Means

A **Contrast** is any linear function of treatment means such that the coefficients sum to 0. Thus, for instance $D = \mu_i - \mu_{i'}$ is a contrast among treatment means that only involves the difference between treatments i and i'. In general, we define contrasts as follows.

$$L = \sum_{i=1}^{r} c_i\mu_i \qquad \text{such that:} \ \sum_{i=1}^{r} c_i = 0$$

The estimator for L is $\hat{L}$ that replaces the $\{\mu_i\}$ with the treatment means $\left\{\overline{Y}_{i\bullet}\right\}$.

$$\hat{L} = \sum_{i=1}^{r} c_i\overline{Y}_{i\bullet} \qquad E\left\{\hat{L}\right\} = \sum_{i=1}^{r} c_i\mu_i = L$$

$$\sigma^2\left\{\hat{L}\right\} = \sum_{i=1}^{r} c_i^2\frac{\sigma^2}{n_i} \qquad\qquad s^2\left\{\hat{L}\right\} = MSE\sum_{i=1}^{r}\frac{c_i^2}{n_i}$$

We can make use of the t-distribution to obtain confidence intervals and tests regarding contrasts. Also, we can use an F-test as well.

$$\frac{\hat{L} - L}{s\left\{\hat{L}\right\}} \sim t_{n_T-r} \quad \Rightarrow \quad (1-\alpha)100\% \text{ CI for L:} \quad \hat{L} \pm t_{1-\alpha/2;n_T-r} s\left\{\hat{L}\right\}$$

For testing $H_0 : l = 0$ versus $H_A : L \neq 0$, we can use either a t-test or F-test, which provide the same conclusions and P-values. The t-test is given here.

$$\text{Test Statistic:} \quad t^* = \frac{\hat{L}}{s\left\{\hat{L}\right\}} \qquad \text{Rejection Region:} \quad |t^*| \geq t_{1-\alpha/2;n_T-r}$$

For the F-test, we first define the sum of squares for the contrast, then construct the F-statistic.

$$\text{Contrast Sum of Squares:} \quad SSL = \frac{\left(\hat{L}\right)^2}{\sum_{i=1}^{r} \frac{c_i^2}{n_i}}$$

$$n_1 = \cdots = n_r = n \quad \Rightarrow \quad SSL = n\frac{\left(\hat{L}\right)^2}{\sum_{i=1}^{r} c_i^2}$$

$$\text{Test Statistic:} \quad F^* = \frac{SSL}{MSE} \qquad \text{Rejection Region:} \quad F^* \geq F_{1-\alpha;1,n_T-r}$$

Two contrasts $L_1 = \sum_{i=1}^{r} a_i\mu_i$ and $L_2 = \sum_{i=1}^{r} b_i\mu_i$ where $\sum_{i=1}^{r} a_i = \sum_{i=1}^{r} b_i = 0$ are said to be **Orthogonal Contrasts** if the product of their coefficients divided by their sample sizes is 0.

$$L_1, L_2 \text{ are orthogonal if:} \quad \sum_{i=1}^{r} \frac{a_i b_i}{n_i} = 0$$

When the experiment is balanced, this simplifies to $\sum_{i=1}^{r} a_i b_i = 0$. Among r treatments/groups, we can construct $r-1$ pairwise orthogonal contrasts, whose sums of squares and degrees of freedom sum to $SSTR$ and df_{TR}, respectively.

Example 3.1 - Virtual Training for a Lifeboat Launching Task

Here, we re-describe the virtual training study to set up some interesting contrasts among the training methods. A study in South Korea compared $r = 4$ methods of training to conduct a lifeboat launching task [Jung and Ahn, 2018]. The treatments and their labels from the paper are given below and the response Y was a procedural knowledge score for subjects post training.

- **Lecture/Materials** - Traditional Lecture with no computer component (LEC/MAT)
- **Monitor/Keyboard** - Trained virtually with a monitor, keyboard, and mouse (MON/KEY)
- **Head-Mounted Display/Joypad** - Trained virtually with HMD and joypad (HMD/JOY)
- **Head-Mounted Display/Wearable Sensors** - (HMD/WEA)

There were a total of $n_T = 64$ subjects and they were randomized so that 16 subjects received each treatment ($n_1 = n_2 = n_3 = n_4 = 16$).

Consider the following 3 contrasts.

- **Lecture Materials versus Virtual Training** $a_1 = 3, a_2 = a_3 = a_4 = -1$
- **Monitor/Keyboard versus Head Mounted Display** $b_1 = 0, b_2 = 2, b_3 = b_4 = -1$
- **Head Mounted Diplays: Joypad versus Wearables** $c_1 = c_2 = 0, c_3 = 1, c_4 = -1$

All of these are contrasts, as their coefficients sum to zero. Further, they are pairwise orthogonal.

$$\sum_{i=1}^{r} a_i b_i = 3(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0 - 2 + 1 + 1 = 0$$

$$\sum_{i=1}^{r} a_i c_i = 3(0) + 2(0) + (-1)(1) + (-1)(-1) = 0 + 0 - 1 + 1 = 0$$

Table 3.1: Lifeboat Virtual Training Contrast Calculations

	n	Mean	a	b	d	L1	L2	L3	a^2/n	b^2/n	d^2/n
LEC/MAT (i=1)	16	4.931	3	0	0	14.793	0	0	0.5625	0	0
MON/KEY (i=2)	16	7.708	-1	2	0	-7.708	15.416	0	0.0625	0.25	0
HMD/JOY (i=3)	16	6.736	-1	-1	1	-6.736	-6.736	6.736	0.0625	0.0625	0.0625
HMD/WEA (i=4)	16	6.875	-1	-1	-1	-6.875	-6.875	6.875	0.0625	0.0625	0.0625
Sum	64		0	0	0	-6.526	1.805	-0.139	0.75	0.375	0.125

$$\sum_{i=1}^{r} b_i c_i = 0(0) + 2(0) + (-1)(1) + (-1)(-1) = 0 + 0 - 1 + 1 = 0$$

Table 3.1 gives calculations necessary to obtain the Confidence Intervals, t-tests, sums of squares and F-tests for the three contrasts. Recall from the last chapter, the following quantities.

$$SSTR = 65.628 \quad df_{TR} = 3 \qquad SSE = 265.815 \quad df_E = 60 \quad MSE = 4.430$$

The critical t-value for $\alpha = .05$ and $df_E = 60$ is $t_{.975,60} = 2.000$.

For contrast L_1, comparing the virtual training methods to the "control" treatment has the following values.

$$\hat{L}_1 = 3\overline{Y}_{1\bullet} - \overline{Y}_{2\bullet} - \overline{Y}_{3\bullet} - \overline{Y}_{4\bullet} = -6.526$$

$$s\left\{\hat{L}_1\right\} = \sqrt{MSE\sum_{i=1}^{4}\frac{a_i^2}{n_i}} = \sqrt{4.430(0.75)} = 1.823$$

$$95\% \text{ CI for } L_1: \quad -6.526 \pm 2.000(1.823) \equiv -6.526 \pm 3.646 \equiv (-10.172, -2.880)$$

$$\text{Test Statistic:} \quad t_1^* = \frac{-6.526}{1.823} = -3.580$$

$$\text{Rejection Region:} \quad |t_1^*| \geq 2.000 \qquad P = 2P\left(t_{60} \geq |-3.580|\right) = .0007$$

$$SSL_1 = \frac{\left(\hat{L}\right)^2}{\sum_{i=1}^{r}\frac{a_i^2}{n_i}} = \frac{(-6.526)^2}{0.75} = 56.785$$

$$\text{Test Statistic:} \quad F_1^* = \frac{SSL_1}{MSE} = \frac{56.785}{4.430} = 12.818$$

$$\text{Rejection Region:} \quad F_1^* \geq F_{.95,1,60} = 4.001 \qquad P = P\left(F_{1,60} \geq 12.818\right) = .0007$$

Without going through the calculations, we obtain the following values for L_2 and L_3.

$$L_2: \quad \hat{L}_2 = -1.805 \quad s\left\{\hat{L}_2\right\} = 1.289 \quad SSL_2 = 8.688$$

$$L_3: \quad \hat{L}_3 = -0.139 \quad s\left\{\hat{L}_3\right\} = 0.744 \quad SSL_3 = 0.155$$

The sums of squares for the three pairwise orthogonal contrasts sum up to the treatment sum of squares.

$$SSL_1 + SSL_2 + SSL_3 = 56.785 + 8.688 + 0.155 = 65.628 = SSTR$$

The main take away is that the three virtual training methods (as a group) have significantly higher scores than the Lecture/Material control group. There appears to be no significant differences among the virtual training methods.

While it is possible to construct contrasts directly in R, it is very easy to compute them directly, as we show below.

```
##   grp.trt procKnow
## 1       1   4.5614
## 2       1   6.6593
## 3       1   5.6427
## 4       1   6.4394
## 5       1   4.8635
## 6       1   0.3268
```

```
##      vt.n  vt.mean   vt.var
## [1,]   16 4.930550 3.763633
## [2,]   16 7.708337 2.044874
## [3,]   16 6.736106 7.952409
## [4,]   16 6.874994 3.960078
```

```
##                    Lhat Std Err     t*      LB     UB     SS     F* P(>F*)
## L1:(3,-1,-1,-1) -6.528   1.823 -3.581 -10.174 -2.882 56.816 12.825  0.001
## L2:(0,2,-1,-1)   1.806   1.289  1.401  -0.773  4.384  8.694  1.962  0.166
## L3:(0,0,1,-1)   -0.139   0.744 -0.187  -1.627  1.350  0.154  0.035  0.853
```

∇

3.3 Simultaneous Comparisons

Often when fitting models, we wish to make multiple comparisons simultaneously. If we have k independent comparisons and we make each one with significance level of α, the chances that all conclusions are correct is $(1-\alpha)^k$ which is less than $1-\alpha$.

Other issues arise when researchers wait until observing the data to choose which comparisons to make (like comparing the maximum and minimum observed means). This practice has the effect of inflating the Type I error rate, α, and is referred to as **data snooping**.

There are various methods for making simultaneous corrections. One common method is to do each comparison with $\alpha^* = \alpha/k$, and then $(1-\alpha^*)^k \geq (1-\alpha)$.

For instance, if we wish to compute $k = 5$ confidence intervals, and we want them all to contain their true parameter with 95% confidence, $\alpha = 0.05$ and $\alpha^* = \alpha/k = 0.01$, then we should construct 99% confidence intervals for the individual parameters. Note that $(1-.01)^5 = .9510 > .95$.

In this section, several procedures are described for making simultaneous comparisons.

3.3.1 Tukey's Honest Significant Difference (HSD)

This method is widely used for ANOVA models. It makes use of the **Studentized Range Distribution** with critical values given in many textbooks, online, and on the course slides. In R, the **qtukey** function can be used to obtain critical values and the **ptukey** function can be used to obtain adjusted P-values.

The basis for the distribution is as follows, where $Y_1, \ldots, Y_r$ are independent and normally distributed with mean μ and variance σ^2. Further, let s^2 be an unbiased estimator of σ^2 that is independent of $Y_1, \ldots, Y_r$, with degrees of freedom ν. Note That s^2 cannot be computed from $Y_1, \ldots, Y_r$ in this situation, but it does work when we are comparing treatment/group means below. Let w be the range of $Y_1, \ldots, Y_r$ and w/s be the studentized range, indexed by the sample size, r and the degrees of freedom for s, ν. We will use $q_{1-\alpha;r,\nu}$ represent the $1-\alpha$ quantile of the Studentized Range distribution.

$$w = \max(Y_1, \ldots, Y_r) - \min(Y_1, \ldots, Y_r) \qquad \frac{w}{s} = q(r, \nu)$$

$$P\left(\frac{w}{s} = q(r, \nu) \leq q_{1-\alpha;r,\nu}\right) = 1-\alpha$$

$$\Rightarrow \quad P\left(\frac{|Y_i - Y_{i'}|}{s} \leq q_{1-\alpha;r,\nu}\right) = 1-\alpha \quad \text{for all } i, i'$$

Now, consider making all pairwise comparisons among means under the assumption of equal means ($\mu_1 = \cdots = \mu_r = \mu$) and equal sample sizes ($n_1 = \cdots = n_r = n$).

Then, we have the following results for the sampling distribution of $\overline{Y}_{i\bullet}$ and its estimated variance, which is unbiased with $n_T - r$ degrees of freedom. Furthermore the sample means $\{\overline{Y}_{i\bullet}\}$ and MSE are independent.

$$\overline{Y}_{i\bullet} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \qquad s^2\left\{\overline{Y}_{i\bullet}\right\} = \frac{MSE}{n}$$

$$\Rightarrow \quad P\left(\frac{|\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}|}{\sqrt{MSE/n}} \leq q_{1-\alpha;r,\nu}\right) = 1-\alpha \quad \text{for all } i, i'$$

Then we conclude $\mu_i \neq \mu_{i'}$ if the following criteria is met.

$$\frac{|\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}|}{\sqrt{MSE/n}} \geq q_{1-\alpha;r,\nu} \quad \Rightarrow \quad |\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}| \geq q_{1-\alpha;r,\nu}\sqrt{\frac{MSE}{n}}$$

Subsequently this method was generalized to allow for unequal sample sizes with the following decision rule. Conclude $\mu_i \neq \mu_{i'}$ if the following result holds.

$$|\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}| \geq \frac{q_{1-\alpha;r,\nu}}{\sqrt{2}}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)}$$

An equivalent way of writing this test, in terms of $D = \mu_i - \mu_{i'}$ and $\hat{D} = \overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}$ is to reject $H_0 : D = 0$ based on the following test.

$$\text{Test Statistic:} \quad q^* = \frac{\sqrt{2}\hat{D}}{s\left\{\hat{D}\right\}} \qquad \text{Rejection Region:} \quad |q^*| \geq q_{1-\alpha;r,\nu}$$

The adjusted P-value is the area above $|q^*|$ in the Studentized Range distribution.

To obtain simultaneous $(1-\alpha)100\%$ Confidence Intervals for all $r(r-1)/2$ mean differences we can simply invert the test as follows.

$$(1-\alpha)100\% \text{ for } \mu_i - \mu_j : \quad \left(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}\right) \pm \frac{1}{\sqrt{2}} q_{1-\alpha;r,n_T-r}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)}$$

Example 3.2 - Virtual Training for a Lifeboat Launching Task

First, we obtain $r = 4$, $\nu = 64 - 4 = 60$, and $q_{.95;4,60} = 3.737$ for this study. Using calculations from before, we have the following results.

$$n_1 = n_2 = n_3 = n_4 = 16 \qquad \overline{y}_{1\bullet} = 4.931 \quad \overline{y}_{2\bullet} = 7.708 \quad \overline{y}_{3\bullet} = 6.736 \quad \overline{y}_{4\bullet} = 6.875$$

$$MSE = 4.430 \qquad s\left\{\hat{D}\right\} = \sqrt{4.430\left(\frac{1}{16} + \frac{1}{16}\right)} = 0.744$$

Next we compute the Tukey HSD for all pairs of means (as this is a balanced design).

$$HSD_{ii'} = \frac{3.737}{\sqrt{2}}(0.744) = 2.642(0.744) = 1.966$$

That is, any pairs of means that differ by more than 1.966 in absolute value will be declared significantly different. We also obtain simultaneous 95% Confidence Intervals by adding and subtracting 1.966 to each difference $\hat{D}$. Here we compute the values for the comparison of treatments 1 (LEC/MAT) and 2 (MON/KEY). Then we will use the R function **TukeyHSD** after fitting the model with an **aov** object.

$$\hat{D}_{21} = \overline{Y}_{2\bullet} - \overline{Y}_{1\bullet} = 7.708 - 4.931 = 2.777$$

$$95\% \text{ Simultaneous CI for } \mu_2 - \mu_1: \quad 2.777 \pm 1.966 \quad \equiv \quad (0.811, 4.743)$$

Thus, we can conclude that the Monitor/Keyboard condition provides higher scores on average than the Lecture/Material condition. That is $\mu_2 > \mu_1$. We next obtain the adjusted P-value, making use of the **ptukey** function in R. First, though, we have to scale it back to a studentized range, q^*.

$$q^* = \frac{\sqrt{2}\hat{D}}{s\left\{\hat{D}\right\}} = \frac{\sqrt{2}(2.777)}{0.744} = 5.279 \qquad P = P\left(q_{4,60} \geq |5.279|\right) = .0023$$

Now, we use R to run all possible comparisons. Note that we have to fit an **aov** object and not a **lm** object to use the **TukeyHSD** function.

```
##   grp.trt procKnow
## 1       1   4.5614
## 2       1   6.6593
## 3       1   5.6427
## 4       1   6.4394
## 5       1   4.8635
## 6       1   0.3268
```

```
##
## Call:
## aov(formula = procKnow ~ factor(grp.trt), data = vt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0831 -1.4444  0.5774  1.5495  3.1900
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)        4.9305     0.5262   9.370 2.37e-13 ***
## factor(grp.trt)2   2.7778     0.7442   3.733 0.000423 ***
## factor(grp.trt)3   1.8056     0.7442   2.426 0.018277 *
## factor(grp.trt)4   1.9444     0.7442   2.613 0.011329 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.105 on 60 degrees of freedom
## Multiple R-squared:  0.1981, Adjusted R-squared:  0.158
## F-statistic: 4.941 on 3 and 60 DF,  p-value: 0.003931
```

```
## Analysis of Variance Table
##
## Response: procKnow
##                 Df  Sum Sq Mean Sq F value   Pr(>F)
## factor(grp.trt)  3  65.664 21.8880  4.9406 0.003931 **
## Residuals       60 265.815  4.4302
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = procKnow ~ factor(grp.trt), data = vt)
##
## $`factor(grp.trt)`
##           diff        lwr       upr     p adj
## 2-1  2.7777875  0.8113167 4.7442583 0.0023332
## 3-1  1.8055563 -0.1609146 3.7720271 0.0829543
## 4-1  1.9444438 -0.0220271 3.9109146 0.0537097
## 3-2 -0.9722313 -2.9387021 0.9942396 0.5624876
## 4-2 -0.8333438 -2.7998146 1.1331271 0.6788308
## 4-3  0.1388875 -1.8275833 2.1053583 0.9976681
```

Note that only treatments 2 and 1 are significantly different. The Confidence Interval does not contain 0 and the adjusted P-value is below 0.05. Treatments 1 and 4 are close to being significantly different, but the Confidence Interval contains 0 and the P-value is above 0.05.

∇

3.3.2 Scheffe's Method for Multiple Comparisons

This method is very conservative, but due to its procedure, it can be applied to all possible contrasts among treatment means. Using previous results on contrasts, we have the following set up for the procedure.

$$L = \sum_{i=1}^{r} c_i \mu_i \qquad \text{such that } \sum_{i=1}^{r} c_i = 0$$

$$\hat{L} = \sum_{i=1}^{r} c_i \overline{Y}_{i\bullet} \qquad s\left\{\hat{L}\right\} = \sqrt{MSE \sum_{i=1}^{r} \frac{c_i^2}{n_i}}$$

Then we construct simultaneous intervals of the following form.

$$(1-\alpha)100\% \text{ Simultaneous CI for } L: \quad \hat{L} \pm \left(\sqrt{(r-1)F_{1-\alpha;r-1,n_T-r}}\right) s\left\{\hat{L}\right\}$$

The critical F-value is the same critical value for testing $H_0 : \mu_1 = \cdots = \mu_r$ for the Analysis of Variance. Note that when this hypothesis is true, all contrasts are 0.

We can also simultaneously test whether a set of contrasts are 0.

$H_0 : L = \sum_{i=1}^{r} c_i \mu_i = 0$ vs $H_A : L = \sum_{i=1}^{r} c_i \mu_i \neq 0$

$$\text{Test Statistic:} \quad F^* = \frac{\left(\hat{L}\right)^2}{(r-1)s^2\left\{\hat{L}\right\}} \qquad \text{Rejection Region:} \quad F^* \geq F_{1-\alpha;r-1,n_T-r}$$

Example 3.3 - Virtual Training for a Lifeboat Launching Task

Consider the three pairwise orthogonal contrasts that we previously estimated.

- L_1: Lecture/Materials vs Virtual Training
- L_2: Monitor/Keyboard versus Head Mounted Displays
- L_3: HMD/Joypad versus HMD/Wearables

The critical F-value we need is $F_{.95;3,60} = 2.758$, leading to the following critical value for the Scheffe simultaneous confidence intervals.

$$\sqrt{(r-1)F_{1-\alpha;r-1,n_T-r}} = \sqrt{(4-1)2.758} = 2.876$$

Next, we give the estimated contrasts, their standard errors, and simultaneous 95% Confidence Intervals for the population based contrasts.

$$\hat{L}_1 = -6.526 \qquad s\left\{\hat{L}_1\right\} = 1.823$$

$$\hat{L}_2 = 1.805 \qquad s\left\{\hat{L}_2\right\} = 1.289$$

$$\hat{L}_3 = -0.139 \qquad s\left\{\hat{L}_3\right\} = 0.744$$

$$95\% \text{ CI for: } L_1 = 3\mu_1 - \mu_2 - \mu_3 - \mu_4 : \quad -6.526 \pm 2.876(1.823)$$

$$\equiv -6.526 \pm 5.243 \equiv (-11.769, -1.283)$$

$$95\% \text{ CI for: } L_2 = 2\mu_2 - \mu_3 - \mu_4 : \quad 1.805 \pm 2.876(1.289)$$

$$\equiv 1.805 \pm 3.707 \equiv (-1.902, 5.512)$$

$$95\% \text{ CI for: } L_3 = \mu_3 - \mu_4 : \quad -0.139 \pm 2.876(0.744)$$

$$\equiv -0.139 \pm 2.140 \equiv (-2.779, 2.001)$$

As a comparison between no adjustment for multiple Confidence Intervals and the Scheffe adjustment, consider the case of L_1. When there was no adjustment, the interval was $(10.172, -2.880)$. When the Scheffe adjustment was used, it was $(11.769, -1.283)$. Again, Scheffe's method can be used for all possible contrasts simultaneously.

∇

3.3.3 Bonferroni Method for Multiple Comparisons

The Bonferroni method is widely used in many situations where multiple tests are conducted (not just among treatment means). If we have g pre-planned comparisons or contrasts to be made, we make each comparison at the $\alpha^* = \alpha/g$ significance level. Note that a special case of this is making $g = r(r-1)/2$ comparisons among all pairs of treatment means. In that case, Tukey's HSD will provide more powerful tests than the Bonferroni method (unless r=2, and there is only one comparison).

$$L = \sum_{i=1}^{r} c_i \mu_i \qquad \text{such that } \sum_{i=1}^{r} c_i = 0$$

$$\hat{L} = \sum_{i=1}^{r} c_i \overline{Y}_{i\bullet} \qquad s\left\{\hat{L}\right\} = \sqrt{MSE \sum_{i=1}^{r} \frac{c_i^2}{n_i}}$$

Then we construct simultaneous confidence intervals of the following form.

$$(1-\alpha)100\% \text{ Simultaneous CI for } L : \quad \hat{L} \pm t_{1-\alpha/(2g);n_T-r}\, s\left\{\hat{L}\right\}$$

Tables that give critical t-values for the Bonferroni method are available in textbooks, online, and on the course slides.

Simultaneous tests for g pre-planned contrasts can be conducted to test whether the contrast is equal to zero.

$$H_0 : L = \sum_{i=1}^{r} c_i \mu_i = 0 \qquad H_A : L = \sum_{i=1}^{r} c_i \mu_i \neq 0$$

$$\text{Test Statistic:} \quad t^* = \frac{\hat{L}}{s\left\{\hat{L}\right\}} \qquad \text{Rejection Region:} \quad |t^*| \geq t_{1-\alpha/(2g), n_T - r}$$

$$\text{Adjusted P-value:} \quad P = \min\left(1, g \times 2P\left(t_{n_T - r} \geq |t^*|\right)\right)$$

Example 3.4 - Virtual Training for a Lifeboat Launching Task

Consider the $g = 3$ pairwise orthogonal contrasts that we previously estimated.

- L_1 Lecture/Materials vs Virtual Training
- L_2 Monitor/Keyboard versus Head Mounted Displays
- L_3 HMD/Joypad versus HMD/Wearables

The critical t-value we need is $t_{1-.05/(2(3));60} = 2.463$. We will conduct the three simultaneous t-tests and obtain their adjusted P-values.

Next, we give the estimated contrasts, their standard errors, and simultaneous t-tests with an overall experimentwise error rate of $\alpha = 0.05$ for the population based contrasts. Note that the individual tests are being conducted at $\alpha^* = .05/3 = .0167$ error rate.

$$\hat{L}_1 = -6.526 \qquad s\left\{\hat{L}_1\right\} = 1.823 \qquad t_1^* = \frac{-6.526}{1.823} = -3.580$$

$$\hat{L}_2 = 1.805 \qquad s\left\{\hat{L}_2\right\} = 1.289 \qquad t_2^* = \frac{1.805}{1.289} = 1.400$$

$$\hat{L}_3 = -0.139 \qquad s\left\{\hat{L}_3\right\} = 0.744 \qquad t_3^* = \frac{-0.139}{0.744} = -0.187$$

Thus, again, only the first contrast is significant. We compute the 3 adjusted P-values by first obtaining the 2-sided tail areas for the $g = 3$ t-statistics, then applying the Bonferroni adjustments.

$$2P\left(t_{60} \geq |3.580|\right) = 2(.00034) = .0007 \qquad P_1 = \min\left(1, 3(.0007)\right) = .0021$$

$$2P\left(t_{60} \geq |1.400|\right) = 2(.0833) = .1666 \qquad P_2 = \min\left(1, 3(.1666)\right) = .4998$$

$$2P\left(t_{60} \geq |-0.187|\right) = 2(.4261) = .8522 \qquad P_3 = \min\left(1, 3(.8522)\right) = 1$$

Note that if we used the Bonferroni method for comparing all $g = 4(4-1)/2 = 6$ pairs of treatment means, the critical t-value would be $t_{1-.05/(2(6));60} = 2.729$. Since the standard error for the difference in two mean for this example is $s\left\{\hat{D}\right\} = 0.744$, the simultaneous confidence intervals would be of the following form.

$$\hat{D}_{ii'} \pm 2.729(0.744) \quad \equiv \quad \hat{D}_{ii'} \pm 2.030$$

These intervals are slightly wider than the Tukey based intervals $\left(\hat{D}_{ii'} \pm 1.966\right)$. The same conclusions among treatment means are made.

∇

3.3.4 Multiple Range Methods

In this section, we briefly describe two methods that use multiple ranges for comparisons: **Student-Newman-Keuls'** (often shortened to SNK) and **Duncan's** methods. They both are based on ordering the observed means. Neither can be used to construct simultaneous confidence intervals.

$$\overline{Y}_{(1)\bullet} \leq \cdots \leq \overline{Y}_{(r)\bullet} \qquad n_{(i)} \text{ is the sample size for the ordered groups}$$

Each method uses different critical values for testing pairs of treatment means, depending on how many means are between them. They both begin by comparing the extreme means, which encompass r means.

For the SNK method, we reject $H_0 : \mu_{(i)} - \mu_{(i')} = 0$ where $i > i'$ if the following result holds, where $q_{1-\alpha;k,n_T-r}$ is obtained from the studentized range distribution.

$$\overline{Y}_{(i)\bullet} - \overline{Y}_{(i')\bullet} \geq \frac{q_{1-\alpha;k,n_T-r}}{\sqrt{2}} \sqrt{MSE\left(\frac{1}{n_{(i)}} + \frac{1}{n_{(i')}}\right)} \qquad k = i - i' + 1$$

If we fail to reject this hypothesis, no means between these are compared.

For Duncan's method, the error rates are increased as the means are farther apart in terms of the number of means between them. This makes it "easier" to detect differences at the cost of increasing overall error rates. In this case we reject $H_0 : \mu_{(i)} = \mu_{(i')} = 0$ if the following result holds. As with the SNK method, if two means are not significantly different, no means between them are compared.

$$\overline{Y}_{(i)\bullet} - \overline{Y}_{(i')\bullet} \geq \frac{q_{(1-\alpha)^{k-1};k,n_T-r}}{\sqrt{2}} \sqrt{MSE\left(\frac{1}{n_{(i)}} + \frac{1}{n_{(i')}}\right)} \qquad k = i - i' + 1$$

Example 3.5 - Virtual Training for a Lifeboat Launching Task

For this example, we have the following values.

$$r = 4 \qquad n_1 = \cdots = n_r = 16 \qquad MSE = 4.430 \qquad s\left\{\hat{D}\right\} = \sqrt{4.430\left(\frac{1}{16} + \frac{1}{16}\right)} = 0.744$$

$$\overline{Y}_{(1)\bullet} = 4.931 \quad \overline{Y}_{(2)\bullet} = 6.736 \quad \overline{Y}_{(3)\bullet} = 6.875 \quad \overline{Y}_{(4)\bullet} = 7.708$$

We first test $\mu_{(4)} - \mu_{(1)} =$ with range $k = 4 - 1 + 1 = 4$. Using the **qtukey** function in R, we obtain the following critical values and rejection regions based on the studentized range distribution with overall experimentwise error rate $\alpha = .05$.

$$\text{SNK:} \quad q(.95; 4, 60) = 3.737 \qquad \text{Rejection Region:} \quad \overline{Y}_{(4)\bullet} - \overline{Y}_{(1)\bullet} \geq \frac{3.737}{\sqrt{2}}(0.744) = 1.966$$

$$\text{Duncan:} \quad q((.95)^{4-1}; 4, 60) = 3.073 \qquad \text{Rejection Region:} \quad \overline{Y}_{(4)\bullet} - \overline{Y}_{(1)\bullet} \geq \frac{3.073}{\sqrt{2}}(0.744) = 1.617$$

The observed difference is $\overline{y}_{(4)\bullet} - \overline{y}_{(1)\bullet} = 7.708 - 4.931 = 2.777$. Both methods conclude these means are significantly different.

Next, we obtain the critical values and rejection regions for testing $H_0 : \mu_{(4)} - \mu_{(2)} = 0$ and $H_0 : \mu_{(3)} - \mu_{(1)} = 0$, each with a range of $k = 3$.

$$\text{SNK:} \quad q(.95; 3, 60) = 3.399 \qquad \text{Rejection Region:} \quad \overline{Y}_{(4)\bullet} - \overline{Y}_{(1)\bullet} \geq \frac{3.399}{\sqrt{2}}(0.744) = 1.788$$

$$\text{Duncan:} \quad q((.95)^{3-1}; 3, 60) = 2.976 \qquad \text{Rejection Region:} \quad \overline{Y}_{(i)\bullet} - \overline{Y}_{(j)\bullet} \geq \frac{2.976}{\sqrt{2}}(0.744) = 1.566$$

The observed differences are as follow.

$$\overline{y}_{(4)\bullet} - \overline{y}_{(2)\bullet} = 7.708 - 6.736 = 0.972 \qquad \overline{y}_{(3)\bullet} - \overline{y}_{(1)\bullet} = 6.875 - 4.931 = 1.944$$

Neither test rejects $H_0 : \mu_{(4)} - \mu_{(2)} = 0$ and both reject $H_0 : \mu_{(3)} - \mu_{(1)} = 0$. Thus, we do not compare $\mu_{(4)}$ and $\mu_{(3)}$. We will conduct the following two tests though, each with $k = 2$.

$$H_0 : \mu_{(3)} - \mu_{(2)} = 0 \qquad H_0 : \mu_{(2)} - \mu_{(1)} = 0$$

$$\text{SNK:} \quad q(.95, 2, 60) = 2.829 \qquad \text{Rejection Region:} \quad \overline{Y}_{(i)\bullet} - \overline{Y}_{(i')\bullet} \geq \frac{2.829}{\sqrt{2}}(0.744) = 1.488$$

$$\text{Duncan:} \quad q((.95)^{2-1}, 2, 60) = 2.829 \qquad \text{Rejection Region:} \quad \overline{Y}_{(i)\bullet} - \overline{Y}_{(i')\bullet} \geq \frac{2.829}{\sqrt{2}}(0.744) = 1.488$$

The observed differences are as follow.

$$\overline{y}_{(3)\bullet} - \overline{y}_{(2)\bullet} = 6.875 - 6.736 = 0.139 \qquad \overline{y}_{(2)\bullet} - \overline{y}_{(1)\bullet} = 6.736 - 4.931 = 1.805$$

Thus, we fail to reject $H_0 : \mu_{(3)} - \mu_{(2)} = 0$, but we do reject $H_0 : \mu_{(2)} - \mu_{(1)} = 0$.

Based on both the SNK and Duncan methods, we conclude that all three virtual reality methods perform better than the control (Lecture/Material) method. These methods are available in the **agricolae** package in R. Note that the treatment factor must be defined as a factor outside of **aov** or **lm** object to be used in the **SNK.test** and **duncan.test** functions.

```
##   grp.trt procKnow
## 1       1   4.5614
## 2       1   6.6593
## 3       1   5.6427
## 4       1   6.4394
## 5       1   4.8635
## 6       1   0.3268
```

```
##
## Call:
## lm(formula = procKnow ~ grp.trt.f, data = vt)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0831 -1.4444  0.5774  1.5495  3.1900
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.9305     0.5262   9.370 2.37e-13 ***
## grp.trt.f2    2.7778     0.7442   3.733 0.000423 ***
## grp.trt.f3    1.8056     0.7442   2.426 0.018277 *
## grp.trt.f4    1.9444     0.7442   2.613 0.011329 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.105 on 60 degrees of freedom
## Multiple R-squared:  0.1981, Adjusted R-squared:  0.158
## F-statistic: 4.941 on 3 and 60 DF,  p-value: 0.003931
```

```
## Analysis of Variance Table
##
## Response: procKnow
##           Df  Sum Sq Mean Sq F value   Pr(>F)
## grp.trt.f  3  65.664 21.8880  4.9406 0.003931 **
## Residuals 60 265.815  4.4302
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Study: vt.mod1 ~ "grp.trt.f"
##
## Student Newman Keuls Test
## for procKnow
##
## Mean Square Error:  4.430248
##
## grp.trt.f,  means
##
##   procKnow      std  r        se    Min    Max      Q25     Q50      Q75
## 1 4.930550 1.940008 16 0.5262039 0.3268 6.7509 4.230175 5.67085 6.375725
## 2 7.708337 1.429991 16 0.5262039 5.0733 9.7848 6.682275 8.33320 8.707800
## 3 6.736106 2.820002 16 0.5262039 1.6530 9.9261 5.298875 7.04235 9.101975
## 4 6.874994 1.989995 16 0.5262039 3.2286 9.8754 5.344875 6.84850 8.346350
##
## Alpha: 0.05 ; DF Error: 60
##
## Critical Range
##        2        3        4
## 1.488551 1.788389 1.966471
##
## Means with the same letter are not significantly different.
##
##   procKnow groups
## 2 7.708337      a
## 4 6.874994      a
## 3 6.736106      a
## 1 4.930550      b

##
## Study: vt.mod1 ~ "grp.trt.f"
##
## Duncan's new multiple range test
## for procKnow
##
## Mean Square Error:  4.430248
##
## grp.trt.f,  means
##
##   procKnow      std  r        se    Min    Max      Q25     Q50      Q75
## 1 4.930550 1.940008 16 0.5262039 0.3268 6.7509 4.230175 5.67085 6.375725
## 2 7.708337 1.429991 16 0.5262039 5.0733 9.7848 6.682275 8.33320 8.707800
## 3 6.736106 2.820002 16 0.5262039 1.6530 9.9261 5.298875 7.04235 9.101975
## 4 6.874994 1.989995 16 0.5262039 3.2286 9.8754 5.344875 6.84850 8.346350
##
## Alpha: 0.05 ; DF Error: 60
##
```

```
## Critical Range
##        2        3        4
## 1.488551 1.565916 1.616955
##
## Means with the same letter are not significantly different.
##
##   procKnow groups
## 2 7.708337      a
## 4 6.874994      a
## 3 6.736106      a
## 1 4.930550      b
```

∇

```
library(tidyverse)
library(kableExtra)
library(effectsize)
library(agricolae)
library(car)
```

```
## Warning: package 'car' was built under R version 4.1.3
## Loading required package: carData
## Warning: package 'carData' was built under R version 4.1.3
##
## Attaching package: 'car'
## The following object is masked from 'package:dplyr':
##
##     recode
## The following object is masked from 'package:purrr':
##
##     some
```

```
library(PMCMRplus)
```

```
## Warning: package 'PMCMRplus' was built under R version 4.1.3
```

Chapter 4

Alternative Tests for Treatment Effects

In this chapter, we consider tests for the equality of variances among treatments/groups, as well as alternatives to the F-test that are used in practice.

The tests for equal variances are **Levene's Test**, **Hartley's Test**, and **Bartlett's Test**.

Alternatives to the F-test are **Randomization/Permutation Tests**, the **Kruskal-Wallis Test**, and **Welch's Test**. Also, **Weighted Least Squares** can be implemented with unequal variances.

Three departures of the model assumptions are briefly described below.

- **Non-normal Errors** - This is generally not too problematic for the F-test as long as data are not too far from normal with reasonable sample sizes
- **Unequal Error Variances** - As long as the sample sizes are approximately equal this is not a large problem for the F-test, but can be an issue with very unbalanced designs.
- **Non-independence of Error Terms** This can cause issues for the F-test. If each unit receives each treatment, can get a stronger test based on a Repeated Measures Design.

4.1 Randomization/Permutation Tests

These tests are based on resampling, and make no assumptions on the distribution of the error terms. The test treats the units observed in the study as a finite population of units with fixed error terms ϵ_{ij}. The model and hypothesis are of the following forms.

$$Y_{ij} = \mu_{\bullet} + \tau_i + \epsilon_{ij} \qquad\qquad H_0 : \tau_1 = \cdots = \tau_r = 0$$

The idea is that if this null hypothesis true, the observed values are due only to the units themselves and not to treatment/group differences. A statistic, such as F^* or $SSTR$ is computed and saved. Then, many permutations of the n_T responses are randomly assigned to the treatment labels of sizes $n_1, \dots, n_r$. The statistic is computed for each permutation and saved. The P-value is the number of more extreme values than the observed statistic plus 1, divided by the number of permutations plus 1. That is, the observed value counts in both the numerator and denominator. Often the number of permutations is 9999 in practice.

This test is applied to the lifeboat training data here.

Example 4.1 - Virtual Training for a Lifeboat Launching Task

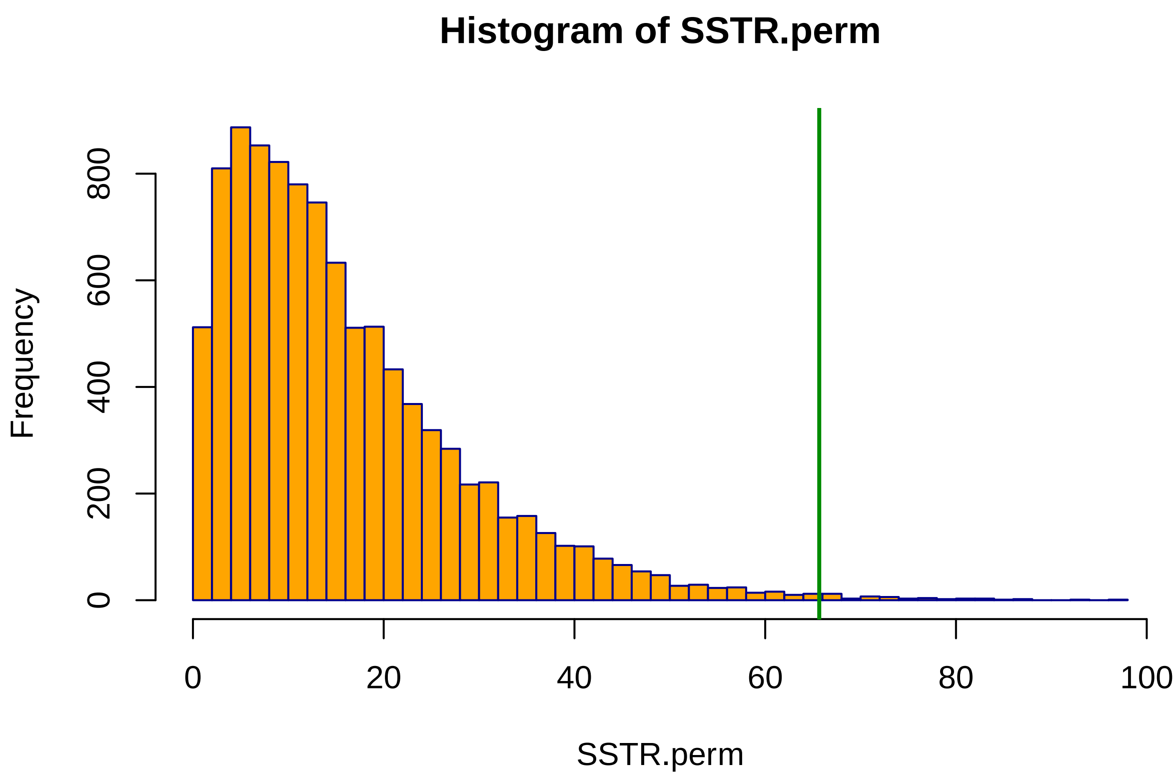
Here we will apply the Randomization test to the Lifeboat training data. For the original sample, $SSTR = 65.664$. Note that for each permutation, the total sum of squares will be the same, $SSTO = 331.479$, so we can simply save $SSTR$ from each permuation, as opposed to computing F^*. The R code and output are given below.

```
##   grp.trt procKnow
## 1       1   4.5614
## 2       1   6.6593
## 3       1   5.6427
```

```
## 4       1   6.4394
## 5       1   4.8635
## 6       1   0.3268

## [1] 65.66393

## [1] 0.005
```



The P-value is 0.005, which is well below 0.05. Only 49 permutations out of 9999 had larger values of $SSTR$ than the observed data. There is strong evidence of treatment effects.

∇

4.2 Tests for Constant Error Variances

In this section, we consider three widely used tests of equal variances among the treatments/groups. They are based on various assumptions and do not always provide the same conclusion. The three tests covered here are **Hartley's Test**, **Levene's Test**, and **Bartlett's Test**. All tests are testing the following hypotheses.

$$H_0 : \sigma_1^2 = \cdots = \sigma_r^2 \qquad H_A : \text{ Not all } \sigma_i^2 \text{ are equal}$$

4.2.1 Hartley's Test

This test is dependent on the errors being normally distributed. Further, its derivation is based on equal sample sizes. It uses a special table of critical values available in many textbooks and online. There is a R package that purports to generate the critical values, but I have had issues with certain sample sizes/numbers of treatments in obtaining results.

The test statistic (H) is simply the ratio of the largest to smallest variance (not standard deviation), and the critical values are indexed by the number of treatments/groups (r) and the number of degrees of freedom for each sample variance ($n - 1$), where n is the number of units per treatment.

Example 4.2 - Virtual Training for a Lifeboat Launching Task

For the lifeboat training experiment, there were $r = 4$ treatments and each sample size was $n = 16$. The sample standard deviations are given below. The critical value for this test with $\alpha = 0.05$, $r = 4$, and $n - 1 = 15$ is $H_{.95;4,15} = 4.01$.

$$s_1 = 1.94 \quad s_2 = 1.43 \quad s_3 = 2.82 \quad s_4 = 1.99 \qquad H^* = \frac{\max\left(s_i^2\right)}{\min\left(s_i^2\right)} = \frac{(2.82)^2}{(1.43)^2} = 3.89$$

The test statistic is just below the critical value, so we fail to reject the hypothesis of equal variances.

∇

4.2.2 Levene's Test

This test, also known as the Brown-Forsythe test is robust to outliers and does not require equal sample sizes. There are various versions, based on whether deviations from group means or medians are used. For each observation, we compute d_{ij} as described here.

$$d_{ij} = \left|Y_{ij} - \tilde{Y}_i\right| \qquad \tilde{Y}_i = \text{median}\left(Y_{i1}, \dots, Y_{in_i}\right)$$

Then an Analysis of Variance is applied to the d_{ij}, with $\overline{d}_{i\bullet}$ and $\overline{d}_{\bullet\bullet}$ being the treatment and overall means.

$$SSTR_L = \sum_{i=1}^{r} n_i \left(\overline{d}_{i\bullet} - \overline{d}_{\bullet\bullet}\right)^2 \qquad df_{TR} = r - 1$$

$$SSE_L = \sum_{i=1}^{r} \sum_{j=1}^{n_i} \left(d_{ij} - \overline{d}_{i\bullet}\right)^2 \qquad df_E = n_T - r$$

Then, just as when we tested for equality of means, we compute the F-statistic and compare it with $F_{1-\alpha;r-1,n_T-r}$.

$$\text{Test Statistic:} \quad F_L^* = \frac{\left[\frac{SSTR_L}{r-1}\right]}{\left[\frac{SSE_L}{n_T-r}\right]} \qquad \text{Rejection Region:} \quad F_L^* \geq F_{1-\alpha;r-1,n_T-r}$$

Example 4.3 - Virtual Training for a Lifeboat Launching Task

We will first create the d_{ij} "brute-force", then use the **leveneTest** function in the **car** package.

```
##   grp.trt procKnow med.trt
## 1       1   4.5614 5.67085
## 2       1   6.6593 5.67085
## 3       1   5.6427 5.67085
## 4       1   6.4394 5.67085
## 5       1   4.8635 5.67085
## 6       1   0.3268 5.67085
```

```
##    grp.trt procKnow med.trt
## 59       4   7.5014  6.8485
## 60       4   8.2456  6.8485
## 61       4   4.2465  6.8485
## 62       4   3.2286  6.8485
## 63       4   6.5520  6.8485
## 64       4   9.8754  6.8485
```

```
## Analysis of Variance Table
##
## Response: d
##                 Df  Sum Sq Mean Sq F value Pr(>F)
## factor(grp.trt)  3  11.108  3.7026  2.1463 0.1038
## Residuals       60 103.506  1.7251
```

```
## Levene's Test for Homogeneity of Variance (center = "median")
##       Df F value Pr(>F)
## group  3  2.1463 0.1038
##       60
```

We fail to reject the null hypothesis of equal variances. Note that we ran the **leveneTest** function on the Y_{ij} values, not on the with the d_{ij} values (it creates them internally).

∇

4.2.3 Bartlett's Test

Bartlett's test is based on the residuals being normally distributed and can be used in applications beyond ANOVA models, including regression models with categorical predictors. It makes use of the individual variances $\{s_i^2\}$ and the pooled MSE to obtain a chi-square statistic that is defined here.

$$\text{Test Statistic:} \quad X_B^2 = \frac{1}{C}\left[(n_T - r)\ln(MSE) - \sum_{i=1}^{r}(n_i - 1)\ln\left(s_i^2\right)\right]$$

$$\text{where:} \quad C = 1 + \frac{1}{3(r-1)}\left[\left(\sum_{i=1}^{r}\frac{1}{n_i - 1}\right) - \left(\frac{1}{n_T - r}\right)\right]$$

$$\text{Rejection Region:} \quad X_B^2 \geq \chi^2_{1-\alpha;r-1} \qquad P = P\left(\chi^2_{r-1} \geq X_B^2\right)$$

Example 4.4 - Virtual Training for a Lifeboat Launching Task

For the lifeboat training experiment, there were $r = 4$ treatments and each sample size was $n = 16$. The sample standard deviations are given below. The critical value for this test with $\alpha = 0.05$ and $r = 4$ is $\chi_{.95;4-1} = 7.815$.

$$s_1^2 = 3.764 \quad s_2^2 = 2.045 \quad s_3^2 = 7.952 \quad s_4^2 = 3.960 \qquad MSE = 4.430$$

$$(n_T - r)\ln(MSE) = 89.304 \qquad \sum_{i=1}^{r}(n_i - 1)\ln\left(s_i^2\right) = 82.358$$

$$C = 1 + \frac{1}{3(r-1)}\left[\left(\sum_{i=1}^{r}\frac{1}{n_i - 1}\right) - \left(\frac{1}{n_T - r}\right)\right] = 1 + \frac{1}{9}\left[\frac{4}{16-1} - \frac{1}{64-4}\right] = 1.028$$

$$\text{Test Statistic:} \quad X_B^2 = \frac{1}{1.028}(89.304 - 82.358) = 6.757 \qquad P = P\left(\chi_3^2 \geq 6.757\right) = .0801$$

Finally, we use the **bartlett.test** function in R to conduct the test.

```
##
##  Bartlett test of homogeneity of variances
##
## data:  procKnow by factor(grp.trt)
## Bartlett's K-squared = 6.7626, df = 3, p-value = 0.07986
```

∇

4.3 Remedial Measures

In this section, several procedures are considered for issues of normality and/or unequal variances among the error terms. Methods will include **Weighted Least Squares** and **Welch's Test** when the errors are normal, but the error variances are unequal (heteroscedastic).

When variances are non-normal and heteroscedastic, transformations can be applied to Y that may solve both problems. These include the **Box-Cox Transformation**, and were covered in the Linear Regression course.

When the distributions are highly skewed, non-parametric tests can be conducted based on ranks. We will describe the **Kruskal-Wallis Test**.

4.3.1 Weighted Least Squares and Welch's Test

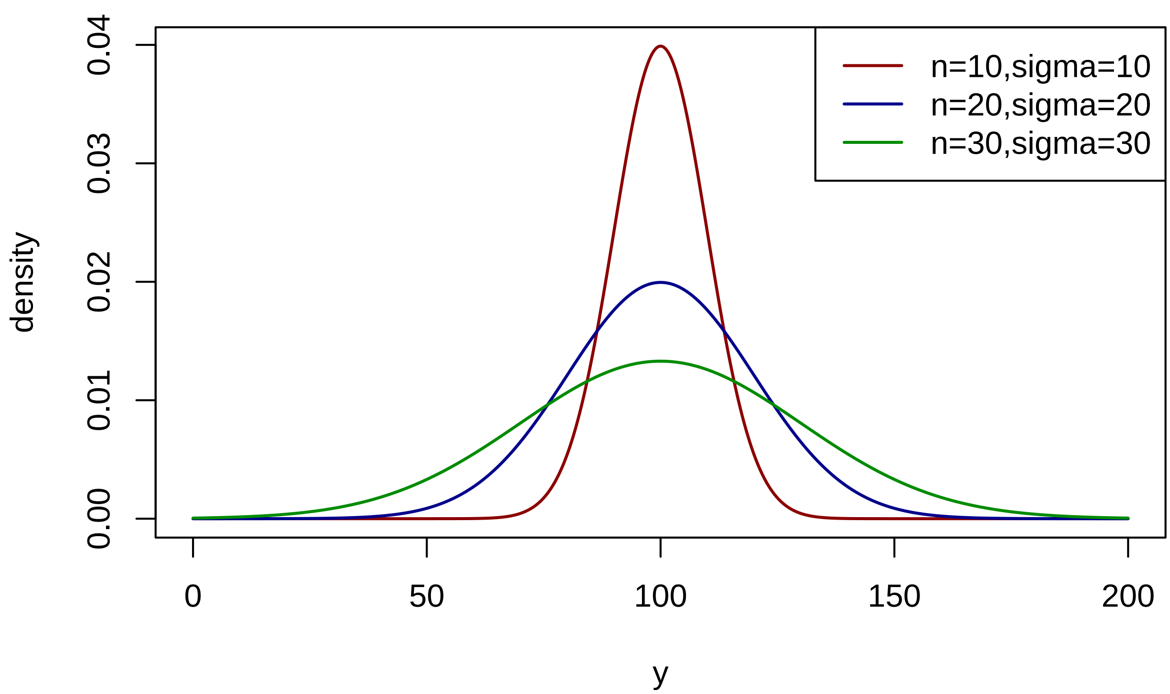
Consider the following scenario and simulation.

- There are $r = 3$ Treatments with equal means ($\mu_1 = \mu_2 = \mu_3$)
- The Treatments have unequal standard deviations (and thus variances)
- The Treatments have equal or unequal sample sizes

Note that the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$ holds, so optimally, the rejection rate based on the F-test with significance level α should be approximately α. We will simulate datasets for the following scenarios, all with $\mu = 100$ for all treatments, and normal error terms.

- **Scenario 1** - $n_1 = n_2 = n_3 = 20 \quad \sigma_1 = \sigma_2 = \sigma_3 = 20$ (standard assumptions)
- **Scenario 2** - $n_1 = n_2 = n_3 = 20 \quad \sigma_1 = 10, \sigma_2 = 20, \sigma_3 = 30$
- **Scenario 3** - $n_1 = 10, n_2 = 20, n_3 = 30 \quad \sigma_1 = 10, \sigma_2 = 20, \sigma_3 = 30$
- **Scenario 4** - $n_1 = 30, n_2 = 20, n_3 = 10 \quad \sigma_1 = 10, \sigma_2 = 20, \sigma_3 = 30$

Each scenario will be run, and the rejection rate will be reported in each case. The program below is the same for each scenario with only the inputs for **n** and **sigma** changing.



```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##  0.000047  0.160073  0.414482  0.691802  0.904428 13.217528
```

```
## [1] 0.022
```

Each scenario was run on 10000 simulated data sets with $\alpha = 0.05$. The following results were obtained based on the seed 32611.

- **Scenario 1** - Equal variances, Equal sample sizes. Rejection Rate = .0513
- **Scenario 2** - Unequal variances, Equal sample sizes. Rejection Rate = .0573
- **Scenario 3** - Increasing variances, Increasing sample sizes. Rejection Rate = .0220.
- **Scenario 4** - Increasing variances, Decreasing sample sizes. Rejection Rate = .1637

The primary result is that as long as sample sizes are equal (or very similar), the F-test performs well whether the variances are equal or not (both Scenarios 1 and 2 have rejection rates close to α).

When the sample sizes are positively correlated with the variance (Scenario 3), the rejection rate was well below α ($.0220 < .05$). Thus, when the larger (smaller) samples correspond to the larger (smaller) variances the F-test rejects less often than it should.

When the sample sizes are negatively correlated with the variance (Scenario 4), the rejection rate was well above α ($.1637 > .05$). Thus, when the smaller (larger) samples correspond to the larger (smaller) variances the F-test rejects more often than it should.

Estimated Weighted Least Squares can be used when the error variances are unequal. The estimated weights for the individual measurements is the reciprocal of the variance of the measurements within that treatment/group, $w_{ij} = 1/s_i^2$. It is much easier doing direct computations in matrix form than in scalar form. A sketch of the method is given below for the cell means model with balanced data: $n_1 = \cdots = n_r = n$. Let I_n be the $n \times n$ identity matrix and 0_n be the $n \times n$ matrix of 0^s. The $n_T \times r$ matrix X and the $n_T \times 1$ vector Y are described in Chapter 2.

$$\hat{V} = \begin{bmatrix} s_1^2 I_n & 0_n & \cdots & 0_n \\ 0_n & s_2^2 I_n & \cdots & 0_n \\ \vdots & \vdots & \ddots & \vdots \\ 0_n & 0_n & \cdots & s_r^2 I_n \end{bmatrix} \qquad X^* = \hat{V}^{-1/2} X \qquad Y^* = \hat{V}^{-1/2} Y$$

$$H^* = X^* \left(X^{*'} X^*\right)^{-1} X^{*'} \qquad SSE_W(F) = Y^{*'} \left(I_{n_T} - H^*\right) Y^* \quad df_E(F) = n_T - r$$

Under the null hypothesis $H_0 : \mu_1 = \cdots \mu_r = \mu$, the X_0 matrix is the $n_T \times 1$ vector of 1^s. In this (null) case we obtain the reduced SSE below.

$$X_0^* = \hat{V}^{-1/2} X_0 \quad H_0^* = X_0^* \left(X_0^{*'} X_0^*\right)^{-1} X_0^{*'} \qquad SSE_W(R) = Y^{*'} \left(I_{n_T} - H_0^*\right) Y^* \quad df_E(R) = n_T - 1$$

The test statistic and rejection region for the general linear test are given below.

$$\text{Test Statistic:} \quad F_{EWLS}^* = \frac{\left[\frac{(SSE_W(R) - SSE_W(F))}{(n_T - 1) - (n_T - r)}\right]}{\left[\frac{SSE_W(F)}{n_T - r}\right]} \qquad \text{Rejection Region:} F_{EWLS}^* \geq F_{1-\alpha, r-1, n_T - r}$$

Example 4.5 - Virtual Training for a Lifeboat Launching Task

For the lifeboat training experiment, there were $r = 4$ treatments and each sample size was $n = 16$. The following R code runs the matrix form of the general linear test and obtains the sample variance treatment and creates the weights for the individual subjects in the experiment and uses the **weight** command used in the **aov** function.

```
##      df(F) df(R) SSE(F)  SSE(R)     F* F(.95) P(>F*)
## EWLS    60    63     60 81.3196 7.1065 2.7581  4e-04
```

```
##   grp.trt procKnow
## 1       1   4.5614
## 2       1   6.6593
## 3       1   5.6427
```

```
## 4       1   6.4394
## 5       1   4.8635
## 6       1   0.3268
```

```
## Analysis of Variance Table
##
## Response: procKnow
##                 Df Sum Sq Mean Sq F value    Pr(>F)
## factor(grp.trt)  3  21.32  7.1065  7.1065 0.0003655 ***
## Residuals       60  60.00  1.0000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Analysis of Variance Table
##
## Response: procKnow
##                 Df  Sum Sq Mean Sq F value   Pr(>F)
## factor(grp.trt)  3  65.664 21.8880  4.9406 0.003931 **
## Residuals       60 265.815  4.4302
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

∇

Welch's Test was developed to handle the issues that arise in Scenarios 3 and 4 above. The method involves computing a weighted F-statistic for testing for treatment effects. The weights for the treatment means are the reciprocals of their variances, $w_i = n_i/s_i^2$. The procedure (which is easy to run in a spreadsheet or statistical package) computes the weighted F-statistic and generates multiples of that statistic and the error degrees of freedom to achieve an approximate F-distribution for the test statistic.

A sketch of the calculations is given below. Note that all calculations are based on the treatment/group sample sizes, means, and variances.

$$w_i = \frac{n_i}{s_i^2} \quad w_\bullet = \sum_{i=1}^{r} w_i \quad C_W = \sum_{i=1}^{r}\left[\frac{1}{n_i-1}\left(1-\frac{w_i}{w_\bullet}\right)\right] \quad m_W = \left[1+\frac{2(r-2)}{r^2-1}C_W\right]^{-1}$$

$$F^* = \frac{1}{r-1}\left[\sum_{i=1}^{r} w_i\left(\overline{y}_{i\bullet}\right)^2 - \frac{\left(w_i\overline{y}_{i\bullet}\right)^2}{w_\bullet}\right] \qquad \nu_W = \frac{r^2-1}{3C_W}$$

$$F_W = m_W F^* \dot{\sim} F_{r-1,\nu_W}$$

Example 4.6 - Virtual Training for a Lifeboat Launching Task

For the lifeboat training experiment, there were $r = 4$ treatments and each sample size was $n = 16$. The following R code produces Welch's F_W-test directly from the treatment means, variances, and sample sizes as well as using the **oneway.test** function in R.

```
##                df1     df2  F_W* F(.95) P(>F_W*)
## Welch's F-test   3 32.5622 6.827 2.8957   0.0011
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  procKnow and factor(grp.trt)
## F = 6.827, num df = 3.000, denom df = 32.562, p-value = 0.001072
```

4.3.2 Post-hoc Comparisons with Unequal Variances

The **Games-Howell** method can be used to make pairwise comparisons among all pairs of treatments when variances are unequal. Like Tukey's method, it makes use of the studentized range distribution. It makes use of **Satterthwaite's approximation** for the degrees of freedom for each pair of means.

$$\nu_{ii'} = \frac{\left(\frac{s_i^2}{n_i} + \frac{s_{i'}^2}{n_{i'}}\right)^2}{\frac{\left(\frac{s_i^2}{n_i}\right)^2}{n_i - 1} + \frac{\left(\frac{s_{i'}^2}{n_{i'}}\right)^2}{n_{i'} - 1}} \qquad GH_{ii'} = \frac{q_{1-\alpha;r,\nu_{ii'}}}{\sqrt{2}} \sqrt{\frac{s_i^2}{n_i} + \frac{s_{i'}^2}{n_{i'}}}$$

Conclude $\mu_i \neq \mu_{i'}$ if $\left|\overline{y}_{i\bullet} - \overline{y}_{i'\bullet}\right| \geq GH_{ii'}$. Simultaneous $(1-\alpha)100\%$ Confidence Intervals for $\mu_i - \mu_{i'}$ are of the form $(\overline{y}_{i\bullet} - \overline{y}_{i'\bullet}) \pm GH_{ii'}$.

Example 4.7 - Virtual Training for a Lifeboat Launching Task

For the lifeboat training experiment, there were $r = 4$ treatments and each sample size was $n = 16$. The following R code computes Simultaneous 95% Confidence Intervals and adjusted P-values based on the Games-Howell method.

```
##   grp.trt procKnow
## 1       1   4.5614
## 2       1   6.6593
## 3       1   5.6427
## 4       1   6.4394
## 5       1   4.8635
## 6       1   0.3268
```

```
##      Trt i Trt i' Ybar diff      df    MSD      LL     UL  Adj P
## [1,]     2      1    2.7778 27.5847 1.6466  1.1312 4.4244 0.0005
## [2,]     3      1    1.8056 26.5999 2.3440 -0.5384 4.1495 0.1759
## [3,]     3      2   -0.9722 22.2357 2.1931 -3.1654 1.2209 0.6150
## [4,]     4      1    1.9444 29.9806 1.8893  0.0552 3.8337 0.0418
## [5,]     4      2   -0.8333 27.2301 1.6756 -2.5089 0.8422 0.5340
## [6,]     4      3    0.1389 26.9707 2.3614 -2.2226 2.5003 0.9985
```

```
##
##  Pairwise comparisons using Games-Howell test
```

```
## data: procKnow by grp.trt
```

```
##   1       2       3
## 2 0.00046 -       -
## 3 0.17589 0.61503 -
## 4 0.04183 0.53405 0.99848
```

```
##
## P value adjustment method: none
```

```
## alternative hypothesis: two.sided
```

Treatment 2 (Monitor/Keyboard) is significantly higher than Treatment 1 (Lecture/Materials) and Treatment 4 (Head Mounted Display/Wearable sensors) is significantly higher than Treatment 1.

∇

Finally, we simulate Welch's Test for Scenario 4 described previously where $\mu_1 = \mu_2 = \mu_3 = 100$, $\sigma_1 = 10, \sigma_2 = 20, \sigma_3 = 30$, and $n_1 = 30, n_2 = 20, n_3 = 10$.

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##  0.00079  0.41930  1.05392  1.75428  2.30132 37.07031
```

```
## [1] 0.1637
```

```
## [1] 0.0562
```

The ANOVA F-test rejected the null hypothesis in 16.37% of the samples, while Welch's F_W-test rejected it in 5.62% of the samples, much closer to the target of 5%.

4.4 Nonparametric Test for Non-normal Data

The **Kruskal-Wallis Test** is a rank based test used in cases where the errors are not normally distributed. In particular, outlying observations, which have large impacts on the means and standard deviations have smaller impact when measurements are replaced by ranks.

The n_T observations are ranked from 1 (smallest) to n_T (largest), with ties being given the average of the ranks they would have received if not exactly equal. For instance, if two observations tie for the smallest value, they would each receive the rank of 1.5, and the next smallest observation would get the rank of 3. The sum of the ranks is $1 + \cdots + n_T = n_T\left(n_T + 1\right)/2$.

Once the individual measurements are ranked, rank sums are obtained for each treatment/group, as well as the Kruskal-Wallis H-statistic, which under the null hypothesis of equal medians $H_0 : M_1 = \cdots = M_r$ follows a chi-square distribution with $r - 1$ degrees of freedom approximately.

$$R_{ij} = \text{rank}\left(Y_{ij}\right) \text{ among } Y_{11}, \ldots Y_{rn_r} \qquad R_{i\bullet} = \sum_{j=1}^{n_i} R_{ij}$$

$$\text{Test Statistic:} \quad H^* = \left[\frac{12}{n_T\left(n_T+1\right)} \sum_{i=1}^{r} \frac{R_{i\bullet}^2}{n_i}\right] - 3\left(n_T + 1\right)$$

$$\text{Rejection Region:} \quad H^* \geq \chi^2_{r-1} \qquad P = P\left(\chi^2_{r-1} \geq H^*\right)$$

When there are ties, an adjustment can be made to H^*, that rarely makes a large effect. Let the number of groups with ties be g and let the k^{th} group have t_k tied observations. Then, the adjusted test statistic is given below.

$$\text{Test Statistic:} \quad H^{*'} = \frac{H^*}{\left[1 - \frac{\sum_{k=1}^{g}(t_k-1)t_k(t_k+1)}{(n_T-1)n_T(n_T+1)}\right]}$$

When the null hypothesis of equal medians is rejected, approximate Confidence Intervals for the differences in mean ranks among the $g = r(r-1)/2$ pairs of treatments are computed as follows.

$$\left(\overline{R}_{i\bullet} - \overline{R}_{i'\bullet}\right) \pm z_{1-\alpha/(2g)} \sqrt{\frac{n_T\left(n_T+1\right)}{2}\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)}$$

Example 4.8 - Virtual Training for a Lifeboat Launching Task

For the lifeboat training experiment, there were $r = 4$ treatments and each sample size was $n = 16$. We will directly compute the ranks, test, and multiple comparisons. Then the **kruskal.test** function will be used. Note that the **rank** function in R correctly adjusts ranks for ties. The null hypothesis is that the population medians are equal for the $r = 4$ treatments.

```
##   grp.trt procKnow
## 1       1   4.5614
## 2       1   6.6593
## 3       1   5.6427
## 4       1   6.4394
## 5       1   4.8635
## 6       1   0.3268
```

```
##                     KW stat df X2(.95) P(>KW)
## Kruskal-Wallis Test 12.9151  3  7.8147 0.0048
```

```
##      Trt i Trt i' Rank diff     MSD       LL      UL  Adj P
## [1,]     2      1    22.625 17.3671   5.2579 39.9921 0.0035
## [2,]     3      1    16.875 17.3671  -0.4921 34.2421 0.0622
## [3,]     3      2    -5.750 17.3671 -23.1171 11.6171 1.0000
## [4,]     4      1    15.250 17.3671  -2.1171 32.6171 0.1231
## [5,]     4      2    -7.375 17.3671 -24.7421  9.9921 1.0000
## [6,]     4      3    -1.625 17.3671 -18.9921 15.7421 1.0000
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  procKnow by grp.trt
## Kruskal-Wallis chi-squared = 12.915, df = 3, p-value = 0.004824
```

The test concludes that the medians are not all equal. Based on the simultaneous comparisons, conclude that the median for treatment 2 (Monitor/Keyboard) is significantly higher than that for treatment 1 (Lecture/Materials). No other medians are significantly different.

∇

```
library(tidyverse)
library(kableExtra)
library(effectsize)
library(agricolae)
library(car)
library(PMCMRplus)
```

Chapter 5

Balanced Two Factor Designs

In this chapter, we introduce two factor designs with equal numbers of replicates (n) per treatment (factor combination). The factors will be generically labelled as **A** with a levels and **B** with b levels. The total number of observations is $n_T = abn$.

5.1 Introduction

In a controlled experiment, experimental units are obtained and randomly assigned to the ab treatments, with n units per treatment. In observational studies, n units are sampled from each of the ab populations/subpopulations.

The **1-factor-at-a-time** method is used in some fields of study. This involves selecting a particular level of factor A and obtain the best level of factor B. Then, at the "best" level of factor B, obtain the best level of factor A. This method has problems, in particular it does not permit studying interactions between the factors. It is best (when feasible) to conduct the experiment at all ab treatment levels.

We first consider the population mean structures for **additive** and **interaction** models based on hypothetical values for an experiment of the "halo effect".

Example 5.1 - Halo Effect

In this study, researchers were interested in observing the "halo effect" where individuals who are known to be good/bad in a particular dimension and assumed to also be good/bad in another dimension. This phenomenon has been studied involving people, consumer products, and in many other subject areas. This hypothetical data is taken from a psychology experiment where students were given a picture of a student who had written an essay (Good, Bad, No Picture), and an essay that was supposedly written by the student (Good, Poor). The response measured was the score that the student assigned to the essay. Each student rater was assigned to one Picture/Essay Quality condition [Landy and Sigall, 1974].

For this example, factor A is the Picture condition with $a = 3$ levels and factor B is Essay quality with $b = 2$ levels. Let μ_{ij} be the mean when factor A is at level i and factor B is at level j. Further, let $\mu_{i\bullet}$ be the mean when factor A is at level i (across levels of factor B) and $\mu_{\bullet j}$ be the mean when factor B is at level j. Finally $\mu_{\bullet\bullet}$ is the overall mean across all levels of factors A and B.

Consider the following hypothetical means.

```
means <- matrix(c(25,18,20,21, 17,10,12,13, 21,14,16,17), ncol=3)
colnames(means) <- c("j=1", "j=2", "Row Mean")
rownames(means) <- c("i=1", "i=2", "i=3", "Col Mean")
means
```

```
##          j=1 j=2 Row Mean
## i=1       25  17       21
## i=2       18  10       14
## i=3       20  12       16
```

```
## Col Mean  21  13       17
```

For the **additive effects** model, we define the following parameters based on the population means.

$$\mu_{ij} = \mu_{\bullet\bullet} + \alpha_i + \beta_j \qquad \text{s.t.} \sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = 0 \qquad \mu_{\bullet\bullet} = \frac{1}{ab} \sum_{i=1}^{a} \sum_{j=1}^{b} \mu_{ij}$$

$$\mu_{i\bullet} = \frac{1}{b} \sum_{j=1}^{b} \mu_{ij} = \frac{1}{b} \sum_{j=1}^{b} (\mu_{\bullet\bullet} + \alpha_i + \beta_j) = \frac{1}{b} \left[b\mu_{\bullet\bullet} + b\alpha_i + \sum_{j=1}^{b} \beta_j \right] = \mu_{\bullet\bullet} + \alpha_i \qquad \mu_{\bullet j} = \mu_{\bullet\bullet} + \beta_j$$

$$\alpha_i = \mu_{i\bullet} - \mu_{\bullet\bullet} \qquad \beta_j = \mu_{\bullet j} - \mu_{\bullet\bullet} \qquad \mu_{\bullet\bullet} = \frac{1}{a} \sum_{i=1}^{a} \mu_{i\bullet} = \frac{1}{b} \sum_{ij1}^{b} \mu_{\bullet j}$$

For the halo effect values, we obtain the following parameters.

$$\alpha_1 = \mu_{1\bullet} - \mu_{\bullet\bullet} = 21 - 17 = 4 \quad \alpha_2 = 14 - 17 = -3 \quad \alpha_3 = 16 - 17 = -1 \qquad \alpha_1 + \alpha_2 + \alpha_3 = 0$$

$$\beta_1 = \mu_{\bullet 1} - \mu_{\bullet\bullet} = 21 - 17 = 4 \qquad \beta_2 = 13 - 17 = -4 \qquad \beta_1 + \beta_2 = 0$$

Plots of the means for each cell are given in Figure 5.1. The first plot gives means versus Picture with separate lines for each Essay Quality. The second plot gives means versus Essay Quality with separate lines for each Picture. In both cases lines are parallel, consistent with an additive model. This is an example of an **interaction plot** (though for these means, there is no interaction).

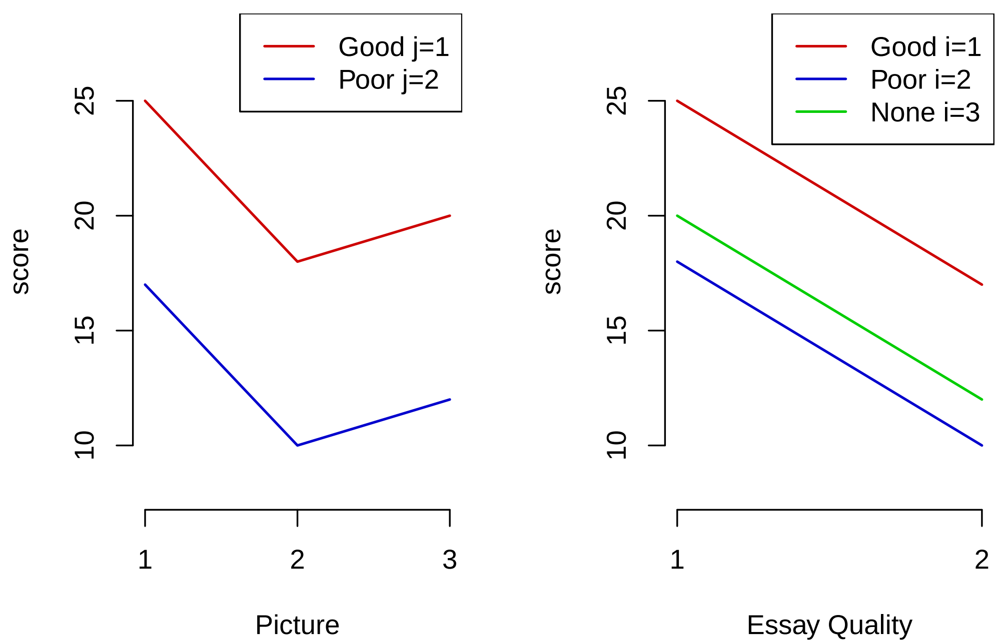


Figure 5.1: Halo effect means - Interaction plots with additive effects

The effect of Good versus Poor Essay Quality is the same for each Picture condition and the effects of the Picture conditions are the same for each Essay Quality.

Now, consider the following mean structure, consistent with an **interaction** model.

```
##           j=1 j=2 Row Mean
## i=1        23  19       21
## i=2        20   8       14
## i=3        20  12       16
## Col Mean   21  13       17
```

For the **interaction effects** model, we define the following parameters based on the population means.

$$\mu_{ij} = \mu_{\bullet\bullet} + \alpha_i + \beta_j + (\alpha\beta)_{ij} \qquad \text{s.t.} \sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = \sum_{i=1}^{a} (\alpha\beta)_{ij} = \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0$$

$$(\alpha\beta)_{ij} = \mu_{ij} - \mu_{\bullet\bullet} - \alpha_i - \beta_j = \mu_{ij} - \mu_{\bullet\bullet} - (\mu_{i\bullet} - \mu_{\bullet\bullet}) - (\mu_{\bullet j} - \mu_{\bullet\bullet})) = \mu_{ij} + \mu_{i\bullet} + \mu_{\bullet j} - \mu_{\bullet\bullet}$$

For the halo effect example, we obtain the following values for the interaction effects.

$$(\alpha\beta)_{11} = 23 - 21 - 21 + 17 = -2 \qquad (\alpha\beta)_{12} = 19 - 21 - 13 + 17 = 2$$
$$(\alpha\beta)_{21} = 20 - 14 - 21 + 17 = 2 \qquad (\alpha\beta)_{22} = 8 - 14 - 13 + 17 = -2$$

$$(\alpha\beta)_{31} = 20 - 16 - 21 + 17 = 0 \qquad (\alpha\beta)_{32} = 12 - 16 - 13 + 17 = 0$$

Interaction plots are given in Figure 5.2. In this case, the lines are not parallel.

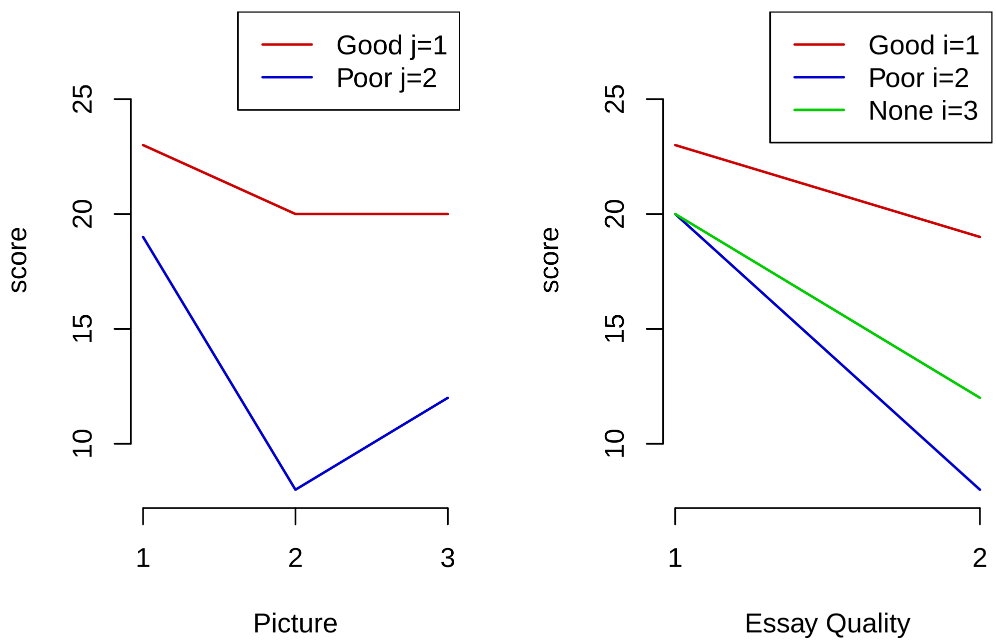


Figure 5.2: Halo effect means - Interaction plots with non-additive effects

∇

In practice, we don't know the structure of the population means, and must estimate them and make inferences regarding them based on sample data. Before describing the data model, we make a few comments regarding interactions.

- In many situations, the interaction among effects is small relative to main effects and can be ignored.
- Transformations (as in the Box-Cox transformation) can be applied which may remove an interaction.
- In many research settings, interactions may be hypothesized and have interesting theoretical interpretations.
- When factors are ordinal or quantitative, interactions that are synergistic or antagonistic can be observed.

5.2 Two Factor Analysis of Variance

In this section, we describe the 2-factor Analysis of Variance. We will first consider the **Cell Means** model, where interest is only on the ab means, typically trying to determine the "best" treatment. Then we consider the **Factor Effects** model, which measures main effects and interactions among the treatments. In each model, Y_{ijk} represents a random outcome when factor A is at level i, factor B is at level j, and replicate is k.

5.2.1 Cell Means Model

In this model, the focus is on the ab cell means among the combination of levels of factors A and B.

$$Y_{ijk} = \mu_{ij} + \epsilon_{ijk} \quad i = 1, \ldots, a; \quad j = 1, \ldots, b; \quad k = 1, \ldots, n \quad n_T = abn$$

$$\mu_{ij} \equiv \text{ mean when factor A is at level } i \text{ and factor B is at level } j \qquad \epsilon_{ijk} \sim NID\left(0, \sigma^2\right)$$

The model can be written in matrix form with Y being $n_T \times 1$, X being $n_T \times ab$, β being $ab \times 1$, and ϵ being $n_T \times 1$. For the case where $a = b = n = 2$, we have the following structure.

$$Y = \begin{bmatrix} Y_{111} \\ Y_{112} \\ Y_{121} \\ Y_{122} \\ Y_{211} \\ Y_{212} \\ Y_{221} \\ Y_{222} \end{bmatrix} \quad X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \beta = \begin{bmatrix} \mu_{11} \\ \mu_{12} \\ \mu_{21} \\ \mu_{22} \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_{111} \\ \epsilon_{112} \\ \epsilon_{121} \\ \epsilon_{122} \\ \epsilon_{211} \\ \epsilon_{212} \\ \epsilon_{221} \\ \epsilon_{222} \end{bmatrix}$$

Under this model, $\sigma^2\{Y\} = \sigma^2\{\epsilon\} = \sigma^2 I_{n_T}$.

The **Factor Effects** model decomposes μ_{ij} into main effects among the levels of factors A and B, as well as the interactions among the combinations of levels of factors A and B.

$$Y_{ijk} = \mu_{\bullet\bullet} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \quad i = 1, \ldots, a; \quad j = 1, \ldots, b; \quad k = 1, \ldots, n$$

$$\mu_{\bullet\bullet} = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\mu_{ij} \qquad \mu_{i\bullet} = \frac{1}{b}\sum_{j=1}^{b}\mu_{ij} \qquad \mu_{\bullet j} = \frac{1}{a}\sum_{i=1}^{a}\mu_{ij}$$

$$\text{Main effect of level } i \text{ factor A:} \quad \alpha_i = \mu_{i\bullet} - \mu_{\bullet\bullet} \qquad \sum_{i=1}^{a}\alpha_i = 0$$

$$\text{Main effect of level } j \text{ factor B:} \quad \beta_j = \mu_{\bullet j} - \mu_{\bullet\bullet} \qquad \sum_{j=1}^{b}\beta_j = 0$$

$$\text{Interaction effect of A at level } i \text{ and B at level } j: \quad (\alpha\beta)_{ij} = \mu_{ij} - \mu_{i\bullet} - \mu_{\bullet j} + \mu_{\bullet\bullet}$$

$$\sum_{i=1}^{a}(\alpha\beta)_{ij} = \sum_{j=1}^{b}(\alpha\beta)_{ij} = 0$$

$$Y_{ijk} \sim N\left(\mu_{\bullet\bullet} + \alpha_i + \beta_j + (\alpha\beta)_{ij}, \sigma^2\right) \quad \text{independent}$$

5.2.2 Least Squares/Maximum Likelihood Estimators

First, we define the following means.

$$\overline{Y}_{ij\bullet} = \frac{1}{n}\sum_{k=1}^{n} Y_{ijk} \quad \overline{Y}_{i\bullet\bullet} = \frac{1}{bn}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk} \quad \overline{Y}_{\bullet j\bullet} = \frac{1}{an}\sum_{i=1}^{a}\sum_{k=1}^{n} Y_{ijk}$$

$$\overline{Y}_{\bullet\bullet\bullet} = \frac{1}{abn}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} Y_{ijk}$$

For the cell means model, the least squares estimator for μ_{ij} is simply the sample mean for that cell, $\overline{Y}_{ij\bullet}$.

$$Q = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\epsilon_{ijk}^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\mu_{ij}\right)^2 \quad \text{Set: } \frac{\partial Q}{\partial \mu_{ij}} = 0 \quad \Rightarrow \quad \hat{\mu}_{ij} = \overline{Y}_{ij\bullet}$$

The fitted values and residuals for the cell means model are given below.

$$\hat{Y}_{ijk} = \hat{\mu}_{ij} = \overline{Y}_{ij\bullet} \qquad e_{ijk} = Y_{ijk} - \hat{Y}_{ijk} = Y_{ijk} - \overline{Y}_{ij\bullet}$$

For the factor effects model, we obtain the following least squares estimators.

$$\hat{\mu}_{\bullet\bullet} = \overline{Y}_{\bullet\bullet\bullet} \qquad \hat{\alpha}_i = \overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet\bullet\bullet} \qquad \hat{\beta}_j = \overline{Y}_{\bullet j\bullet} - \overline{Y}_{\bullet\bullet\bullet}$$

$$\widehat{(\alpha\beta)}_{ij} = \overline{Y}_{ij\bullet} - \overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet j\bullet} + \overline{Y}_{\bullet\bullet\bullet}$$

The fitted values and the residuals for the factor effects model are the same as for the cell means model.

$$\hat{Y}_{ijk} = \hat{\mu}_{\bullet\bullet} + \hat{\alpha}_i + \hat{\beta}_j + \widehat{(\alpha\beta)}_{ij} = \overline{Y}_{ij\bullet} \qquad e_{ijk} = Y_{ijk} - \hat{Y}_{ijk} = Y_{ijk} - \overline{Y}_{ij\bullet}$$

5.2.3 Sums of Squares and the Analysis of Variance

For the cell means model, the 2-Way ANOVA simplifies to a 1-Way ANOVA with $r = ab$ treatments.

$$Y_{ijk} - \overline{Y}_{\bullet\bullet\bullet} = \left(Y_{ijk} - \overline{Y}_{ij\bullet}\right) + \left(\overline{Y}_{ij\bullet} - \overline{Y}_{\bullet\bullet\bullet}\right)$$

This identity leads to the sums of squares, as the cross-products sum to zero.

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk} - \overline{Y}_{\bullet\bullet\bullet}\right)^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk} - \overline{Y}_{ij\bullet}\right)^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(\overline{Y}_{ij\bullet} - \overline{Y}_{\bullet\bullet\bullet}\right)^2$$

$$SSTO = SSE + SSTR \qquad df_{TO} = n_T - 1 = (n_T - ab) + (ab - 1) = df_E + df_{TR}$$

For the factor effects model, we decompose the treatment deviations and sums of squares into main effects and interactions among levels of factors A and B.

$$Y_{ijk} - \overline{Y}_{\bullet\bullet\bullet} =$$

$$\left(Y_{ijk} - \overline{Y}_{ij\bullet}\right) + \left(\overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet\bullet\bullet}\right) + \left(\overline{Y}_{\bullet j\bullet} - \overline{Y}_{\bullet\bullet\bullet}\right) + \left(\overline{Y}_{ij\bullet} - \overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet j\bullet} + \overline{Y}_{\bullet\bullet\bullet}\right)$$

Again, sums of cross-product terms are zero and we partition the sums of squares as follows.

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\overline{Y}_{\bullet\bullet\bullet}\right)^2 =$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\overline{Y}_{ij\bullet}\right)^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(\overline{Y}_{i\bullet\bullet}-\overline{Y}_{\bullet\bullet\bullet}\right)^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(\overline{Y}_{\bullet j\bullet}-\overline{Y}_{\bullet\bullet\bullet}\right)^2 +$$

$$+\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(\overline{Y}_{ij\bullet}-\overline{Y}_{i\bullet\bullet}-\overline{Y}_{\bullet j\bullet}+\overline{Y}_{\bullet\bullet\bullet}\right)^2$$

$$SSTO = SSE + SSA + SSB + SSAB$$

$$df_{TO} = n_T - 1 = (n_T - ab) + (a-1) + (b-1) + (a-1)(b-1) =$$
$$= df_E + df_A + df_B + df_{AB}$$

For the factor effects model, we summarize the following sums of squares, degrees of freedom, and expected mean squares.

$$\text{Factor A:} \quad SSA = bn\sum_{i=1}^{a}\left(\overline{Y}_{i\bullet\bullet}-\overline{Y}_{\bullet\bullet\bullet}\right)^2 \qquad df_A = a-1 \qquad MSA = \frac{SSA}{a-1}$$

$$E\{MSA\} = \sigma^2 + \frac{bn\sum_{i=1}^{a}\alpha_i^2}{a-1}$$

$$\text{Factor B:} \quad SSB = an\sum_{j=1}^{b}\left(\overline{Y}_{\bullet j\bullet}-\overline{Y}_{\bullet\bullet\bullet}\right)^2 \qquad df_B = b-1 \qquad MSB = \frac{SSB}{b-1}$$

$$E\{MSB\} = \sigma^2 + \frac{an\sum_{j=1}^{b}\beta_j^2}{b-1}$$

$$\text{AB Interaction:} \quad SSAB = n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{Y}_{ij\bullet}-\overline{Y}_{i\bullet\bullet}-\overline{Y}_{\bullet j\bullet}+\overline{Y}_{\bullet\bullet\bullet}\right)^2$$

$$df_{AB} = (a-1)(b-1) \qquad MSAB = \frac{SSAB}{(a-1)(b-1)}$$

$$E\{MSAB\} = \sigma^2 + \frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}(\alpha\beta)_{ij}^2}{(a-1)(b-1)}$$

$$\text{Error:} \quad SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\overline{Y}_{ij\bullet}\right)^2 \qquad df_E = ab(n-1) = n_T - ab$$

$$MSE = \frac{SSE}{ab(n-1)} \qquad E\{MSE\} = \sigma^2$$

When working with published results, the error sum of squares can be obtained from the treatment standard deviation as in the 1-Way ANOVA.

$$SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk}-\overline{Y}_{ij\bullet}\right)^2 = (n-1)\sum_{i=1}^{a}\sum_{j=1}^{b}s_{ij}^2$$

When testing for the effects of the factors, begin with a test for interaction, then test for main effects. If the interaction is important, the tests for the main effects may be deceiving, as the effects of factors A and B depend of the level of the other factor.

The test for interaction is given here.

$$H_0^{AB}: (\alpha\beta)_{11} = \cdots = (\alpha\beta)_{ab} = 0 \qquad H_A^{AB}: \text{ Not all } (\alpha\beta)_{ij} = 0$$

$$\text{Test Statistic:} \quad F_{AB}^* = \frac{MSAB}{MSE} \qquad \text{Rejection Region:} \quad F_{AB}^* \geq F_{1-\alpha;(a-1)(b-1),ab(n-1)}$$

The tests for the main effects of factors A and B are conducted as follow.

$$H_0^A: \alpha_1 = \cdots = \alpha_a = 0 \qquad H_A^A: \text{ Not all } \alpha_i = 0$$

$$\text{Test Statistic:} \quad F_A^* = \frac{MSA}{MSE} \qquad \text{Rejection Region:} \quad F_A^* \geq F_{1-\alpha;a-1,ab(n-1)}$$

$$H_0^B: \beta_1 = \cdots = \beta_b = 0 \qquad H_A^B: \text{ Not all } \beta_j = 0$$

$$\text{Test Statistic:} \quad F_B^* = \frac{MSB}{MSE} \qquad \text{Rejection Region:} \quad F_B^* \geq F_{1-\alpha;b-1,ab(n-1)}$$

Effect sizes for the main effects and interaction can be obtained from the sums of squares in the Analysis of Variance. The effect size η^2 is obtained by dividing each sum of squares by the total sum of squares. Partial η^2 is obtained by dividing each sum of squares by the sum of it and the error sum of squares.

$$\eta_A^2 = \frac{SSA}{SSTO} \qquad \eta_B^2 = \frac{SSB}{SSTO} \qquad \eta_{AB}^2 = \frac{SSAB}{SSTO}$$

$$\text{Partial-}\eta_A^2 = \frac{SSA}{SSA + SSE} \qquad \text{Partial-}\eta_B^2 = \frac{SSB}{SSA + SSE} \qquad \text{Partial-}\eta_{AB}^2 = \frac{SSAB}{SSAB + SSE}$$

Example 5.2 - Drumstick Weights of Broiler Chickens

A study was conducted to compare weights of chicken parts under 4 diets [Aksu et al., 2007]. The 4 diets were combinations of two factors, each with two levels ($a = b = 2$). Factor A was base diet (Sorghum (i=1) and Corn (i=2)). Factor B was methionine supplement (Absent (j=1) and Present (j=2)). There were $n = 60$ chickens receiving each diet for a total of ($n_T = abn = 2(2)(60) = 240$).

The observed means and standard deviations for the diets are given below.

$$\overline{y}_{11\bullet} = 106.08 \quad \overline{y}_{12\bullet} = 93.67 \quad \overline{y}_{21\bullet} = 101.17 \quad \overline{y}_{22\bullet} = 108.83$$

$$s_{11} = 15.04 \quad s_{12} = 12.29 \quad s_{21} = 16.74 \quad s_{22} = 20.93$$

Here we compute the marginal means for each level of factors A and B, as well as the overall mean, and then obtain least squares estimates and the Analysis of Variance.

$$\overline{y}_{1\bullet\bullet} = \frac{106.08 + 93.67}{2} = 99.875 \quad \overline{y}_{2\bullet\bullet} = \frac{101.17 + 108.83}{2} = 105.000$$

$$\overline{y}_{\bullet 1\bullet} = \frac{106.08 + 101.17}{2} = 103.625 \quad \overline{y}_{\bullet 2\bullet} = \frac{93.67 + 108.83}{2} = 101.250$$

$$\overline{y}_{\bullet\bullet\bullet} = \frac{106.08 + 93.67 + 101.17 + 108.83}{4} = 102.4375$$

The main effect and interacion effects are computed from the means above.

$$\hat{\alpha}_1 = 99.875 - 102.4375 = -2.5625 \quad \hat{\alpha}_2 = 105.000 - 102.4375 = 2.5625$$

$$\hat{\beta}_1 = 103.625 - 102.4375 = 1.1875 \quad \hat{\beta}_2 = 101.250 - 102.4375 = -1.1875$$

$$(\hat{\alpha\beta})_{11} = 106.08 - 99.875 - 103.625 + 102.4375 = 5.0175$$

$$(\hat{\alpha\beta})_{12} = 93.67 - 99.875 - 101.250 + 102.4375 = -5.0175$$

$$(\hat{\alpha\beta})_{21} = 101.17 - 105.000 - 103.625 + 102.4375 = -5.0175$$

$$(\hat{\alpha\beta})_{22} = 108.83 - 105.000 - 101.250 + 102.4375 = 5.0175$$

Next, we compute the sums of squares, making use of the estimated effects for SSA, SSB, and $SSAB$.

$$SSA = bn\sum_{i=1}^{a} \hat{\alpha}_i^2 = 2(60)\left[(-2.5625)^2 + (2.5625)^2\right] = 1575.94$$

$$df_A = 2 - 1 \qquad MSA = 1575.94$$

$$SSB = an\sum_{j=1}^{b} \hat{\beta}_j^2 = 2(60)\left[(1.1875)^2 + (-1.1875)^2\right] = 338.44$$

$$df_B = 2 - 1 \qquad MSB = 338.44$$

$$SSAB = n\sum_{i=1}^{a}\sum_{j=1}^{b} (\hat{\alpha\beta})_{ij}^2 =$$

$$60\left[(5.0175)^2 + (-5.0175)^2 + (-5.0175)^2 + (5.0175)^2\right] = 6445.58$$

$$df_{AB} = (2-1)(2-1) = 1 \qquad MSAB = 6445.58$$

$$SSE = (n-1)\sum_{i=1}^{a}\sum_{j=1}^{b} s_{ij}^2 = 59\left[(15.04)^2 + (12.29)^2 + (16.74)^2 + (20.93)^2\right] = 64636.75$$

$$df_E = 2(2)(60-1) = 236 \qquad MSE = 273.88$$

First, test for an interaction between base diet and methionine supplement.

$$H_0^{AB} : (\alpha\beta)_{11} = \cdots = (\alpha\beta)_{22} = 0$$

$$\text{Test Statistic:} \quad F_{AB}^* = \frac{MSAB}{MSE} = \frac{6445.58}{273.88} = 23.534$$

$$\text{Rejection Region:} \quad F_{AB}^* \geq F_{.95;1,236} = 3.881 \qquad P = P\left(F_{1,236} \geq 23.534\right) < .0001$$

As there is strong evidence of an interaction (adding methionine has a large negative effect when sorghum is used, and has a large positive effect when corn is used), we will conduct the main effects test for completeness.

$$H_0^A : \alpha_1 = \alpha_2 = 0 \qquad H_A^A : \text{ Not all } \alpha_i = 0$$

$$\text{Test Statistic:} \quad F_A^* = \frac{MSA}{MSE} = \frac{1575.94}{273.88} = 5.754$$

$$\text{Rejection Region:} \quad F_A^* \geq F_{.95;1,236} = 3.881 \qquad P = P\left(F_{1,236} \geq 5.754\right) = .0172$$

$$H_0^B : \beta_1 = \beta_2 = 0 \qquad H_A^B : \text{ Not all } \beta_j = 0$$

$$\text{Test Statistic:} \quad F_B^* = \frac{MSB}{MSE} = \frac{338.44}{273.88} = 1.236$$

$$\text{Rejection Region:} \quad F_A^* \geq F_{.95;1,236} = 3.881 \qquad P = P\left(F_{1,236} \geq 1.236\right) = .2674$$

The test for factor A is significant (Corn does better marginally than Sorghum). The test for factor B is not significant (marginally Methionine Absent and Present give similar results). This last result is mis-leading as Methionine has a large positve effect for Corn (7.66 gram increase) but a large negative effect for Sorghum (12.41 gram decrease), so marginally they "cancel out."

The effect sizes and partial effect sizes for the base diet, methionine, and their interaction are obtained below, along with the total sum of squares.

$$SSTO = SSA + SSB + SSAB + SSE = 1575.94 + 338.44 + 6445.58 + 64636.75 = 72996.71$$

$$\eta_A^2 = \frac{1575.94}{72996.71} = 0.0216 \qquad \eta_B^2 = \frac{338.44}{72996.71} = 0.0046 \qquad \eta_{AB}^2 = \frac{6445.58}{72996.71} = 0.0883$$

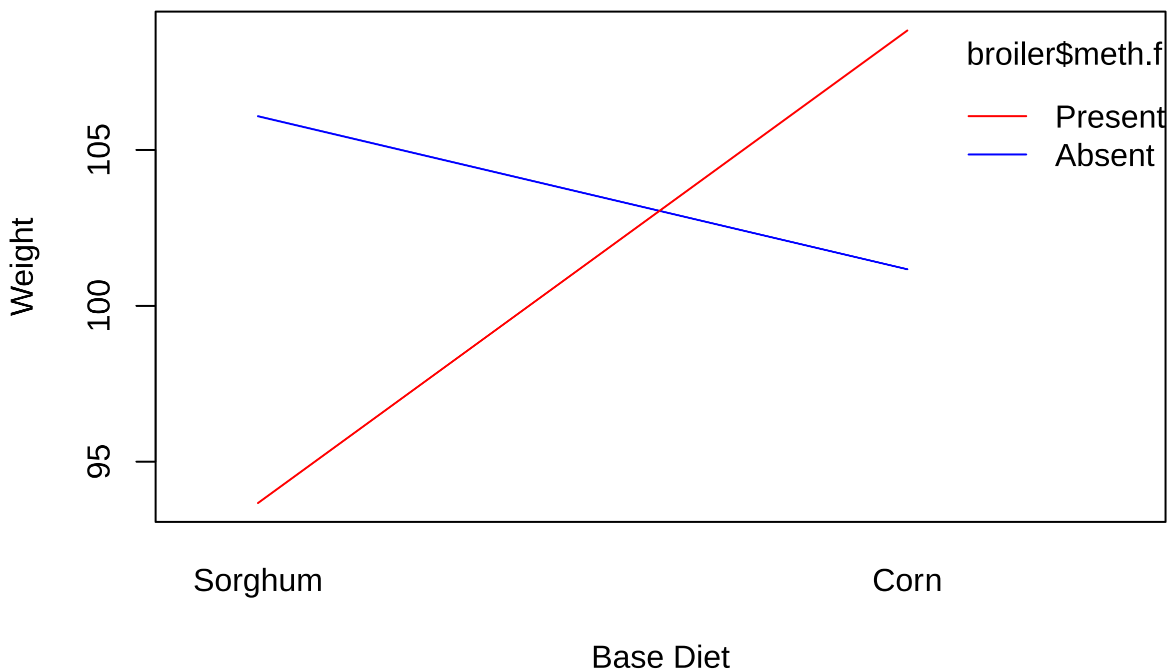
$$\text{Partial-}\eta_A^2 = \frac{1575.94}{1575.94 + 64636.75} = 0.0238 \qquad \text{Partial-}\eta_B^2 = \frac{338.44}{338.44 + 64636.75} = 0.0052$$

$$\text{Partial-}\eta_{AB}^2 = \frac{6445.58}{6445.58 + 64636.75} = 0.0907$$

The interaction has a much larger effect size than either of the main effects. Based on η_{AB}^2, approximately 9% of the total variance is attributable to the interaction.

The following R code uses the **aov** function. As before, we set the options to have the effects sum to zero.

```
##   diet base meth  ds.wt
## 1    1    1    1 116.33
## 2    1    1    1  99.43
## 3    1    1    1 106.58
## 4    1    1    1 109.64
## 5    1    1    1  78.58
## 6    1    1    1  93.18
```




```
##
## Call:
## aov(formula = ds.wt ~ base.f * meth.f, data = broiler)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -59.16 -10.94   0.78   9.58  57.21
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       102.437      1.068  95.892  < 2e-16 ***
## base.f1            -2.562      1.068  -2.399   0.0172 *
## meth.f1             1.187      1.068   1.112   0.2674
## base.f1:meth.f1     5.017      1.068   4.697 4.49e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.55 on 236 degrees of freedom
## Multiple R-squared:  0.1096, Adjusted R-squared:  0.09828
## F-statistic: 9.683 on 3 and 236 DF,  p-value: 4.731e-06
```

```
## Analysis of Variance Table
##
## Response: ds.wt
##                 Df Sum Sq Mean Sq F value    Pr(>F)
## base.f           1   1576  1575.8  5.7535   0.01723 *
## meth.f           1    338   338.4  1.2356   0.26745
## base.f:meth.f    1   6042  6041.8 22.0596 4.488e-06 ***
## Residuals      236  64636   273.9
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## # Effect Size for ANOVA (Type I)
##
## Parameter     | Eta2 (partial) |       95% CI
## ---------------------------------------------
## base.f        |           0.02 | [0.00, 1.00]
## meth.f        |       5.21e-03 | [0.00, 1.00]
## base.f:meth.f |           0.09 | [0.04, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

```
## # Effect Size for ANOVA (Type I)
##
## Parameter     |     Eta2 |       95% CI
## ---------------------------------------
## base.f        |     0.02 | [0.00, 1.00]
## meth.f        | 4.66e-03 | [0.00, 1.00]
## base.f:meth.f |     0.08 | [0.04, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

∇

Example 5.3 - Reading Times on Electronic Readers

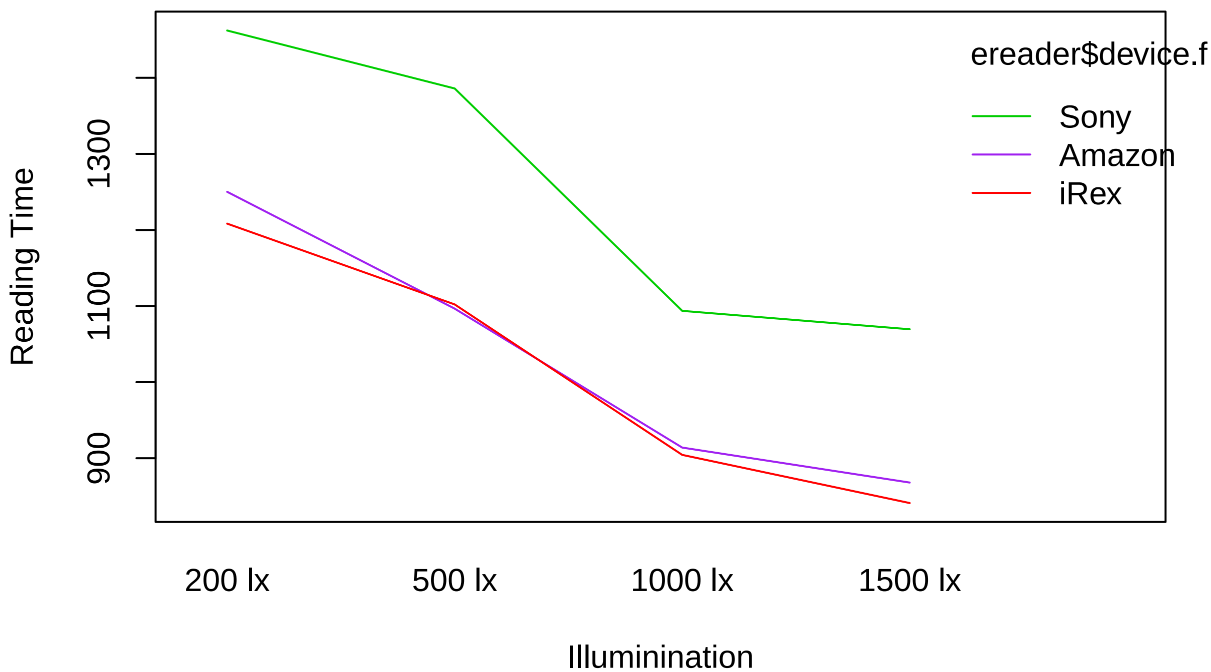
Next, consider a study that compared $a = 3$ electronic reader models, each at $b = 4$ illuminance levels [Chang et al., 2013]. The three reader models are described as follows.

- Sony PRS 700 - 6” diagonal screen size ($i = 1$)
- Amazon Kindle DX - 9.7” diagonal screen size ($i = 2$)
- iRex 1000S - 10.2” diagonal screen size ($i = 3$)

The illuminance levels studied were 200 lx ($j = 1$), 500 lx ($j = 2$), 1000 lx ($j = 3$), and 1500 lx ($j = 4$). A total study group of $n_T = 60$ subjects was obtained with $n = 5$ being measured on the $ab = 3(4) = 12$ combinations of device and lighting level. The response variable Y is the reading time in seconds.

R code and output are given below.

```
##   device light readtime
## 1      1     1  1656.26
## 2      1     1  1405.92
## 3      1     1  1797.21
## 4      1     1  1155.96
## 5      1     1  1295.44
## 6      1     2  1022.32
```




```
##          200 lx   500 lx  1000 lx 1500 lx
## Sony   1462.158 1385.960 1093.694 1069.49
## Amazon 1250.188 1096.594  914.006  868.01
## iRex   1208.350 1102.146  904.400  841.10
```

```
##
## Call:
## aov(formula = readtime ~ device.f * light.f, data = ereader)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -497.41 -187.20  -20.66  206.88  463.53
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         1099.675     35.601  30.889  < 2e-16 ***
## device.f1            153.151     50.347   3.042  0.00380 **
## device.f2            -67.475     50.347  -1.340  0.18649
## light.f1             207.224     61.663   3.361  0.00153 **
## light.f2              95.225     61.663   1.544  0.12909
## light.f3            -128.975     61.663  -2.092  0.04179 *
## device.f1:light.f1     2.108     87.204   0.024  0.98081
## device.f2:light.f1    10.765     87.204   0.123  0.90227
## device.f1:light.f2    37.909     87.204   0.435  0.66572
## device.f2:light.f2   -30.831     87.204  -0.354  0.72523
## device.f1:light.f3   -30.157     87.204  -0.346  0.73099
## device.f2:light.f3    10.781     87.204   0.124  0.90212
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 275.8 on 48 degrees of freedom
## Multiple R-squared:  0.3771, Adjusted R-squared:  0.2343
## F-statistic: 2.641 on 11 and 48 DF,  p-value: 0.01001
```

```
## Analysis of Variance Table
##
## Response: readtime
##                  Df  Sum Sq Mean Sq F value    Pr(>F)
## device.f          2  706968  353484  4.6483 0.0142790 *
## light.f           3 1481064  493688  6.4920 0.0008906 ***
## device.f:light.f  6   21543    3591  0.0472 0.9995253
## Residuals        48 3650203   76046
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = readtime ~ device.f * light.f, data = ereader)
##
## $device.f
##                  diff       lwr        upr     p adj
## Amazon-Sony -220.6260 -431.5285  -9.723488 0.0384849
## iRex-Sony   -238.8265 -449.7290 -27.923988 0.0230557
## iRex-Amazon  -18.2005 -229.1030 192.702012 0.9762840
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = readtime ~ device.f * light.f, data = ereader)
##
## $light.f
##                      diff       lwr          upr     p adj
## 500 lx-200 lx   -111.9987 -379.9852  155.9878183 0.6838249
## 1000 lx-200 lx  -336.1987 -604.1852  -68.2121817 0.0085849
## 1500 lx-200 lx  -380.6987 -648.6852 -112.7121817 0.0023697
## 1000 lx-500 lx  -224.2000 -492.1865   43.7864849 0.1306514
## 1500 lx-500 lx  -268.7000 -536.6865   -0.7135151 0.0491639
## 1500 lx-1000 lx  -44.5000 -312.4865  223.4864849 0.9708523
```

```
## # Effect Size for ANOVA (Type I)
##
## Parameter        | Eta2 (partial) |       95% CI
## -----------------------------------------------
## device.f         |           0.16 | [0.02, 1.00]
## light.f          |           0.29 | [0.10, 1.00]
## device.f:light.f |       5.87e-03 | [0.00, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

```
## # Effect Size for ANOVA (Type I)
##
## Parameter        |     Eta2 |       95% CI
## -----------------------------------------
## device.f         |     0.12 | [0.00, 1.00]
## light.f          |     0.25 | [0.07, 1.00]
## device.f:light.f | 3.68e-03 | [0.00, 1.00]
##
```

```
## - One-sided CIs: upper bound fixed at [1.00].
```

Unlike the chicken broilers example, there is no evidence (whatsoever) of a significant interaction between device and light level $F^*_{AB} = 0.0472$, $P_{AB} = .9995$. There is evidence of device differences $F^*_A = 4.648$, $P_A = .0143$ and lighting differences $F^*_B = 6.492$, $P_B = .0009$.

The residual standard error, which estimates σ, is 275.8 seconds, a little over 4.5 minutes. The variation in people's reading times (within treatments) is quite large. Based on η^2, the illumination levels explains 25% of the total variation in reading times, while device explains 12%.

Based on Tukey comparisons among devices and among lighting levels we obtain the following significant differences. Formulas will be given in the next section.

- The larger Amazon and iRex readers have significantly lower reading times than the smaller Sony model.
- 1000 lx and 1500 lx have significantly lower reading times than 200 lx
- 1500 lx has significantly lower reading times than 500 lx

∇

Finally, some guidelines are given for testing and modelling strategy.

If the interaction is significant, and the goal was to demonstrate interactions (as is often the case in behavioral studies), describe the interaction effects in terms of cell means, as in the Chicken Diet example. If the interaction is significant, and the goal is to simplify the model, try some power (Box-Cox) transformations on Y and see if that simplifies the model.

If the interaction isn't significant or important, test for main effects for factors A and B. Also, make comparisons among the marginal means of factors A and B, as in the E-reader example.

5.3 Factor Effect Contrasts

In this section, we consider contrasts among means when interaction is absent and when it is present.

5.3.1 Contrasts when there is No Interaction

This is the simpler case, where contrasts are made among levels of factors A and B, as was done previously for single factor models.

For contrasts among levels for factor A, we summarize the results from the 1-Way ANOVA. Keep in mind that the marginal means for the levels of factor A are based on an levels.

$$\text{Contrasts among Factor A Levels:} \quad L = \sum_{i=1}^{a} c_i \mu_{i\bullet} \qquad \sum_{i=1}^{a} c_i = 0$$

$$\text{Estimator:} \quad \hat{L} = \sum_{i=1}^{a} c_i \overline{Y}_{i\bullet\bullet} \qquad \text{Standard Error:} \quad s\left\{\hat{L}\right\} = \sqrt{\frac{MSE}{bn}\sum_{i=1}^{a} c_i^2}$$

$$(1-\alpha)100\% \text{ Confidence Interval for } L: \quad \hat{L} \pm t_{1-\alpha/2;ab(n-1)} s\left\{\hat{L}\right\}$$

$$\text{Scheffe (all contrasts):} \quad \hat{L} \pm \left(\sqrt{(a-1)F_{1-\alpha;a-1,ab(n-1)}}\right) s\left\{\hat{L}\right\}$$

$$\text{Bonferroni (g pre-planned) contrasts:} \quad \hat{L} \pm t_{1-\alpha/(2g);ab(n-1)} s\left\{\hat{L}\right\}$$

$$\text{Tukey (All pairs):} \quad \hat{L} \pm \frac{q_{1-\alpha;a,ab(n-1)}}{\sqrt{2}} s\left\{\hat{L}\right\}$$

Similar results for factor B occur, with the following adjustments, as well as critical values for Scheffe's and Tukey's methods.

$$L = \sum_{j=1}^{b} c_j \mu_{\bullet j} \qquad \sum_{j=1}^{b} c_j = 0 \qquad \hat{L} = \sum_{j=1}^{b} c_j \overline{Y}_{\bullet j \bullet} \qquad s\left\{\hat{L}\right\} = \sqrt{\frac{MSE}{an} \sum_{j=1}^{b} c_j^2}$$

5.3.2 Contrasts when Interaction is Present

In this case, contrasts are made among cell means as opposed to marginal means.

$$\text{Contrasts:} \quad L = \sum_{i=1}^{a} \sum_{j=1}^{b} c_{ij} \mu_{ij} \qquad \sum_{i=1}^{a} \sum_{j=1}^{b} c_{ij} = 0$$

$$\text{Estimator:} \quad \hat{L} = \sum_{i=1}^{a} \sum_{j=1}^{b} c_{ij} \overline{Y}_{ij\bullet} \qquad \text{Std Error:} \quad s\left\{\hat{L}\right\} = \sqrt{\frac{MSE}{n} \sum_{i=1}^{a} \sum_{j=1}^{b} c_{ij}^2}$$

$$(1-\alpha)100\% \text{ Confidence Interval for } L: \quad \hat{L} \pm t_{1-\alpha/2;ab(n-1)} s\left\{\hat{L}\right\}$$

$$\text{Scheffe (all contrasts):} \quad \hat{L} \pm \left(\sqrt{(ab-1)F_{1-\alpha;ab-1,ab(n-1)}}\right) s\left\{\hat{L}\right\}$$

$$\text{Bonferroni (g pre-planned) contrasts:} \quad \hat{L} \pm t_{1-\alpha/(2g);ab(n-1)} s\left\{\hat{L}\right\}$$

$$\text{Tukey (All pairs):} \quad \hat{L} \pm \frac{q_{1-\alpha;ab,ab(n-1)}}{\sqrt{2}} s\left\{\hat{L}\right\}$$

```
library(tidyverse)
library(kableExtra)
library(effectsize)
library(agricolae)
library(car)
library(PMCMRplus)
library(additivityTests)
```

Chapter 6

Two-Factor Designs with 1 Observation per Treatment

In some cases, there is only a single observation within each combination of factor levels. This causes a problem when considering models with interaction effects. The error degrees of freedom are $df_E = ab(n-1) = ab(1-1) = 0$ and the error sum of squares is $SSE = 0$. Several methods have been developed for testing for interaction under particular restrictions. In this chapter, we describe **Tukey's One Degree of Freedom for Non-Additivity Test** (ODOFNA). For this test, the form of the interaction and the resulting test are given below.

$$(\alpha\beta)_{ij} = D\alpha_i\beta_j \qquad H_0: D = 0 \quad H_A: D \neq 0$$

Intuitively, the test involves estimating μ, as well as α_i and β_j from the marginal means for factors A and B. Then D is estimated by fitting a regression through the origin relating Y^* (Y minus the sum of the mean and the main effects) to X^* where X^* is the product of the estimates of α_i, and β_j.

$$\hat{\mu} = \overline{Y}_{\bullet\bullet} \qquad \hat{\alpha}_i = \overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet} \qquad \hat{\beta}_j = \overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet} \qquad Y^*_{ij} = Y_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j \qquad X^*_{ij} = \hat{\alpha}_i\hat{\beta}_j$$

For regression through the origin, we obtain the estimate and test whether $D = 0$ as follows, where the model with D is the complete model.

$$\hat{D} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b} X^*_{ij}Y^*_{ij}}{\sum_{i=1}^{a}\sum_{j=1}^{b} X^{*2}_{ij}} \qquad \hat{Y}_{ij}(C) = \hat{\mu}_{\bullet\bullet} + \hat{\alpha}_i + \hat{\beta}_j + \hat{D}\hat{\alpha}_i\hat{\beta}_j \qquad i = 1, \dots, a; \quad j = 1, \dots, b$$

$$SSE(C) = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(Y_{ij} - \hat{Y}_{ij}(C)\right)^2 \qquad df_E(C) = ab - (a-1) - (b-1) - 1 = ab - a - b + 1$$

Under the reduced model, $D = 0$, and we obtain the following fitted values and error sum of squares and degrees of freedom.

$$\hat{Y}_{ij}(R) = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j \qquad SSE(R) = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(Y_{ij} - \hat{Y}_{ij}(R)\right)^2 \qquad df_E(R) = ab - a - b + 2$$

Then Tukey's ODOFNA test is a special case of the general linear test with $H_0: D = 0$ vs $H_A: D \neq 0$.

$$\text{Test Statistic:} \quad F^* = \frac{\left[\frac{SSE(R) - SSE(C)}{df_E R - df_E(C)}\right]}{\left[\frac{SSE(C)}{df_E(C)}\right]} \qquad \text{Rejection Region:} \quad F^* \geq F_{1-\alpha;1,ab-a-b}$$

Example 6.1 - Economic Indices for 8 Sources Over 18 Years

A study considered U.S. economic indices for $a = 8$ business news sources over a $b = 18$ year period [Smith, 1969]. The sources were: DJIA, Poors, NYSE, GNP, CPI, Forbes, Business Week, and Money Magazine. The years were 1948-1965. There was one index per source per year. The following R code runs calculations directly, then structures the data in matrix form and uses the **additivityTests** package. Note that we order the data frame first by source, then by year within source, as **tapply** output is sorted by the alphanumeric levels of the grouping variable(s).

```
biz1 <- read.table("http://www.stat.ufl.edu/~winner/data/jb42.dat",header=F,
    col.names=c("source","year","Y"))
head(biz1)
```

```
##   source year     Y
## 1   DJIA 1965 1.103
## 2   DJIA 1964 1.145
## 3   DJIA 1963 1.169
## 4   DJIA 1962 0.890
## 5   DJIA 1961 1.207
## 6   DJIA 1960 0.896
```

```
## Create a new data frame that orders first by source then year within source
biz <- biz1[order(biz1$source, biz1$year),]
head(biz)
```

```
##     source year     Y
## 126  BWEEK 1948 1.019
## 125  BWEEK 1949 0.995
## 124  BWEEK 1950 1.213
## 123  BWEEK 1951 1.012
## 122  BWEEK 1952 1.141
## 121  BWEEK 1953 0.952
```

```
tail(biz)
```

```
##    source year     Y
## 24   POOR 1960 0.953
## 23   POOR 1961 1.231
## 22   POOR 1962 0.872
## 21   POOR 1963 1.201
## 20   POOR 1964 1.131
## 19   POOR 1965 1.099
```

```
all.mean <- mean(biz$Y)
source.mean <- as.vector(tapply(biz$Y, biz$source, mean))
year.mean <- as.vector(tapply(biz$Y, biz$year, mean))
a <- length(source.mean)
b <- length(year.mean)

biz$mu <- rep(all.mean, a*b)
biz$alpha <- rep(source.mean-all.mean, each=b)
biz$beta <- rep(year.mean-all.mean, times=a)

biz$Ystar <- biz$Y - biz$mu - biz$alpha - biz$beta
biz$Xstar <- biz$alpha * biz$beta

(Dhat <- sum(biz$Xstar*biz$Ystar) / sum(biz$Xstar^2))
```

```
## [1] 23.6517
```

```
biz$Yhat.C <- biz$mu + biz$alpha + biz$beta + Dhat*biz$Xstar
biz$Yhat.R <- biz$mu + biz$alpha + biz$beta
```

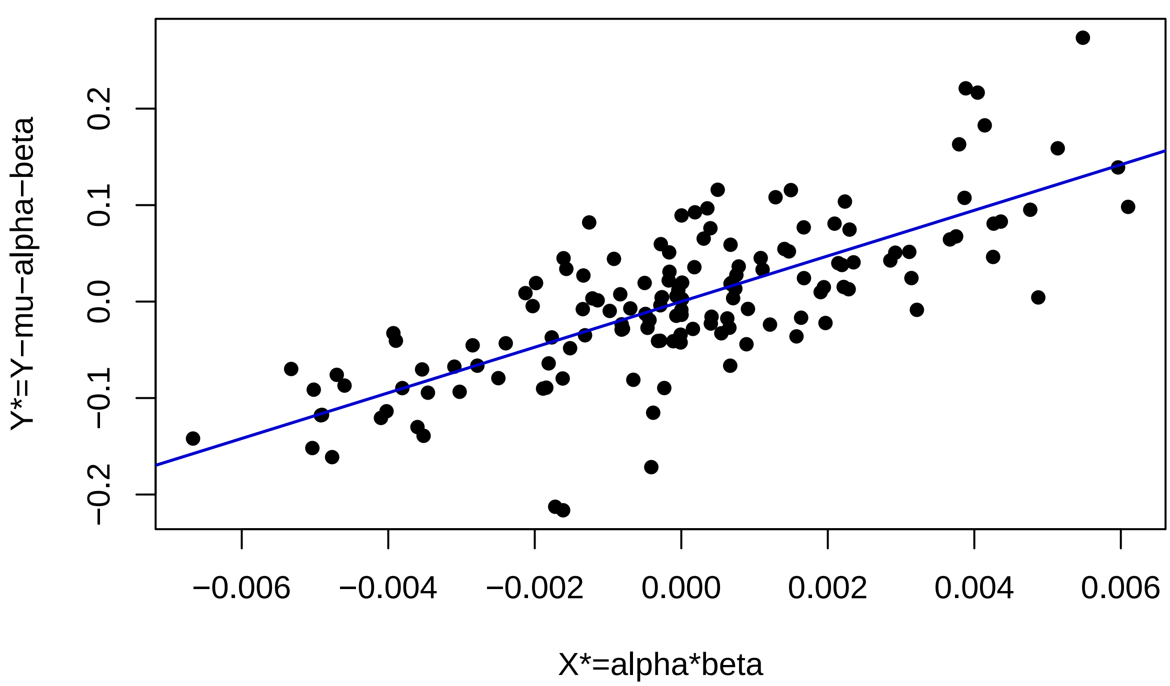
```
SSE.C <- sum((biz$Y-biz$Yhat.C)^2)
SSE.R <- sum((biz$Y-biz$Yhat.R)^2)
df_E.C <- a*b - a - b
df_E.R <- a*b - a - b + 1

Fstar <- ((SSE.R-SSE.C)/(df_E.R-df_E.C)) / (SSE.C/df_E.C)

odofna.out <- cbind(df_E.R, df_E.C, SSE.R, SSE.C, Fstar, qf(.95,1,df_E.C),
                    1-pf(Fstar,1,df_E.C))
colnames(odofna.out) <- c("df(R)", "df(C)", "SSE(R)", "SSE(C)", "F*", "F(.95)", "P(>F*)")
rownames(odofna.out) <- c("ODOFNA")
round(odofna.out, 4)
```

```
##        df(R) df(C) SSE(R) SSE(C)       F* F(.95) P(>F*)
## ODOFNA   119   118 0.9305 0.4075 151.4522 3.9215      0
```

```
plot(biz$Ystar ~ biz$Xstar, pch=16, xlab="X*=alpha*beta", ylab="Y*=Y-mu-alpha-beta")
abline(lm(Ystar ~ Xstar - 1,data=biz), col="blue3", lwd=1.5)
```




```
### biz is sorted by index, then year (stacked COLUMNS from spreadsheet)
(Y.mat <- matrix(biz$Y,byrow=F,ncol=8))
```

```
##       [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]  [,8]
## [1,] 1.019 1.077 0.972 1.041 1.110 0.985 0.972 0.996
## [2,] 0.995 0.990 1.114 0.945 0.996 0.994 1.102 1.091
## [3,] 1.213 1.009 1.179 1.157 1.102 1.057 1.211 1.247
## [4,] 1.012 1.080 1.174 1.085 1.156 1.055 1.132 1.178
## [5,] 1.141 1.022 1.074 1.037 1.055 1.038 1.065 1.109
## [6,] 0.952 1.008 0.965 1.083 1.053 1.010 0.938 0.925
```

```
##  [7,] 1.074 1.004 1.393 0.940 0.994 1.029 1.426 1.497
##  [8,] 1.125 0.997 1.231 1.126 1.095 1.020 1.222 1.301
##  [9,] 1.011 1.015 1.000 1.034 1.055 1.012 1.026 1.034
## [10,] 0.906 1.035 0.833 1.008 1.056 0.993 0.866 0.856
## [11,] 1.099 1.028 1.425 0.930 1.044 1.038 1.366 1.376
## [12,] 1.073 1.008 1.184 1.127 1.086 1.006 1.097 1.094
## [13,] 0.917 1.016 0.896 1.029 1.041 0.924 0.976 0.953
## [14,] 1.154 1.011 1.207 1.099 1.032 1.031 1.240 1.231
## [15,] 1.018 1.012 0.890 1.078 1.072 1.013 0.880 0.872
## [16,] 1.060 1.012 1.169 1.051 1.050 1.038 1.180 1.201
## [17,] 1.073 1.013 1.145 1.064 1.066 1.043 1.143 1.131
## [18,] 1.093 1.017 1.103 1.083 1.086 1.048 1.095 1.099
```

```
tukey.test(Y.mat)
```

```
##
## Tukey test on 5% alpha-level:
##
## Test statistic: 151.5
## Critival value: 3.921
## The additivity hypothesis was rejected.
```

There is strong evidence of an interaction between source and year, of this form.

$$(\alpha\beta)_{ij} = D\alpha_i\beta_j \quad \text{with } \hat{D} = 23.65$$

∇

```
library(tidyverse)
library(kableExtra)
library(effectsize)
library(agricolae)
library(car)
library(PMCMRplus)
library(additivityTests)
```

Chapter 7

Unbalanced Two-Factor Analysis of Variance

When sample sizes are not all equal, the sums of squares cannot be obtained simply among the cell, marginal, and overall means. Even in well planned controlled experiments, observations may not be used due to malfunction or subjects dropping out. In observational studies, it may not be feasible to get equal number of units from the various sub-populations.

The model can be fit based on regression models in scalar and matrix form. In this chapter, we describe the process based on the treatment effects model. The balanced case can be formed in this manner as well.

7.1 Statistical Model and the Analysis of Variance

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \qquad i = 1, \dots, a; \quad j = 1, \dots, b; \quad k = 1, \dots, n_{ij}$$

$$\epsilon_{ijk} \sim NID\left(0, \sigma^2\right)$$

$$n_{i\bullet} = \sum_{j=1}^{b} n_{ij} \quad n_{\bullet j} = \sum_{i=1}^{a} n_{ij} \quad n_T = \sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij} \qquad Y_{ij\bullet} = \sum_{k=1}^{n_{ij}} Y_{ijk} \quad \overline{Y}_{ij\bullet} = \frac{Y_{ij\bullet}}{n_{ij}}$$

$$\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = \sum_{i=1}^{a} (\alpha\beta)_{ij} = \sum_{j=1}^{b} (\alpha\beta)_{ij} = 0$$

$$\alpha_a = -\sum_{i=1}^{a-1} \alpha_i \quad \beta_b = -\sum_{j=1}^{b-1} \beta_j \quad (\alpha\beta)_{aj} = -\sum_{i=1}^{a-1} (\alpha\beta)_{ij} \quad j = 1, \dots, b \quad (\alpha\beta)_{ib} = -\sum_{j=1}^{b-1} (\alpha\beta)_{ij} \quad i = 1, \dots, a$$

To fit this model, construct $a-1$ X variables for the factor A effects, $b-1$ X variables for the factor B effects, and then obtain the $(a-1)(b-1)$ cross-products of these variables for the interaction effects.

$$X_{ijk1}^{A} = \begin{cases} 1 & : \quad i = 1 \\ 0 & : \quad i = 2, \dots, a-1 \\ -1 & : \quad i = a \end{cases} \qquad \cdots \qquad X_{ijk,a-1}^{A} = \begin{cases} 1 & : \quad i = a-1 \\ 0 & : \quad i = 1, \dots, a-2 \\ -1 & : \quad i = a \end{cases}$$

$$X_{ijk1}^{B} = \begin{cases} 1 & : \quad j = 1 \\ 0 & : \quad j = 2, \dots, b-1 \\ -1 & : \quad j = b \end{cases} \qquad \cdots \qquad X_{ijk,b-1}^{B} = \begin{cases} 1 & : \quad j = b-1 \\ 0 & : \quad j = 1, \dots, b-2 \\ -1 & : \quad j = b \end{cases}$$

$$Y_{ijk} = \mu_{\bullet\bullet} + \alpha_1 X_{ijk1}^{A} + \cdots + \alpha_{a-1} X_{ijk,a-1}^{A} + \beta_1 X_{ijk1}^{B} + \cdots + \beta_{b-1} X_{ijk,b-1}^{B} +$$

Table 7.1: Author Age at Peak Data

author	Type	Style	ageDth	cageDth	agePeak
Eliot	0	0	77	8.57	23
Cummings	0	0	68	-0.43	26
Plath	0	0	31	-37.43	30
Pound	0	0	87	18.57	30
Wilber	0	0	86	17.57	34
Williams	0	1	80	11.57	40
Bishop	0	1	68	-0.43	29
Moore	0	1	85	16.57	32
Lowell	0	1	60	-8.43	41
Stevens	0	1	76	7.57	42
Frost	0	1	89	20.57	48
Fitzgerald	1	0	44	-24.43	29
Hemingway	1	0	62	-6.43	30
Melville	1	0	72	3.57	32
Lawrence	1	0	45	-23.43	35
Joyce	1	0	59	-9.43	40
James	1	1	73	4.57	38
Faulkner	1	1	65	-3.43	39
Dickens	1	1	58	-10.43	41
Woolf	1	1	59	-9.43	45
Conrad	1	1	67	-1.43	47
Twain	1	1	75	6.57	50
Hardy	1	1	88	19.57	51

$$+(\alpha\beta)_{11}X^A_{ijk1}X^B_{ijk1} + \cdots + (\alpha\beta)_{a-1,b-1}X^A_{ijk,a-1}X^B_{ijk,b-1} + \epsilon_{ijk}$$

The testing strategy is done as it was in the balanced case. Fit various models and compare them using the general linear test procedure.

- **Model 1** - Include main effects for factors A and B as well as interaction effects.
- **Model 2** - Include main effects for factors A and B, but no interaction effects
- **Model 3** - Include main effects for factor B and interaction effects, but no factor A effects
- **Model 4** - Include main effects for factor A and interaction effects, but no factor B effects

To test for interaction effects, controlling for main effects for factors A and B, we compare Model 1 (Complete) and Model 2 (Reduced).

To test for factor A effects, controlling for factor B effects and interaction effects, we compare Model 1 (Complete) and Model 3 (Reduced).

To test for factor B effects, controlling for factor A effects and interaction effects, we compare Model 1 (Complete) and Model 4 (Reduced).

When the test for interaction is "very non-significant" (high P-value), often tests for factor A and B effects use Model 2 as the Complete Model.

Example 7.1 - Age at Peak for Famous Writers

A study considered the age at career peak for famous writers [Simonson, 2007]. The factors were **Style** (Factor A: Conceptualists ($i = 1$), Experimentalists ($i = 2$)) and Writer **Type** (Factor B:Poets ($j = 1$), Novelists ($j = 2$)). Thus, there are $a = b = 2$ levels for the factors, there will only be $a - 1 = 1$ X variable for factor A and $b - 1 = 1$ X variable for factor B, and $(a - 1)(b - 1) = 1$ cross-product term for the AB interaction. The numbers of writers for the $ab = 4$ conditions are $n_{11} = n_{12} = 5, n_{21} = 6, n_{22} = 7$. The data and the generated X variables are given in Table 7.1. The regression model is fit directly, then with the **aov** function.

```
##
```

```
## Call:
## lm(formula = agePeak ~ X1A + X1B + X1AX1B, data = ap1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.667 -3.814  1.333  2.952  9.333
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  36.2238     1.1402  31.769  < 2e-16 ***
## X1A           5.3238     1.1402   4.669 0.000167 ***
## X1B           2.5905     1.1402   2.272 0.034902 *
## X1AX1B        0.2905     1.1402   0.255 0.801652
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.415 on 19 degrees of freedom
## Multiple R-squared:  0.5978, Adjusted R-squared:  0.5343
## F-statistic: 9.413 on 3 and 19 DF,  p-value: 0.0005033
```

```
##
## Call:
## aov(formula = agePeak ~ factor(Style) + factor(Type) + factor(Style) *
##     factor(Type), data = ap1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.667 -3.814  1.333  2.952  9.333
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   36.2238     1.1402  31.769  < 2e-16 ***
## factor(Style)1                -5.3238     1.1402  -4.669 0.000167 ***
## factor(Type)1                 -2.5905     1.1402  -2.272 0.034902 *
## factor(Style)1:factor(Type)1   0.2905     1.1402   0.255 0.801652
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.415 on 19 degrees of freedom
## Multiple R-squared:  0.5978, Adjusted R-squared:  0.5343
## F-statistic: 9.413 on 3 and 19 DF,  p-value: 0.0005033
```

```
## Analysis of Variance Table
##
## Response: agePeak
##                            Df Sum Sq Mean Sq F value    Pr(>F)
## factor(Style)               1 667.75  667.75 22.7758 0.0001324 ***
## factor(Type)                1 158.26  158.26  5.3979 0.0314109 *
## factor(Style):factor(Type)  1   1.90    1.90  0.0649 0.8016516
## Residuals                  19 557.05   29.32
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We see from the regression coefficients (that control for all other factors) that there is no evidence of an interaction, but there is evidence of main effects for Style and Writer Type. Formal tests are given below. Given that each test has 1 degree of freedom, we use t-tests.

$$H_0^{AB}: (\alpha\beta)_{11} = (\alpha\beta)_{12} = (\alpha\beta)_{21} = (\alpha\beta)_{22} = 0 \qquad t^*_{AB} = \frac{0.2905}{1.1402} = 0.255 \qquad P_{AB} = .8017$$

$$H_0^{A}: \alpha_1 = \alpha_2 = 0 \qquad t^*_{A} = \frac{-5.3238}{1.1402} = -4.669 \qquad P_{A} = .0002$$

$$H_0^{B}: \beta_1 = \beta_2 = 0 \qquad t^*_{B} = \frac{-2.5905}{1.1402} = -2.272 \qquad P_{B} = .0349$$

As there are strong evidence of main effects and no suggestion of an interaction, we will fit Model 2 for making comparisons, but not fit Models 3 or 4.

```
##
## Call:
## lm(formula = agePeak ~ X1A + X1B, data = ap1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.940 -4.027  1.060  2.933  9.060
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   36.234      1.113  32.566  < 2e-16 ***
## X1A            5.334      1.113   4.794 0.000111 ***
## X1B            2.628      1.104   2.380 0.027397 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.287 on 20 degrees of freedom
## Multiple R-squared:  0.5964, Adjusted R-squared:  0.5561
## F-statistic: 14.78 on 2 and 20 DF,  p-value: 0.0001146
```

```
##
## Call:
## aov(formula = agePeak ~ factor(Style) + factor(Type), data = ap1)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.940 -4.027  1.060  2.933  9.060
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      36.234      1.113  32.566  < 2e-16 ***
## factor(Style)1   -5.334      1.113  -4.794 0.000111 ***
## factor(Type)1    -2.628      1.104  -2.380 0.027397 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.287 on 20 degrees of freedom
## Multiple R-squared:  0.5964, Adjusted R-squared:  0.5561
## F-statistic: 14.78 on 2 and 20 DF,  p-value: 0.0001146
```

```
## Analysis of Variance Table
##
## Response: agePeak
##               Df Sum Sq Mean Sq F value    Pr(>F)
## factor(Style)  1 667.75  667.75 23.8930 8.894e-05 ***
## factor(Type)   1 158.26  158.26  5.6627    0.0274 *
## Residuals     20 558.95   27.95
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the additive model (Model 2), we obtain the following estimates and predicted ages at peak.

$$\hat{\mu}_{\bullet\bullet} = 36.234 \qquad \hat{\alpha}_1 = -5.334 \quad \hat{\alpha}_2 = 5.334 \qquad \hat{\beta}_1 = -2.628 \quad \hat{\beta}_2 = 2.628$$

$$\text{Conceptualists/Poets:} \quad \hat{Y}_{11} = \hat{\mu}_{\bullet\bullet} + \hat{\alpha}_1 + \hat{\beta}_1 = 36.234 - 5.334 - 2.628 = 28.272$$

$$\text{Conceptualists/Novelists:} \quad \hat{Y}_{12} = \hat{\mu}_{\bullet\bullet} + \hat{\alpha}_1 + \hat{\beta}_2 = 36.234 - 5.334 + 2.628 = 33.528$$

$$\text{Experimantalists/Poets:} \quad \hat{Y}_{21} = \hat{\mu}_{\bullet\bullet} + \hat{\alpha}_2 + \hat{\beta}_1 = 36.234 + 5.334 - 2.628 = 38.940$$

$$\text{Experimantalists/Novelists:} \quad \hat{Y}_{21} = \hat{\mu}_{\bullet\bullet} + \hat{\alpha}_2 + \hat{\beta}_2 = 36.234 + 5.334 + 2.628 = 44.196$$

Experimentalists take longer on average to reach their peak than conceptualists, and novelists take longer than poets.

∇

A second example with more levels of factor B is included here.

Example 7.2 - Makiwara Punching Boards

An experiment was conducted to compare two types of karate boards, made from 4 wood types [Smith et al., 2010]. The board types were stacked ($i = 1$) and tapered ($i = 2$), the wood types were: cherry ($j = 1$), ash ($j = 2$), fir ($j = 3$), and oak ($j = 4$). Apparently due to breakage, there were unequal numbers of replicates among the $ab = 2(4) = 8$ treatments. We use the **aov** function in R with the treatment effects form to fit Models 1 (Interaction) and 2 (Additive) to test whether there is an interaction between board and wood types. The response Y is the deflection (millimeters) for the Makiwara board, with a total of $n_T = 336$ measurements.

```
##   trt.id wood board id deflect
## 1      1    1     1  1   144.3
## 2      1    1     1  2   125.9
## 3      1    1     1  3   263.2
## 4      1    1     1  4   114.6
## 5      1    1     1  5   242.5
## 6      1    1     1  6   141.9
```

```
##     trt.id wood board id deflect
## 331      8    4     2 40    56.6
## 332      8    4     2 41   123.5
## 333      8    4     2 42    12.0
## 334      8    4     2 43    62.0
## 335      8    4     2 44    73.3
## 336      8    4     2 45    44.9
```

```
##            cherry      ash      fir      oak
## stacked 118.00000 105.9944 93.99706 89.99556
## tapered  77.99556  76.0000 55.00000 49.99333
```

```
##
## Call:
## aov(formula = deflect ~ board.f * wood.f, data = mb1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -97.600 -35.775  -7.246  25.705 281.104
```

```
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)       83.3720     3.0838  27.035  < 2e-16 ***
## board.f1          18.6248     3.0838   6.039 4.19e-09 ***
## wood.f1           14.6258     5.2834   2.768  0.00596 **
## wood.f2            7.6252     5.4085   1.410  0.15953
## wood.f3           -8.8735     5.4678  -1.623  0.10558
## board.f1:wood.f1   1.3775     5.2834   0.261  0.79448
## board.f1:wood.f2  -3.6275     5.4085  -0.671  0.50288
## board.f1:wood.f3   0.8738     5.4678   0.160  0.87314
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 56.2 on 328 degrees of freedom
## Multiple R-squared:  0.1359, Adjusted R-squared:  0.1174
## F-statistic: 7.366 on 7 and 328 DF,  p-value: 3.178e-08
```

```
## Analysis of Variance Table
##
## Response: deflect
##                 Df  Sum Sq Mean Sq F value   Pr(>F)
## board.f          1  115479  115479 36.5595 4.03e-09 ***
## wood.f           3   45955   15318  4.8496 0.002575 **
## board.f:wood.f   3    1444     481  0.1524 0.928115
## Residuals      328 1036045    3159
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Call:
## aov(formula = deflect ~ board.f + wood.f, data = mb1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -96.220 -35.852  -7.682  26.736 282.422
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   83.431      3.067  27.205  < 2e-16 ***
## board.f1      18.683      3.067   6.092 3.09e-09 ***
## wood.f1       14.506      5.252   2.762  0.00607 **
## wood.f2        7.976      5.359   1.488  0.13757
## wood.f3       -9.046      5.407  -1.673  0.09525 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 55.99 on 331 degrees of freedom
## Multiple R-squared:  0.1346, Adjusted R-squared:  0.1242
## F-statistic: 12.88 on 4 and 331 DF,  p-value: 9.383e-10
```

```
## Analysis of Variance Table
##
## Response: deflect
##            Df  Sum Sq Mean Sq F value    Pr(>F)
## board.f     1  115479  115479 36.8425 3.506e-09 ***
## wood.f      3   45955   15318  4.8872  0.002445 **
## Residuals 331 1037489    3134
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

When comparing the mean deflection differences (stacked-tapered) within the 4 wood types, based on the **tapply** function, we observe: 40 (cherry), 30 (ash), 39 (fir), and 41 (oak). While these differences are not exactly the same, they are very consistent. To test for interaction, we directly use the ANOVA table from Model 1.

$$H_0 : (\alpha\beta)_{11} = \cdots = (\alpha\beta)_{42} = 0 \qquad H_A : \text{ Not all } (\alpha\beta)_{ij} = 0$$

$$\text{Test Statistic:} \quad F_{AB}^* = \frac{MSAB}{MSE} = \frac{481}{3159} = 0.1524 \qquad \text{Rejection Region:} \quad F_{AB}^* \geq F_{.95;3,328} = 2.632$$

The P-value is large, $P_{AB} = .9281$, giving no evidence of any interaction effects.

Based on this, we test for main effects for Board and Wood Effects, based on Model 2, the additive model.

For Board effects, $H_0^A : \alpha_1 = \alpha_2 = 0$, we obtain the following test.

$$\text{Test Statistic:} \quad F_A^* = \frac{MSA}{MSE} = \frac{115479}{3134} = 36.8425 \qquad \text{Rejection Region:} \quad F_A^* \geq F_{.95;1,331} = 3.870$$

For Wood effects, $H_0^B : \beta_1 = \cdots = \beta_4 = 0$, we obtain the following test.

$$\text{Test Statistic:} \quad F_B^* = \frac{MSB}{MSE} = \frac{15318}{3134} = 4.8872 \qquad \text{Rejection Region:} \quad F_B^* \geq F_{.95;3,331} = 2.632$$

Both results are significant, with $P_A < .0001$ and $P_B = .0024$.

∇

7.2 Least Squares Estimators and Contrasts/Linear Functions Among Means

In this section, the formulas for estimators and estimated standard errors for means and linear functions are given. For treatment (cell) means, we have the following results.

$$\text{Parameter:} \quad \mu_{ij} \qquad \text{Estimator:} \quad \hat{\mu}_{ij} = \frac{\sum_{k=1}^{n_{ij}} Y_{ijk}}{n_{ij}} = \overline{Y}_{ij\bullet}$$

$$\text{Estimated Standard Error:} \quad s\left\{\hat{\mu}_{ij}\right\} = \sqrt{\frac{MSE}{n_{ij}}}$$

For factor A level means, we have the following results.

$$\text{Parameter:} \quad \mu_{i\bullet} = \frac{\sum_{j=1}^{b} \mu_{ij}}{b} \qquad \text{Estimator:} \quad \hat{\mu}_{i\bullet} = \frac{\sum_{j=1}^{b} \overline{Y}_{ij\bullet}}{b}$$

$$\text{Estimated Standard Error:} \quad s\left\{\hat{\mu}_{i\bullet}\right\} = \sqrt{\frac{MSE}{b^2}\sum_{j=1}^{b}\frac{1}{n_{ij}}}$$

For factor B level means, we have the following results.

$$\text{Parameter:} \quad \mu_{\bullet j} = \frac{\sum_{i=1}^{a} \mu_{ij}}{a} \qquad \text{Estimator:} \quad \hat{\mu}_{\bullet j} = \frac{\sum_{i=1}^{a} \overline{Y}_{ij\bullet}}{a}$$

$$\text{Estimated Standard Error:} \quad s\left\{\hat{\mu}_{\bullet j}\right\} = \sqrt{\frac{MSE}{a^2}\sum_{i=1}^{a}\frac{1}{n_{ij}}}$$

For contrasts or linear functions among factor A and B levels, we have the following parameters, estimators, and estimated standard errors.

$$L_A = \sum_{i=1}^{a} c_i \mu_{i\bullet} \qquad \hat{L}_A = \sum_{i=1}^{a} c_i \hat{\mu}_{i\bullet} \qquad s\left\{\hat{L}_A\right\} = \sqrt{\frac{MSE}{b^2} \sum_{i=1}^{a} c_i^2 \sum_{j=1}^{b} \frac{1}{n_{ij}}}$$

$$L_B = \sum_{j=1}^{b} c_j \mu_{\bullet j} \qquad \hat{L}_B = \sum_{j=1}^{b} c_j \hat{\mu}_{\bullet j} \qquad s\left\{\hat{L}_B\right\} = \sqrt{\frac{MSE}{a^2} \sum_{j=1}^{b} c_j^2 \sum_{i=1}^{a} \frac{1}{n_{ij}}}$$

Finally, the results are given for contrasts and linear functions among treatment (cell) means.

$$L_{AB} = \sum_{i=1}^{a} \sum_{j=1}^{b} c_{ij} \mu_{ij} \qquad \hat{L}_{AB} = \sum_{i=1}^{a} \sum_{j=1}^{b} c_{ij} \hat{\mu}_{ij} \qquad s\left\{\hat{L}_{AB}\right\} = \sqrt{MSE \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{c_{ij}^2}{n_{ij}}}$$

The standard error multipliers for contrasts/linear functions based on the Scheffe (all comparisons), Bonferroni (g pre-planned comparisons), and Tukey (all pairwise comparisons) methods are given here.

For single comparisons, use $t_{1-\alpha/2;n_T-ab}$.

For comparisons among treatment (cell) means, use the following multipliers.

$$\text{Scheffe:} \quad S = \sqrt{(ab-1)F_{1-\alpha;ab-1,n_T-ab}}$$

$$\text{Bonferroni:} \quad B = t_{1-\alpha/(2g);n_T-ab}$$

$$\text{Tukey:} \quad T = \frac{1}{\sqrt{2}} q_{1-\alpha;ab,n_T-ab}$$

Standard error multipliers for contrasts and linear functions among levels of factors A and B are given here.

$$\text{Scheffe:} \qquad \text{A:} \quad S_A = \sqrt{(a-1)F_{1-\alpha;a-1,n_T-ab}} \qquad \text{B:} \quad S_B = \sqrt{(b-1)F_{1-\alpha;b-1,n_T-ab}}$$

$$\text{Bonferroni (Factors A and B):} \quad B = t_{1-\alpha/(2g);n_T-ab}$$

$$\text{Tukey:} \qquad \text{A:} \quad T_A = \frac{1}{\sqrt{2}} q_{1-\alpha;a,n_T-ab} \qquad \text{B:} \quad T_B = \frac{1}{\sqrt{2}} q_{1-\alpha;b,n_T-ab}$$

Example 7.3 - Age at Peak for Famous Writers

Although the interaction was not significant for the Age at Peak of the writers, we will use the Tukey method to compare all pairs of Style/Type means. The Mean Square Error for the interaction model was $MSE = 29.32$ with degrees of freedom $n_T - ab = 23 - 2(2) = 19$.

$$\overline{y}_{11\bullet} = 28.60 \qquad n_{11} = 5 \qquad s\left\{\overline{Y}_{11\bullet}\right\} = \sqrt{\frac{MSE}{n_{11}}} = \sqrt{\frac{29.32}{5}} = 2.42$$

$$\overline{y}_{12\bullet} = 33.20 \qquad n_{12} = 5 \qquad s\left\{\overline{Y}_{12\bullet}\right\} = \sqrt{\frac{MSE}{n_{12}}} = \sqrt{\frac{29.32}{5}} = 2.42$$

$$\overline{y}_{21\bullet} = 38.67 \qquad n_{21} = 6 \qquad s\left\{\overline{Y}_{21\bullet}\right\} = \sqrt{\frac{MSE}{n_{21}}} = \sqrt{\frac{29.32}{6}} = 2.21$$

$$\overline{y}_{22\bullet} = 44.43 \qquad n_{22} = 7 \qquad s\left\{\overline{Y}_{22\bullet}\right\} = \sqrt{\frac{MSE}{n_{22}}} = \sqrt{\frac{29.32}{7}} = 2.05$$

$$T = \frac{1}{\sqrt{2}} q_{.95;4,19} = \frac{1}{\sqrt{2}}(3.977) = 2.812 \qquad s\left\{\overline{Y}_{ij\bullet} - \overline{Y}_{i'j'\bullet}\right\} = \sqrt{MSE\left(\frac{1}{n_{ij}} + \frac{1}{n_{i'j'}}\right)}$$

$$\overline{y}_{11\bullet} - \overline{y}_{12\bullet} = 28.60 - 33.20 = -4.60 \qquad s\left\{\overline{Y}_{11\bullet} - \overline{Y}_{12\bullet}\right\} = \sqrt{29.32\left(\frac{1}{5} + \frac{1}{5}\right)} = 3.43$$

$$\overline{y}_{11\bullet} - \overline{y}_{21\bullet} = 28.60 - 38.67 = -10.07 \qquad s\left\{\overline{Y}_{11\bullet} - \overline{Y}_{21\bullet}\right\} = \sqrt{29.32\left(\frac{1}{5} + \frac{1}{6}\right)} = 3.28$$

$$\overline{y}_{11\bullet} - \overline{y}_{22\bullet} = 28.60 - 44.43 = -15.83 \qquad s\left\{\overline{Y}_{11\bullet} - \overline{Y}_{22\bullet}\right\} = \sqrt{29.32\left(\frac{1}{5} + \frac{1}{7}\right)} = 3.17$$

$$\overline{y}_{12\bullet} - \overline{y}_{21\bullet} = 33.20 - 38.67 = -5.47 \qquad s\left\{\overline{Y}_{12\bullet} - \overline{Y}_{21\bullet}\right\} = \sqrt{29.32\left(\frac{1}{5} + \frac{1}{6}\right)} = 3.28$$

$$\overline{y}_{12\bullet} - \overline{y}_{22\bullet} = 33.20 - 44.43 = -11.23 \qquad s\left\{\overline{Y}_{12\bullet} - \overline{Y}_{22\bullet}\right\} = \sqrt{29.32\left(\frac{1}{5} + \frac{1}{7}\right)} = 3.17$$

$$\overline{y}_{21\bullet} - \overline{y}_{22\bullet} = 38.67 - 44.43 = -5.76 \qquad s\left\{\overline{Y}_{21\bullet} - \overline{Y}_{22\bullet}\right\} = \sqrt{29.32\left(\frac{1}{6} + \frac{1}{7}\right)} = 3.01$$

Conc/Poet-Conc/Novel: $HSD = 2.812(3.43) = 9.65 \qquad \overline{y}_{11\bullet} - \overline{y}_{12\bullet} = -4.60$

Conc/Poet-Exp/Poet: $HSD = 2.812(3.28) = 9.22 \qquad \overline{y}_{11\bullet} - \overline{y}_{21\bullet} = 28.60 - 38.67 = -10.07$

Conc/Poet-Exp/Novel: $HSD = 2.812(3.17) = 8.92 \qquad \overline{y}_{11\bullet} - \overline{y}_{22\bullet} = -15.83$

Conc/Nov-Exp/Poet: $HSD = 2.812(3.28) = 9.22 \qquad \overline{y}_{12\bullet} - \overline{y}_{21\bullet} = -5.47$

Conc/Nov-Exp/Nov: $HSD = 2.812(3.17) = 8.92 \qquad \overline{y}_{12\bullet} - \overline{y}_{22\bullet} = -11.23$

Exp/Poet-Exp/Nov: $HSD = 2.812(3.01) = 8.47 \qquad \overline{y}_{21\bullet} - \overline{y}_{22\bullet} = -5.76$

The groups that are significantly different, based on Tukey's HSD are as follow.

- Conceptualist/Poet is significantly lower than Experimentalist/Poet
- Conceptualist/Poet is significantly lower than Experimentalist/Novelist
- Conceptualist/Novelist is significantly lower than Experimentalist/Novelist

The following R program runs Tukey's HSD on the marginal means for factors A and B, and on the cell means as well.

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = agePeak ~ factor(Style) + factor(Type) + factor(Style) * factor(Type), data = ap1)
##
## $`factor(Style)`
##         diff      lwr      upr     p adj
## 1-0 10.86923 6.102333 15.63613 0.0001324
##
## $`factor(Type)`
##         diff       lwr      upr     p adj
## 1-0 5.247378 0.5167309 9.978024 0.0315205
##
## $`factor(Style):factor(Type)`
##              diff         lwr       upr     p adj
## 1:0-0:0 10.066667   0.8473951 19.285938 0.0293965
## 0:1-0:0  4.600000  -5.0292152 14.229215 0.5481270
## 1:1-0:0 15.828571   6.9136505 24.743492 0.0004327
## 0:1-1:0 -5.466667 -14.6859383  3.752605 0.3672669
## 1:1-1:0  5.761905  -2.7085734 14.232383 0.2559999
## 1:1-0:1 11.228571   2.3136505 20.143492 0.0107243
```

$$\nabla$$

Example 7.4 - Makiwara Punching Boards

For the Karate board experiment, we will conduct tests among levels of factors A and B, respectively. The Mean Square Error was $MSE = 3159$ with degrees of freedom $n_T - ab = 336 - 2(4) = 328$ for Model 1 (with interaction).

For comparing the Board Types (Stacked-Tapered), we obtain the following cell means from the **tapply** function previously.

$$\overline{y}_{11\bullet} = 118 \quad \overline{y}_{12\bullet} = 106 \quad \overline{y}_{13\bullet} = 94 \quad \overline{y}_{14\bullet} = 90 \quad \Rightarrow \quad \hat{\mu}_{1\bullet} = \frac{118+106+94+90}{4} = 102$$

$$\overline{y}_{21\bullet} = 78 \quad \overline{y}_{22\bullet} = 76 \quad \overline{y}_{23\bullet} = 55 \quad \overline{y}_{24\bullet} = 50 \quad \Rightarrow \quad \hat{\mu}_{2\bullet} = \frac{78+76+55+50}{4} = 64.75$$

The standard errors are computed here, where the sample sizes for each cell are given first.

$$n_{11} = 41 \quad n_{12} = 36 \quad n_{13} = 34 \quad n_{14} = n_{21} = n_{22} = n_{23} = n_{24} = 45$$

$$s\left\{\hat{\mu}_{1\bullet}\right\} = \sqrt{\frac{3159}{4^2}\left(\frac{1}{41}+\frac{1}{36}+\frac{1}{34}+\frac{1}{45}\right)} = 4.53$$

$$s\left\{\hat{\mu}_{2\bullet}\right\} = \sqrt{\frac{3159}{4^2}\left(\frac{1}{45}+\frac{1}{45}+\frac{1}{45}+\frac{1}{45}\right)} = 4.19$$

The standard error of the difference $\hat{\mu}_{1\bullet} - \hat{\mu}_{2\bullet}$ is the square root of the sum of their variances, as these are independent samples. Also, we obtain the critical t-value, and the 95% Confidence Interval for $\mu_{1\bullet} - \mu_{2\bullet}$.

$$s\left\{\hat{\mu}_{1\bullet} - \hat{\mu}_{2\bullet}\right\} = \sqrt{(4.53)^2 + (4.19)^2} = 6.17 \qquad t_{.975;328} = 1.967$$

$$\text{95\% Confidence Interval:} \quad (102 - 64.75) \pm 1.967(6.17) \quad \equiv \quad 37.25 \pm 12.14 \quad \equiv \quad (25.11, 49.39)$$

Thus, the mean deflection is higher for the stacked board than the tapered on average (likely by somewhere between 25-50 millimeters). We could also compare all pairs of Wood Types. The computations are similar, but there are $4(3)/2 = 6$ possible pairs. We will use the **TukeyHSD** function along with the **aov** object in R below.

```
##   trt.id wood board id deflect
## 1      1    1     1  1   144.3
## 2      1    1     1  2   125.9
## 3      1    1     1  3   263.2
## 4      1    1     1  4   114.6
## 5      1    1     1  5   242.5
## 6      1    1     1  6   141.9

##     trt.id wood board id deflect
## 331      8    4     2 40    56.6
## 332      8    4     2 41   123.5
## 333      8    4     2 42    12.0
## 334      8    4     2 43    62.0
## 335      8    4     2 44    73.3
## 336      8    4     2 45    44.9

##            cherry      ash      fir      oak
## stacked 118.00000 105.9944 93.99706 89.99556
## tapered  77.99556  76.0000 55.00000 49.99333
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = deflect ~ board.f * wood.f, data = mb1)
##
## $board.f
##                      diff       lwr       upr p adj
## tapered-stacked -37.17265 -49.26685 -25.07845     0
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = deflect ~ board.f * wood.f, data = mb1)
##
## $wood.f
##                  diff       lwr        upr     p adj
## ash-cherry  -6.535911 -29.00664 15.9348190 0.8761668
## fir-cherry -23.560409 -46.17714 -0.9436753 0.0375052
## oak-cherry -27.937478 -49.82197 -6.0529827 0.0059563
## fir-ash    -17.024498 -39.97301  5.9240184 0.2234297
## oak-ash    -21.401567 -43.62878  0.8256437 0.0639197
## oak-fir     -4.377069 -26.75187 17.9977342 0.9578232
```

In terms of the Wood Types, fir has significantly lower deflection than cherry, and oak has significantly lower deflection than cherry.

∇

```
library(tidyverse)
library(kableExtra)
library(effectsize)
library(agricolae)
library(car)
library(PMCMRplus)
library(additivityTests)
```

Chapter 8

Multi-Factor Studies

Many experiments include three or more factors. Main effects and two-way interactions have the same interpretations as in two factor studies. A three-way interaction is when the patterns of the two-way interactions between two factors differ at different levels of the other factor(s).

In screening experiments, there can be any number of factors, typically with two or three levels each. The goal in these experiments is to determine which factors have the larger effects, then conduct new experiments focusing on these factors.

8.1 Three Factor Models - Mean Structure

In this section, we consider balanced three factor models, with labels A, B, and C, and with numbers of levels a, b, and c, respectively. The number of replicates per treatment is n, with an overall sample size of $n_T = abcn$. The same adjustments can be made for unbalanced data as was described in the previous chapter.

The population mean when factor A is at level i, B is at level j, and C is at level k is μ_{ijk}. The marginal and overall means are given below.

$$\mu_{ij\bullet} = \frac{\sum_{k=1}^{c} \mu_{ijk}}{c} \qquad \mu_{i\bullet k} = \frac{\sum_{j=1}^{b} \mu_{ijk}}{b} \qquad \mu_{\bullet jk} = \frac{\sum_{i=1}^{a} \mu_{ijk}}{a}$$

$$\mu_{i\bullet\bullet} = \frac{\sum_{j=1}^{b}\sum_{k=1}^{c} \mu_{ijk}}{bc} \qquad \mu_{\bullet j\bullet} = \frac{\sum_{i=1}^{a}\sum_{k=1}^{c} \mu_{ijk}}{ac} \qquad \mu_{\bullet\bullet k} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b} \mu_{ijk}}{ab}$$

$$\mu_{\bullet\bullet\bullet} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} \mu_{ijk}}{abc}$$

Main effects and interactions are defined below.

$$\alpha_i = \mu_{i\bullet\bullet} - \mu_{\bullet\bullet\bullet} \qquad \beta_j = \mu_{\bullet j\bullet} - \mu_{\bullet\bullet\bullet} \qquad \gamma_k = \mu_{\bullet\bullet k} - \mu_{\bullet\bullet\bullet}$$

$$(\alpha\beta)_{ij} = \mu_{ij\bullet} - \mu_{i\bullet\bullet} - \mu_{\bullet j\bullet} + \mu_{\bullet\bullet\bullet} \qquad (\alpha\gamma)_{ik} = \mu_{i\bullet k} - \mu_{i\bullet\bullet} - \mu_{\bullet\bullet k} + \mu_{\bullet\bullet\bullet}$$

$$(\beta\gamma)_{jk} = \mu_{\bullet jk} - \mu_{\bullet j\bullet} - \mu_{\bullet\bullet k} + \mu_{\bullet\bullet\bullet}$$

$$(\alpha\beta\gamma)_{ijk} = \mu_{ijk} - \mu_{ij\bullet} - \mu_{i\bullet k} - \mu_{\bullet jk} + \mu_{i\bullet\bullet} + \mu_{\bullet j\bullet} + \mu_{\bullet\bullet k} - \mu_{\bullet\bullet\bullet}$$

The algorithm for obtaining the main effects and interaction effects goes as follows.

- Begin with the mean corresponding to the effect.
- Cover each subscript, one-at-a-time and multiply by $(-1)^1 = -1$.
- Cover subscripts two-at-a-time and multiply by $(-1)^2 = 1$.
- Continue until all subscripts have been covered.

- Coefficients ± 1 will sum to 0.

The following constraints are obtained for the effects and their estimators.

$$\sum_{i=1}^{a}\alpha_i = \sum_{j=1}^{b}\beta_j = \sum_{k=1}^{c}\gamma_k = 0$$

$$\sum_{i=1}^{a}(\alpha\beta)_{ij} = \sum_{j=1}^{b}(\alpha\beta)_{ij} = \sum_{i=1}^{a}(\alpha\gamma)_{ik} = \sum_{k=1}^{c}(\alpha\gamma)_{ij} = \sum_{j=1}^{b}(\beta\gamma)_{jk} = \sum_{k=1}^{c}(\beta\gamma)_{jk} = 0$$

$$\sum_{i=1}^{a}(\alpha\beta\gamma)_{ijk} = \sum_{j=1}^{b}(\alpha\beta\gamma)_{ijk} = \sum_{k=1}^{c}(\alpha\beta\gamma)_{ijk} = 0$$

Example 8.1 - Flavonoids in Wine

The following "population means" are based on observed experimental values to apply the formulas above and have been adjusted to have integer values for all effects [Genova et al., 2012]. There were three factors, each at two levels, described below and the response was total flavonoids.

- Factor A - **Grape Type** Sangiovese (Red, $i = 1$) and Muscat (White, $i = 2$)
- Factor B - **Storage Temperature** 4C ($j = 1$) and -20C ($j = 2$)
- Factor C - **Storage Time** 24 hours (1 day, $k = 1$) and 2 weeks (14 days, $k = 2$)

The means are given below.

$$\text{Sangiovese:} \quad \mu_{111} = 78 \quad \mu_{112} = 68 \quad \mu_{121} = 84 \quad \mu_{122} = 90$$

$$\text{Muscat:} \quad \mu_{211} = 22 \quad \mu_{212} = 12 \quad \mu_{221} = 24 \quad \mu_{222} = 22$$

$$\mu_{11\bullet} = \frac{78+68}{2} = 73 \qquad \mu_{12\bullet} = \frac{84+90}{2} = 87$$

$$\mu_{1\bullet 1} = \frac{78+84}{2} = 81 \qquad \mu_{1\bullet 2} = \frac{68+90}{2} = 79$$

$$\mu_{21\bullet} = \frac{22+12}{2} = 17 \qquad \mu_{22\bullet} = \frac{24+22}{2} = 23$$

$$\mu_{2\bullet 1} = \frac{22+24}{2} = 23 \qquad \mu_{2\bullet 2} = \frac{12+22}{2} = 17$$

$$\mu_{\bullet 11} = \frac{78+22}{2} = 50 \qquad \mu_{\bullet 12} = \frac{68+12}{2} = 40$$

$$\mu_{\bullet 21} = \frac{84+24}{2} = 54 \qquad \mu_{\bullet 22} = \frac{90+22}{2} = 56$$

$$\mu_{1\bullet\bullet} = \frac{78+68+84+90}{4} = 80 \quad \mu_{2\bullet\bullet} = \frac{22+12+24+22}{4} = 20$$

$$\mu_{\bullet 1\bullet} = \frac{78+68+22+12}{4} = 45 \quad \mu_{\bullet 2\bullet} = \frac{84+90+24+22}{4} = 55$$

$$\mu_{\bullet\bullet 1} = \frac{78+84+22+24}{4} = 52 \quad \mu_{\bullet\bullet 2} = \frac{68+90+12+22}{4} = 48$$

$$\mu_{\bullet\bullet\bullet} = \frac{78+68+84+90+22+12+24+22}{8} = 50$$

The main effects and interaction effects are obtained below.

$$\text{Grape:} \quad \alpha_1 = \mu_{1\bullet\bullet} - \mu_{\bullet\bullet\bullet} = 80 - 50 = 30 \qquad \alpha_2 = 20 - 50 = -30$$

$$\text{Temperature:} \quad \beta_1 = \mu_{\bullet 1\bullet} - \mu_{\bullet\bullet\bullet} = 45 - 50 = -5 \qquad \beta_2 = 55 - 50 = 5$$

$$\text{Time:} \quad \gamma_1 = \mu_{\bullet\bullet 1} - \mu_{\bullet\bullet\bullet} = 52 - 50 = 2 \qquad \gamma_2 = 48 - 50 = -2$$

$$\text{Grape/Temp:} \quad (\alpha\beta)_{11} = \mu_{11\bullet} - \mu_{1\bullet\bullet} - \mu_{\bullet 1\bullet} + \mu_{\bullet\bullet\bullet} = 73 - 80 - 45 + 50 = -2$$
$$(\alpha\beta)_{12} = 2 \qquad (\alpha\beta)_{21} = 2 \qquad (\alpha\beta)_{22} = -2$$
$$\text{Grape/Time:} \quad (\alpha\gamma)_{11} = \mu_{1\bullet 1} - \mu_{1\bullet\bullet} - \mu_{\bullet\bullet 1} + \mu_{\bullet\bullet\bullet} = 81 - 80 - 52 + 50 = -1$$
$$(\alpha\gamma)_{12} = 1 \qquad (\alpha\gamma)_{21} = 1 \qquad (\alpha\gamma)_{22} = -1$$
$$\text{Temp/Time:} \quad (\beta\gamma)_{11} = \mu_{\bullet 11} - \mu_{\bullet 1\bullet} - \mu_{\bullet\bullet 1} + \mu_{\bullet\bullet\bullet} = 50 - 45 - 52 + 50 = 3$$
$$(\beta\gamma)_{12} = -3 \qquad (\beta\gamma)_{21} = -3 \qquad (\beta\gamma)_{22} = 3$$
$$(\alpha\beta_\gamma)_{111} = 78 - 73 - 81 - 50 + 80 + 45 + 52 - 50 = 1$$
$$(\alpha\beta_\gamma)_{112} = (\alpha\beta_\gamma)_{121} = (\alpha\beta_\gamma)_{211} = (\alpha\beta_\gamma)_{222} = -1 \quad (\alpha\beta_\gamma)_{122} = (\alpha\beta_\gamma)_{212} = (\alpha\beta_\gamma)_{221} = 1$$

Interaction plots are given in Figure 8.1.

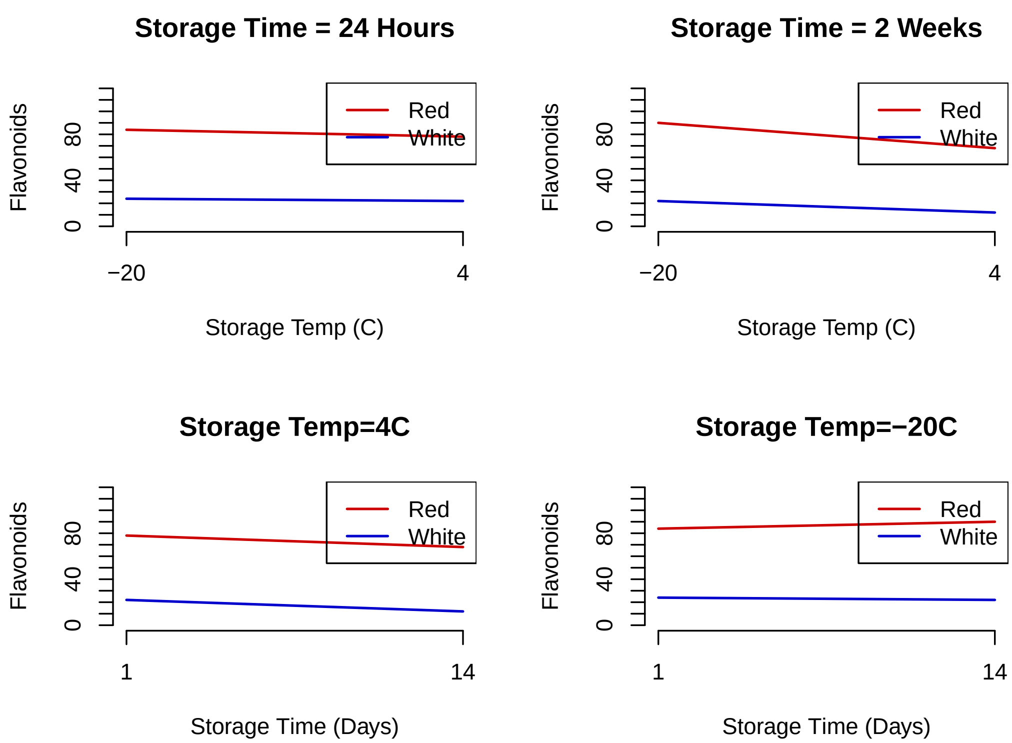


Figure 8.1: Interaction plots for flavonoid mean structure

The interactions are very small. By far the largest effects are red versus white grapes.

∇

8.2 Statistical Model and the Analysis of Variance

In this section, the model is defined, least squares estimators are given, as well as the sums of squares and F-tests. The model for Y_{ijkl} is given below in terms of the parameters described in the previous section.

$$Y_{ijkl} = \mu_{\bullet\bullet\bullet} + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkl}$$
$$\epsilon_{ijkl} \sim N\left(0, \sigma^2\right) \qquad i = 1, \dots, a; \quad j = 1, \dots, b; \quad k = 1, \dots, c; \quad l = 1, \dots, n$$

The parameters and their least squares estimators are given here.

$$\text{Overall Mean:} \quad \mu_{\bullet\bullet\bullet} \qquad \hat{\mu}_{\bullet\bullet\bullet} = \overline{Y}_{\bullet\bullet\bullet\bullet}$$

$$\text{Factor A Effects:} \quad \alpha_i \quad \hat{\alpha}_i = \overline{Y}_{i\bullet\bullet\bullet} - \overline{Y}_{\bullet\bullet\bullet\bullet} \qquad i = 1, \dots, a$$

$$\text{Factor B Effects:} \quad \beta_j \quad \hat{\beta}_j = \overline{Y}_{\bullet j\bullet\bullet} - \overline{Y}_{\bullet\bullet\bullet\bullet} \qquad j = 1, \dots, b$$

$$\text{Factor C Effects:} \quad \gamma_k \quad \hat{\gamma}_k = \overline{Y}_{\bullet\bullet k\bullet} - \overline{Y}_{\bullet\bullet\bullet\bullet} \qquad k = 1, \dots, c$$

$$\text{AB Interactions:} \quad (\alpha\beta)_{ij} \quad \widehat{(\alpha\beta)}_{ij} = \overline{Y}_{ij\bullet\bullet} - \overline{Y}_{i\bullet\bullet\bullet} - \overline{Y}_{\bullet j\bullet\bullet} + \overline{Y}_{\bullet\bullet\bullet\bullet}$$

$$\text{AC Interactions:} \quad (\alpha\gamma)_{ik} \quad \widehat{(\alpha\gamma)}_{ik} = \overline{Y}_{i\bullet k\bullet} - \overline{Y}_{i\bullet\bullet\bullet} - \overline{Y}_{\bullet\bullet k\bullet} + \overline{Y}_{\bullet\bullet\bullet\bullet}$$

$$\text{BC Interactions:} \quad (\beta\gamma)_{jk} \quad \widehat{(\beta\gamma)}_{ij} = \overline{Y}_{\bullet jk\bullet} - \overline{Y}_{\bullet j\bullet\bullet} - \overline{Y}_{\bullet\bullet k\bullet} + \overline{Y}_{\bullet\bullet\bullet\bullet}$$

$$\text{ABC Interactions:} \quad (\alpha\beta\gamma)_{ijk}$$

$$\widehat{(\alpha\beta\gamma)}_{ijk} = \overline{Y}_{ijk\bullet} - \overline{Y}_{ij\bullet\bullet} - \overline{Y}_{i\bullet k\bullet} - \overline{Y}_{\bullet jk\bullet} + \overline{Y}_{i\bullet\bullet\bullet} + \overline{Y}_{\bullet j\bullet\bullet} + \overline{Y}_{\bullet\bullet k\bullet} - \overline{Y}_{\bullet\bullet\bullet\bullet}$$

The predicted values simplify to the following values when the estimated overall mean and corresponding main effects and interactions are added together.

$$\hat{Y}_{ijkl} = \hat{\mu}_{\bullet\bullet\bullet} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \widehat{(\alpha\beta)}_{ij} + \widehat{(\alpha\gamma)}_{ik} + \widehat{(\beta\gamma)}_{jk} + \widehat{(\alpha\beta\gamma)}_{ijk} = \overline{Y}_{ijk\bullet}$$

The residuals are $e_{ijkl} = Y_{ijkl} - \overline{Y}_{ijk\bullet}$.

The sums of squares, degrees of freedom and expected mean squares are given below. To save space they will be given in terms of effects estimates.

$$\text{Total:} \quad SSTO = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n}\left(Y_{ijkl} - \overline{Y}_{\bullet\bullet\bullet\bullet}\right)^2 \qquad df_{TO} = abcn - 1$$

$$\text{A:} \quad SSA = bcn\sum_{i=1}^{a}\hat{\alpha}_i^2 \qquad df_A = a - 1 \qquad E\{MSA\} = \sigma^2 + \frac{bcn\sum_{i=1}^{a}\alpha_i^2}{a-1}$$

$$\text{B:} \quad SSB = acn\sum_{j=1}^{b}\hat{\beta}_j^2 \qquad df_B = b - 1 \qquad E\{MSB\} = \sigma^2 + \frac{acn\sum_{j=1}^{b}\beta_j^2}{b-1}$$

$$\text{C:} \quad SSC = abn\sum_{k=1}^{c}\hat{\gamma}_k^2 \qquad df_C = c - 1 \qquad E\{MSC\} = \sigma^2 + \frac{abn\sum_{k=1}^{c}\gamma_k^2}{c-1}$$

$$\text{AB:} \quad SSAB = cn\sum_{i=1}^{a}\sum_{j=1}^{b}\widehat{(\alpha\beta)}_{ij}^2 \qquad df_{AB} = (a-1)(b-1)$$

$$E\{MSAB\} = \sigma^2 + \frac{cn\sum_{i=1}^{a}\sum_{j=1}^{b}(\alpha\beta)_{ij}^2}{(a-1)(b-1)}$$

$$\text{AC:} \quad SSAC = bn\sum_{i=1}^{a}\sum_{k=1}^{c}\left(\widehat{\alpha\gamma}\right)_{ik}^{2} \qquad df_{AC} = (a-1)(c-1)$$

$$E\{MSAC\} = \sigma^2 + \frac{bn\sum_{i=1}^{a}\sum_{k=1}^{c}(\alpha\gamma)_{ik}^2}{(a-1)(c-1)}$$

$$\text{BC:} \quad SSBC = an\sum_{j=1}^{b}\sum_{k=1}^{c}\left(\widehat{\beta\gamma}\right)_{jk}^{2} \qquad df_{BC} = (b-1)(c-1)$$

$$E\{MSBC\} = \sigma^2 + \frac{an\sum_{j=1}^{b}\sum_{k=1}^{c}(\beta\gamma)_{jk}^2}{(b-1)(c-1)}$$

$$\text{ABC:} \quad SSABC = n\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\left(\widehat{\alpha\beta\gamma}\right)_{ijk}^{2} \qquad df_{ABC} = (a-1)(b-1)(c-1)$$

$$E\{MSABC\} = \sigma^2 + \frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}(\alpha\beta\gamma)_{ijk}^2}{(a-1)(b-1)(c-1)}$$

$$\text{Error:} \quad SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{l=1}^{n}\left(Y_{ijkl} - \overline{Y}_{ijk\bullet}\right)^2 \quad df_E = abc(n-1) \quad E\{MSE\} = \sigma^2$$

To test for main effects and interactions, the following F-tests are obtained (where TS is test statistic and RR is rejection region).

$$H_0^{ABC} : (\alpha\beta\gamma)_{ijk} = 0 \qquad i = 1, \dots, a; \quad j = 1, \dots, b; \quad k = 1, \dots, c$$

$$TS : F_{ABC}^* = \frac{MSABC}{MSE} \qquad RR : F_{ABC}^* \geq F_{1-\alpha;(a-1)(b-1)(c-1),abc(n-1)}$$

$$H_0^{AB} : (\alpha\beta)_{ij} = 0 \qquad i = 1, \dots, a; \quad j = 1, \dots, b;$$

$$TS : F_{AB}^* = \frac{MSAB}{MSE} \qquad RR : F_{AB}^* \geq F_{1-\alpha;(a-1)(b-1),abc(n-1)}$$

$$H_0^{AC} : (\alpha\gamma)_{ik} = 0 \qquad i = 1, \dots, a; \qquad k = 1, \dots, c$$

$$TS : F_{AC}^* = \frac{MSAC}{MSE} \qquad RR : F_{AC}^* \geq F_{1-\alpha;(a-1)(c-1),abc(n-1)}$$

$$H_0^{BC} : (\beta\gamma)_{jk} = 0 \qquad j = 1, \dots, b; \quad k = 1, \dots, c$$

$$TS : F_{BC}^* = \frac{MSBC}{MSE} \qquad RR : F_{BC}^* \geq F_{1-\alpha;(b-1)(c-1),abc(n-1)}$$

$$H_0^{A} : \alpha_i = 0 \qquad i = 1, \dots, a$$

$$TS : F_{A}^* = \frac{MSA}{MSE} \qquad RR : F_{A}^* \geq F_{1-\alpha;a-1,abc(n-1)}$$

$$H_0^{B} : \beta_j = 0 \qquad j = 1, \dots, b$$

$$TS : F_{B}^* = \frac{MSB}{MSE} \qquad RR : F_{B}^* \geq F_{1-\alpha;b-1,abc(n-1)}$$

$$H_0^{C} : \gamma_k = 0 \qquad k = 1, \dots, c$$

$$TS : F_C^* = \frac{MSC}{MSE} \qquad RR : F_C^* \geq F_{1-\alpha;c-1,abc(n-1)}$$

Example 8.2 - Finishing Treatments for Chef Jackets

A study was conducted to observe the effects of three factors for protecting chef jackets from scald injuries [Deverajan et al., 2017]. Three responses were measured: absorbed energy (kj/m^2) at 3 sensors (upper, middle, and lower). This example will analyze the upper sensor measurements. The three factors are as follow, there were $n = 5$ replicates per treatment combination.

- Factor A ($a = 4$) - **Finish**: Regular ($i = 1$), Water ($i = 2$), Soil ($i = 3$), Teflon ($i = 4$)
- Factor B ($b = 2$) - **Structure**: Plain ($j = 1$), Twill ($j = 2$)
- Factor C ($c = 3$) - **Number of Layers**: 1 ($k = 1$), 2 ($k = 2$), 3 ($k = 3$)

The R code and output are given below and the interaction plot is given in Figure 8.2.

```
##   finish structure layers trt absEnrgU absEnrgM absEnrgL
## 1      1         1      1   1  239.846  202.176  204.611
## 2      1         1      1   1  267.275  205.320  231.043
## 3      1         1      1   1  249.861  213.090  215.680
## 4      1         1      1   1  233.576  203.884  206.538
## 5      1         1      1   1  256.942  239.030  230.627
## 6      1         1      2   2  262.962  198.276  182.644
```

```
##     finish structure layers trt absEnrgU absEnrgM absEnrgL
## 115      4         2      2  23  162.693  134.630  117.672
## 116      4         2      3  24   99.663  104.037  107.369
## 117      4         2      3  24  127.183   94.204  107.297
## 118      4         2      3  24  107.016  108.169   98.877
## 119      4         2      3  24  107.863   90.537   80.929
## 120      4         2      3  24  117.774  109.554  104.529
```

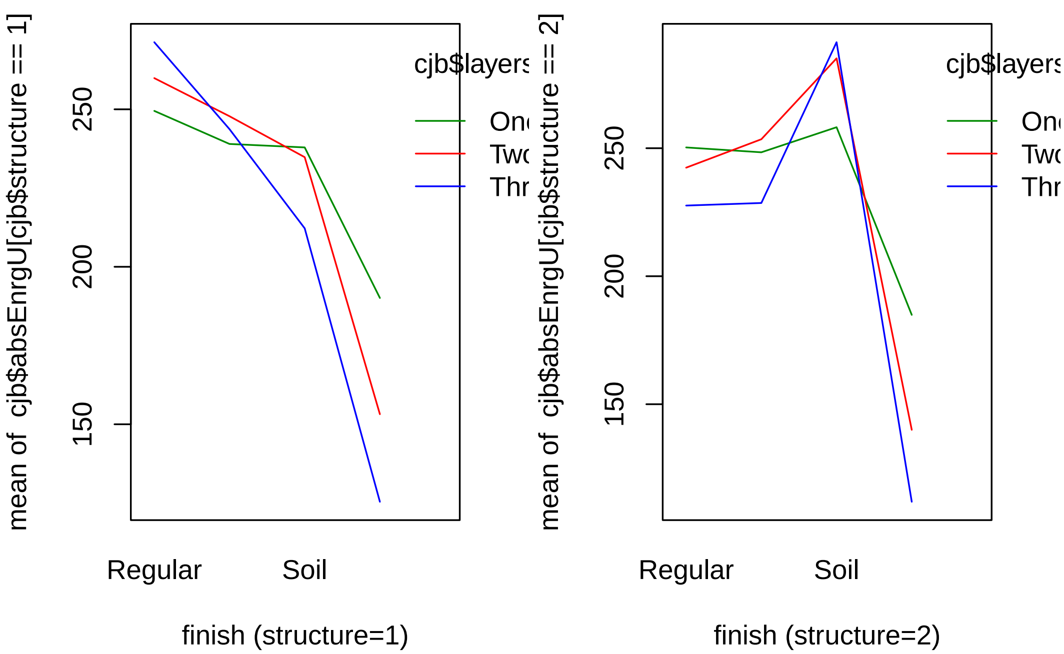


Figure 8.2: Interaction plots for chef jacket study

```
##
## Call:
## aov(formula = absEnrgU ~ finish.f * structure.f * layers.f, data = cjb)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -59.117  -7.176  -0.843   7.294  84.653
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     224.4625     1.5753 142.486  < 2e-16 ***
## finish.f1                        25.7041     2.7285   9.420 2.62e-15 ***
## finish.f2                        19.0374     2.7285   6.977 3.87e-10 ***
## finish.f3                        28.8042     2.7285  10.557  < 2e-16 ***
## structure.f1                     -2.3958     1.5753  -1.521 0.131593
## layers.f1                         7.8250     2.2278   3.512 0.000679 ***
## layers.f2                         2.6250     2.2278   1.178 0.241607
## finish.f1:structure.f1           12.4625     2.7285   4.567 1.47e-05 ***
## finish.f2:structure.f1            2.3958     2.7285   0.878 0.382105
## finish.f3:structure.f1          -22.5709     2.7285  -8.272 7.52e-13 ***
## finish.f1:layers.f1              -8.0917     3.8587  -2.097 0.038624 *
## finish.f2:layers.f1              -7.6250     3.8587  -1.976 0.051021 .
## finish.f3:layers.f1             -13.0416     3.8587  -3.380 0.001050 **
## finish.f1:layers.f2              -1.6415     3.8587  -0.425 0.671491
## finish.f2:layers.f2               4.5249     3.8587   1.173 0.243841
## finish.f3:layers.f2               4.0582     3.8587   1.052 0.295579
## structure.f1:layers.f1           -0.7666     2.2278  -0.344 0.731527
## structure.f1:layers.f2           -0.7667     2.2278  -0.344 0.731477
## finish.f1:structure.f1:layers.f1  -9.7000     3.8587  -2.514 0.013609 *
## finish.f2:structure.f1:layers.f1  -3.9333     3.8587  -1.019 0.310607
## finish.f3:structure.f1:layers.f1  15.5833     3.8587   4.038 0.000108 ***
## finish.f1:structure.f1:layers.f2  -0.5500     3.8587  -0.143 0.886966
## finish.f2:structure.f1:layers.f2  -2.0833     3.8587  -0.540 0.590524
## finish.f3:structure.f1:layers.f2   0.5833     3.8587   0.151 0.880169
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.26 on 96 degrees of freedom
## Multiple R-squared:  0.9055, Adjusted R-squared:  0.8829
## F-statistic: 40.01 on 23 and 96 DF,  p-value: < 2.2e-16
```

```
## Analysis of Variance Table
##
## Response: absEnrgU
##                             Df Sum Sq Mean Sq  F value    Pr(>F)
## finish.f                     3 217854   72618 243.8496 < 2.2e-16 ***
## structure.f                  1    689     689   2.3129 0.1315926
## layers.f                     2   7093    3546  11.9090 2.399e-05 ***
## finish.f:structure.f         3  21899    7300  24.5126 7.328e-12 ***
## finish.f:layers.f            6  18696    3116  10.4634 6.608e-09 ***
## structure.f:layers.f         2    141      71   0.2368 0.7895699
## finish.f:structure.f:layers.f  6   7680    1280   4.2983 0.0007001 ***
## Residuals                   96  28589     298
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## [1] "diff"  "lwr"   "upr"   "p.adj"
```

```
## [1] 137
```

With the exception of the Structure main effects and Structure/Layers interaction effects, all terms are significant. Since the 3-way interaction is significant, Structure and Structure/Layers should be kept in the model. For any pair of treatment means, Tukey's HSD is computed as follows (there are 4(2)(3)=24 means).

$$MSE = 298 \quad df_E = 24(5-1) = 96 \qquad HSD = q_{.95;24,96}\sqrt{\frac{298}{5}} = 5.297(7.720) = 40.893$$

There are 137 (out of 24(23)/2=276) pairs of treatments that differ by more than 40.893.

```
library(tidyverse)
library(kableExtra)
library(effectsize)
library(agricolae)
library(car)
library(PMCMRplus)
library(additivityTests)
```

Chapter 9

Block Designs

In many experimental settings, it is possible to group experimental units into homogeneous groups or **blocks**, allowing each treatment to be assigned to individual units within each block. It is often possible to evaluate each treatment on the same individual unit or subject. The goal is to remove block-to-block variation and obtain more powerful tests for treatment effects. When individual units receive each treatment (preferably in random order), these are often referred to as **Repeated Measures**.

In other situations, it may be necessary to block on two or three factors to obtain better precision for studying treatment effects. These designs are **Latin Squares** for two blocking factors and **Graeco-Latin Squares** for three blocking factors. In many situations, these designs are used to obtain "fair" comparisons of treatments, even if the factors are not concluded in the analysis. This helps remove biases that can occur in the implementation of the experiment, particularly when individual subjects receive each treatment.

Some examples containing a single blocking factor are given below.

- **Agriculture** - An experiment to compare r varieties of fertilizer is conducted on b blocks of land. Within each block, r plots are obtained and one variety of fertilizer is applied to each plot.
- **Industrial** - An experiment to compare r methods (or machines) to manufacture a product is conducted on b batches of raw material. Each batch is broken into r sub-batches and each treatment is applied to one sub-batch within the batch.
- **Pharmaceutical** - An experiment is conducted to compare bioavailability among r formulations of a drug product, by having each product measured on b subjects. Subjects are given a particular sequence, and availability is measured for each formulation.

In many cases, blocks are simply individual replications of an experiment on different days or at different locations. In other cases, individuals may be ordered based on an external measurement (e.g. severity, ability) and blocks may be formed by that criteria and the various treatments may be assigned to the group of individuals within these blocks.

Block designs can have a single treatment factor or combinations of multiple treatment factors. The larger the number of treatment levels, the less homogeneous will be the units within a block. In these cases **Incomplete Block Designs** can be utilized.

So far, when considering treatment factors, we have implicitly treated them as **fixed factors**, the levels observed in the experiment being the specific treatments of interest to researchers. In block designs, the blocks are very often **random factors**, representing a sample of all levels that could have been selected. These include: plots of land, batches of raw material, subjects in bioequivalence studies, and in virtually all repeated measures settings. The primary goal in all of these experiments is to compare the treatment effects as we have done in previous chapters. Fortunately, the F-test and methods for making contrasts among treatments are the same whether blocks are fixed or random.

9.1 Randomized Block and Repeated Measures Designs

In this section we introduce the model for the Randomized Complete Block Design. This same model is used for the Repeated Measures Design, when each subject receives each treatment. This is not the same as a Repeated Measures Design, where each subject receives only one treatment, but is measured longitudinally at multiple time points.

For this model, each of r treatments are observed, once within each of b blocks or subjects, for a total of $n_T = rb$ observations. For this model, when treatments are applied to subjects in some time order, we are assuming there are no order or carryover effects. More complex designs can be used when this is the case. In this section, we will use **blocks** and **subjects** interchangeably. In virtually all cases, subjects are treated as random, except in settings where the only subjects of interest are included in the experiment.

Note: This notation is NOT consistent with Kutner, et al.

9.1.1 Model Structure and Estimators

$$Y_{ij} = \mu_{\bullet\bullet} + \tau_i + \beta_j + \epsilon_{ij} \qquad i = 1, \ldots, r; \quad j = 1, \ldots, b \qquad \epsilon_{ij} \sim NID\left(0, \sigma^2\right)$$

We assume fixed treatment effects as before, and allow for either fixed or random block effects.

$$\text{Overall Mean:} \quad \mu_{\bullet\bullet} \qquad \hat{\mu}_{\bullet\bullet} = \frac{\sum_{i=1}^{r}\sum_{j=1}^{b} Y_{ij}}{rb} = \overline{Y}_{\bullet\bullet}$$

$$\text{Treatment Effects:} \quad \tau_i = \mu_{i\bullet} - \mu_{\bullet\bullet} \qquad \hat{\tau}_i = \overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet} \qquad \sum_{i=1}^{r} \tau_i = \sum_{i=1}^{r} \hat{\tau}_i = 0$$

We consider the block effects separately, considering the fixed and random effects cases.

$$\text{Fixed Blocks:} \quad \beta_j = \mu_{\bullet j} - \mu_{\bullet\bullet} \qquad \hat{\beta}_j = \overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet} \qquad \sum_{j=1}^{b} \beta_j = \sum_{j=1}^{b} \hat{\beta}_j = 0$$

For the fixed block case, we obtain the following model for Y_{ij}.

$$E\left\{Y_{ij}\right\} = \mu + \tau_i + \beta_j \qquad \sigma^2\left\{Y_{ij}\right\} = \sigma^2 \qquad \sigma\left\{Y_{ij}, Y_{i'j'}\right\} = 0 \quad \forall i \neq i', j \neq j'$$

$$\text{Random Blocks:} \quad \beta_j \sim NID\left(0, \sigma_b^2\right) \qquad \sigma\left\{\beta_j, \epsilon_{ij}\right\} = 0 \qquad \hat{\beta}_j = \overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet} \qquad \sum_{j=1}^{b} \hat{\beta}_j = 0$$

The fitted values and residuals are similar to other models described previously.

$$\hat{Y}_{ij} = \hat{\mu}_{\bullet\bullet} + \hat{\tau}_i + \hat{\beta}_j = \overline{Y}_{i\bullet} + \overline{Y}_{\bullet j} - \overline{Y}_{\bullet\bullet} \qquad e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \overline{Y}_{i\bullet} - \overline{Y}_{\bullet j} + \overline{Y}_{\bullet\bullet}$$

The random block effects and the random errors are assumed independent. By the structure of the estimators the estimated β_j sum to zero, but it's the population of β_j that has mean 0. For the random block case, we obtain the following model for Y_{ij}.

$$E\left\{Y_{ij}\right\} = \mu + \tau_i \qquad \sigma^2\left\{Y_{ij}\right\} = \sigma_b^2 + \sigma^2 \qquad \sigma\left\{Y_{ij}, Y_{i'j'}\right\} = \begin{cases} \sigma_b^2 + \sigma^2 & : \quad i = i', j = j' \\ \sigma_b^2 & : \quad i \neq i', j = j' \\ 0 & : \quad \forall i, i', j \neq j' \end{cases}$$

Two measurements (for different treatments) within the same block are correlated. Measurements across different blocks are independent.

Before setting up the Analysis of Variance, we state the mean and variance structure for the treatment means for fixed and random blocks. In each case $E\left\{\overline{Y}_{i\bullet}\right\} = \mu + \tau_i$.

$$\text{Fixed Blocks:} \quad \sigma^2\left\{\overline{Y}_{i\bullet}\right\} = \frac{\sigma^2}{b} \qquad i \neq i': \quad \sigma\left\{\overline{Y}_{i\bullet}, \overline{Y}_{i'\bullet}\right\} = 0 \qquad \sigma^2\left\{\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}\right\} = \frac{2\sigma^2}{b}$$

$$\text{Random Blocks:} \quad \sigma^2\left\{\overline{Y}_{i\bullet}\right\} = \frac{\sigma_b^2 + \sigma^2}{b} \qquad i \neq i': \quad \sigma\left\{\overline{Y}_{i\bullet}, \overline{Y}_{i'\bullet}\right\} = \frac{\sigma_b^2}{b} \qquad \sigma^2\left\{\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}\right\} = \frac{2\sigma^2}{b}$$

Thus, while the variances of the treatment means differ depending on whether blocks are fixed or random, the variances of the **difference in the treatment means is the same**.

Example 9.1 - Sensory Analysis of Soy Sauce Recipes

A study compared $r = 4$ soy sauce recipes, each judged by a panel of $b = 8$ trained raters [Fidaleo et al., 2012]. The $r = 4$ soy sauce recipes are given below, this being treated as a fixed factor (only recipes of interest).

- **Recipe 1** - Original Soy Sauce
- **Recipe 2** - Electrodialyzed desalted Soy Sauce
- **Recipe 3** - Recipe 1 diluted at electric conductivity = 2.3 S/m
- **Recipe 4** - Recipe 2 re-salted to original level

In sensory studies such as this, there is a debate among researchers whether raters (blocks) should be treated as a fixed or random factor.

- **Fixed Blocks** - These are the only raters of interest to the researchers (possibly the only raters employed by this company).
- **Random Blocks** - The researchers are interested in generalizing these findings to a population of trained raters (these 8 raters being a sample from that population).

Based on the sample values reported in this paper, we will construct population models for the fixed and random blocks cases. These were ratings on a 4-point scale (all numbers have been multiplied by 10 to keep variances from being so small). The response these are based on was appearance/color. Parameter values used are given here.

$$\mu_{\bullet\bullet} = 32 \qquad \tau_1 = 4 \quad \tau_2 = 3 \quad \tau_3 = -13 \quad \tau_4 = 6 \qquad \sigma^2 = 11.56 \quad \sigma = 3.4$$

$$\text{Fixed Blocks:} \quad \beta_1, \dots, \beta_8 = \text{ -5.30 to 5.30 by 1.514} \qquad \text{Random Blocks:} \quad \sigma_b^2 = 13.76 \quad \sigma_b = 3.71$$

The variance among block effects was chosen so that the following equality holds. This makes the models comparable with respect to variability in the block effects (see $E\{MSBL\}$ in the next sub-section).

$$\sigma_b^2 = \frac{\sum_{j=1}^{8} \beta_j^2}{8-1} = \frac{(-5.300)^2 + \cdots + (5.300)^2}{7} = 13.76$$

We will generate 100000 samples from this model (one for fixed block effects, one for random block effects). The algorithms are described here.

Fixed Effects

- Generate pseudo $\epsilon_{11}, \dots, \epsilon_{48} \sim NID(0, 3.4^2)$
- Assign $Y_{ij}^F = \mu_{\bullet\bullet} + \tau_i + \beta_j^F + \epsilon_{ij} \qquad i = 1, \dots, 4; \quad j = 1, \dots, 8$
- Compute and save $\overline{Y}_{1\bullet}^F, \dots, \overline{Y}_{4\bullet}^F$
- Obtain (across samples) the mean and variance of $\overline{Y}_{1\bullet}^F, \dots, \overline{Y}_{4\bullet}^F$
- Obtain (across samples) the mean and variance of the differences among $\overline{Y}_{1\bullet}^F, \dots, \overline{Y}_{4\bullet}^F$

Random Effects

- Generate pseudo $\epsilon_{11}, \dots, \epsilon_{48} \sim NID\left(0, 3.4^2\right)$
- Generate pseudo $\beta_1^R, \dots, \beta_8^R \sim NID\left(0, 3.71^2\right)$
- Assign $Y_{ij}^R = \mu_{\bullet\bullet} + \tau_i + \beta_j^R + \epsilon_{ij} \qquad i = 1, \dots, 4; \quad j = 1, \dots, 8$
- Compute and save $\overline{Y}_{1\bullet}^R, \dots, \overline{Y}_{4\bullet}^R$
- Obtain (across samples) the mean and variance of $\overline{Y}_{1\bullet}^R, \dots, \overline{Y}_{4\bullet}^R$
- Obtain (across samples) the mean and variance of the differences among $\overline{Y}_{1\bullet}^R, \dots, \overline{Y}_{4\bullet}^R$

For each model (fixed and random blocks), the theoretical means for the recipe means and for differences among the means are given below.

$$E\{\overline{Y}_{i\bullet}\} = \mu_i = \mu_{\bullet\bullet} + \tau_i \qquad E\{\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}\} = \mu_i - \mu_{i'} = \tau_i - \tau_{i'}$$

For fixed block effects, the theoretical variances of the recipe means and for the differences among the means are given below.

$$\sigma^2\{\overline{Y}_{i\bullet}\} = \frac{\sigma^2}{b} = \frac{11.56}{8} = 1.445 \qquad \sigma^2\{\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}\} = \frac{2\sigma^2}{b} = \frac{2(11.56)}{8} = 2.90$$

For random block effects, the theoretical variances of the recipe means and for the differences among the means are given below.

$$\sigma^2\{\overline{Y}_{i\bullet}\} = \frac{\sigma_b^2 + \sigma^2}{b} = \frac{13.76 + 11.56}{8} = 3.165 \qquad \sigma^2\{\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}\} = \frac{2\sigma^2}{b} = \frac{2(11.56)}{8} = 2.90$$

```
## [1] -5.3000000 -3.7857143 -2.2714286 -0.7571429  0.7571429  2.2714286
## [7]  3.7857143  5.3000000

## [1] 13.75837

## [1] 3.709227

##             Mean(Fixed) Mean(Random) Var(Fixed) Var(Random)
## Ybar1            36.006       36.006      1.438       3.167
## Ybar2            34.998       34.997      1.439       3.154
## Ybar3            19.001       19.001      1.441       3.148
## Ybar4            37.997       37.997      1.446       3.157
## Ybar2-Ybar1      -1.009       -1.009      2.867       2.867
## Ybar3-Ybar1     -17.005      -17.005      2.892       2.892
## Ybar3-Ybar2     -15.997      -15.997      2.888       2.888
## Ybar4-Ybar1       1.990        1.990      2.878       2.878
## Ybar4-Ybar2       2.999        2.999      2.874       2.874
## Ybar4-Ybar3      18.996       18.996      2.889       2.889
```

All of the empirical results are very close to the theoretical results (as they should be). The difference arises when considering $\sigma^2\left\{\overline{Y}_{i\bullet}\right\}$ when block effects are generated as fixed effects versus when they are generated as random effects.

∇

9.2 Analysis of Variance and the F-test

We use the sums of squares as previously set up, with the Sum of Squares Error SSE, technically being the Sum of Squares for the Treatment by Block Interaction as there is only one replicate per combination of Treatment and Block. The sums of squares, degrees of freedom, and expected mean squares are given here. In this model, we assume that there is no Treatment/Block interaction. Tukey's One-Degree of Freedom Test for Non-Additivity can be used to check this assumption.

$$\text{Total:} \quad SSTO = \sum_{i=1}^{r}\sum_{j=1}^{b}\left(Y_{ij} - \overline{Y}_{\bullet\bullet}\right)^2 \qquad df_{TO} = rb - 1$$

$$\text{Treatments:} \quad SSTR = b\sum_{i=1}^{r}\left(Y_{i\bullet} - \overline{Y}_{\bullet\bullet}\right)^2 \qquad df_{TR} = r - 1$$

$$E\left\{MSTR\right\} = \sigma^2 + \frac{b\sum_{i=1}^{r}\tau_i^2}{r-1}$$

$$\text{Blocks:} \quad SSBL = r\sum_{j=1}^{b}\left(Y_{\bullet j} - \overline{Y}_{\bullet\bullet}\right)^2 \qquad\qquad df_{BL} = b-1$$

$$\text{Fixed Blocks:} \quad E\left\{MSBL\right\} = \sigma^2 + \frac{r\sum_{j=1}^{b}\beta_j^2}{b-1} \qquad \text{Random Blocks:} \quad E\left\{MSBL\right\} = \sigma^2 + r\sigma_b^2$$

$$\text{Error (TRxBL):} \quad SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(Y_{ij} - \overline{Y}_{i\bullet} - \overline{Y}_{\bullet j} + \overline{Y}_{\bullet\bullet}\right)^2 \quad df_E = (r-1)(b-1) \quad E\left\{MSE\right\} = \sigma^2$$

To test for treatment effects, we compute the ratio of $MSTR$ to MSE with the F-test. Software will automatically compute the F-statistic for blocks as well, but the primary interest is typically treatment effects. However, when blocks are random it may be of interest to estimate σ_b^2, the variance of the block effects.

$$H_0 : \tau_1 = ... = \tau_r = 0 \qquad TS : F_{TR}^* = \frac{MSTR}{MSE} \qquad RR : F_{TR}^* \geq F_{1-\alpha;r-1,(r-1)(b-1)}$$

Any contrasts or pairwise comparisons can be carried out as in the Completely Randomized Design. For Tukey's HSD for all possible comparisons, we compute HSD, which will be the same for all pairs ($n_i = b$), and can obtain simultaneous $(1-\alpha)100$ Confidence Intervals for $\tau_i - \tau_{i'}$ as follows.

$$HSD = q_{1-\alpha;r,(r-1)(b-1)}\sqrt{\frac{MSE}{b}}$$

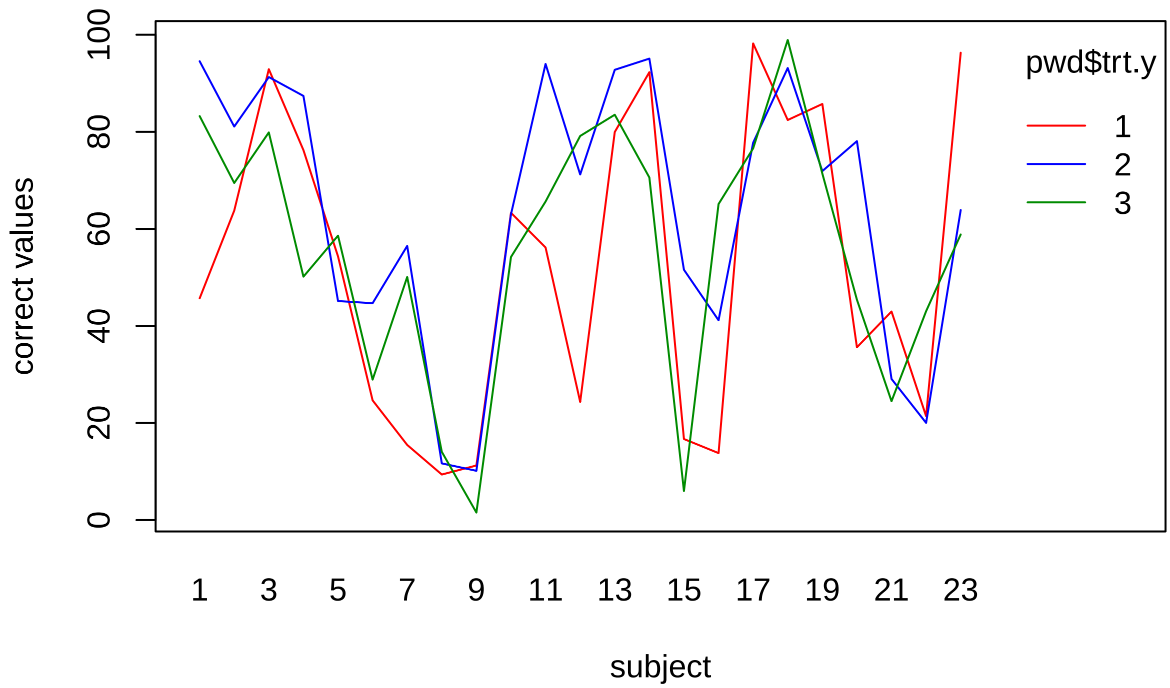
$$(1-\alpha)100\% \text{ CI for } \tau_i - \tau_{i'} : \qquad \left(\overline{Y}_{i\bullet} - \overline{Y}_{i'\bullet}\right) \pm HSD$$

Example 9.2 - Methods of Presenting Weather Information to Pilots A study compared $r = 3$ cockpit weather displays (treatments) for pilots in terms of information recall [O'Hare and Stenhouse, 2009] There were $b = 23$ pilots (blocks) who were exposed to each display on a flight simulator. The three displays are described below. The response Y, is the percent of information correctly recalled by the pilot.

- Ordinary text - English phrases ($i = 1$)
- Redesigned graphical display ($i = 2$)
- Old graphical display ($i = 3$)

Using R, we will set up the Analysis of Variance directly, then use the **aov** function along with **TukeyHSD** function to confirm results.

```
##   trt.y blk.y corrRecall
## 1     1     1   45.68415
## 2     1     2   63.76311
## 3     1     3   92.90009
## 4     1     4   76.22723
## 5     1     5   54.33312
## 6     1     6   24.66595
```




```
##             df       SS       MS    F* F(.95) P(>F*)
## Treatments  2  1582.86  791.430 3.330  3.209  0.045
## Blocks     22 42814.57 1946.117 8.188  1.116  0.000
## Error      44 10457.33  237.667    NA     NA     NA
## Total      68 54854.76       NA    NA     NA     NA

##       i i' ybar_i ybar_i' diff      LB     UB p adj
## [1,] 2  1   63.7    52.3 11.4   0.374 22.426 0.041
## [2,] 3  1   55.6    52.3  3.3  -7.726 14.326 0.750
## [3,] 3  2   55.6    63.7 -8.1 -19.126  2.926 0.187

## Analysis of Variance Table
##
## Response: corrRecall
##                Df Sum Sq Mean Sq F value    Pr(>F)
## factor(trt.y)  2   1583  791.43  3.3300   0.04501 *
## factor(blk.y) 22  42815 1946.12  8.1884 2.253e-09 ***
## Residuals     44  10457  237.67
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = corrRecall ~ factor(trt.y) + factor(blk.y), data = pwd)
##
## $`factor(trt.y)`
##     diff         lwr       upr     p adj
## 2-1 11.4   0.3736119 22.426388 0.0413298
## 3-1  3.3  -7.7263883 14.326388 0.7495471
```

```
## 3-2 -8.1 -19.1263883  2.926388 0.1874886
```

Thus, there are differences among the three cockpit weather displays. In particular, pilots had higher recall with the redesigned graphical display ($i = 2$) than the ordinary text display ($i = 1$).

∇

9.2.1 Checking Model Assumptions

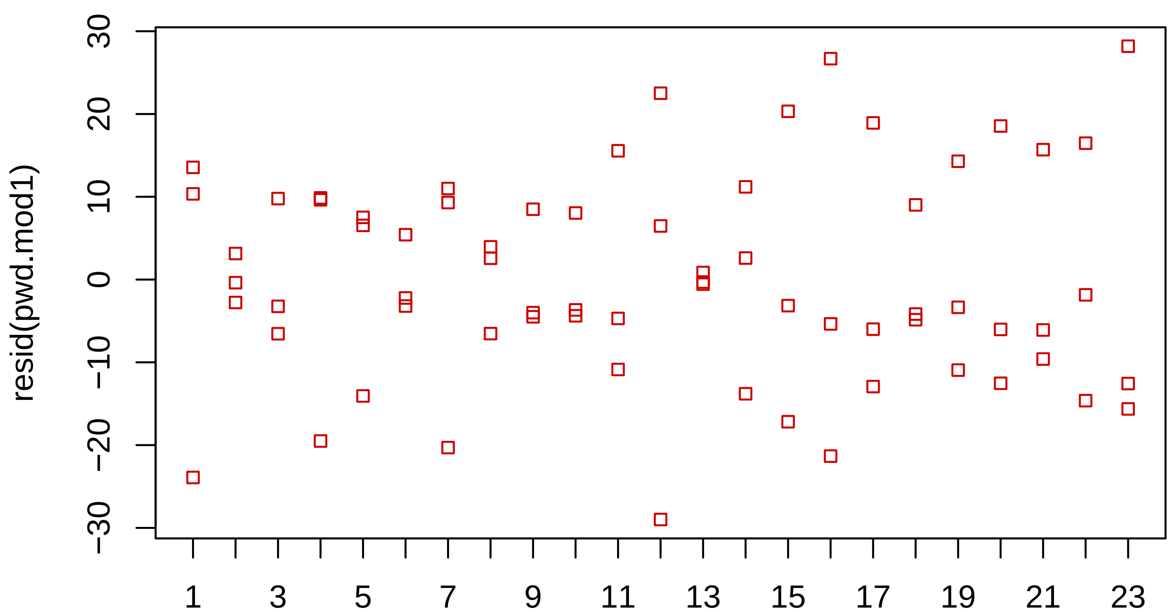
Some methods for checking model assumptions are given in this sub-section, involving the error terms for the model.

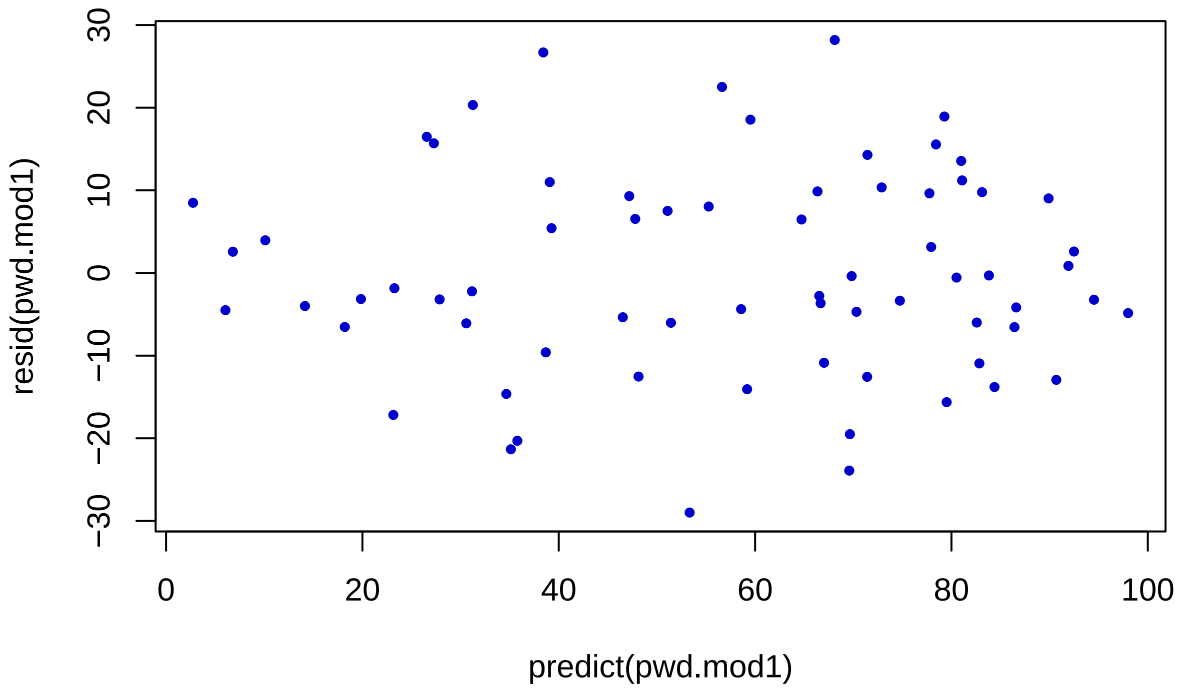
- Stripchart of residuals versus blocks to visually observe whether the variance of errors is constant across blocks (each block received each treatment)
- Plots of residuals versus fitted values to observe whether residual variance is related to mean
- Plot of residuals versus time order when experiment is conducted sequentially or when blocks are days to observe whether serial correlation is present
- Block x Treatment Interactions - Can be tested using Tukey's ODOFNA test. If significant, the remainder MSE^* can be used after removing the 1 degree of freedom SS for interaction

Example 9.3 - Methods of Presenting Weather Information to Pilots

Here we give examples of the model checks, with the exception of residuals versus time order, as that does not pertain to this study.

```
##   trt.y blk.y corrRecall
## 1     1     1   45.68415
## 2     1     2   63.76311
## 3     1     3   92.90009
## 4     1     4   76.22723
## 5     1     5   54.33312
## 6     1     6   24.66595
```






```
##
## Tukey test on 5% alpha-level:
##
## Test statistic: 0.009099
## Critival value: 4.067
## The additivity hypothesis cannot be rejected.
```

While a few blocks (pilots) have little spread in his/her residuals based on the stripchart, the spread is reasonably equal across the $b = 23$ pilots. Plus, there are only $r = 3$ residuals per pilot.

The plot of the residuals versus predicted values does not show and evidence of "funneling" out as the fitted values increase, so constant variance seems reasonable.

Tukey's ODOFNA test gives a very small test statistic, well below its critical value. This provides no evidence of an interaction between treatment (display of weather information) by block (pilot) interaction.

∇

9.2.2 Relative Efficiency and Within-Subject Variance-Covariance Matrix

In this sub-section, we consider two measures that come up in Randomized Block and Repeated Measures Designs.

- **Relative Efficiency** - A measure of the ratio of the error variance of the Completely Randomized Design (CRD) to the Randomized Block Design (RBD)
- **Within Subject Variance-Covariance Matrix** - When blocks/subjects are random, the measurements within subjects are correlated. The assumptions are of equal variances for the r treatments and equal covariances among the $r(r-1)/2$ pairs of treatments.

For **Relative Efficiency**, we label σ^2_{CRD} to be the error variance for a model based on a Completely Randomized Design and σ^2_{RBD} the error variance for the Randomized Block Design. Recall that the goal of the RBD is to reduce the error variance, thus making more precise estimates of contrasts (smaller standard errors).

We define the following ratios in terms of true (unknown) variances and estimated variances from the RBD.

$$E = \frac{\sigma^2_{CRD}}{\sigma^2_{RBD}} \qquad \hat{E} = \frac{s^2_{CRD}}{s^2_{RBD}} = \frac{(b-1)MSBL + b(r-1)MSE}{(br-1)MSE}$$

Note that $(b-1)+b(r-1) = br-1$, so that $\hat{E}$ is a weighted average of $MSBL/MSE$ and 1. The larger the ratio of $MSBL/MSE$, the larger will be $\hat{E}$. Then, experiments with large variability among blocks relative to error variability are more efficient.

The Relative Efficiency measures the multiple of the number of blocks from the RBD that would need to be the treatment sample size in a CRD to have comparable standard errors for contrasts.

Suppose that an experiment run as a RBD had $b = 5$ blocks had a Relative Efficiency of $\hat{E} = 2$. Then an experiment run as a CRD would need $n = 2(5) = 10$ units per treatments to have comparable standard errors for treatment contrasts.

Some authors describe a degrees of freedom adjustment for the estimated Relative Efficiency. In many cases, it makes little difference, the formula is given below, with r being the number of treatments and b being the number of blocks.

$$df_{CRD} = r(b-1) \quad df_{RBD} = (r-1)(b-1) \qquad \hat{E}^* = \frac{(df_{RBD}+1)(df_{CRD}+3)}{(df_{RBD}+3)(df_{CRD}+1)}\hat{E}$$

When blocks/subjects are random, then the responses Y within blocks are correlated as described at the beginning of the chapter.

$$\sigma^2\left\{Y_{ij}\right\} = \sigma^2_b + \sigma^2 = \sigma^2_i \qquad i \neq i' : \quad \sigma\left\{Y_{ij}, Y_{i'j}\right\} = \sigma^2_b = \sigma_{ii'}$$

For this model to be appropriate we need the following assumptions.

$$\sigma^2_1 = \cdots = \sigma^2_r \qquad \sigma_{12} = \cdots = \sigma_{r-1,r}$$

In matrix form, the within-subject Variance-Covariance matrix can be written as follows, with a similar set of assumptions.

$$\Sigma = \begin{bmatrix} \sigma^2_1 & \sigma_{12} & \cdots & \sigma_{1r} \\ \sigma_{21} & \sigma^2_2 & \cdots & \sigma_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{r1} & \sigma_{r2} & \cdots & \sigma^2_r \end{bmatrix} = \begin{bmatrix} \sigma^2 & \sigma_{ii'} & \cdots & \sigma_{ii'} \\ \sigma_{ii'} & \sigma^2 & \cdots & \sigma_{ii'} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{ii'} & \sigma_{ii'} & \cdots & \sigma^2 \end{bmatrix}$$

The sample version of the Variance-Covariance matrix is computed as follows.

$$S = \begin{bmatrix} s^2_1 & s_{12} & \cdots & s_{1r} \\ s_{21} & s^2_2 & \cdots & s_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ s_{r1} & s_{r2} & \cdots & s^2_r \end{bmatrix}$$

$$s^2_i = \frac{\sum_{j=1}^{b}\left(Y_{ij} - \overline{Y}_{i\bullet}\right)^2}{b-1} \qquad s_{ii'} = \frac{\sum_{j=1}^{b}\left(Y_{ij} - \overline{Y}_{i\bullet}\right)\left(Y_{i'j} - \overline{Y}_{i'\bullet}\right)}{b-1}$$

There are tests of whether the Population Variance-Covariance meets this assumption based on the Sample Variance-Covariance matrix, but are beyond the scope of this course. A visual inspection of S should be able to detect large discrepancies. Other structures (including an unstructured form) can be implemented using mixed model software packages. Also, transformations on Y can be implemented (such as the Box-Cox transformation).

We will use the **cov** function by looping through the treatments to obtain the Variance-Covariance matrix, keeping in mind that $s\{Y_{ij}, Y_{ij}\} = s^2\{Y_{ij}\}$.

Table 9.1: Latin Square Design (r=6)

Row\Col	1	2	3	4	5	6
1	A	B	C	D	E	F
2	B	C	D	E	F	A
3	C	D	E	F	A	B
4	D	E	F	A	B	C
5	E	F	A	B	C	D
6	F	A	B	C	D	E

Example 9.4 - Methods of Presenting Weather Information to Pilots Here we compute the Relative Efficiency and the Sample Variance-Covariance matrix for the pilot weather information study based on direct calculations.

```
##   trt.y blk.y corrRecall
## 1     1     1   45.68415
## 2     1     2   63.76311
## 3     1     3   92.90009
## 4     1     4   76.22723
## 5     1     5   54.33312
## 6     1     6   24.66595
```

```
##       b r     MSBL     MSE E-hat df_CRD df_RBD E-hat*
## [1,] 23 3 1946.117 237.667 3.326     46     44   3.32
```

```
##         [,1]    [,2]    [,3]
## [1,] 975.803 607.759 532.015
## [2,] 607.759 764.172 568.676
## [3,] 532.015 568.676 681.475
```

The Relative Efficiency is $\hat{E} = 3.32$, which can be interpreted as, if this had been run as a Completely Randomized Design, with each pilot being only observed in one condition, there would need to be $b\hat{E} = 23(3.32) = 76.4 \approx 77$ pilots per treatment ($n_T = 3(77) = 231$ total) to have comparable standard errors for contrasts. This is an efficient design.

While the variances range from 681 to 976, and the covariances range from 532 to 607, there is no strong evidence that the Variance-Covariance is far from the assumed model (standard deviations range from 26 to 31).

∇

9.3 Latin Square Designs

When there are two blocking factors, a **Latin Square Design** can be used. There will be a row blocking factor, a column blocking factor, and the treatment factor. In a "true" Latin Square, each factor will have the same number of levels (r). When r is small, this will cause very low error degrees of freedom ($df_E = (r-1)(r-2)$), so typically either the row and/or column blocking factor may be replicated and have kr levels, where k is an integer value.

The key idea in the Latin Square is that each treatment appears once in each row, and once in each column. This way, we can directly remove row and column effects to get more precise estimates of the treatment effects and contrasts among them. An example of a Latin Square is given below with $r = 6$, where rows and columns are blocking factors and labels $A, \ldots, F$ are the treatments.

Keep in mind that while there are three factors, each with r levels, there are only r^2 observations. Once the row and columns have been indexed, there is only one treatment that appears. This can make writing the model confusing with an observation being Y_{ijk}, because you cannot cycle through all three subscripts.

However, there are still r means for each treatment, row, and column, each based on r observations. Let T_{kl} represent the measurement for the l^{th} replicate when treatment k is assigned, R_{il} represent the l^{th} replicate in row

i, and C_{jl} represent the l^{th} replicate in column j. Further let $\overline{Y}$ represent the overall mean and $SSTO$ be the usual total sum of squares. We obtain the Analysis of Variance as follows (note that when using software packages, this is a trivial extension of the Randomized Block Design).

$$\text{Total:} \quad SSTO = \sum_{\text{all data}} \left(Y - \overline{Y}\right)^2 \qquad df_{TO} = r^2 - 1$$

$$\text{Treatments:} \quad \overline{T}_k = \frac{\sum_{l=1}^{r} T_{kl}}{r} \qquad SSTR = r \sum_{k=1}^{r} \left(\overline{T}_k - \overline{Y}\right)^2 \qquad df_{TR} = r - 1$$

$$\text{Rows:} \quad \overline{R}_i = \frac{\sum_{l=1}^{r} R_{il}}{r} \qquad SSR = r \sum_{i=1}^{r} \left(\overline{R}_i - \overline{Y}\right)^2 \qquad df_R = r - 1$$

$$\text{Columns:} \quad \overline{C}_j = \frac{\sum_{l=1}^{r} C_{jl}}{r} \qquad SSC = r \sum_{j=1}^{r} \left(\overline{C}_j - \overline{Y}\right)^2 \qquad df_C = r - 1$$

$$\text{Error:} \quad SSE = SSTO - SSTR - SSR - SSC$$

$$df_E = (r^2 - 1) - 3(r - 1) = (r + 1)(r - 1) - 3(r - 1) = (r - 1)[(r + 1) - 3] = (r - 1)(r - 2)$$

The usual F-test for treatment effects and contrasts among treatment levels can be conducted as before. For instance, for Tukey's HSD, we would compute the following values for the Honest Significant Difference and for simultaneous $(1 - \alpha)100\%$ Confidence Intervals. As in the RBD, these will all be based on common sample sizes (r), and τ_k is the treatment effect for treatment k.

$$HSD = q(1 - \alpha; r, (r - 1)(r - 2))\sqrt{\frac{MSE}{r}}$$

$$(1 - \alpha)100\% \text{ Confidence Interval for } \tau_k - \tau_{k'} : \left(\overline{T}_k - \overline{T}_{k'}\right) \pm HSD$$

Example 9.5 - Comparison of Medium (Psychic) Readings A study compared $r = 5$ Psychic Mediums (treatments) in a Latin Square experiment [O'Keeffe and Wiseman, 2005]. We will treat Medium as a fixed factor (only interest is comparing these specific ones). There were 5 Sitters (row factor) and 5 Sessions (column factor). Each Medium read each Sitter once, and each Medium read in each Session once. This design would look like the previous example, with one less row, one less column, and "F" removed. We will do computations directly, then use the **aov** and **TukeyHSD** functions for the analysis. The response was Y, the number of statements by the Medium on the Sitter.

```
##   session medium sitter statements rating
## 1       1      1      1         55   3.33
## 2       2      2      1         92   3.72
## 3       3      3      1          6   1.52
## 4       4      4      1         24   3.67
## 5       5      5      1         80   5.24
## 6       2      1      2         62   2.88
```

```
##         df       SS      MS     F* F(.95) P(>F*)
## Medium   4 17280.56 4320.14 55.006  3.259  0.000
## Sitter   4   205.76   51.44  0.655  3.259  0.635
## Session  4   231.76   57.94  0.738  3.259  0.584
## Error   12   942.48   57.94     NA     NA     NA
## Total   24 18660.56      NA     NA     NA     NA
```

```
##       k k' ybar_i ybar_i'  diff       LB       UB  p adj
##  [1,] 2  1   75.4    59.8  15.6  -2.2656  33.4656 0.0983
##  [2,] 3  1    8.0    59.8 -51.8 -69.6656 -33.9344 0.0000
##  [3,] 3  2    8.0    75.4 -67.4 -85.2656 -49.5344 0.0000
##  [4,] 4  1   23.4    59.8 -36.4 -54.2656 -18.5344 0.0002
##  [5,] 4  2   23.4    75.4 -52.0 -69.8656 -34.1344 0.0000
```

```
##  [6,] 4  3   23.4     8.0  15.4  -2.4656  33.2656 0.1042
##  [7,] 5  1   67.2    59.8   7.4 -10.4656  25.2656 0.6848
##  [8,] 5  2   67.2    75.4  -8.2 -26.0656   9.6656 0.6026
##  [9,] 5  3   67.2     8.0  59.2  41.3344  77.0656 0.0000
## [10,] 5  4   67.2    23.4  43.8  25.9344  61.6656 0.0000
```

```
## Analysis of Variance Table
##
## Response: statements
##                 Df  Sum Sq Mean Sq F value   Pr(>F)
## factor(medium)   4 17280.6  4320.1 55.0056 1.28e-07 ***
## factor(sitter)   4   205.8    51.4  0.6550   0.6346
## factor(session)  4   231.8    57.9  0.7377   0.5840
## Residuals       12   942.5    78.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = statements ~ factor(medium) + factor(sitter) + factor(session), data = mr)
##
## $`factor(medium)`
##      diff        lwr        upr     p adj
## 2-1  15.6  -2.265551  33.465551 0.0983090
## 3-1 -51.8 -69.665551 -33.934449 0.0000069
## 4-1 -36.4 -54.265551 -18.534449 0.0002343
## 5-1   7.4 -10.465551  25.265551 0.6847850
## 3-2 -67.4 -85.265551 -49.534449 0.0000004
## 4-2 -52.0 -69.865551 -34.134449 0.0000066
## 5-2  -8.2 -26.065551   9.665551 0.6026044
## 4-3  15.4  -2.465551  33.265551 0.1042146
## 5-3  59.2  41.334449  77.065551 0.0000016
## 5-4  43.8  25.934449  61.665551 0.0000387
```

There are highly significant differences among the mediums in terms of mean number of statements made. Of the 5(4)/2=10 pairs, six are significantly different based on Tukey's HSD.

∇

```
library(tidyverse)
library(kableExtra)
library(effectsize)
library(agricolae)
library(car)
library(PMCMRplus)
library(additivityTests)
library(lmerTest)
```

```
## Loading required package: lme4
```

```
## Loading required package: Matrix
```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
##
## Attaching package: 'lmerTest'
## The following object is masked from 'package:lme4':
##
##     lmer
## The following object is masked from 'package:stats':
##
##     step
```

Chapter 10

Random and Mixed Effects Models

In some situations, experiments include random treatment factors. In these cases, the treatment/group levels observed in the study represent a sample from a larger population of such levels. Interest is in the overall mean response, as well as in variation in responses across and within treatments/groups.

In this chapter, we consider the **1-Way Random Effects**, the **2-Way Random Effects** and the **2-Way Mixed Effects** Models.

10.1 1-Way Random Effects Model

This model has a single factor, with levels that represent a sample from a population of such levels. The goal is to make inferences regarding the overall mean and variances between groups and within groups. The statistical model for the balanced case is given below.

$$Y_{ij} = \mu_i + \epsilon_{ij} \qquad i = 1, \ldots, r; \quad j = 1, \ldots, n$$

$$\mu_i \sim N\left(\mu_\bullet, \sigma_\mu^2\right) \text{ independent} \qquad \epsilon_{ij} \sim N\left(\mu, \sigma^2\right) \text{ independent} \qquad \{\mu_i\}, \{\epsilon_{ij}\} \text{ independent}$$

For this model, $\mu_\bullet$ is the overall population mean response, σ_μ^2 is the between group variance, and σ^2 is the within group variance. The experiment is conducted as follows.

- Sample r groups from a population of groups.
- Sample n units from within each group.
- Observe the $n_T = rn$ responses.

The mean and variance-covariance structure of the observations are given below.

$$E\left\{Y_{ij}\right\} = E\left\{\mu_i + \epsilon_{ij}\right\} = \mu_\bullet + 0 = \mu_\bullet$$

$$\sigma^2\left\{Y_{ij}\right\} = \sigma^2\left\{\mu_i + \epsilon_{ij}\right\} = \sigma^2\left\{\mu_i\right\} + \sigma^2\left\{\epsilon_{ij}\right\} + 2\sigma\left\{\mu_i, \epsilon_{ij}\right\} =$$
$$= \sigma_\mu^2 + \sigma^2 + 2(0) = \sigma_\mu^2 + \sigma^2 = \sigma_Y^2$$

For two measurements within the same group ($j \neq j'$), we have the following covariance, which differs from the fixed effects case (where the observations are independent).

$$\sigma\left\{Y_{ij}, Y_{ij'}\right\} = \sigma\left\{\mu_i + \epsilon_{ij}, \mu_i + \epsilon_{ij'}\right\} =$$

$$\sigma\left\{\mu_i, \mu_i\right\} + \sigma\left\{\mu_i, \epsilon_{ij'}\right\} + \sigma\left\{\epsilon_{ij}, \mu_i\right\} + \sigma\left\{\epsilon_{ij}, \epsilon_{ij'}\right\} = \sigma_\mu^2 + 0 + 0 + 0 = \sigma_\mu^2$$
$$\sigma\left\{Y_{ij}, Y_{i'j'}\right\} = 0 \quad i \neq i' \quad \forall j, j'$$

For the case, with $r = n = 2$, the mean and variance-covariance structure for Y are given in the following matrix forms.

$$Y = \begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{21} \\ Y_{22} \end{bmatrix} \qquad E\{Y\} = \begin{bmatrix} \mu_\bullet \\ \mu_\bullet \\ \mu_\bullet \\ \mu_\bullet \end{bmatrix} \qquad \sigma^2\{Y\} = \begin{bmatrix} \sigma_Y^2 & \sigma_\mu^2 & 0 & 0 \\ \sigma_\mu^2 & \sigma_Y^2 & 0 & 0 \\ 0 & 0 & \sigma_Y^2 & \sigma_\mu^2 \\ 0 & 0 & \sigma_\mu^2 & \sigma_Y^2 \end{bmatrix}$$

The **Intraclass Correlation Coefficient** measures the proportion of the total variation in responses that is due to variation in group means.

$$\rho_I = \frac{\sigma\left\{Y_{ij}, Y_{ij'}\right\}}{\sigma\left\{Y_{ij}\right\}\sigma\left\{Y_{ij'}\right\}} = \frac{\sigma_\mu^2}{\sigma_Y^2}$$

10.1.1 Analysis of Variance

In this section, we set up the Analysis of Variance and derive the expected mean squares for $MSTR$ and MSE. Note that the quantities are the same for the random effects model as for the fixed effects model, but the expectations differ for the two models.

$$SSTR = n\sum_{i=1}^{r}\left(\overline{Y}_{i\bullet} - \overline{Y}_{\bullet\bullet}\right)^2 = n\sum_{i=1}^{r}\overline{Y}_{i\bullet}^2 - nr\overline{Y}_{\bullet\bullet}^2$$

$$SSE = \sum_{i=1}^{r}\sum_{j=1}^{n}\left(Y_{ij} - \overline{Y}_{i\bullet}\right)^2 = \sum_{i=1}^{r}\sum_{j=1}^{n}Y_{ij}^2 - n\sum_{i=1}^{r}\overline{Y}_{i\bullet}^2$$

$$E\left\{Y_{ij}\right\} = E\left\{\overline{Y}_{i\bullet}\right\} = E\left\{\overline{Y}_{\bullet\bullet}\right\} = \mu_\bullet$$

$$\sigma^2\left\{Y_{ij}\right\} = \sigma_\mu^2 + \sigma^2 = \sigma_Y^2$$

$$\sigma^2\left\{\overline{Y}_{i\bullet}\right\} = \frac{1}{n^2}\left[\sum_{j=1}^{n}\sigma^2\left\{Y_{ij}\right\} + 2\sum_{j=1}^{n-1}\sum_{j'=2}^{n}\sigma\left\{Y_{ij}, Y_{ij'}\right\}\right] =$$

$$= \frac{1}{n^2}\left[n\left(\sigma_\mu^2 + \sigma^2\right) + 2\frac{n(n-1)}{2}\sigma_\mu^2\right] = \frac{n\sigma_\mu^2 + \sigma^2}{n}$$

Note that the sample means for the various groups are independent, which is helpful in deriving the variance of the overall mean.

$$\sigma^2\left\{\overline{Y}_{\bullet\bullet}\right\} = \sigma^2\left\{\frac{1}{r}\sum_{i=1}^{r}\overline{Y}_{i\bullet}\right\} = \frac{1}{r^2}r\frac{n\sigma_\mu^2 + \sigma^2}{n} = \frac{n\sigma_\mu^2 + \sigma^2}{rn}$$

Now, we obtain the expected sums of squares and mean squares.

$$E\left\{Y_{ij}^2\right\} = \sigma_\mu^2 + \sigma^2 + \mu_\bullet^2 \quad \Rightarrow \quad \sum_{i=1}^{r}\sum_{j=1}^{n}E\left\{Y_{ij}^2\right\} = rn\left(\sigma_\mu^2 + \sigma^2 + \mu_\bullet^2\right)$$

$$E\left\{\overline{Y}_{i\bullet}^2\right\} = \frac{n\sigma_\mu^2 + \sigma^2}{n} + \mu_\bullet^2 \quad \Rightarrow \quad n\sum_{i=1}^{r}E\left\{\overline{Y}_{i\bullet}^2\right\} = nr\sigma_\mu^2 + r\sigma^2 + nr\mu_\bullet^2$$

$$E\left\{\overline{Y}_{\bullet\bullet}^2\right\} = \frac{n\sigma_\mu^2 + \sigma^2}{rn} + \mu_\bullet^2 \quad \Rightarrow \quad rnE\left\{\overline{Y}_{\bullet\bullet}^2\right\} = n\sigma_\mu^2 + \sigma^2 + rn\mu_\bullet^2$$

$$E\{SSTR\} = \left(nr\sigma_\mu^2 + r\sigma^2 + nr\mu_\bullet^2\right) - \left(n\sigma_\mu^2 + \sigma^2 + rn\mu_\bullet^2\right) = n(r-1)\sigma_\mu^2 + (r-1)\sigma^2$$

$$E\{SSE\} = \left(rn\left(\sigma_\mu^2 + \sigma^2 + \mu_\bullet^2\right)\right) - \left(nr\sigma_\mu^2 + r\sigma^2 + nr\mu_\bullet^2\right) = r(n-1)\sigma^2$$

The expected mean squares are the expected sums of squares divided by the degrees of freedom with $df_{TR} = r - 1$ and $df_E = r(n-1)$.

$$E\{MSTR\} = E\left\{\frac{SSTR}{df_{TR}}\right\} = \sigma^2 + n\sigma_\mu^2 \qquad\qquad E\{MSE\} = E\left\{\frac{SSE}{df_E}\right\} = \sigma^2$$

The F-test is used to test whether the variance of the population of group means is 0.

$$H_0 : \sigma_\mu^2 = 0 \quad H_A : \sigma_\mu^2 > 0 \qquad TS : F^* = \frac{MSTR}{MSE} \qquad RR : F^* \geq F_{1-\alpha;r-1,r(n-1)}$$

Example 10.1 - Alpha Acids in Varieties of Beer

A study was conducted to measure alpha acids in a sample of $r = 10$ varieties of beer [Meilgaard, 1960]. There were $n = 10$ replicates per variety. The following R code obtains the ANOVA based on direct computations and a plot of the measurements is given in Figure 10.1.

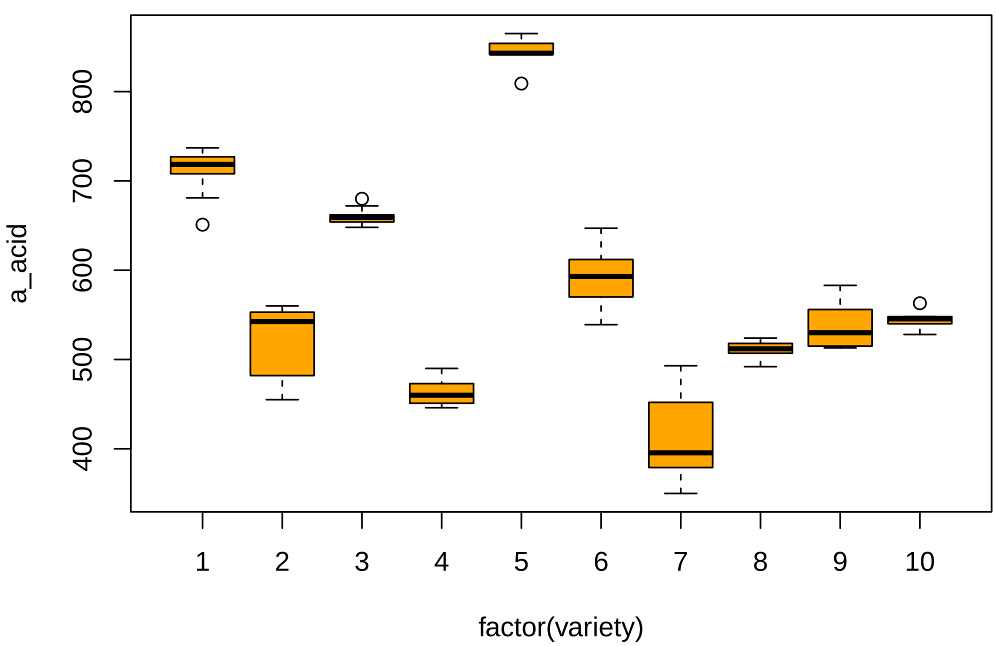


Figure 10.1: Interaction plots for chef jacket study

```
##          df        SS          MS      F* F(.95) P(>F*)
## Variety   9 1463915.9 162657.3211 235.728 1.9856      0
## Error    90   62101.9    690.0211      NA     NA     NA
## Total    99 1526017.8          NA      NA     NA     NA
```

There is strong evidence that the population means (of all varieties) are not all equal, that is $\sigma^2_\mu > 0$.

∇

10.1.2 Estimating the Population Mean $\mu_\bullet$

In this section, we consider estimating the population mean response. The expectations of the individual measurements, group means, and overall mean are all $\mu_\bullet$. However, the overall mean has the smallest variance. Note also that $E\{MSTR\} = \sigma^2 + n\sigma^2_\mu$.

$$\hat{\mu}_\bullet = \overline{Y}_{\bullet\bullet} \qquad \sigma^2\left\{\overline{Y}_{\bullet\bullet}\right\} = \frac{n\sigma^2_\mu + \sigma^2}{rn} \qquad s^2\left\{\overline{Y}_{\bullet\bullet}\right\} = \frac{MSTR}{rn} \qquad s\left\{\overline{Y}_{\bullet\bullet}\right\} = \sqrt{\frac{MSTR}{rn}}$$

$$t = \frac{\overline{Y}_{\bullet\bullet} - \mu_\bullet}{s\left\{\overline{Y}_{\bullet\bullet}\right\}} \sim t_{r-1} \quad \Rightarrow \quad (1-\alpha)100\% \text{ CI for } \mu_\bullet \equiv \overline{Y}_{\bullet\bullet} \pm t_{1-\alpha/2;r-1}\sqrt{\frac{MSTR}{rn}}$$

Example 10.2 - Alpha Acids in Varieties of Beer The following R code obtains the point estimate of the overall mean and 95% Confidence Interval for $\mu_\bullet$ based on direct computations.

```
##      Estimate Std Err df t(.975)       LB       UB
## Mean   579.61 40.3308  9  2.2622 488.3754 670.8446
```

The point estimate is about 580, with a 95% Confidence Interval of 488 to 671. Even though there were $rn = 10(10) = 100$ observations, the between treatment mean square is very large, yielding a wide confidence interval.

∇

10.1.3 Estimating the Intraclass Correlation Coefficient

The Intraclass Correlation Coefficient is $\rho_I = \sigma^2_\mu/\sigma^2_Y$, the ratio of the between group variance to the total variance. To obtain a Confidence Interval for ρ_I, we make use of the following distributional results and the fact that SSE and $SSTR$ are independent.

$$E\{MSE\} = \sigma^2 \qquad \frac{SSE}{\sigma^2} = \frac{r(n-1)MSE}{\sigma^2} \sim \chi^2_{r(n-1)}$$

$$E\{MSTR\} = \sigma^2 + n\sigma^2_\mu \qquad \frac{SSTR}{\sigma^2 + n\sigma^2_\mu} = \frac{(r-1)MSTR}{\sigma^2 + n\sigma^2_\mu} \sim \chi^2_{r-1}$$

$$\frac{\left[\frac{\frac{SSTR}{\sigma^2+n\sigma^2_\mu}}{r-1}\right]}{\left[\frac{\frac{SSE}{\sigma^2}}{r(n-1)}\right]} = \frac{\left[\frac{MSTR}{\sigma^2+n\sigma^2_\mu}\right]}{\left[\frac{MSE}{\sigma^2}\right]} = F^*\left(\frac{\sigma^2}{\sigma^2+n\sigma^2_\mu}\right) \sim F_{r-1,r(n-1)}$$

$$\Rightarrow \quad P\left(F_{\alpha/2;r-1,r(n-1)} \leq F^*\left(\frac{\sigma^2}{\sigma^2+n\sigma^2_\mu}\right) \leq F_{1-\alpha/2;r-1,r(n-1)}\right)$$

The goal is to isolate $\rho_I = \sigma^2_\mu/\sigma^2_Y$ in the middle of the probability statement. Define L and U as follow.

$$L = \frac{1}{n}\left[\frac{F^*}{F_{1-\alpha/2;r-1,r(n-1)}} - 1\right] \qquad U = \frac{1}{n}\left[\frac{F^*}{F_{\alpha/2;r-1,r(n-1)}} - 1\right]$$

Then the $(1-\alpha)100\%$ Confidence Interval for ρ_I is obtained as follows.

$$(1-\alpha)100\% \text{ Confidence Interval for } \rho_I = \frac{\sigma_\mu^2}{\sigma_\mu^2+\sigma^2}: \quad \left[L^* = \frac{L}{1+L}, U^* = \frac{U}{1+U}\right]$$

Example 10.3 - Alpha Acids in Varieties of Beer The following R code obtains the 95% Confidence Interval for ρ_I based on direct computations.

```
##            F* F(.025) F(.975)       L       U     LB     UB
## rho_I 235.728  0.2932  2.2588 10.3361 80.3097 0.9118 0.9877
```

∇

10.1.4 Estimating the Within Group Variance

The withing group variance, σ^2 can be estimated by the mean square error, MSE, as it is an unbiased estimator. Further, a $(1-\alpha)100\%$ Confidence Interval can be obtained as SSE/σ^2 is distributed as Chi-square with $r(n-1)$ degrees of freedom.

$$E\{MSE\} = \sigma^2 \qquad \frac{SSE}{\sigma^2} = \frac{r(n-1)MSE}{\sigma^2} \sim \chi^2_{r(n-1)}$$

$$\Rightarrow \quad P\left(\chi^2_{\alpha/2;r(n-1)} \leq \frac{SSE}{\sigma^2} \leq \chi^2_{1-\alpha/2;r(n-1)}\right) = 1-\alpha$$

$$\Rightarrow \quad P\left(\frac{\chi^2_{\alpha/2;r(n-1)}}{SSE} \leq \frac{1}{\sigma^2} \leq \frac{\chi^2_{1-\alpha/2;r(n-1)}}{SSE}\right) = 1-\alpha$$

$$\Rightarrow \quad P\left(\frac{SSE}{\chi^2_{1-\alpha/2;r(n-1)}} \leq \sigma^2 \leq \frac{SSE}{\chi^2_{\alpha/2;r(n-1)}}\right) = P\left(\frac{r(n-1)MSE}{\chi^2_{1-\alpha/2;r(n-1)}} \leq \sigma^2 \leq \frac{r(n-1)MSE}{\chi^2_{\alpha/2;r(n-1)}}\right) = 1-\alpha$$

The Confidence Interval is simply from SSE divided by the upper critical value to SSE divided by the lower critical value.

Example 10.4 - Alpha Acids in Varieties of Beer The following R code obtains the 95% Confidence Interval for σ^2 based on direct computations.

```
##              MSE X2(.025) X2(.975)       LB      UB
## sigma^2 690.0211  65.6466 118.1359 525.6819 946.003
```

The point estimate is 690, while the 95% Confidence Interval ranges from 525 to 946. In terms of the within group standard deviation, the 95% Confidence Interval ranges from 23 to 31.

∇

10.1.5 Estimating the Between Group Variance

Recall the expected mean squares derived previously.

$$E\{MSTR\} = \sigma^2 + n\sigma_\mu^2 \qquad E\{MSE\} = \sigma^2$$

$$\Rightarrow \quad E\left\{\frac{MSTR-MSE}{n}\right\} = \left(\frac{1}{n}\right)E\{MSTR\} + \left(-\frac{1}{n}\right)E\{MSE\} = \sigma_\mu^2$$

This represents a linear combination of mean squares, each with a known degrees of freedom. **Satterthwaite's Approximation** is based on obtaining an approximation to a chi-square distribution based on linear combinations of mean squares. The approximation is for the degrees of freedom for the (approximate) chi-square distribution. Note that this estimator can be negative, and is usually treated as if it is 0 in practice.

Let $MS_1, \dots, MS_h$ be h mean squares, with degrees of freedom $df_1, \dots, df_h$, respectively. Further, let $c_1, \dots, c_h$, be fixed constants. Consider the following quantities.

$$L = \sum_{i=1}^{h} c_i E\{MS_i\} \qquad \hat{L} = \sum_{i=1}^{h} c_i MS_i \qquad \frac{(df)\,\hat{L}}{L} \stackrel{\bullet}{\sim} \chi^2_{df}$$

This method matches the first two moments of the approximate chi-square distribution to solve for the approximate degrees of freedom. For the chi-square distribution with ν degrees of freedom, the mean is ν and the variance is 2ν. Also, the mean squares are independent.

$$E\left\{\hat{L}\right\} = L \quad \Rightarrow \quad E\left\{\frac{(df)\,\hat{L}}{L}\right\} = df$$

$$\frac{(df_i)\,MS_i}{E\left\{MS_i\right\}} \sim \chi^2_{df_i} \quad \Rightarrow \quad E\left\{\frac{(df_i)\,MS_i}{E\left\{MS_i\right\}}\right\} = df_i \qquad \sigma^2\left\{\frac{(df_i)\,MS_i}{E\left\{MS_i\right\}}\right\} = 2df_i$$

$$\Rightarrow \quad \sigma^2\left\{MS_i\right\} = \frac{2\,(df_i)\,(E\left\{MS_i\right\})^2}{(df_i)^2} \quad \Rightarrow \quad \sigma^2\left\{\hat{L}\right\} = 2\sum_{i=1}^{h} c_i^2 \frac{(E\left\{MS_i\right\})^2}{df_i}$$

$$\sigma^2\left\{\frac{(df)\,\hat{L}}{L}\right\} = 2df \quad \Rightarrow \quad \sigma^2\left\{\hat{L}\right\} = \frac{2(df)L^2}{(df)^2} = \frac{2L^2}{df}$$

Now, equating the two versions of $\sigma^2\left\{\hat{L}\right\}$ and solving for df gives the following result. The unknown expected mean squares in L are replaced with the observed mean squares.

$$\frac{2L^2}{df} = 2\sum_{i=1}^{h} c_i^2 \frac{(E\left\{MS_i\right\})^2}{df_i} \quad \Rightarrow \quad df = \frac{\hat{L}^2}{\sum_{i=1}^{h} c_i^2 \frac{(MS_i)^2}{df_i}} = \frac{\left(\sum_{i=1}^{h} c_i MS_i\right)^2}{\sum_{i=1}^{h} c_i^2 \frac{(MS_i)^2}{df_i}}$$

The approximate $(1-\alpha)100\%$ Confidence Interval is computed as follows.

$$(1-\alpha)100\% \text{ Confidence Interval for } \sigma^2_\mu: \quad \left[\frac{(df)\hat{L}}{\chi_{1-\alpha/2;df}}, \frac{(df)\hat{L}}{\chi_{\alpha/2;df}}\right]$$

Example 10.5 - Alpha Acids in Varieties of Beer The following R code obtains the approximate 95% Confidence Interval for σ^2_μ based on direct computations. Note that $c_1 = 1/n$ and $c_2 = -1/n$.

```
##              s2_mu     df X2(.025) X2(.975)       LB       UB
## sigma2_mu 16196.73 8.9238   2.6597  18.9104 7643.216 54342.17
```

The point estimate is 16197, and the approximate 95% Confidence Interval goes from 7643 to 54342. In terms of the between group standard deviation, the range is approximately from 87 to 233.

We will use the **lmerTest** package and the **lmer** function to fit the model. With this function, we must include a "1" for the fixed mean, and we include "(1|factor(variety))" to include random effects for varieties.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: a_acid ~ 1 + (1 | variety.f)
##    Data: hops
##
## REML criterion at convergence: 981.9
##
## Scaled residuals:
```

```
##      Min       1Q   Median       3Q      Max
## -2.63193 -0.46012  0.00626  0.56199  3.06408
##
## Random effects:
##  Groups    Name        Variance Std.Dev.
##  variety.f (Intercept) 16197    127.27
##  Residual                690     26.27
## Number of obs: 100, groups:  variety.f, 10
##
## Fixed effects:
##             Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)   579.61      40.33   9.00   14.37 1.64e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

∇

10.2 Two-Way Random Effects Model

Some studies include two (or more) random factors. In these cases we want to generalize across populations of units for the two factors and their interactions. In **Measurement Systems Analysis** variation in parts and operators who measure the parts are studied. The experiments are often referred to as **Gage Repeatability & Reproducibility (R&R)** studies. Each of a parts are measured by b operators n times in random order. The variation in part sizes, operator measurements, and interactions between parts and operators are of interest. Repeatability refers to the variation when the same part is measured by the same operator multiple times.

The statistical model for the balanced case is given below.

$$Y_{ijk} = \mu_{\bullet\bullet} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \qquad i = 1, \dots, a; \quad j = 1, \dots, b; \quad k = 1, \dots, n$$

It is assumed that $\{\alpha_i\}$, $\{\beta_j\}$, $\{(\alpha\beta)_{ij}\}$, and $\{\epsilon_{ijk}\}$ are independent and normally distributed with means 0 and variances σ^2_α, σ^2_β, $\sigma^2_{\alpha\beta}$, and σ^2, respectively. The mean, variance and covariance structure are given below.

$$E\left\{Y_{ijk}\right\} = E\left\{\mu_{\bullet\bullet} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}\right\} = \mu_{\bullet\bullet} + 0 + 0 + 0 + 0 = \mu_{\bullet\bullet}$$

$$\sigma^2\left\{Y_{ijk}\right\} = \sigma^2\left\{\mu_{\bullet\bullet} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}\right\} = \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_{\alpha\beta} + \sigma^2$$

$$\sigma\left\{Y_{ijk}, Y_{i'j'k'}\right\} = \begin{cases} \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_{\alpha\beta} + \sigma^2 & : \quad i = i', j = j', k = k' \\ \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_{\alpha\beta} & : \quad i = i', j = j', k \neq k' \\ \sigma^2_\alpha & : \quad i = i', j \neq j', \forall k, k' \\ \sigma^2_\beta & : \quad i \neq i', j = j', \forall k, k' \\ 0 & : \quad i \neq i', j \neq j', \forall k, k' \end{cases}$$

The expected mean squares for each source of variation can be derived from this covariance structure. The results are given here.

$$\text{Factor A:} \quad E\{MSA\} = \sigma^2 + bn\sigma^2_\alpha + n\sigma^2_{\alpha\beta} \qquad \text{Factor B:} \quad E\{MSB\} = \sigma^2 + an\sigma^2_\beta + n\sigma^2_{\alpha\beta}$$

$$\text{AB Interaction:} \quad E\{MSAB\} = \sigma^2 + n\sigma^2_{\alpha\beta} \qquad \text{Error:} \quad E\{MSE\} = \sigma^2$$

The tests for interaction and main effects are obtained from the expected mean squares. Note that the "error" terms for factors A and B are the AB interaction, unlike the fixed effects model, which uses MSE for the error term.

$$H_0^{AB}: \sigma^2_{\alpha\beta} = 0 \qquad H_A^{AB}: \sigma^2_{\alpha\beta} > 0 \qquad TS: F_{AB} = \frac{MSAB}{MSE} \qquad RR: F_{AB} \geq F_{1-\alpha;(a-1)(b-1),ab(n-1)}$$

$$H_0^A : \sigma_\alpha^2 = 0 \qquad H_A^A : \sigma_\alpha^2 > 0 \qquad TS : F_A = \frac{MSA}{MSAB} \qquad RR : F_A \geq F_{1-\alpha;a-1,(a-1)(b-1)}$$

$$H_0^B : \sigma_\beta^2 = 0 \qquad H_A^B : \sigma_\beta^2 > 0 \qquad TS : F_B = \frac{MSB}{MSAB} \qquad RR : F_B \geq F_{1-\alpha;b-1,(a-1)(b-1)}$$

Point estimates for the variance components are obtained from the expected mean squares. In practice, some of these can be negative, and in practice are typically replaced by 0.

$$s^2 = MSE \qquad s_{\alpha\beta}^2 = \frac{MSAB - MSE}{n} \qquad s_\alpha^2 = \frac{MSA - MSAB}{bn} \qquad s_\beta^2 = \frac{MSB - MSAB}{an}$$

As these are all linear functions of mean squares, Satterthwaite's approximation can be used to obtain approximate degrees' of freedom and Confidence Intervals for $\sigma_{\alpha\beta}^2$, σ_α^2, and σ_β^2.

$$df_{AB} = \frac{\left(\frac{1}{n}MSAB + \frac{-1}{n}MSE\right)^2}{\frac{\left(\frac{1}{n}MSAB\right)^2}{(a-1)(b-1)} + \frac{\left(\frac{-1}{n}MSE\right)^2}{ab(n-1)}}$$

$$df_A = \frac{\left(\frac{1}{bn}MSA + \frac{-1}{bn}MSAB\right)^2}{\frac{\left(\frac{1}{bn}MSA\right)^2}{a-1} + \frac{\left(\frac{-1}{bn}MSAB\right)^2}{(a-1)(b-1)}} \qquad df_B = \frac{\left(\frac{1}{an}MSB + \frac{-1}{an}MSAB\right)^2}{\frac{\left(\frac{1}{an}MSB\right)^2}{b-1} + \frac{\left(\frac{-1}{an}MSAB\right)^2}{(a-1)(b-1)}}$$

Example 10.6 - Measurements of Skulls by Observers

An anthropological study was conducted to measure variation in frontal chord measurements of skulls by various observers [Hanihara et al., 1999]. There were $a = 10$ skulls, $b = 6$ observers, and $n = 2$ replicates per skull/observer. The R code for the analysis and output are included here.

```
##   observer skull time chord
## 1        1     1    1 100.9
## 2        1     1    2 100.8
## 3        1     2    1  91.1
## 4        1     2    2  91.2
## 5        1     3    1 102.5
## 6        1     3    2 102.7
```

```
##     observer skull time chord
## 115        6     8    1  99.0
## 116        6     8    2  99.0
## 117        6     9    1  91.7
## 118        6     9    2  91.8
## 119        6    10    1  93.7
## 120        6    10    2  93.8
```

```
## Analysis of Variance Table
##
## Response: chord
##                   Df  Sum Sq Mean Sq   F value    Pr(>F)
## skull.f            9 1612.01 179.112 3074.8786 < 2.2e-16 ***
## observer.f         5    7.32   1.463   25.1207 1.449e-13 ***
## skull.f:observer.f 45    5.26   0.117    2.0065  0.005956 **
## Residuals         60    3.49   0.058
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                  MS        F* F(.95) P(>F*)     s^2 Satt df     LB      UB
## sigma_a^2  179.1117 1532.4942 2.0958  0.000 14.9162  8.9883 7.0543 49.7643
## sigma_b^2    1.4633   12.5200 2.4221  0.000  0.0673  4.2302 0.0247  0.5102
## sigma_ab^2   0.1169    2.0065 1.5749  0.006  0.0293  9.5444 0.0141  0.0934
## sigma^2      0.0582        NA     NA     NA  0.0582 60.0000 0.0420  0.0863
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## chord ~ 1 + (1 | skull.f) + (1 | observer.f) + (1 | skull.f:observer.f)
##    Data: sm
##
## REML criterion at convergence: 123.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.90775 -0.49619  0.01282  0.41430  2.21863
##
## Random effects:
##  Groups             Name        Variance Std.Dev.
##  skull.f:observer.f (Intercept)  0.02931 0.1712
##  skull.f            (Intercept) 14.91283 3.8617
##  observer.f         (Intercept)  0.06732 0.2595
##  Residual                        0.05825 0.2414
## Number of obs: 120, groups:
## skull.f:observer.f, 60; skull.f, 10; observer.f, 6
##
## Fixed effects:
##             Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)   96.464      1.226  9.138   78.67 2.97e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The variance of the skulls is by far the largest (not surprising) while the variances among observers, interactions and error are much smaller.

In the measurement systems analysis, the variances are broken into various components.

- **Repeatability** - Variation within same part/operator (skull/observer): σ^2
- **Reproducibility** - Variation across parts/operators: $\sigma^2_\beta + \sigma^2_{\alpha\beta}$
- **Parts** - Variation across parts: σ^2_α
- **Gage** - Repeatability + Reproducibility
- **Total** - Parts + Gage
- **% R&R** - $100 * \sigma_{\text{Gage}}/\sigma_{\text{Total}}$ in terms of standard deviations, not variances.

For this example (based on the lmer results), we get the following estimates.

$$s^2_{\text{Repeatability}} = s^2 = 0.05825 \qquad s^2_{\text{Reproducibility}} = s^2_\beta + s^2_{\alpha\beta} = 0.06732 + 0.02931 = 0.09663$$

$$s^2_{\text{Parts}} = 14.92183 \qquad s^2_{\text{Gage}} = 0.05825 + 0.09963 = 0.015488$$

$$s^2_{\text{Total}} = 14.92183 + 0.015488 = 15.07671 \qquad \%\ \text{R\&R:}\quad 100\sqrt{\frac{0.015488}{15.07671}} = 3.21\%$$

Only about 3.2% of the total variation (standard deviation) is due to measurement error.

∇

10.3 Two-Way Mixed Effects Model

Models can include both fixed and random effects. In this section, we describe the **unrestricted** two-way mixed model, with factor A fixed and factor B random. The **restricted** model has the two-factor interactions sum to 0 over the fixed factor. Statistical software packages typically fit the unrestricted model that allows for unbalanced data. The model is given below.

$$Y_{ijk} = \mu_{\bullet\bullet} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk} \qquad i = 1, \dots, a; \quad j = 1, \dots, b; \quad k = 1, \dots, n$$

$$\sum_{i=1}^{a} \alpha_i = 0 \qquad \beta_j \sim NID\left(0, \sigma_\beta^2\right) \qquad (\alpha\beta)_{ij} \sim NID\left(0, \sigma_{\alpha\beta}^2\right) \qquad \epsilon_{ijk} \sim NID\left(0, \sigma^2\right)$$

The random effects $\left\{\beta_j\right\}$, $\left\{(\alpha\beta)_{ij}\right\}$, and error terms $\left\{\epsilon_{ijk}\right\}$ are assumed pairwise independent.

The mean, covariance structure, expected mean squares and estimators for the unrestricted mixed model are given below.

$$E\left\{Y_{ijk}\right\} = \mu_{\bullet\bullet} + \alpha_i \qquad \sigma\left\{Y_{ijk}, Y_{i'j'k'}\right\} = \begin{cases} \sigma_\beta^2 + \sigma_{\alpha\beta}^2 + \sigma^2 & : \quad i = i', j = j', k = k' \\ \sigma_\beta^2 + \sigma_{\alpha\beta}^2 & : \quad i = i', j = j', k \neq k' \\ \sigma_\beta^2 & : \quad i \neq i', j = j', \forall k, k' \\ 0 & : \quad \forall i, i', j \neq j', \forall k, k' \end{cases}$$

$$E\{MSA\} = \sigma^2 + n\sigma_{\alpha\beta}^2 + \frac{bn\sum_{i=1}^{a}\alpha_i^2}{a-1} \qquad E\{MSB\} = \sigma^2 + n\sigma_{\alpha\beta}^2 + an\sigma_\beta^2$$
$$E\{MSAB\} = \sigma^2 + n\sigma_{\alpha\beta}^2 \qquad E\{MSE\} = \sigma^2$$

$$s^2 = MSE \qquad s_{\alpha\beta}^2 = \frac{MSAB - MSE}{n} \qquad s_\beta^2 = \frac{MSB - MSAB}{an}$$

$$\sigma^2\left\{\overline{Y}_{i\bullet\bullet}\right\} = \frac{\sigma^2 + n\sigma_{\alpha\beta}^2}{bn} \quad \Rightarrow \quad s\left\{\overline{Y}_{i\bullet\bullet}\right\} = \sqrt{\frac{MSAB}{bn}} \qquad s\left\{\overline{Y}_{i\bullet\bullet} - \overline{Y}_{i'\bullet\bullet}\right\} = \sqrt{\frac{2MSAB}{bn}}$$
$$\sigma^2\left\{\overline{Y}_{\bullet\bullet\bullet}\right\} = \frac{\sigma^2 + n\sigma_{\alpha\beta}^2}{abn} \quad \Rightarrow \quad s\left\{\overline{Y}_{\bullet\bullet\bullet}\right\} = \sqrt{\frac{MSAB}{abn}}$$

Tukey's method can be used to compare all levels of factor A.

$$HSD = q_{1-\alpha;a,(a-1)(b-1)}\sqrt{\frac{MSAB}{bn}}$$
$$(1-\alpha)100\% \text{ Confidence Interval for } \alpha_i - \alpha_{i'}: \quad \left(\overline{Y}_{i\bullet\bullet} - \overline{Y}_{i'\bullet\bullet}\right) \pm HSD$$

Example 10.7 - Breath Alcohol Concentration Measurement Comparison

A study compared $a = 6$ (fixed) instruments among $b = 3$ (random) subjects in measuring breath alcohol levels [Gullberg, 2008]. Each subject was measured by each machine $n = 10$ times. Scores have been multiplied by 100 to avoid very small variances.

R code for direct calculations and the **lmer** function are used for the analysis. The confidence level for the **difflsmeans** has been adjusted by the Bonferroni method to construct simultaneous 95% Confidence Intervals among the $6(6-1)/2 = 15$ pairs of instruments.

$$s\left\{\overline{Y}_{i\bullet\bullet} - \overline{Y}_{i'\bullet\bullet}\right\} = \sqrt{\frac{2MSAB}{bn}} = \sqrt{\frac{2(0.3083)}{3(10)}} = 0.1434 \qquad t_{1-.05/(2(15));10} = 3.827$$

$$95\% \text{ CI for } \alpha_i - \alpha_{i'}: \quad \left(\overline{Y}_{i\bullet\bullet} - \overline{Y}_{i'\bullet\bullet}\right) \pm 3.827(0.1434)$$
$$\equiv \quad 95\% \text{ CI for } \alpha_i - \alpha_{i'}: \quad \left(\overline{Y}_{i\bullet\bullet} - \overline{Y}_{i'\bullet\bullet}\right) \pm 0.549$$

Note that the P-values are not adjusted for the multiple tests.

```
##   instrument subject rep breathAlc
## 1          1       1   1    0.0938
## 2          1       1   2    0.0946
## 3          1       1   3    0.0944
## 4          1       1   4    0.0943
## 5          1       1   5    0.0922
## 6          1       1   6    0.0924

##     instrument subject rep breathAlc
## 175          6       2   5    0.0612
## 176          6       2   6    0.0616
## 177          6       2   7    0.0603
## 178          6       2   8    0.0597
## 179          6       2   9    0.0599
## 180          6       2  10    0.0590

##   instrument subject rep breathAlc
## 1          1       1   1    0.0938
## 2          1       1   2    0.0946
## 3          1       1   3    0.0944
## 4          1       1   4    0.0943
## 5          1       1   5    0.0922
## 6          1       1   6    0.0924

##     instrument subject rep breathAlc
## 165          6       3   5    0.0738
## 166          6       3   6    0.0717
## 167          6       3   7    0.0713
## 168          6       3   8    0.0703
## 169          6       3   9    0.0697
## 170          6       3  10    0.0745

##             df       SS       MS       F* F(.95) P(>F*)
## Instrument   5  19.0059   3.8012  12.3290 3.3258  5e-04
## Subject      2 219.0440 109.5220 355.2307 4.1028  0e+00
## InstxSubj   10   3.0831   0.3083   5.5595 1.8896  0e+00
## Error      162   8.9840   0.0555       NA     NA     NA
## Total      179 250.1171       NA       NA     NA     NA

##              s^2  Satt df     LB      UB
## Subject   1.8202   1.9888 0.4922 73.1909
## InstxSubj 0.0253   6.7127 0.0109  0.1091
## Error     0.0555 162.0000 0.0451  0.0698

##       i i' Ybar_i Ybar_i'    Diff      LB      UB  p adj
##  [1,] 2  1 7.3547  7.8340 -0.4793 -0.9773  0.0186 0.0609
##  [2,] 3  1 7.6247  7.8340 -0.2093 -0.7073  0.2886 0.6941
##  [3,] 3  2 7.6247  7.3547  0.2700 -0.2280  0.7680 0.4621
##  [4,] 4  1 6.8353  7.8340 -0.9987 -1.4966 -0.5007 0.0004
##  [5,] 4  2 6.8353  7.3547 -0.5193 -1.0173 -0.0214 0.0399
##  [6,] 4  3 6.8353  7.6247 -0.7893 -1.2873 -0.2914 0.0026
##  [7,] 5  1 7.5943  7.8340 -0.2397 -0.7376  0.2583 0.5760
##  [8,] 5  2 7.5943  7.3547  0.2397 -0.2583  0.7376 0.5760
##  [9,] 5  3 7.5943  7.6247 -0.0303 -0.5283  0.4676 0.9999
## [10,] 5  4 7.5943  6.8353  0.7590  0.2610  1.2570 0.0034
## [11,] 6  1 7.2090  7.8340 -0.6250 -1.1230 -0.1270 0.0131
## [12,] 6  2 7.2090  7.3547 -0.1457 -0.6436  0.3523 0.9023
## [13,] 6  3 7.2090  7.6247 -0.4157 -0.9136  0.0823 0.1188
## [14,] 6  4 7.2090  6.8353  0.3737 -0.1243  0.8716 0.1818
```

```
## [15,] 6  5 7.2090  7.5943 -0.3853 -0.8833  0.1126 0.1618
```

```
## Analysis of Variance Table
##
## Response: 100 * breathAlc
##                                    Df  Sum Sq Mean Sq   F value    Pr(>F)
## factor(instrument)                  5  19.006   3.801   68.5430 < 2.2e-16
## factor(subject)                     2 219.044 109.522 1974.9028 < 2.2e-16
## factor(instrument):factor(subject)  10   3.083   0.308    5.5595 4.177e-07
## Residuals                         162   8.984   0.055
##
## factor(instrument)                 ***
## factor(subject)                    ***
## factor(instrument):factor(subject) ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## 100 * breathAlc ~ instrument.f + (1 | subject.f) + (1 | instrument.f:subject.f)
##    Data: ba
##
## REML criterion at convergence: 46.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.7063 -0.5975  0.0655  0.5337  3.5359
##
## Random effects:
##  Groups                 Name        Variance Std.Dev.
##  instrument.f:subject.f (Intercept) 0.02529  0.1590
##  subject.f              (Intercept) 1.82025  1.3492
##  Residual                           0.05546  0.2355
## Number of obs: 180, groups:  instrument.f:subject.f, 18; subject.f, 3
##
## Fixed effects:
##               Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)    7.40867    0.78004  1.99996   9.498 0.010905 *
## instrument.f1  0.42533    0.09254  9.99985   4.596 0.000986 ***
## instrument.f2 -0.05400    0.09254  9.99985  -0.584 0.572476
## instrument.f3  0.21600    0.09254  9.99985   2.334 0.041762 *
## instrument.f4 -0.57333    0.09254  9.99985  -6.195 0.000102 ***
## instrument.f5  0.18567    0.09254  9.99985   2.006 0.072627 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) inst.1 inst.2 inst.3 inst.4
## instrmnt.f1  0.000
## instrmnt.f2  0.000 -0.200
## instrmnt.f3  0.000 -0.200 -0.200
## instrmnt.f4  0.000 -0.200 -0.200 -0.200
## instrmnt.f5  0.000 -0.200 -0.200 -0.200 -0.200
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF  DenDF F value    Pr(>F)
```

```
## instrument.f 3.4186 0.68372     5 9.9999  12.329 0.0005164 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Least Squares Means table:
##
##                                Estimate Std. Error df t value     lower
## instrument.f1 - instrument.f2  0.479333   0.143368 10  3.3434 -0.069387
## instrument.f1 - instrument.f3  0.209333   0.143368 10  1.4601 -0.339387
## instrument.f1 - instrument.f4  0.998667   0.143368 10  6.9658  0.449947
## instrument.f1 - instrument.f5  0.239667   0.143368 10  1.6717 -0.309053
## instrument.f1 - instrument.f6  0.625000   0.143368 10  4.3594  0.076280
## instrument.f2 - instrument.f3 -0.270000   0.143368 10 -1.8833 -0.818720
## instrument.f2 - instrument.f4  0.519333   0.143368 10  3.6224 -0.029387
## instrument.f2 - instrument.f5 -0.239667   0.143368 10 -1.6717 -0.788387
## instrument.f2 - instrument.f6  0.145667   0.143368 10  1.0160 -0.403053
## instrument.f3 - instrument.f4  0.789333   0.143368 10  5.5056  0.240613
## instrument.f3 - instrument.f5  0.030333   0.143368 10  0.2116 -0.518387
## instrument.f3 - instrument.f6  0.415667   0.143368 10  2.8993 -0.133053
## instrument.f4 - instrument.f5 -0.759000   0.143368 10 -5.2941 -1.307720
## instrument.f4 - instrument.f6 -0.373667   0.143368 10 -2.6063 -0.922387
## instrument.f5 - instrument.f6  0.385333   0.143368 10  2.6877 -0.163387
##                                   upper  Pr(>|t|)
## instrument.f1 - instrument.f2  1.028053 0.0074481 **
## instrument.f1 - instrument.f3  0.758053 0.1749512
## instrument.f1 - instrument.f4  1.547387 3.873e-05 ***
## instrument.f1 - instrument.f5  0.788387 0.1255364
## instrument.f1 - instrument.f6  1.173720 0.0014225 **
## instrument.f2 - instrument.f3  0.278720 0.0890443 .
## instrument.f2 - instrument.f4  1.068053 0.0046713 **
## instrument.f2 - instrument.f5  0.309053 0.1255364
## instrument.f2 - instrument.f6  0.694387 0.3335656
## instrument.f3 - instrument.f4  1.338053 0.0002597 ***
## instrument.f3 - instrument.f5  0.579053 0.8366871
## instrument.f3 - instrument.f6  0.964387 0.0158528 *
## instrument.f4 - instrument.f5 -0.210280 0.0003505 ***
## instrument.f4 - instrument.f6  0.175053 0.0262049 *
## instrument.f5 - instrument.f6  0.934053 0.0227889 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##   Confidence level: 100%
##   Degrees of freedom method: Satterthwaite
```

∇

10.4 Three-Way Mixed Effects Models

In this section, the Three-Way mixed model with two fixed factors and a single random factor (typically subjects) is described. In many cases, there is a single replicate per combination of factors ($n = 1$) and this can be thought of as an extension of a Randomized Block/Repeated Measures Design. Here is the case when $n = 1$, factors A and B are fixed, and factor C is random (Subject).

$$SSA = bc\sum_{i=1}^{a}\left(\overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet\bullet\bullet}\right)^2 \qquad df_A = a - 1$$

$$E\{MSA\} = \sigma^2 + \sigma^2_{\alpha\beta\gamma} + b\sigma^2_{\alpha\gamma} + \frac{bc\sum_{i=1}^{a}\alpha_i^2}{a-1} = E\{MSAC\} + \frac{bc\sum_{i=1}^{a}\alpha_i^2}{a-1}$$

$$SSB = ac\sum_{j=1}^{b}\left(\overline{Y}_{\bullet j\bullet} - \overline{Y}_{\bullet\bullet\bullet}\right)^2 \qquad df_B = b-1$$

$$E\{MSB\} = \sigma^2 + \sigma^2_{\alpha\beta\gamma} + a\sigma^2_{\beta\gamma} + \frac{ac\sum_{j=1}^{b}\beta_j^2}{b-1} = E\{MSBC\} + \frac{ac\sum_{j=1}^{b}\beta_j^2}{b-1}$$

$$SSAB = c\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{Y}_{ij\bullet} - \overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet j\bullet} + \overline{Y}_{\bullet\bullet\bullet}\right)^2 \qquad df_{AB} = (a-1)(b-1)$$

$$E\{MSAB\} = \sigma^2 + \sigma^2_{\alpha\beta\gamma} + \frac{c\sum_{i=1}^{a}\sum_{j=1}^{b}(\alpha\beta)^2_{ij}}{(a-1)(b-1)} = E\{MSABC\} + \frac{c\sum_{i=1}^{a}\sum_{j=1}^{b}(\alpha\beta)^2_{ij}}{(a-1)(b-1)}$$

The error terms for the fixed main effects and their 2-way interaction are the interactions between each term and the random factor C.

Tukey's HSD or the Bonferroni method can be used to compare the levels of factors A and B, or the ab treatment means if the interaction is important.

Example 10.8 - Navigational Techniques for Web Maps A study was conducted to measure effects of two factors on subjects' skill on reading web maps [Wu et al., 2011]. The factors included are given below.

- **Navigational Technique** - Combined Panning Buttons (CPB, $i = 1$), Distributed Panning Buttons (DPB, $i = 2$), Enhanced Navigaor with Continuous Control (ENCC, $i = 3$), and Grab & Drag (GD, $i = 4$)
- **Input Method** - Touch Screen (DT, $j = 1$), Mouse (M, $j = 2$)
- **Subject** - 36 subjects who each performed once on the 8 combinations of Input Method and Navigational Technique

One response was task completion time (seconds). We will analyze the data directly and with the **lmer** function below.

```
##   navTech inputMetg subject taskTime
## 1       1         1       1   163.30
## 2       1         1       2   214.95
## 3       1         1       3   179.73
## 4       1         1       4   164.35
## 5       1         1       5   184.68
## 6       1         1       6   165.21
```

```
##     navTech inputMetg subject taskTime
## 283       4         2      31   155.26
## 284       4         2      32   172.36
## 285       4         2      33   147.22
## 286       4         2      34   105.90
## 287       4         2      35    73.22
## 288       4         2      36    78.10
```

```
##            df       SS        MS Err df    Err MS      F* F(.95) P(>F*)
## Navigation  3 66995.51 22331.838    105 1417.1179 15.7586 2.6911  0e+00
## Input       1 30635.76 30635.757     35 1960.1337 15.6294 4.1213  4e-04
## Inp/Nav     3 18710.29  6236.762    105  926.4971  6.7316 2.6911  3e-04
```

```
## Analysis of Variance Table
##
## Response: taskTime
##                        Df Sum Sq Mean Sq F value    Pr(>F)
## navTech.f               3  66996 22331.8 24.1035 6.033e-12 ***
## inputMet.f              1  30636 30635.8 33.0662 8.840e-08 ***
## subject.f              35  84008  2400.2  2.5906 0.0001025 ***
```

```
## navTech.f:inputMet.f   3  18710  6236.8  6.7316 0.0003376 ***
## navTech.f:subject.f  105 148797  1417.1  1.5295 0.0152294 *
## inputMet.f:subject.f  35  68605  1960.1  2.1156 0.0018646 **
## Residuals            105  97282   926.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In this case the main effects and the interaction are all very significant, so we will use Tukey's HSD and the Bonferroni minimum significant difference method to compare the 8 treatments (8(8-1)/2=28 comparisons). These will use $MSABC$ as the "error term" with $df = (4-1)(2-1)(36-1) = 105$.

$$\text{Tukey:} \quad HSD = q_{.95,8,105}\sqrt{\frac{MSABC}{c}} = 4.374\sqrt{\frac{926.4971}{36}} = 22.19$$

$$\text{Bonferroni:} \quad MSD = t_{1-.05/(2(28));105}\sqrt{\frac{2MSABC}{c}} = 3.206\sqrt{\frac{2(926.4971)}{36}} = 23.00$$

```
##   navTech inputMetg subject taskTime
## 1       1         1       1   163.30
## 2       1         1       2   214.95
## 3       1         1       3   179.73
## 4       1         1       4   164.35
## 5       1         1       5   184.68
## 6       1         1       6   165.21
```

```
##     navTech inputMetg subject taskTime
## 283       4         2      31   155.26
## 284       4         2      32   172.36
## 285       4         2      33   147.22
## 286       4         2      34   105.90
## 287       4         2      35    73.22
## 288       4         2      36    78.10
```

```
## boundary (singular) fit: see ?isSingular
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## taskTime ~ navTech.f * inputMet.f + (1 | subject.f) + (1 | navTech.f:subject.f) +
##     (1 | inputMet.f:subject.f)
##    Data: ntm
##
## REML criterion at convergence: 2851.3
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.14082 -0.60108  0.01166  0.59746  2.75578
##
## Random effects:
##  Groups               Name        Variance Std.Dev.
##  navTech.f:subject.f  (Intercept) 243.0    15.59
##  inputMet.f:subject.f (Intercept) 253.6    15.92
##  subject.f            (Intercept)   0.0     0.00
##  Residual                         927.9    30.46
## Number of obs: 288, groups:
## navTech.f:subject.f, 144; inputMet.f:subject.f, 72; subject.f, 36
##
## Fixed effects:
```

```
##                         Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)              147.964      2.904  78.527  50.957  < 2e-16 ***
## navTech.f1                15.511      3.838 112.449   4.042 9.75e-05 ***
## navTech.f2                14.616      3.838 112.449   3.809 0.000228 ***
## navTech.f3               -18.413      3.838 112.449  -4.798 4.96e-06 ***
## inputMet.f1               10.314      2.597  54.542   3.972 0.000211 ***
## navTech.f1:inputMet.f1     9.371      3.109 107.587   3.014 0.003214 **
## navTech.f2:inputMet.f1     5.476      3.109 107.587   1.762 0.080985 .
## navTech.f3:inputMet.f1   -11.414      3.109 107.587  -3.671 0.000378 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) nvTc.1 nvTc.2 nvTc.3 inpM.1 nT.1:M nT.2:M
## navTech.f1   0.000
## navTech.f2   0.000 -0.333
## navTech.f3   0.000 -0.333 -0.333
## inputMet.f1  0.000  0.000  0.000  0.000
## nvTch.1:M.1  0.000  0.000  0.000  0.000  0.000
## nvTch.2:M.1  0.000  0.000  0.000  0.000  0.000 -0.333
## nvTch.3:M.1  0.000  0.000  0.000  0.000  0.000 -0.333 -0.333
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see ?isSingular

## Type III Analysis of Variance Table with Satterthwaite's method
##                       Sum Sq Mean Sq NumDF   DenDF F value    Pr(>F)
## navTech.f              43964 14654.6     3 112.449 15.7940 1.228e-08 ***
## inputMet.f             14636 14635.9     1  54.542 15.7738 0.0002109 ***
## navTech.f:inputMet.f   18710  6236.8     3 107.587  6.7216 0.0003360 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

∇

```
library(tidyverse)
library(kableExtra)
library(effectsize)
library(agricolae)
library(car)
library(PMCMRplus)
library(additivityTests)
library(lmerTest)
library(nlme)
```

```
##
## Attaching package: 'nlme'

## The following object is masked from 'package:lme4':
##
##     lmList

## The following object is masked from 'package:dplyr':
##
##     collapse
```

Chapter 11

Nested Designs

So far, factorial designs have contained **crossed** factors. The levels of factor B were the same within the levels of factor A, and vice versa. In this chapter the levels of factor B will be different under the various levels of factor A, and are said to be **nested**. These models are often referred to as hierarchical or multilevel.

In a nested model, with factor A at the "top," the levels of factor B will differ within the levels of factor A. Consider a study to compare NCAA conferences, with the conference as factor A, and schools are factor B, with different schools being in the different conferences (although they are changing constantly).

11.1 Estimators and the Analysis of Variance

The factors can be fixed (all levels included), random (levels are sampled), or mixed (one factor fixed, the other random). We will write out the general model, and whether effects are fixed or random should be clear from the setting of the experiment. In the two-factor balanced design, factor A will have a levels, within each level of A, factor B will have b levels, and there will be n replicates within each of the ab treatments. The model is given here.

$$\text{Factor A:} \quad \mu_{i\bullet} = \mu_{\bullet\bullet} + \alpha_i \qquad i = 1, \dots, a$$

$$\text{Factor B within A:} \quad \mu_{ij} = \mu_{\bullet\bullet} + \alpha_i + \beta_{j(i)} \qquad i = 1, \dots, a; \quad j = 1, \dots, b$$

$$\Rightarrow \quad \beta_{j(i)} = \mu_{ij} - (\mu_{\bullet\bullet} + \alpha_i) = \mu_{ij} - \mu_{i\bullet}$$

When A and B are both fixed factors, all levels of interest are included and $\{\alpha_i\}$ and $\left\{\beta_{j(i)}\right\}$ are all fixed (unknown) parameters and we can use the following restrictions and model.

$$\sum_{i=1}^{a} \alpha_i = 0 \qquad \sum_{j=1}^{b} \beta_{j(i)} = 0 \quad i = 1, \dots, a$$

$$Y_{ijk} = \mu_{\bullet\bullet} + \alpha_i + \beta_{j(i)} + \epsilon_{ijk} \qquad i = 1, \dots, a; \quad j = 1, \dots, b; \quad k = 1, \dots, n$$

$$\epsilon_{ijk} \sim NID\left(0, \sigma^2\right) \qquad E\left\{Y_{ijk}\right\} = \mu_{\bullet\bullet} + \alpha_i + \beta_{j(i)} \qquad \sigma^2\left\{Y_{ijk}\right\} = \sigma^2$$

When factor A is fixed and factor B is random, the model changes as follows.

$$\sum_{i=1}^{a} \alpha_i = 0 \qquad \beta_{j(i)} \sim NID\left(0, \sigma_\beta^2\right) \qquad \left\{\beta_{j(i)}\right\} \perp \left\{\epsilon_{ijk}\right\}$$

$$E\left\{Y_{ijk}\right\} = \mu_{\bullet\bullet} + \alpha_i \qquad \sigma^2\left\{Y_{ijk}\right\} = \sigma_\beta^2 + \sigma^2$$

$$\sigma\left\{Y_{ijk}, Y_{i'j'k'}\right\} = \begin{cases} \sigma_\beta^2 + \sigma^2 & : \quad i = i', j = j', k = k' \\ \sigma_\beta^2 & : \quad i = i', j = j', k \neq k' \\ 0 & : \quad \text{otherwise} \end{cases}$$

When factors A and B are both random, the model changes as follows.

$$\alpha_i \sim NID\left(0, \sigma_\alpha^2\right) \qquad \beta_{j(i)} \sim NID\left(0, \sigma_\beta^2\right) \qquad \{\alpha_i\} \perp \left\{\beta_{j(i)}\right\} \perp \{\epsilon_{ijk}\}$$

$$E\left\{Y_{ijk}\right\} = \mu_{\bullet\bullet} \qquad \sigma^2\left\{Y_{ijk}\right\} = \sigma_\alpha^2 + \sigma_\beta^2 + \sigma^2$$

$$\sigma\left\{Y_{ijk}, Y_{i'j'k'}\right\} = \begin{cases} \sigma_\alpha^2 + \sigma_\beta^2 + \sigma^2 & : \quad i = i', j = j', k = k' \\ \sigma_\alpha^2 + \sigma_\beta^2 & : \quad i = i', j = j', k \neq k' \\ \sigma_\alpha^2 & : \quad i = i', j \neq j', \forall k, k' \\ 0 & : \quad i \neq i', \forall j, j', \forall k, k' \end{cases}$$

Estimators, fitted values, residuals, sums of squares and expected mean squares are given below.

$$\hat{\mu}_{\bullet\bullet} = \overline{Y}_{\bullet\bullet\bullet} \qquad \hat{\alpha}_i = \overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet\bullet\bullet} \qquad \hat{\beta}_{j(i)} = \overline{Y}_{ij\bullet} - \overline{Y}_{i\bullet\bullet}$$

$$\hat{Y}_{ijk} = \hat{\mu}_{\bullet\bullet} + \hat{\alpha}_i + \hat{\beta}_{j(i)} = \overline{Y}_{ij\bullet} \qquad e_{ijk} = Y_{ijk} - \hat{Y}_{ijk} = Y_{ijk} - \overline{Y}_{ij\bullet}$$

$$\text{Error:} \quad SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}\left(Y_{ijk} - \overline{Y}_{ij\bullet}\right)^2 \qquad df_E = ab(n-1)$$

$$\text{Factor A:} \quad SSA = bn\sum_{i=1}^{a}\left(Y_{i\bullet\bullet} - Y_{\bullet\bullet\bullet}\right)^2 \qquad df_A = a-1$$

$$\text{Factor B within A:} \quad SSB(A) = n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(Y_{ij\bullet} - Y_{i\bullet\bullet}\right)^2 \qquad df_{B(A)} = a(b-1)$$

The expected mean squares depend on whether the factors are fixed or random. In all cases, $E\{MSE\} = \sigma^2$. When both A and B are fixed, we obtain the following results.

$$E\{MSA\} = \sigma^2 + \frac{bn\sum_{i=1}^{a}\alpha_i^2}{a-1} \qquad E\{MSB(A)\} = \sigma^2 + \frac{n\sum_{i=1}^{a}\sum_{j=1}^{b}\beta_{j(i)}^2}{a(b-1)}$$

The denominator for the F-tests for factors A and B in the fixed case is MSE.

When factor A is fixed and factor B is random, the following expected mean squares are obtained.

$$E\{MSA\} = \sigma^2 + n\sigma_{\beta(\alpha)}^2 + \frac{bn\sum_{i=1}^{a}\alpha_i^2}{a-1} \qquad E\{MSB(A)\} = \sigma^2 + n\sigma_{\beta(\alpha)}^2$$

In this mixed case, the error term for factor B is MSE, but the error term for factor A is $MSB(A)$ due to the covariance structure.

When factors A and B are both random, we obtain the following expected mean squares.

$$E\{MSA\} = \sigma^2 + n\sigma_\beta^2 + bn\sigma_\alpha^2 \qquad E\{MSB(A)\} = \sigma^2 + n\sigma_{\beta(\alpha)}^2$$

The error terms are the same in the random effects case are the same as in the mixed effects case. Variance components can be estimated as in the previous chapter.

11.2 Tests, Contrasts, and Pairwise Comparisons

When both factors are fixed, MSE is an unbiased estimator of $\sigma^2\left\{Y_{ijk}\right\}$ and we can make estimate contrasts among levels of factor A and B within A.

First, we can test for effects among the levels of factor A.

$$H_0^A : \alpha_1 = \cdots = \alpha_a = 0 \qquad TS : F_A^* = \frac{MSA}{MSE} \qquad RR : F_A^* \geq F_{1-\alpha;a-1,ab(n-1)}$$

The estimators of the population means for levels of factor A and their properties are given below.

$$E\left\{\overline{Y}_{i\bullet\bullet}\right\} = \mu_i = \mu_{\bullet\bullet} + \alpha_i \qquad \sigma^2\left\{\overline{Y}_{i\bullet\bullet}\right\} = \frac{\sigma^2}{bn} \qquad s^2\left\{\overline{Y}_{i\bullet\bullet}\right\} = \frac{MSE}{bn}$$

Any contrast among the levels of factor A can be written as follows, with $\sum_{i=1}^{a} c_i = 0$.

$$L_A = \sum_{i=1}^{a} c_i \mu_i \qquad \hat{L}_A = \sum_{i=1}^{a} c_i \overline{Y}_{i\bullet\bullet} \qquad s^2\left\{\hat{L}_A\right\} = \frac{MSE}{bn}\sum_{i=1}^{a} c_i^2$$

$$(1-\alpha)100\% \text{ Confidence Interval for } L_A : \quad \hat{L}_A \pm t_{1-\alpha/2;ab(n-1)}\sqrt{\frac{MSE}{bn}\sum_{i=1}^{a} c_i^2}$$

$$\text{Tukey's:} \quad HSD = q_{1-\alpha;a,ab(n-1)}\sqrt{\frac{MSE}{bn}} \qquad \text{Simultaneous CI's:} \quad \left(\overline{Y}_{i\bullet\bullet} - \overline{Y}_{i'\bullet\bullet}\right) \pm HSD$$

Next, consider inference for factor B when both factors are fixed.

$$H_0^B(A) : \beta_{1(1)} = \cdots = \beta_{b(a)} = 0 \qquad TS : F_{B(A)}^* = \frac{MSB(A)}{MSE}$$
$$RR : F_{B(A)}^* \geq F_{1-\alpha;a(b-1),ab(n-1)}$$

$$E\left\{\overline{Y}_{ij\bullet}\right\} = \mu_{\bullet\bullet} + \alpha_i + \beta_{j(i)} \qquad \sigma^2\left\{\overline{Y}_{ij\bullet}\right\} = \frac{\sigma^2}{n} \qquad s^2\left\{\overline{Y}_{ij\bullet}\right\} = \frac{MSE}{n}$$

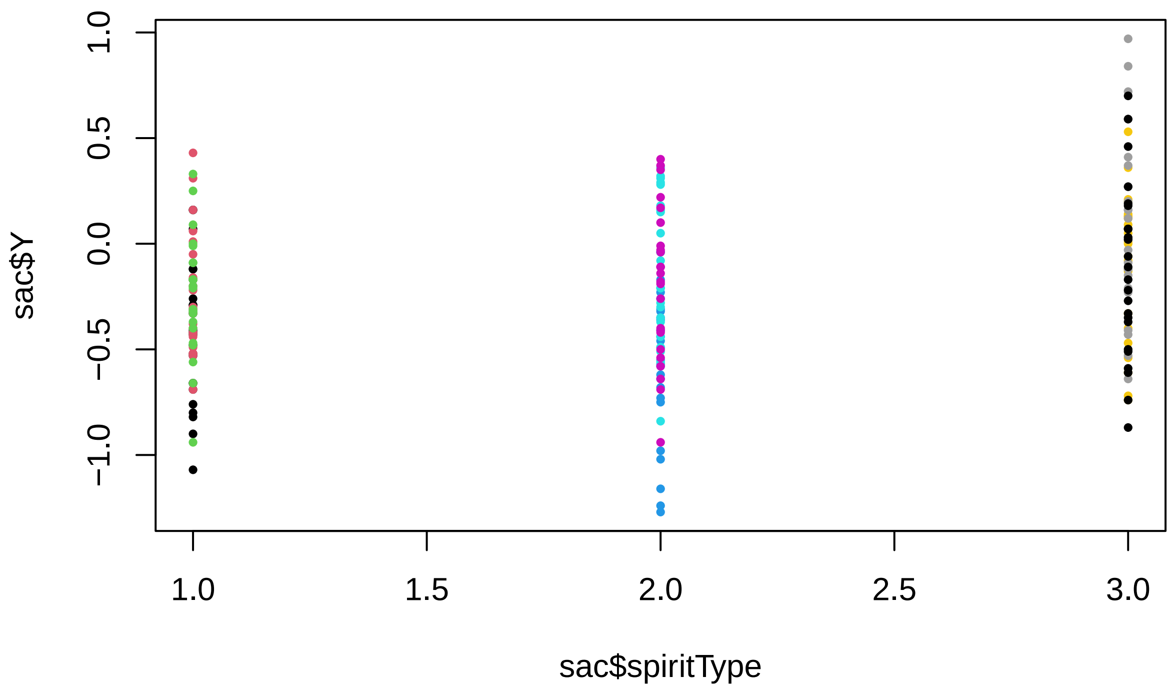
Tukey's method can be applied to compare the b means within each level of factor A as well.

$$HSD_B = q_{1-\alpha;b,ab(n-1)}\sqrt{\frac{MSE}{n}} \qquad \text{Simultaneous CI's:} \quad \left(\overline{Y}_{ij\bullet} - \overline{Y}_{ij'\bullet}\right) \pm HSD_B$$

Example 11.1 - Determination of Alcohol Content in Liquor A study was conducted to measure alcohol content in 3 types of spirits [Oliveira et al., 2017]. The researchers bought 3 brands each of Vodka, Whisky, and Cachaca. Factor A with $a = 3$ is the spirit type (Vodka, Whisky, Cachaca). The brands are nested within the spirit types with $b = 3$. Each of the brands ($ab = 3(3) = 9$) was measured $n = 24$ times. The response is Y, the difference between the measured alcohol content and the content provided on the bottle's label. We will treat both factors as fixed in this analysis. Note that the **TukeyHSD** function does not work well with the nested (brand) factor.

```
##   spiritType brandSprt labelAC alcCntnt
## 1          1         1    37.5    36.81
## 2          1         1    37.5    36.43
## 3          1         1    37.5    37.02
## 4          1         1    37.5    37.38
## 5          1         1    37.5    36.70
## 6          1         1    37.5    36.97
```

```
##     spiritType brandSprt labelAC alcCntnt
## 211          3         9      39    38.89
## 212          3         9      39    38.78
## 213          3         9      39    38.41
## 214          3         9      39    39.27
## 215          3         9      39    39.03
## 216          3         9      39    38.94
```




```
##              df      SS     MS      F* F(.95) P(>F*)  eta^2 Part eta^2
## Type          2  3.1967 1.5983 13.2669 3.0395      0 0.0992     0.1136
## Brand(Type)   6  4.0872 0.6812  5.6543 2.1426      0 0.1268     0.1408
## Error       207 24.9385 0.1205      NA     NA     NA     NA         NA
## Total       215 32.2224     NA      NA     NA     NA     NA         NA
```

```
##      i i'  Ybar_i Ybar_i'    Diff      LB     UB p adj
## [1,] 2  1 -0.3062 -0.3037 -0.0025 -0.1391 0.1341 0.999
## [2,] 3  1 -0.0469 -0.3037  0.2568  0.1202 0.3934 0.000
## [3,] 3  2 -0.0469 -0.3062  0.2593  0.1227 0.3959 0.000
```

```
##       i j j' Ybar_j(i) Ybar_j'(i)    Diff      LB     UB  p adj
##  [1,] 1 2  1   -0.2400    -0.4308  0.1908 -0.0457 0.4274 0.1399
##  [2,] 1 3  1   -0.2404    -0.4308  0.1904 -0.0461 0.4270 0.1411
##  [3,] 1 3  2   -0.2404    -0.2400 -0.0004 -0.2370 0.2361 1.0000
##  [4,] 2 5  4   -0.1388    -0.5892  0.4504  0.2139 0.6870 0.0000
##  [5,] 2 6  4   -0.1908    -0.5892  0.3983  0.1618 0.6349 0.0003
##  [6,] 2 6  5   -0.1908    -0.1388 -0.0521 -0.2886 0.1845 0.8618
##  [7,] 3 8  7    0.0792    -0.0900  0.1692 -0.0674 0.4057 0.2121
##  [8,] 3 9  7   -0.1300    -0.0900 -0.0400 -0.2765 0.1965 0.9159
##  [9,] 3 9  8   -0.1300     0.0792 -0.2092 -0.4457 0.0274 0.0949
```

```
## Analysis of Variance Table
```

```
##
## Response: Y
##                          Df  Sum Sq Mean Sq F value    Pr(>F)
## spiritType.f               2  3.1967 1.59834 13.2669 3.791e-06 ***
## spiritType.f:brandSprt.f   6  4.0872 0.68120  5.6543 1.850e-05 ***
## Residuals                207 24.9385 0.12048
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Y ~ spiritType.f + spiritType.f/brandSprt.f, data = sac)
##
## $spiritType.f
##          diff        lwr       upr     p adj
## 2-1 -0.0025000 -0.1390648 0.1340648 0.9989709
## 3-1  0.2568056  0.1202407 0.3933704 0.0000435
## 3-2  0.2593056  0.1227407 0.3958704 0.0000362
```

```
## # Effect Size for ANOVA (Type I)
##
## Parameter                | Eta2 (partial) |       95% CI
## ----------------------------------------------------------
## spiritType.f             |           0.11 | [0.05, 1.00]
## spiritType.f:brandSprt.f |           0.14 | [0.06, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

```
## # Effect Size for ANOVA (Type I)
##
## Parameter                | Eta2 |       95% CI
## ------------------------------------------------
## spiritType.f             | 0.10 | [0.04, 1.00]
## spiritType.f:brandSprt.f | 0.13 | [0.05, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

There is evidence for differences among the $a = 3$ types of spirits. The differences between measured and reported proofs are significantly higher for Cachaca ($i = 3$) than for Whisky ($i = 2$) and for Vodka ($i = 1$).

There are also differences among brands within spirits types. In particular, brands 5 and 6 (within Whisky) are significantly higher than brand 4.

∇

When factor A is fixed and B is random, we would like to compare the levels of factor A, and estimate the variance for factor B. The error therm for factor A is now $MSB(A)$ as opposed to MSE.

$$H_0^A : \alpha_1 = \cdots = \alpha_a = 0 \qquad TS : F_A^* = \frac{MSA}{MSB(A)} \qquad RR : F_A^* \geq F_{1-\alpha;a-1,a(b-1)}$$

The estimators of the population means for levels of factor A and their properties are given below.

$$E\left\{\overline{Y}_{i\bullet\bullet}\right\} = \mu_i = \mu_{\bullet\bullet} + \alpha_i \qquad \sigma^2\left\{\overline{Y}_{i\bullet\bullet}\right\} = \frac{\sigma^2 + \sigma^2_{\beta(\alpha)}}{bn} \qquad s^2\left\{\overline{Y}_{i\bullet\bullet}\right\} = \frac{MSB(A)}{bn}$$

For Tukey's method, relative to the fixed effects case, we replace MSE with $MSB(A)$ and adjust the degrees of freedom.

$$HSD = q_{1-\alpha;a,a(b-1)}\sqrt{\frac{MSB(A)}{bn}}$$

$$(1-\alpha)100\% \text{ CI for } \quad \alpha_i - \alpha_{i'}: \qquad \left(\overline{Y}_{i\bullet\bullet} - \overline{Y}_{i'\bullet\bullet}\right) \pm HSD$$

A point estimate for the variance among the levels of factor A and its corresponding approximate Confidence Interval based on Satterthwaite's approximation are given below. First, we test whether the variance component for factor B is 0.

$$H_0^{B(A)}: \sigma^2_{\beta(\alpha)} = 0 \qquad TS: F^*_{B(A)} = \frac{MSB(A)}{MSE} \qquad RR: F^*_{B(A)} \geq F_{1-\alpha;a(b-1),ab(n-1)}$$

$$E\{MSB(A)\} = \sigma^2 + n\sigma^2_{\beta(\alpha)} \qquad E\{MSE\} = \sigma^2 \quad \Rightarrow \quad s^2_{\beta(\alpha)} = \frac{MSB(A) - MSE}{n}$$

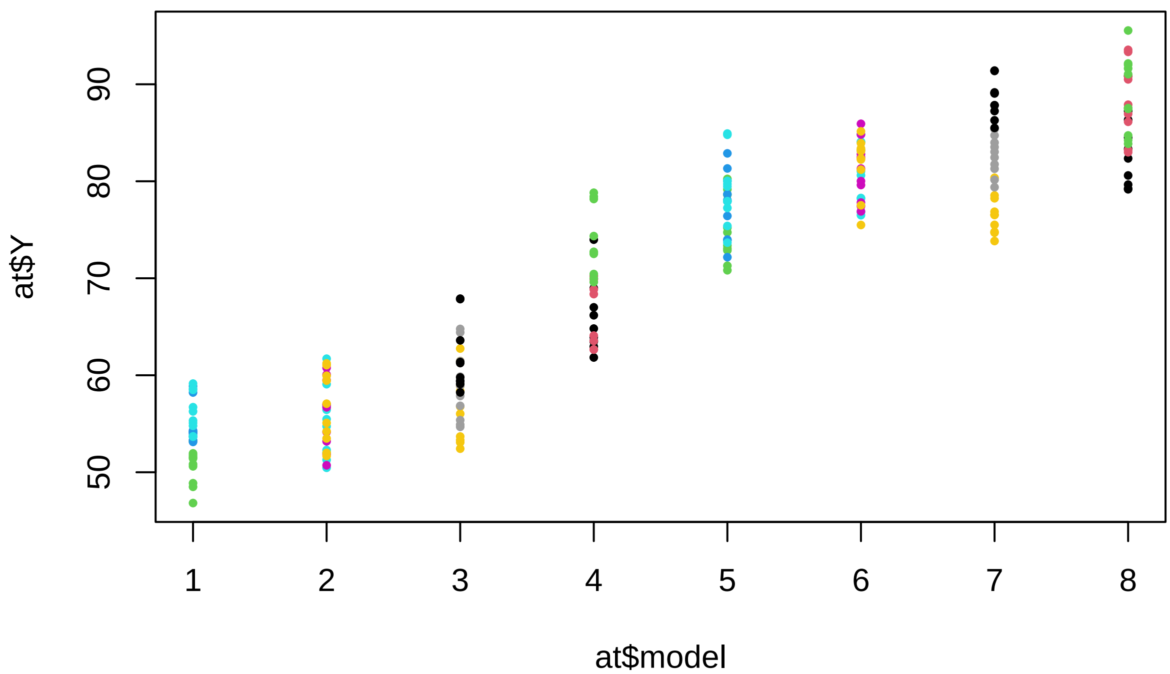
$$df_{\beta(\alpha)} = \frac{\left(s^2_{\beta(\alpha)}\right)^2}{\left[\frac{\left(\frac{1}{n}MSB(A)\right)^2}{a(b-1)} + \frac{\left(-\frac{1}{n}MSE\right)^2}{ab(n-1)}\right]}$$

$$\text{Approximate } (1-\alpha)100\% \text{ CI for } \sigma^2_{\beta(\alpha)}: \qquad \left[\frac{df_{\beta(\alpha)}s^2_{\beta(\alpha)}}{\chi^2_{1-\alpha/2;df_{\beta(\alpha)}}}, \frac{df_{\beta(\alpha)}s^2_{\beta(\alpha)}}{\chi^2_{\alpha/2;df_{\beta(\alpha)}}}\right]$$

Example 11.2 - Momentum of Animal Traps A study compared $a = 8$ models of animal traps in terms of momentum [Cook and Proulx, 1989]. Within each model, $b = 3$ traps were built. Each trap was measured $n = 10$ times. We will treat the models as a fixed factor, and the traps within models as a random factor (any number could have been assembled for each model). The response measured was the momentum at HDISP (when both jaws displayed half way). The response has been multiplied by 100 to avoid small variance estimates.

```
##   trapGrp model trapModel momentum
## 1       1     1        11  0.51423
## 2       1     1        11  0.53202
## 3       1     1        11  0.51550
## 4       1     1        11  0.46825
## 5       1     1        11  0.48496
## 6       1     1        11  0.51743
```

```
##     trapGrp model trapModel momentum
## 235      24     8        83  0.92003
## 236      24     8        83  0.91039
## 237      24     8        83  0.95544
## 238      24     8        83  0.84232
## 239      24     8        83  0.92145
## 240      24     8        83  0.87547
```




```
##               df        SS        MS       F* F(.95) P(>F*)   eta^2 Part eta^2
## Model          7 35845.161 5120.7372 493.3742 2.0522      0 0.9013     0.9411
## Trap(Model)   16  1684.028  105.2518  10.1408 1.6905      0 0.0423     0.4290
## Error        216  2241.867   10.3790       NA     NA     NA     NA         NA
## Total        239 39771.056        NA       NA     NA     NA     NA         NA
```

```
##         i i'  Ybar_i Ybar_i'    Diff      LB      UB  p adj
##   [1,] 2  1 55.8800 53.8400  2.0400 -7.1310 11.2110 0.2220
##   [2,] 3  1 59.1267 53.8400  5.2866 -3.8843 14.4576 0.0000
##   [3,] 3  2 59.1267 55.8800  3.2466 -5.9243 12.4176 0.0031
##   [4,] 4  1 68.8299 53.8400 14.9899  5.8189 24.1609 0.0000
##   [5,] 4  2 68.8299 55.8800 12.9499  3.7789 22.1209 0.0000
##   [6,] 4  3 68.8299 59.1267  9.7033  0.5323 18.8742 0.0000
##   [7,] 5  1 77.1966 53.8400 23.3566 14.1856 32.5276 0.0000
##   [8,] 5  2 77.1966 55.8800 21.3166 12.1456 30.4876 0.0000
##   [9,] 5  3 77.1966 59.1267 18.0700  8.8990 27.2409 0.0000
##  [10,] 5  4 77.1966 68.8299  8.3667 -0.8043 17.5377 0.0000
##  [11,] 6  1 81.1433 53.8400 27.3033 18.1323 36.4742 0.0000
##  [12,] 6  2 81.1433 55.8800 25.2633 16.0923 34.4342 0.0000
##  [13,] 6  3 81.1433 59.1267 22.0166 12.8457 31.1876 0.0000
##  [14,] 6  4 81.1433 68.8299 12.3134  3.1424 21.4843 0.0000
##  [15,] 6  5 81.1433 77.1966  3.9467 -5.2243 13.1176 0.0001
##  [16,] 7  1 82.5467 53.8400 28.7066 19.5357 37.8776 0.0000
##  [17,] 7  2 82.5467 55.8800 26.6666 17.4957 35.8376 0.0000
##  [18,] 7  3 82.5467 59.1267 23.4200 14.2490 32.5910 0.0000
##  [19,] 7  4 82.5467 68.8299 13.7167  4.5458 22.8877 0.0000
##  [20,] 7  5 82.5467 77.1966  5.3500 -3.8209 14.5210 0.0000
##  [21,] 7  6 82.5467 81.1433  1.4034 -7.7676 10.5743 0.6956
##  [22,] 8  1 86.8833 53.8400 33.0433 23.8723 42.2143 0.0000
```

```
## [23,] 8  2 86.8833 55.8800 31.0033 21.8323 40.1743 0.0000
## [24,] 8  3 86.8833 59.1267 27.7567 18.5857 36.9276 0.0000
## [25,] 8  4 86.8833 68.8299 18.0534  8.8824 27.2244 0.0000
## [26,] 8  5 86.8833 77.1966  9.6867  0.5157 18.8577 0.0000
## [27,] 8  6 86.8833 81.1433  5.7400 -3.4309 14.9110 0.0000
## [28,] 8  7 86.8833 82.5467  4.3367 -4.8343 13.5076 0.0000
```

```
##                   df     s^2     LB      UB
## Trap(Model)  12.9907  9.4873 4.9852 24.6343
## Error       216.0000 10.3790 8.6693 12.6524
```

```
## Analysis of Variance Table
##
## Response: Y
##                      Df Sum Sq Mean Sq F value    Pr(>F)
## model.f               7  35845  5120.7 493.374 < 2.2e-16 ***
## model.f:trapModel.f  16   1684   105.3  10.141 < 2.2e-16 ***
## Residuals           216   2242    10.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Y ~ model.f + model.f/trapModel.f, data = at)
##
## $model.f
##          diff        lwr       upr     p adj
## 2-1  2.040000 -0.5061298  4.586130 0.2219660
## 3-1  5.286633  2.7405035  7.832763 0.0000000
## 4-1 14.989900 12.4437702 17.536030 0.0000000
## 5-1 23.356600 20.8104702 25.902730 0.0000000
## 6-1 27.303267 24.7571369 29.849396 0.0000000
## 7-1 28.706633 26.1605035 31.252763 0.0000000
## 8-1 33.043300 30.4971702 35.589430 0.0000000
## 3-2  3.246633  0.7005035  5.792763 0.0031454
## 4-2 12.949900 10.4037702 15.496030 0.0000000
## 5-2 21.316600 18.7704702 23.862730 0.0000000
## 6-2 25.263267 22.7171369 27.809396 0.0000000
## 7-2 26.666633 24.1205035 29.212763 0.0000000
## 8-2 31.003300 28.4571702 33.549430 0.0000000
## 4-3  9.703267  7.1571369 12.249396 0.0000000
## 5-3 18.069967 15.5238369 20.616096 0.0000000
## 6-3 22.016633 19.4705035 24.562763 0.0000000
## 7-3 23.420000 20.8738702 25.966130 0.0000000
## 8-3 27.756667 25.2105369 30.302796 0.0000000
## 5-4  8.366700  5.8205702 10.912830 0.0000000
## 6-4 12.313367  9.7672369 14.859496 0.0000000
## 7-4 13.716733 11.1706035 16.262863 0.0000000
## 8-4 18.053400 15.5072702 20.599530 0.0000000
## 6-5  3.946667  1.4005369  6.492796 0.0001021
## 7-5  5.350033  2.8039035  7.896163 0.0000000
## 8-5  9.686700  7.1405702 12.232830 0.0000000
## 7-6  1.403367 -1.1427631  3.949496 0.6956345
## 8-6  5.740033  3.1939035  8.286163 0.0000000
## 8-7  4.336667  1.7905369  6.882796 0.0000119
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
```

```
## lmerModLmerTest]
## Formula: Y ~ model.f + (1 | model.f:trapModel.f)
##    Data: at
##
## REML criterion at convergence: 1269.7
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.92671 -0.74501 -0.08267  0.72379  2.48273
##
## Random effects:
##  Groups              Name        Variance Std.Dev.
##  model.f:trapModel.f (Intercept)  9.487   3.080
##  Residual                        10.379   3.222
## Number of obs: 240, groups:  model.f:trapModel.f, 24
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)  70.6808     0.6622  16.0000 106.731  < 2e-16 ***
## model.f1    -16.8408     1.7521  16.0000  -9.612 4.75e-08 ***
## model.f2    -14.8008     1.7521  16.0000  -8.447 2.72e-07 ***
## model.f3    -11.5542     1.7521  16.0000  -6.594 6.17e-06 ***
## model.f4     -1.8509     1.7521  16.0000  -1.056  0.30648
## model.f5      6.5158     1.7521  16.0000   3.719  0.00187 **
## model.f6     10.4625     1.7521  16.0000   5.971 1.96e-05 ***
## model.f7     11.8658     1.7521  16.0000   6.772 4.48e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) mdl.f1 mdl.f2 mdl.f3 mdl.f4 mdl.f5 mdl.f6
## model.f1  0.000
## model.f2  0.000 -0.143
## model.f3  0.000 -0.143 -0.143
## model.f4  0.000 -0.143 -0.143 -0.143
## model.f5  0.000 -0.143 -0.143 -0.143 -0.143
## model.f6  0.000 -0.143 -0.143 -0.143 -0.143 -0.143
## model.f7  0.000 -0.143 -0.143 -0.143 -0.143 -0.143 -0.143

## Type III Analysis of Variance Table with Satterthwaite's method
##         Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## model.f 3534.7  504.96     7    16  48.652 1.319e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There are many differences among the 28 pairs of trap models. Also, within models there is relatively high variation among the individual traps.

∇

When factors A and B are both random, we are generally interested in estimating their variance components and testing whether they are 0.

$$H_0^A : \sigma_\alpha^2 = 0 \qquad TS : F_A^* = \frac{MSA}{MSB(A)} \qquad RR : F_A^* \geq F_{1-\alpha;a-1,a(b-1)}$$

$$E\{MSA\} = \sigma^2 + n\sigma_{\beta(\alpha)}^2 + bn\sigma_\alpha^2 \qquad E\{MSB(A)\} = \sigma^2 + n\sigma_{\beta(\alpha)}^2$$

$$s_\alpha^2 = \frac{MSA - MSB(A)}{bn} \qquad df_\alpha = \frac{\left(s_\alpha^2\right)^2}{\left[\frac{\left(\frac{1}{bn}MSA\right)^2}{a-1} + \frac{\left(-\frac{1}{bn}MSB(A)\right)^2}{a(b-1)}\right]}$$

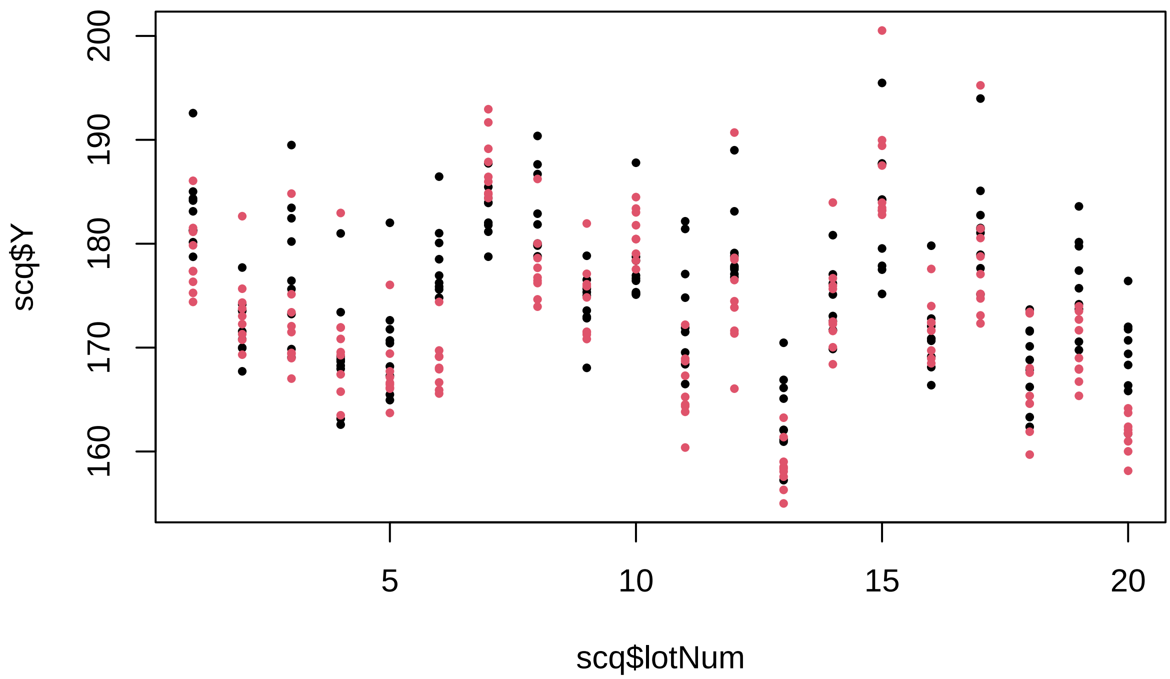
$$\text{Approximate } (1-\alpha)100\% \text{ CI for } \sigma_\alpha^2 : \quad \left[\frac{df_\alpha s_\alpha^2}{\chi^2_{1-\alpha/2;df_\alpha}}, \frac{df_\alpha s_\alpha^2}{\chi^2_{\alpha/2;df_\alpha}}\right]$$

The test and Confidence Interval for $\sigma^2_{\beta(\alpha)}$ are the same for the random effects model as for the mixed effects model.

Example 11.3 - Measurements on Semiconductor Wafers An instructional paper described making measurements on semiconducter wafers [Jensen, 2002]. There were $a = 20$ lots (batches) sampled, within each lot, there were $b = 2$ wafers sampled, and each wafer was measured at $n = 9$ locations (replicates). The response variable was not given for proprietary reasons. In this example, we treat lot and wafer(lot) as random factors. There are two wafer variables, **wafer1** takes on the values 1 and 2 within each lot, **wafer2** gives each wafer a unique value (1-40) and is less risky of creating an error in computing.

```
##   lotNum wafer1 wafer2 repNum       Y
## 1      1      1      1      1 181.247
## 2      1      1      1      2 181.280
## 3      1      1      1      3 185.021
## 4      1      1      1      4 180.144
## 5      1      1      1      5 192.570
## 6      1      1      1      6 178.741
```

```
##     lotNum wafer1 wafer2 repNum       Y
## 355     20      2     40      4 161.783
## 356     20      2     40      5 160.013
## 357     20      2     40      6 160.981
## 358     20      2     40      7 162.384
## 359     20      2     40      8 163.720
## 360     20      2     40      9 161.716
```




```
##              df        SS       MS     F* F(.95) P(>F*)  eta^2 Part eta^2
## Lot          19 14050.871 739.5195 7.5475 2.1370      0 0.6358     0.6976
## Wafer(Lot)   20  1959.633  97.9817 5.1488 1.6035      0 0.0887     0.2435
## Error       320  6089.584  19.0299     NA     NA     NA     NA         NA
## Total       359 22100.088       NA     NA     NA     NA     NA         NA
```

```
##                   df     s^2      LB      UB
## Lot          14.0642 35.6410 19.1268 88.4214
## Wafer(Lot)   12.9551  8.7724  4.6061 22.8150
## Error       320.0000 19.0299 16.3941 22.3605
```

```
## Analysis of Variance Table
##
## Response: Y
##                    Df  Sum Sq Mean Sq F value    Pr(>F)
## lotNum.f           19 14050.9  739.52 38.8608 < 2.2e-16 ***
## lotNum.f:wafer2.f  20  1959.6   97.98  5.1488 3.708e-11 ***
## Residuals         320  6089.6   19.03
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Y ~ 1 + (1 | lotNum.f) + (1 | lotNum.f:wafer2.f)
##    Data: scq
##
## REML criterion at convergence: 2184.6
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -2.2897 -0.5826 -0.1729  0.4287  3.7529
##
## Random effects:
##  Groups            Name        Variance Std.Dev.
##  lotNum.f:wafer2.f (Intercept)  8.772   2.962
##  lotNum.f          (Intercept) 35.641   5.970
##  Residual                      19.030   4.362
## Number of obs: 360, groups:  lotNum.f:wafer2.f, 40; lotNum.f, 20
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  174.342      1.433  19.000   121.6   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The largest source of variation is among lots $\left(s_\alpha^2 = 35.64\right)$, followed by error (within wafer) $\left(s^2 = 19.03\right)$, followed by between wafers within lots $\left(s_{\beta(\alpha)}^2 = 8.77\right)$.

∇

11.3 Repeated Measures Design

In this section we cover a second type of repeated measures design where subjects are randomized to treatments as in the Completely Randomized Design, then are measured at multiple time points. Recall in the chapter on block designs, there was a repeated measures design where each subject (block) received each treatment. In this case, each subject receives only one of the treatments.

Note that the notation used here is not the same as Kutner, et al.

11.3.1 Model Structure

The elements are described below for the case with a single treatment factor (it can be extended to multiple factors) and a balanced design.

- **Treatments** - The primary factor of interest, with a (fixed) levels.
- **Subjects** - Individuals included in the experiment b (random) units per treatment.
- **Time Points** - When the measurements are obtained on each subject, with c (fixed) time points per individual.

We will use i to represent treatment, j represent the subject within treatment and k to represent time point. The model is given below.

$$Y_{ijk} = \mu_{\bullet\bullet\bullet} + \alpha_i + \beta_{j(i)} + \gamma_k + (\alpha\gamma)_{ik} + \epsilon_{ijk} \quad i = 1, \dots, a; \quad j = 1, \dots, b; \quad k = 1, \dots, c$$

$$\sum_{i=1}^{a} \alpha_i = \sum_{k=1}^{c} \gamma_k = \sum_{i=1}^{a} (\alpha\gamma)_{ik} = \sum_{k=1}^{c} (\alpha\gamma)_{ik} = 0$$

$$\beta_{j(i)} \sim NID\left(0, \sigma_{\beta(\alpha)}^2\right) \qquad \epsilon_{ijk} \sim NID\left(0, \sigma^2\right) \qquad \left\{\beta_{j(i)}\right\} \perp \left\{\epsilon_{ijk}\right\}$$

$$E\left\{Y_{ijk}\right\} = \mu_{\bullet\bullet\bullet} + \alpha_i + \gamma_k + (\alpha\gamma)_{ik} \qquad \sigma^2\left\{Y_{ijk}\right\} = \sigma_{\beta(\alpha)}^2 + \sigma^2$$

$$\sigma\left\{Y_{ijk}, Y_{i'j'k'}\right\} = \begin{cases} \sigma_{\beta(\alpha)}^2 + \sigma^2 & : \quad i = i', j = j', k = k' \\ \sigma_{\beta(\alpha)}^2 & : \quad i = i', j = j', k \neq k' \\ 0 & : \quad \text{otherwise} \end{cases}$$

The variance-covariance within subjects has $\sigma^2_{\beta(\alpha)} + \sigma^2$ on the main diagonal and $\sigma^2_{\beta(\alpha)}$ on the off diagonals. In practice, particularly when there are many time points, this structure is too simple and a more complex structure can be fit.

$$\sigma\left\{\begin{bmatrix} Y_{ij1} \\ Y_{ij2} \\ \vdots \\ Y_{ijc} \end{bmatrix}\right\} = \begin{bmatrix} \sigma^2_{\beta(\alpha)} + \sigma^2 & \sigma^2_{\beta(\alpha)} & \cdots & \sigma^2_{\beta(\alpha)} \\ \sigma^2_{\beta(\alpha)} & \sigma^2_{\beta(\alpha)} + \sigma^2 & \cdots & \sigma^2_{\beta(\alpha)} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2_{\beta(\alpha)} & \sigma^2_{\beta(\alpha)} & \cdots & \sigma^2_{\beta(\alpha)} + \sigma^2 \end{bmatrix}$$

11.3.2 Analysis of Variance and F-tests

For the various factors, we have he following sums of squares, degrees of freedom, and expected mean squares.

$$\text{Treatments:} \quad SSA = bc\sum_{i=1}^{a}\left(\overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet\bullet bullet}\right)^2 \qquad df_A = a-1$$

$$E\{MSA\} = \sigma^2 + c\sigma^2_{\beta(\alpha)} + \frac{bc\sum_{i=1}^{a}\alpha_i^2}{a-1}$$

$$\text{Subjects(Trts):} \quad SSB(A) = c\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{Y}_{ij\bullet} - \overline{Y}_{i\bullet\bullet}\right)^2 \qquad df_{B(A)} = a(b-1)$$

$$E\{MSB(A)\} = \sigma^2 + c\sigma^2_{\beta(\alpha)}$$

$$\text{Time:} \quad SSC = ab\sum_{k=1}^{c}\left(\overline{Y}_{\bullet\bullet k} - \overline{Y}_{\bullet\bullet\bullet}\right)^2 \qquad df_C = c-1$$

$$E\{MSC\} = \sigma^2 + \frac{ab\sum_{k=1}^{c}\gamma_k^2}{c-1}$$

$$\text{TreatmentxTime:} \quad SSAC = b\sum_{i=1}^{a}\sum_{k=1}^{c}\left(\overline{Y}_{i\bullet k} - \overline{Y}_{i\bullet\bullet} - \overline{Y}_{\bullet\bullet k} + \overline{Y}_{\bullet\bullet\bullet}\right)^2$$

$$df_{AC} = (a-1)(c-1) \qquad E\{MSAC\} = \sigma^2 + \frac{b\sum_{i=1}^{a}\sum_{k=1}^{c}(\alpha\gamma)_{ik}^2}{(a-1)(c-1)}$$

The error sum of squares represents the Subject(Treatment)xTime interaction and has expectation $E\{MSE\} = \sigma^2$.

$$SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\left(Y_{ijk} - \overline{Y}_{ij\bullet} - \overline{Y}_{i\bullet k} + \overline{Y}_{i\bullet\bullet}\right)^2 \qquad df_E = a(b-1)(c-1)$$

To test for treatment effects, compare MSA to $MSB(A)$. To test for Time effects and TimexTreatment interaction, compare their mean squares to MSE.

$$H_0^A : \alpha_1 = \cdots = \alpha_a \qquad TS: F_A^* = \frac{MSA}{MSB(A)} \qquad RR: F_A^* \geq F_{1-\alpha;a-1,a(b-1)}$$

$$H_0^C : \gamma_1 = \cdots = \gamma_k = 0 \qquad TS: F_C^* = \frac{MSC}{MSE} \qquad RR: F_C^* \geq F_{1-\alpha;c-1,a(b-1)(c-1)}$$

$$H_0^{AC} : (\alpha\gamma)_{11} = \cdots = (\alpha\gamma)_{ac} = 0 \qquad TS: F_{AC}^* = \frac{MSAC}{MSE} \qquad RR: F_{AC}^* \geq F_{1-\alpha;(a-1)(c-1),a(b-1)(c-1)}$$

Depending on whether the TreatmentxTime interaction is important, treatments can be compared within time points (interaction important) or marginally (interaction unimportant).

11.3.3 Treatment Comparisons

First, consider the case where the interaction is not important, when the "time effect" is consistent across treatments. Then we can use Tukey's method for comparing all pairs of treatments, keeping in mind that for treatments, $MSB(A)$ is the error term.

$$HSD = q_{1-\alpha;a,a(b-1)}\sqrt{\frac{MSB(A)}{bc}}$$

$$(1-\alpha)100\% \text{ CI for } \alpha_i - \alpha_{i'}: \quad \left(\overline{Y}_{i\bullet\bullet} - \overline{Y}_{i\bullet\bullet}\right) \pm HSD$$

When the TreatmentxTime interaction is important, we may wish to compare the treatments within the time periods. This involves taking the standard error for the difference between means within the time periods as a linear function of $MSB(A)$ and MSE and using Satterthwaite's approximation for the degrees of freedom.

$$\sigma^2\left\{\overline{Y}_{i\bullet k} - \overline{Y}_{i'\bullet k}\right\} = \frac{2\left(\sigma^2 + \sigma^2_{\beta(\alpha)}\right)}{b} \qquad E\{MSB(A) + (c-1)MSE\} = c\left(\sigma^2 + \sigma^2_{\beta(\alpha)}\right)$$

So if we divide $MSB(A) + (c-1)MSE$ by bc and multiply by 2, this has expectation equal to the variance of the mean difference.

$$s^2\left\{\overline{Y}_{i\bullet k} - \overline{Y}_{i'\bullet k}\right\} = \frac{2\left(MSB(A) + (c-1)MSE\right)}{bc}$$

$$df^* = \frac{\left(MSB(A) + (c-1)MSE\right)^2}{\left[\frac{(MSB(A))^2}{a(b-1)} + \frac{((c-1)MSE)^2}{a(b-1)(c-1)}\right]}$$

Then we can use Tukey's method to compare all pairs of treatments within specific time points.

$$HSD_k = q_{1-\alpha;a,df^*}\sqrt{\frac{MSB(A) + (c-1)MSE}{bc}}$$

$$(1-\alpha)100\% \text{ Confidence Intervals:} \quad \left(\overline{Y}_{i\bullet k} - \overline{Y}_{i'\bullet k}\right) \pm HSD_k$$

Example 11.4 - Treating Cats with Anxiety A study was conducted to test a drug zylkene versus placebo in cats with anxiety [Beata et al., 2007]. There were $a = 2$ treatments (zylkene $i = 1$ and placebo $i = 2$), with $b = 17$ cats randomized to each treatment (34 total) and each cat was observed at each of $c = 5$ time points. The response was a clinical globel impression (higher scores mean less anxiety and are better). R code and output are given below.

```
##   id weight age gender trt timepnt y
## 1  2    4.0  67      1   0       1 9
## 2  2    4.0  67      1   0       2 9
## 3  2    4.0  67      1   0       3 9
## 4  2    4.0  67      1   0       4 9
## 5  2    4.0  67      1   0       5 9
## 6  4    3.5  78      2   0       1 9

##     id weight age gender trt timepnt  y
## 165 31    3.1  90      1   1       5 23
## 166 32    4.0  56      1   1       1  8
## 167 32    4.0  56      1   1       2 11
## 168 32    4.0  56      1   1       3 13
## 169 32    4.0  56      1   1       4 14
## 170 32    4.0  56      1   1       5 17
```

```
##            df        SS       MS      F* F(.95) P(>F*)  eta^2
## Drug        1  382.5000 382.5000  5.7412 4.1491 0.0226 0.1167
## Cat(Drug)  32 2131.9765  66.6243 21.9512 1.5341 0.0000 0.6505
## Time        4  324.1176  81.0294 26.6974 2.4425 0.0000 0.0989
## DrugxTime   4   50.5882  12.6471  4.1669 2.4425 0.0033 0.0154
## Error     128  388.4941   3.0351      NA     NA     NA     NA
## Total     169 3277.6765       NA      NA     NA     NA     NA
```

```
## Analysis of Variance Table
##
## Response: y
##                  Df  Sum Sq Mean Sq  F value    Pr(>F)
## trt.f             1  382.50  382.50 126.0251 < 2.2e-16 ***
## timepnt.f         4  324.12   81.03  26.6974 4.137e-16 ***
## trt.f:id1.f      32 2131.98   66.62  21.9512 < 2.2e-16 ***
## trt.f:timepnt.f   4   50.59   12.65   4.1669  0.003316 **
## Residuals       128  388.49    3.04
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Error: id1.f
##           Df Sum Sq Mean Sq F value Pr(>F)
## trt.f      1  382.5   382.5   5.741 0.0226 *
## Residuals 32 2132.0    66.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: Within
##                  Df Sum Sq Mean Sq F value   Pr(>F)
## timepnt.f         4  324.1   81.03  26.697 4.14e-16 ***
## trt.f:timepnt.f   4   50.6   12.65   4.167  0.00332 **
## Residuals       128  388.5    3.04
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## # Effect Size for ANOVA (Type I)
##
## Parameter       | Eta2 (partial) |       95% CI
## -------------------------------------------------
## trt.f           |           0.50 | [0.40, 1.00]
## timepnt.f       |           0.45 | [0.34, 1.00]
## trt.f:id1.f     |           0.85 | [0.80, 1.00]
## trt.f:timepnt.f |           0.12 | [0.03, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

```
## # Effect Size for ANOVA (Type I)
##
## Parameter       | Eta2 |       95% CI
## ---------------------------------------
## trt.f           | 0.12 | [0.04, 1.00]
## timepnt.f       | 0.10 | [0.02, 1.00]
## trt.f:id1.f     | 0.65 | [0.53, 1.00]
## trt.f:timepnt.f | 0.02 | [0.00, 1.00]
##
## - One-sided CIs: upper bound fixed at [1.00].
```

Based on the ANOVA model, fit based on the formulas described in this section, we observe a significant interaction between Treatment and Time ($F^*_{AC} = 4.167, P = .0033$), as well as significant main effects for Treatment and Time. Here, we will compare the two treatments within the five time points using 95% Confidence Intervals for $\mu_Z - \mu_P$ within each time point (1 comparison per time). We will use $i = 1$ to represent Zylkene and $i = 2$ to represent Placebo.

$$s^2\left\{\overline{Y}_{1\bullet k} - \overline{Y}_{2\bullet k}\right\} = \frac{2\left(MSB(A) + (c-1)MSE\right)}{bc} = \frac{2\left(66.62 + (5-1)(3.04)\right)}{17(5)} = 1.854$$

$$df^* = \frac{\left(66.62 + 4(3.04)\right)^2}{\left[\frac{(66.62)^2}{32} + \frac{(4(3.04))^2}{128}\right]} = \frac{6206.29}{139.85} = 44.4$$

$$t_{.975;44.4} = 2.015 \quad \Rightarrow \quad \left(\overline{Y}_{1\bullet k} - \overline{Y}_{2\bullet k}\right) \pm 2.015\sqrt{1.854} \equiv \left(\overline{Y}_{1\bullet k} - \overline{Y}_{2\bullet k}\right) \pm 2.74$$

For the 5 time points, we compute the following Confidence Intervals.

$$k = 1: \quad (10.94 - 9.06) \pm 2.74 \quad \equiv \quad 1.88 \pm 2.74 \equiv (-0.86, 4.62)$$
$$k = 2: \quad (12.12 - 10.18) \pm 2.74 \equiv 1.94 \pm 2.74 \equiv (-0.80, 4.68)$$
$$k = 3: \quad (13.59 - 10.88) \pm 2.74 \equiv 2.71 \pm 2.74 \equiv (-0.03, 5.45)$$
$$k = 4: \quad (15.18 - 11.41) \pm 3.77 \equiv 3.77 \pm 2.74 \equiv (1.03, 6.51)$$
$$k = 5: \quad (16.12 - 11.41) \pm 4.71 \equiv 4.71 \pm 2.74 \equiv (1.97, 7.45)$$

At time point 3, they are very close to being significantly different as the interval just contains 0. At time points 4 and 5, Zylkene gives significantly higher scores than placebo.

Next, we consider two other covariance structures among measurements within cats: **unstructured** and **AR(1)** which is autoregressive of order 1. We will use the **lme** function in the **nlme** package to fit them.

The unstructured case has no restrictions on the variances and covariances.

$$\sigma\left\{\begin{bmatrix} Y_{ij1} \\ Y_{ij2} \\ \vdots \\ Y_{ijc} \end{bmatrix}\right\} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1c} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1c} & \sigma_{2c} & \cdots & \sigma_c^2 \end{bmatrix}$$

For the AR(1) case, the variances are all equal, and the covariances (and correlations) decrease multiplicatively as the observations are further apart in time. This can be extended to allow for heterogeneity of variances, which is not covered here. In the following, we assume $|\phi| < 1$.

$$\sigma\left\{\begin{bmatrix} Y_{ij1} \\ Y_{ij2} \\ \vdots \\ Y_{ijc} \end{bmatrix}\right\} = \begin{bmatrix} \sigma^2 & \phi\sigma^2 & \cdots & \phi^{c-1}\sigma^2 \\ \phi\sigma^2 & \sigma^2 & \cdots & \phi^{c-2}\sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \phi^{c-1}\sigma^2 & \phi^{c-2}\sigma^2 & \cdots & \sigma^2 \end{bmatrix}$$

```
##   id weight age gender trt timepnt y
## 1  2    4.0  67      1   0       1 9
## 2  2    4.0  67      1   0       2 9
## 3  2    4.0  67      1   0       3 9
## 4  2    4.0  67      1   0       4 9
## 5  2    4.0  67      1   0       5 9
## 6  4    3.5  78      2   0       1 9
```

```
##     id weight age gender trt timepnt  y
## 165 31    3.1  90      1   1       5 23
## 166 32    4.0  56      1   1       1  8
## 167 32    4.0  56      1   1       2 11
## 168 32    4.0  56      1   1       3 13
## 169 32    4.0  56      1   1       4 14
## 170 32    4.0  56      1   1       5 17
```

```
## Linear mixed-effects model fit by REML
##   Data: caz
##        AIC      BIC    logLik
##   798.2436 838.2209 -386.1218
##
## Random effects:
##  Formula: ~trt.f - 1 | id
##  Structure: Compound Symmetry
##              StdDev   Corr
## trt.fplacebo 3.566192
## trt.fzylkene 3.566192 0
## Residual     1.742159
##
## Fixed effects:  y ~ trt.f * timepnt.f
##                        Value Std.Error  DF   t-value p-value
## (Intercept)        12.088235 0.6260226 128 19.309583  0.0000
## trt.f1             -1.500000 0.6260226  32 -2.396080  0.0226
## timepnt.f1         -2.088235 0.2672349 128 -7.814231  0.0000
## timepnt.f2         -0.941176 0.2672349 128 -3.521907  0.0006
## timepnt.f3          0.147059 0.2672349 128  0.550298  0.5831
## timepnt.f4          1.205882 0.2672349 128  4.512443  0.0000
## trt.f1:timepnt.f1  0.558824 0.2672349 128  2.091132  0.0385
## trt.f1:timepnt.f2  0.529412 0.2672349 128  1.981073  0.0497
## trt.f1:timepnt.f3  0.147059 0.2672349 128  0.550298  0.5831
## trt.f1:timepnt.f4 -0.382353 0.2672349 128 -1.430775  0.1549
##  Correlation:
##                   (Intr) trt.f1 tmpn.1 tmpn.2 tmpn.3 tmpn.4 t.1:.1 t.1:.2
## trt.f1             0.00
## timepnt.f1         0.00   0.00
## timepnt.f2         0.00   0.00  -0.25
## timepnt.f3         0.00   0.00  -0.25  -0.25
## timepnt.f4         0.00   0.00  -0.25  -0.25  -0.25
## trt.f1:timepnt.f1  0.00   0.00   0.00   0.00   0.00   0.00
## trt.f1:timepnt.f2  0.00   0.00   0.00   0.00   0.00   0.00  -0.25
## trt.f1:timepnt.f3  0.00   0.00   0.00   0.00   0.00   0.00  -0.25  -0.25
## trt.f1:timepnt.f4  0.00   0.00   0.00   0.00   0.00   0.00  -0.25  -0.25
##                   t.1:.3
## trt.f1
## timepnt.f1
## timepnt.f2
## timepnt.f3
## timepnt.f4
## trt.f1:timepnt.f1
## trt.f1:timepnt.f2
## trt.f1:timepnt.f3
## trt.f1:timepnt.f4 -0.25
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -3.21774790 -0.51423747 -0.04419265  0.71213362  2.31966939
##
## Number of Observations: 170
## Number of Groups: 34

##                 numDF denDF  F-value p-value
## (Intercept)         1   128 372.8600  <.0001
## trt.f               1    32   5.7412  0.0226
```

```
## timepnt.f           4   128  26.6973  <.0001
## trt.f:timepnt.f     4   128   4.1669  0.0033

## [1] 798.2436

## id 1
## Marginal variance covariance matrix
##        1      2      3      4      5
## 1 15.753 12.718 12.718 12.718 12.718
## 2 12.718 15.753 12.718 12.718 12.718
## 3 12.718 12.718 15.753 12.718 12.718
## 4 12.718 12.718 12.718 15.753 12.718
## 5 12.718 12.718 12.718 12.718 15.753
##   Standard Deviations: 3.969 3.969 3.969 3.969 3.969

## Linear mixed-effects model fit by REML
##   Data: caz
##        AIC      BIC    logLik
##   749.8307 829.7852 -348.9153
##
## Random effects:
##  Formula: ~1 | id1.f
##         (Intercept) Residual
## StdDev:    3.073923 1.391268
##
## Correlation Structure: General
##  Formula: ~1 | id1.f
##  Parameter estimate(s):
##  Correlation:
##   1      2      3      4
## 2  0.031
## 3 -0.011  0.621
## 4  0.219  0.697  0.886
## 5 -0.041  0.550  0.805  0.929
## Variance function:
##  Structure: Different standard deviations per stratum
##  Formula: ~1 | timepnt.f
##  Parameter estimates:
##        1        2        3        4        5
## 1.000000 1.139761 1.712557 2.374562 2.335903
## Fixed effects:  y ~ trt.f * timepnt.f
##                       Value Std.Error  DF   t-value p-value
## (Intercept)       12.088235 0.6264483 128 19.296463  0.0000
## trt.f1            -1.500000 0.6264483  32 -2.394452  0.0227
## timepnt.f1        -2.088235 0.3723377 128 -5.608445  0.0000
## timepnt.f2        -0.941176 0.2328305 128 -4.042325  0.0001
## timepnt.f3         0.147059 0.1790910 128  0.821140  0.4131
## timepnt.f4         1.205882 0.2300381 128  5.242098  0.0000
## trt.f1:timepnt.f1  0.558824 0.3723377 128  1.500851  0.1359
## trt.f1:timepnt.f2  0.529412 0.2328305 128  2.273808  0.0246
## trt.f1:timepnt.f3  0.147059 0.1790910 128  0.821140  0.4131
## trt.f1:timepnt.f4 -0.382353 0.2300381 128 -1.662129  0.0989
##  Correlation:
##                   (Intr) trt.f1 tmpn.1 tmpn.2 tmpn.3 tmpn.4 t.1:.1 t.1:.2
## trt.f1             0.000
## timepnt.f1        -0.421  0.000
## timepnt.f2        -0.325  0.000  0.380
## timepnt.f3         0.091  0.000 -0.415 -0.198
```

```
## timepnt.f4         0.533  0.000 -0.742 -0.683  0.090
## trt.f1:timepnt.f1  0.000 -0.421  0.000  0.000  0.000  0.000
## trt.f1:timepnt.f2  0.000 -0.325  0.000  0.000  0.000  0.000  0.380
## trt.f1:timepnt.f3  0.000  0.091  0.000  0.000  0.000  0.000 -0.415 -0.198
## trt.f1:timepnt.f4  0.000  0.533  0.000  0.000  0.000  0.000 -0.742 -0.683
##                   t.1:.3
## trt.f1
## timepnt.f1
## timepnt.f2
## timepnt.f3
## timepnt.f4
## trt.f1:timepnt.f1
## trt.f1:timepnt.f2
## trt.f1:timepnt.f3
## trt.f1:timepnt.f4  0.090
##
## Standardized Within-Group Residuals:
##        Min         Q1        Med         Q3        Max
## -2.3359672 -0.6444285 -0.1169039  0.6566682  2.7544055
##
## Number of Observations: 170
## Number of Groups: 34
```

```
##                 numDF denDF  F-value p-value
## (Intercept)         1   128 385.9196  <.0001
## trt.f               1    32   3.9783  0.0547
## timepnt.f           4   128   9.7978  <.0001
## trt.f:timepnt.f     4   128   2.7241  0.0323
```

```
## [1] 749.8307
```

```
## id1.f 1
## Conditional variance covariance matrix
##           1        2         3       4        5
## 1  1.935600 0.067913 -0.037752  1.0086 -0.18467
## 2  0.067913 2.514500  2.348000  3.6492  2.83350
## 3 -0.037752 2.348000  5.676900  6.9764  6.23720
## 4  1.008600 3.649200  6.976400 10.9140  9.97360
## 5 -0.184670 2.833500  6.237200  9.9736 10.56200
##   Standard Deviations: 1.3913 1.5857 2.3826 3.3037 3.2499
```

```
## Linear mixed-effects model fit by REML
##   Data: caz
##        AIC      BIC    logLik
##   744.0077 783.9849 -359.0038
##
## Random effects:
##  Formula: ~1 | id1.f
##         (Intercept) Residual
## StdDev:  0.00118003 3.963392
##
## Correlation Structure: AR(1)
##  Formula: ~1 | id1.f
##  Parameter estimate(s):
##       Phi
## 0.9047053
## Fixed effects:  y ~ trt.f * timepnt.f
##                      Value Std.Error  DF   t-value p-value
```

```
## (Intercept)       12.088235 0.6296157 128 19.199388  0.0000
## trt.f1            -1.500000 0.6296157  32 -2.382406  0.0233
## timepnt.f1        -2.088235 0.3074310 128 -6.792533  0.0000
## timepnt.f2        -0.941176 0.2261102 128 -4.162469  0.0001
## timepnt.f3         0.147059 0.1916021 128  0.767522  0.4442
## timepnt.f4         1.205882 0.2261102 128  5.333163  0.0000
## trt.f1:timepnt.f1  0.558824 0.3074310 128  1.817720  0.0714
## trt.f1:timepnt.f2  0.529412 0.2261102 128  2.341389  0.0208
## trt.f1:timepnt.f3  0.147059 0.1916021 128  0.767522  0.4442
## trt.f1:timepnt.f4 -0.382353 0.2261102 128 -1.691003  0.0933
##  Correlation:
##                   (Intr) trt.f1 tmpn.1 tmpn.2 tmpn.3 tmpn.4 t.1:.1 t.1:.2
## trt.f1             0.000
## timepnt.f1        -0.075  0.000
## timepnt.f2         0.051  0.000  0.414
## timepnt.f3         0.120  0.000 -0.310 -0.003
## timepnt.f4         0.051  0.000 -0.677 -0.640 -0.003
## trt.f1:timepnt.f1  0.000 -0.075  0.000  0.000  0.000  0.000
## trt.f1:timepnt.f2  0.000  0.051  0.000  0.000  0.000  0.000  0.414
## trt.f1:timepnt.f3  0.000  0.120  0.000  0.000  0.000  0.000 -0.310 -0.003
## trt.f1:timepnt.f4  0.000  0.051  0.000  0.000  0.000  0.000 -0.677 -0.640
##                   t.1:.3
## trt.f1
## timepnt.f1
## timepnt.f2
## timepnt.f3
## timepnt.f4
## trt.f1:timepnt.f1
## trt.f1:timepnt.f2
## trt.f1:timepnt.f3
## trt.f1:timepnt.f4 -0.003
##
## Standardized Within-Group Residuals:
##         Min          Q1         Med          Q3         Max
## -2.37467405 -0.60851025 -0.04452508  0.72724395  2.41919926
##
## Number of Observations: 170
## Number of Groups: 34
```

```
##                 numDF denDF  F-value p-value
## (Intercept)         1   128 369.4090  <.0001
## trt.f               1    32   6.7869  0.0138
## timepnt.f           4   128  12.4477  <.0001
## trt.f:timepnt.f     4   128   2.0486  0.0914
```

```
## [1] 744.0077
```

```
## id1.f 1
## Conditional variance covariance matrix
##        1      2      3      4      5
## 1 15.708 14.212 12.857 11.632 10.524
## 2 14.212 15.708 14.212 12.857 11.632
## 3 12.857 14.212 15.708 14.212 12.857
## 4 11.632 12.857 14.212 15.708 14.212
## 5 10.524 11.632 12.857 14.212 15.708
##   Standard Deviations: 3.9634 3.9634 3.9634 3.9634 3.9634
```

The AR(1) model allows the covariance of the observations within cats to decrease as they are farther apart in time.

Based on having the smallest AIC value, it gives the best fit of the three models without having excess variance parameters.

$$s^2\left\{Y_{ijk}, Y_{ijk'}\right\} = 15.708(0.905)^{|k-k'|}$$

```
library(tidyverse)
library(kableExtra)
library(effectsize)
library(agricolae)
library(car)
library(PMCMRplus)
library(additivityTests)
library(lmerTest)
library(nlme)
```

Chapter 12

Analysis of Covariance (ANCOVA)

In various cases, researchers are interested in comparing treatments after adjusting for a numeric predictor that is associated with an experimental unit. That is often, but not necessarily, a baseline pre-treatment score on the unit. The goal is to determine whether treatment effects are present after controlling for the covariate.

When the Analysis of Covariance was first developed, the computations used were very complicated. Now, it's just as simple as fitting a regression with one or more predictors and dummy variables for the treatment conditions. Models can have one or more factors. We will consider the case with one treatment factor and one covariate.

12.1 Additive Model

In the additive model, the slope relating the covariate to the response is assumed to be the same for each treatment. As with the 1-Way ANOVA model, we will use r as the number of treatments with n_i replicates for treatment i. We will use centered X values for the covariate for ease of interpretation when describing adjusted means. We will also include $r-1$ indicators for treatments $1,\dots,r-1$ assuming again that treatment effects sum to 0.

$$Y_{ij} = \mu_{\bullet} + \tau_1 I_{ij1} + \cdots + \tau_{r-1} I_{ij,r-1} + \gamma\left(X_{ij} - \overline{X}_{\bullet\bullet}\right) + \epsilon_{ijk}$$

$$\epsilon_{ijk} \sim NID\left(0,\sigma^2\right) \qquad \sum_{i=1}^{r} \tau_i = 0 \quad \Rightarrow \quad \tau_r = -\sum_{i=1}^{r-1} \tau_i$$

$$I_{ij1} = \begin{cases} 1 & : \quad i = 1 \\ 0 & : \quad i = 2,\dots,r-1 \\ -1 & : \quad i = r \end{cases} \qquad \cdots \qquad I_{ij,r-1} = \begin{cases} 1 & : \quad i = r-1 \\ 0 & : \quad i = 1,\dots,r-2 \\ -1 & : \quad i = r \end{cases}$$

To test for treatment effects, controlling for the predictor, we test $H_0 : \tau_1 = \cdots = \tau_r = 0$. This easily conducted as a general linear test, where the full model contains the $r-1$ treatment indicator variables and the centered covariate. The reduced model contains only the centered covariate. The test is as follows.

$$TS : F^* = \frac{\left[\frac{SSE(R)-SSE(F)}{df_R - df_F}\right]}{\left[\frac{SSE(F)}{df_F}\right]} \qquad RR : F^* \geq F_{1-\alpha; df_R - df_F, df_F}$$

Note that for the current case, $df_R = n_T - 2$ and $df_F = n_T - 2 - (r-1)$.

Example 12.1 - Comparison of Skin Softeners A study compared two treatments and a control $r = 3$ for skin softening [Ma'Or et al., 1997]. The active treatments were as follows, and the covariate was the pre-treatment softness score for the subjects. There were 20 subjects per treatment, so $n_T = 3(20) = 60$. The response is skin roughness (lower scores are better).

- Treatment 1 - Gel Formulation

- Treatment 2 - Gel Formulation + 1% Dead Sea Concentrate
- Treatment 3 - Placebo

The R code and output are given below.

```
##   trt_Grp  pre_x post_y gel gelDS
## 1       1 121.10  76.30   1     0
## 2       1 375.40 265.10   1     0
## 3       1 138.05  82.92   1     0
## 4       1 114.85  91.75   1     0
## 5       1 263.91 175.32   1     0
## 6       1 255.88 152.54   1     0
```

```
##    trt_Grp  pre_x post_y gel gelDS
## 55       3 127.57 112.80   0     0
## 56       3 138.97 116.75   0     0
## 57       3 269.69 255.30   0     0
## 58       3 138.39 132.16   0     0
## 59       3 138.59 107.22   0     0
## 60       3 134.89 125.20   0     0
```

```
##
## Call:
## lm(formula = post_y ~ I1 + I2 + Xc, data = dsm)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -31.13 -10.35  -3.93  12.29  35.34
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 137.53667    2.08924  65.831  < 2e-16 ***
## I1           -2.66382    2.95473  -0.902    0.371
## I2          -26.91735    2.95595  -9.106 1.22e-12 ***
## Xc            0.70732    0.03257  21.716  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.18 on 56 degrees of freedom
## Multiple R-squared:  0.9121, Adjusted R-squared:  0.9073
## F-statistic: 193.6 on 3 and 56 DF,  p-value: < 2.2e-16
```

```
## Analysis of Variance Table
##
## Response: post_y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## I1         1   9853    9853  37.622 9.205e-08 ***
## I2         1  18744   18744  71.572 1.361e-11 ***
## Xc         1 123501  123501 471.566 < 2.2e-16 ***
## Residuals 56  14666     262
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Call:
## lm(formula = post_y ~ Xc, data = dsm)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -50.220 -19.668  -4.885  21.923  63.420
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 137.53667    3.66612   37.52   <2e-16 ***
## Xc            0.69686    0.05713   12.20   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 28.4 on 58 degrees of freedom
## Multiple R-squared:  0.7195, Adjusted R-squared:  0.7147
## F-statistic: 148.8 on 1 and 58 DF,  p-value: < 2.2e-16
```

```
## Analysis of Variance Table
##
## Response: post_y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Xc         1 119991  119991  148.79 < 2.2e-16 ***
## Residuals 58  46773     806
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Analysis of Variance Table
##
## Model 1: post_y ~ Xc
## Model 2: post_y ~ I1 + I2 + Xc
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1     58 46773
## 2     56 14666  2     32107 61.297 7.888e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -150.41  -48.35  -19.85    0.00   42.88  188.51
```

The F^*-statistic is 61.297 with 2 and 56 degrees of freedm and $P < .0001$. There is strong evidence of differences among the treatments, controlling for baseline score.

$$\nabla$$

12.1.1 Adjusted means

Where the lines cross at the centered X level of 0 are the **adjusted meabs**. That is, they are the predicted scores for the treatments at the "average" value of the uncentered covariate. Based on this additive model, we obtain the following adjusted means.

$$i = 1, \ldots, r-1: \quad \hat{\mu}_i = \hat{\mu}_\bullet + \hat{\tau}_i \qquad \hat{\mu}_r = \hat{\mu}_\bullet - \sum_{i=1}^{r-1} \hat{\tau}_i$$

The goal is to typically compare the adjusted means, as would be done for the 1-Way ANOVA. The standard errors depend on the variance-covariance matrix of the regression coefficients, which is obtained with the **vcov** function from the regression model fit.

$$i, i' < r \quad s\left\{\hat{\tau}_i - \hat{\tau}_{i'}\right\} = \sqrt{s^2\left\{\hat{\tau}_i\right\} + s^2\left\{\hat{\tau}_{i'}\right\} - 2s\left\{\hat{\tau}_i, \hat{\tau}_{i'}\right\}}$$

When comparing the first $r-1$ treatments with the r^{th}, we use the follwing result.

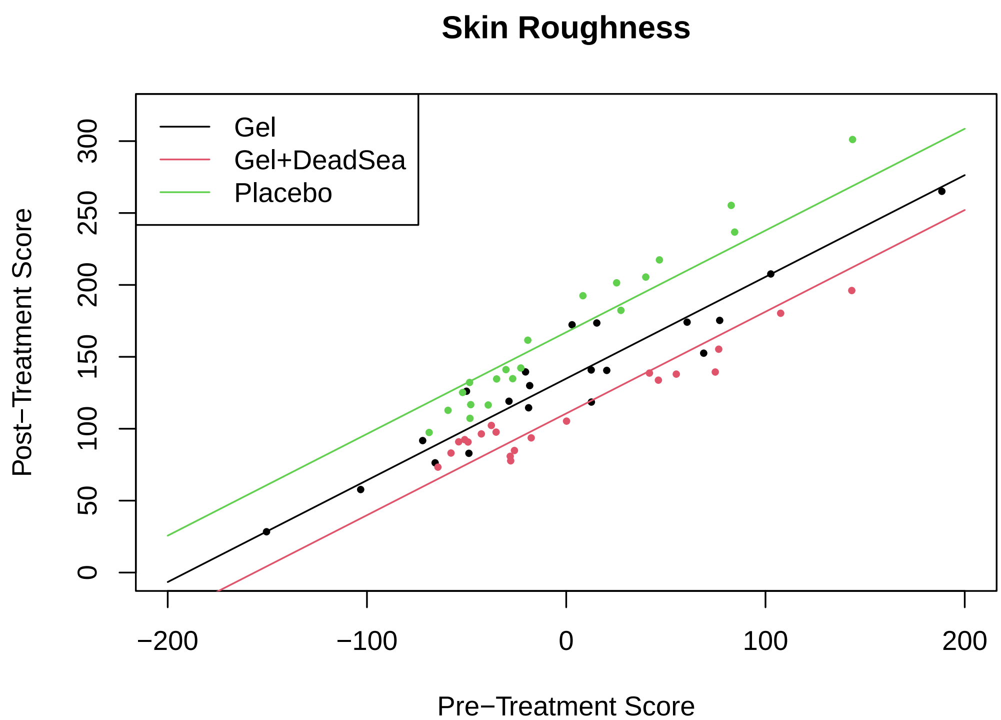


Figure 12.1: Analysis of Covariance for skin softener study - Additive Model

$$\hat{\tau}_i - \hat{\tau}_r = \hat{\tau}_i - \left(-\sum_{i=1}^{r-1}\hat{\tau}_i\right) = 2\hat{\tau}_i + \sum_{\substack{i'=1 \\ i'\neq i}}^{r-1}\hat{\tau}_{i'}$$

Then the (messy) standard errors for these differences are as follow.

$$s\left\{\hat{\tau}_i - \hat{\tau}_r\right\} = \sqrt{4s^2\left\{\hat{\tau}_i\right\} + \sum_{\substack{i'=1 \\ i'\neq i}}^{r-1} s^2\left\{\hat{\tau}_{i'}\right\} + 4\sum_{\substack{i'=1 \\ i'\neq i}}^{r-1} s\left\{\hat{\tau}_i, \hat{\tau}_{i'}\right\} + 2\sum_{\substack{i'=1 \\ i'\neq i}}^{r-2}\ \sum_{\substack{i''=i'+1 \\ i''\neq i}}^{r-1}}$$

Example 12.2 - Comparison of Skin Softeners We obtain the adjusted means and their differences along with multiple comparisons with the following R code.

```
##   trt_Grp  pre_x post_y gel gelDS
## 1       1 121.10  76.30   1     0
## 2       1 375.40 265.10   1     0
## 3       1 138.05  82.92   1     0
## 4       1 114.85  91.75   1     0
## 5       1 263.91 175.32   1     0
## 6       1 255.88 152.54   1     0
```

```
##    trt_Grp  pre_x post_y gel gelDS
## 55       3 127.57 112.80   0     0
## 56       3 138.97 116.75   0     0
## 57       3 269.69 255.30   0     0
## 58       3 138.39 132.16   0     0
## 59       3 138.59 107.22   0     0
```

```
## 60       3 134.89 125.20   0     0
```

```
##
## Call:
## lm(formula = post_y ~ I1 + I2 + Xc, data = dsm)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -31.13 -10.35  -3.93  12.29  35.34
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 137.53667    2.08924  65.831  < 2e-16 ***
## I1           -2.66382    2.95473  -0.902    0.371
## I2          -26.91735    2.95595  -9.106 1.22e-12 ***
## Xc            0.70732    0.03257  21.716  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.18 on 56 degrees of freedom
## Multiple R-squared:  0.9121, Adjusted R-squared:  0.9073
## F-statistic: 193.6 on 3 and 56 DF,  p-value: < 2.2e-16
```

```
## Analysis of Variance Table
##
## Response: post_y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## I1         1   9853    9853  37.622 9.205e-08 ***
## I2         1  18744   18744  71.572 1.361e-11 ***
## Xc         1 123501  123501 471.566 < 2.2e-16 ***
## Residuals 56  14666     262
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## [1] 134.8728 110.6193 167.1178
```

```
##      i i'  AdjMn_i AdjMn_i'     Diff       LB      UB p adj
## [1,] 2  1 110.6193 134.8728 -24.2535 -41.6821 -6.8250     0
## [2,] 3  1 167.1178 134.8728  32.2450  14.8202 49.6698     0
## [3,] 3  2 167.1178 110.6193  56.4985  39.0665 73.9306     0
```

All three pairs of treatments are significantly different. Treatment 2 (Gel+DeadSea) has significantly lower roughness scores than the other treatments and Gel has significantly lower scores than Placebo.

∇

12.2 Interaction Model

When the slope relating the response to the covariate differs among treatments, the model contains interactions. This can be tested by including cross-product terms for the treatment indicator variables and the centered covariate. Then the full model containing these extra terms can be compared with the additive (reduced) model. The intraction model is given below.

$$Y_{ij} = \mu_{\bullet} + \sum_{k=1}^{r-1} \tau_k I_{ijk} + \gamma \left(X_{ij} - \overline{X}_{\bullet\bullet} \right) + \sum_{k=1}^{r-1} \beta_k I_{ijk} \left(X_{ij} - \overline{X}_{\bullet\bullet} \right) + \epsilon_{ijk}$$

Comparing adjusted means is more difficult when there is an interaction, as the difference among treatment means differs at different X levels (unlike the additive model). A method was developed by Johnson and Neyman to find regions of X values where two treatments are significantly different and where they are not. We will not pursue that method here, but will run the general linear test for interactions.

Example 12.3 - Comparison of Skin Softeners We obtain the adjusted means and their differences along with multiple comparisons with the following R code.

```
##   trt_Grp  pre_x post_y gel gelDS
## 1       1 121.10  76.30   1     0
## 2       1 375.40 265.10   1     0
## 3       1 138.05  82.92   1     0
## 4       1 114.85  91.75   1     0
## 5       1 263.91 175.32   1     0
## 6       1 255.88 152.54   1     0
```

```
##    trt_Grp  pre_x post_y gel gelDS
## 55       3 127.57 112.80   0     0
## 56       3 138.97 116.75   0     0
## 57       3 269.69 255.30   0     0
## 58       3 138.39 132.16   0     0
## 59       3 138.59 107.22   0     0
## 60       3 134.89 125.20   0     0
```

```
##
## Call:
## lm(formula = post_y ~ I1 + I2 + Xc, data = dsm)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -31.13 -10.35  -3.93  12.29  35.34
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 137.53667    2.08924  65.831  < 2e-16 ***
## I1           -2.66382    2.95473  -0.902    0.371
## I2          -26.91735    2.95595  -9.106 1.22e-12 ***
## Xc            0.70732    0.03257  21.716  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 16.18 on 56 degrees of freedom
## Multiple R-squared:  0.9121, Adjusted R-squared:  0.9073
## F-statistic: 193.6 on 3 and 56 DF,  p-value: < 2.2e-16
```

```
## Analysis of Variance Table
##
## Response: post_y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## I1         1   9853    9853  37.622 9.205e-08 ***
## I2         1  18744   18744  71.572 1.361e-11 ***
## Xc         1 123501  123501 471.566 < 2.2e-16 ***
## Residuals 56  14666     262
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Call:
## lm(formula = post_y ~ I1 + I2 + Xc + I(I1 * Xc) + I(I2 * Xc),
##     data = dsm)
```

```
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -28.385  -7.771   0.046   7.394  35.485
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 137.83352    1.64131  83.978  < 2e-16 ***
## I1           -2.99032    2.32058  -1.289   0.2030
## I2          -26.78674    2.32170 -11.538 3.41e-16 ***
## Xc            0.72542    0.02632  27.564  < 2e-16 ***
## I(I1 * Xc)   -0.05744    0.03431  -1.674   0.0999 .
## I(I2 * Xc)   -0.17549    0.03784  -4.638 2.28e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.71 on 54 degrees of freedom
## Multiple R-squared:  0.9477, Adjusted R-squared:  0.9429
## F-statistic: 195.8 on 5 and 54 DF,  p-value: < 2.2e-16
```

```
## Analysis of Variance Table
##
## Response: post_y
##            Df Sum Sq Mean Sq F value    Pr(>F)
## I1          1   9853    9853  61.026 1.985e-10 ***
## I2          1  18744   18744 116.096 4.560e-15 ***
## Xc          1 123501  123501 764.925 < 2.2e-16 ***
## I(I1 * Xc)  1   2475    2475  15.330 0.0002553 ***
## I(I2 * Xc)  1   3472    3472  21.507 2.279e-05 ***
## Residuals  54   8719     161
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Analysis of Variance Table
##
## Model 1: post_y ~ I1 + I2 + Xc
## Model 2: post_y ~ I1 + I2 + Xc + I(I1 * Xc) + I(I2 * Xc)
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     56 14666.1
## 2     54  8718.5  2    5947.5 18.419 7.971e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The interaction is significant ($F^* = 18.42, P < .0001$). Based on the plot, there appears to be larger treatment differences among patients with higher baseline skin roughness.

∇

```
library(tidyverse)
library(kableExtra)
library(effectsize)
library(agricolae)
library(car)
library(PMCMRplus)
library(additivityTests)
library(lmerTest)
library(nlme)
```

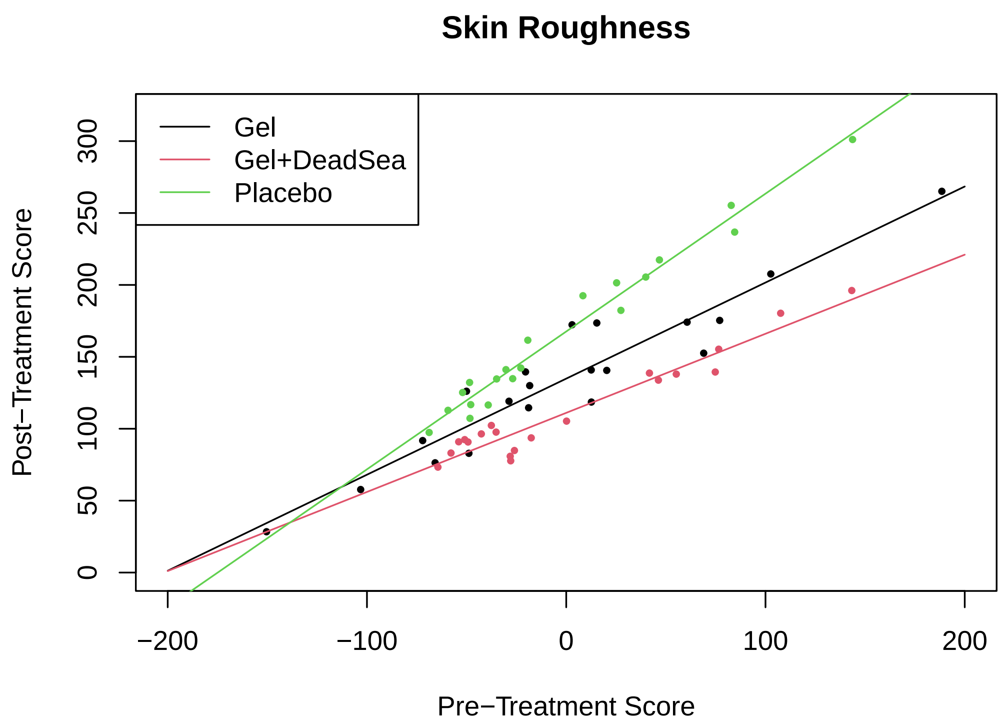


Figure 12.2: Analysis of Covariance for skin softener study - Interaction Model

```
library(rsm)
library(mixexp)
```

```
## Warning: package 'mixexp' was built under R version 4.1.3
```

```
## Loading required package: lattice
```

```
## Loading required package: grid
```

```
## Loading required package: daewr
```

```
library(daewr)
```

Chapter 13

Response Surface and Mixture Designs

Response surface and mixture designs are used in a wide range of engineering fields. The goal is to choose levels of a group of numeric predictor variables that optimize the response variable(s). Response surfaces are based on k predictors at several levels and model based on a second-order model with linear, interaction, and quadratic terms among the predictors. Mixture designs have several inputs, but these are restricted to sum to 1 (or 100%).

13.1 Response Surface Designs

The model for a second order response surface with k factors is given below and can be implemented easily using the **rsm** package in R.

$$Y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \sum_{i=1}^{k-1} \sum_{i'=i+1}^{k} \beta_{ii'} X_i X_i' + \sum_{i=1}^{k} \beta_{ii} X_i^2 + \epsilon$$

The model has $1+k+k(k-1)/2+k$ parameters. Once the regression model is fit, individual parameters and groups (2-factor interactions, quadratic terms) can be tested for significance. The fitted model can be written in scalar or matrix form.

$$\hat{Y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i X_i + \sum_{i=1}^{k-1} \sum_{i'=i+1}^{k} \hat{\beta}_{ii'} X_i X_i' + \sum_{i=1}^{k} \hat{\beta}_{ii} X_i^2 = \hat{\beta}_0 + B_1' x + x' B_2 x$$

where x, B_1, B_2 are defined as follow.

$$x = \begin{bmatrix} X_1 \\ \vdots \\ X_k \end{bmatrix} \qquad B_1 = \begin{bmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \qquad B_2 = \begin{bmatrix} \hat{\beta}_1 1 & \hat{\beta}_{12}/2 & \cdots & \hat{\beta}_{1k}/2 \\ \hat{\beta}_{12}/2 & \hat{\beta}_{22} & \cdots & \hat{\beta}_{2k}/2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\beta}_{1k}/2 & \hat{\beta}_{2k}/2 & \cdots & \hat{\beta}_{kk} \end{bmatrix}$$

Taking the derivative of $\hat{Y}$ with respect to x, setting it equal to zero, ives the optimal set of inputs for the k factors (which may fall outside of possible values). This is printed directly in the summary of the rsm fit.

$$\frac{\partial \hat{Y}}{\partial x} = B_1 + 2B_2 x \quad \stackrel{\text{set}}{=} 0 \quad \Rightarrow \quad x^* = -\frac{1}{2} B_2^{-1} B_1$$

Two commonly used designs for response surfaces are the **Central Composite Design** and the **Box-Behnken Design**.

In the Central Composite Design, there are k factors, each with three equally spaced levels coded $-1, 0, +1$. A 2^k factorial is set up with each factor at $+/-1$. Then $2k$ observations are made with $k-1$ variables at their 0 levels and the remaining factor at $\pm\alpha$ where α is commonly $\sqrt{2}$, $\sqrt{3}$, or 2, these are labelled "axial points." Finally, there are c observations with each of the k factors at their center (0) levels, which permits a goodness of fit test.

Example 13.1 - Solar Drying of Avocado Pulp A study was conducted as k factor response surface with k factors to measure solar drying of avocado pulp [Kowarit et al., 2024]. The 4 factors are described below, and the response was moisture content of the avocado pulp (lower values are better). This is a central composite design with $\alpha = 2$.

- **Hot Air Temperature** (X_1, Celsius): -1=40, 0=55, +1=70 (axial points at 25, 85)
- **Drying Time** (X_2, hours): -1=13, 0=16.5, +1=20 (axial points at 9.5, 23.5)
- **Raw Material Thickness** (X_3, cm): -1=0.5, 0=0.75, +1=1.00 (axial points at 0.25, 1.25)
- **Wind Speed** (X_4, meters/second): -1=0.12, 0=0.19, +1=0.26 (axial points at 0.05, 0.33)

There were $2^4 = 16$ runs with the 4 factors at +/-1, $4(2) = 8$ runs at the axial points for each factor (with the other 3 factors at their 0 level), and 6 runs at the center points, for a total of 30 experimental runs. The R program and output are given below, with response surface and contour plots in Figure 13.1 and Figure 13.2, respectively.

```
##   run hotAirTemp dryTime rawMatThck windSpeed moisture
## 1   1         70    20.0       1.00      0.26    20.33
## 2   2         70    13.0       1.00      0.12    20.15
## 3   3         85    16.5       0.75      0.19    18.26
## 4   4         55    16.5       0.75      0.19    18.90
## 5   5         55    16.5       0.75      0.19    17.99
## 6   6         70    20.0       0.50      0.12    18.54
```

```
##    run hotAirTemp dryTime rawMatThck windSpeed moisture
## 25  25         70    13.0       1.00      0.26    20.29
## 26  26         55     9.5       0.75      0.19    19.03
## 27  27         40    20.0       1.00      0.26    21.61
## 28  28         55    16.5       0.75      0.19    18.90
## 29  29         55    16.5       0.75      0.19    18.90
## 30  30         40    13.0       0.50      0.12    17.73
```

```
##
## Call:
## rsm(formula = moisture ~ SO(hotAirTemp, dryTime, rawMatThck,
##     windSpeed), data = am)
##
##                        Estimate  Std. Error t value Pr(>|t|)
## (Intercept)           2.1996e+01  1.1195e+01  1.9649 0.068230 .
## hotAirTemp            2.6835e-01  1.5944e-01  1.6831 0.113053
## dryTime              -4.8311e-01  7.3777e-01 -0.6548 0.522495
## rawMatThck           -8.5945e+00  9.1563e+00 -0.9386 0.362780
## windSpeed            -6.4616e+01  3.2146e+01 -2.0101 0.062761 .
## hotAirTemp:dryTime    -6.8333e-03  5.4205e-03 -1.2607 0.226691
## hotAirTemp:rawMatThck -5.9667e-02  7.5887e-02 -0.7863 0.443959
## hotAirTemp:windSpeed   2.1786e-01  2.7102e-01  0.8038 0.434049
## dryTime:rawMatThck    -1.4143e-01  3.2523e-01 -0.4349 0.669856
## dryTime:windSpeed      1.5561e+00  1.1615e+00  1.3397 0.200282
## rawMatThck:windSpeed  -2.4071e+01  1.6261e+01 -1.4803 0.159495
## hotAirTemp^2          -1.2204e-03  9.6599e-04 -1.2633 0.225752
## dryTime^2              2.7585e-02  1.7743e-02  1.5547 0.140853
## rawMatThck^2           1.2027e+01  3.4776e+00  3.4584 0.003511 **
## windSpeed^2            1.3274e+02  4.4357e+01  2.9925 0.009110 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Multiple R-squared:  0.7643, Adjusted R-squared:  0.5444
## F-statistic: 3.475 on 14 and 15 DF,  p-value: 0.01123
##
## Analysis of Variance Table
##
```

```
## Response: moisture
##                                              Df  Sum Sq Mean Sq F value
## FO(hotAirTemp, dryTime, rawMatThck, windSpeed)   4 23.1131  5.7783  4.4595
## TWI(hotAirTemp, dryTime, rawMatThck, windSpeed)  6  9.1073  1.5179  1.1715
## PQ(hotAirTemp, dryTime, rawMatThck, windSpeed)   4 30.8123  7.7031  5.9450
## Residuals                                       15 19.4359  1.2957
## Lack of fit                                     10 18.7458  1.8746 13.5823
## Pure error                                       5  0.6901  0.1380
##                                                   Pr(>F)
## FO(hotAirTemp, dryTime, rawMatThck, windSpeed)  0.014231
## TWI(hotAirTemp, dryTime, rawMatThck, windSpeed) 0.371796
## PQ(hotAirTemp, dryTime, rawMatThck, windSpeed)  0.004511
## Residuals
## Lack of fit                                     0.005031
## Pure error
##
## Stationary point of response surface:
## hotAirTemp    dryTime rawMatThck  windSpeed
## 67.1141331 14.0922992  0.7835529  0.1767657
##
## Eigenanalysis:
## eigen() decomposition
## $values
## [1] 133.931086216  10.838362960   0.023680706  -0.002003314
##
## $vectors
##                     [,1]          [,2]          [,3]         [,4]
## hotAirTemp  0.0008310985  0.0017517386 -0.1599582827  0.987121871
## dryTime     0.0058342613 -0.0005625942  0.9871073880  0.159952022
## rawMatThck -0.0982548808 -0.9951596092 -0.0002787034  0.001803565
## windSpeed   0.9951438334 -0.0982546031 -0.0056810736 -0.001584081
```

The linear (first order) terms and polynomial quadratic terms are significant as groups with P-values of .0142 and .0045, respectively. The two-factor interactions are not, with P=.3718. The stationary (optimal) point is x^*=(67.11,14.09,0.78,0.18).

∇

The Box-Behnken design places runs at every combination of ± 1 for each pair of factors, with all other factors at their 0 (central) level. There are multiple runs for each factor at their central level. The analysis is the same as for the Central Composite Design.

Example 13.2 – Gels of Diclofenac and Curcumin for Transdermal Drug Delivery

A study (Chaudhary, et al (2011)) had 3 factors, each at 3 levels [Chaudhary et al., 2011]. The factors were as follow.

- **Polymer Concentration** - (0.5, 1.0, 1.5 % w/w)
- **Ethanol Concentration** (10, 15, 20 % w/w)
- **PG Concentration** (5, 10, 15 %w/w).

There were three response variables: $Y1$ = Flux of DDEA (mg/(cm2 h)), $Y2$ = Flux of CRM (mg/(cm2 h)), and $Y3$ = Viscosity of Gel (cp). The experiment had $n = 17$ runs. The design and data are given below.

∇

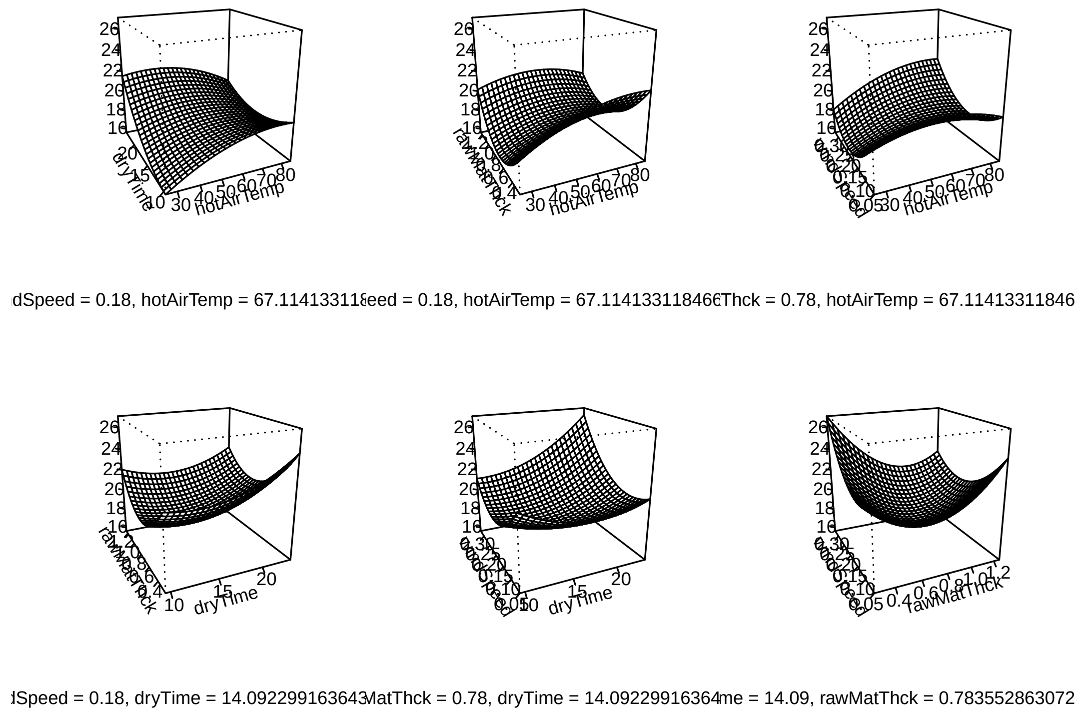


Figure 13.1: Response Surface for avocado pulp study - 3D Plots

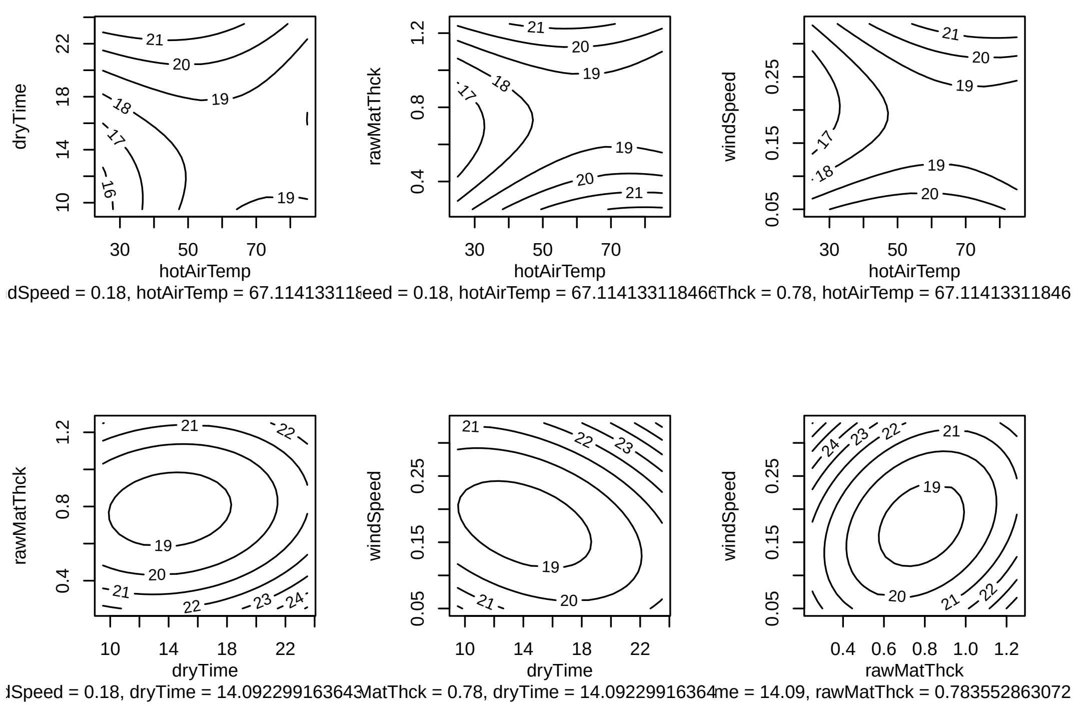


Figure 13.2: Response Surface for avocado pulp study - Contour Plots

Table 13.1: Box-Behnken Design for Drug Delivery Study

runnum	polyconc	ethnconc	pgconc	fluxddea	fluxcrm	viscgel
1	-1	-1	0	0.67	1.48	185
2	0	0	0	0.24	1.90	1924
3	0	1	-1	0.25	3.31	2018
4	0	-1	1	0.22	2.88	2310
5	1	0	-1	0.11	3.30	3227
6	-1	1	0	0.67	1.72	145
7	-1	0	-1	0.69	1.37	143
8	0	0	0	0.23	1.87	1923
9	-1	0	1	0.67	1.52	176
10	0	0	0	0.24	1.91	1800
11	0	0	0	0.24	1.92	1921
12	1	0	1	0.17	2.98	3071
13	0	1	1	0.23	2.01	1783
14	0	0	0	0.24	1.87	1922
15	1	-1	0	0.11	3.07	3320
16	1	1	0	0.16	3.45	2801
17	0	-1	-1	0.20	1.71	2245

13.2 Mixture Models

Mixture models are similar to response surface models, with the restriction that the sum of the X levels is equal to 1, that is the X variables are components of a mixture. There are four widely used models that can be implemented in the **mixexp** package in R. Note that these models do not include an intercept, as the X's sum to 1, and we will use as k as the number of mixture components.

$$\text{Linear:} \quad E\{Y\} = \sum_{i=1}^{k} \beta_i X_i$$

$$\text{Quadratic:} \quad E\{Y\} = \sum_{i=1}^{k} \beta_i X_i + \sum_{i=1}^{k-1} \sum_{i'=i+1}^{k} \beta_{ii'} X_i X_{i'}$$

$$\text{Full Cubic:} \quad E\{Y\} = \sum_{i=1}^{k} \beta_i X_i + \sum_{i=1}^{k-1} \sum_{i'=i+1}^{k} \beta_{ii'} X_i X_{i'} + \sum_{i=1}^{k-1} \sum_{i'=i+1}^{k} \delta_{ii'} X_i X_{i'} (X_i - X_{i'}) +$$

$$+ \sum_{i=1}^{k-2} \sum_{i'=i+1}^{k-1} \sum_{i''=i'+1}^{k} \beta_{ii'i''} X_i X_{i'} X_{i''}$$

$$\text{Special Cubic:} \quad E\{Y\} = \sum_{i=1}^{k} \beta_i X_i + \sum_{i=1}^{k-1} \sum_{i'=i+1}^{k} \beta_{ii'} X_i X_{i'} + \sum_{i=1}^{k-2} \sum_{i'=i+1}^{k-1} \sum_{i''=i'+1}^{k} \beta_{ii'i''} X_i X_{i'} X_{i''}$$

The goal is to choose the mixture that optimizes the output.

Example 13.3 - Breaking Strength of Lipstick Blends A mixture experiment considered $k = 3$ inputs for lipstick [Hui et al., 2017]. One response was breaking strength (grams), with the following inputs, this was Stage 1C in the paper, and the Full Cubic model was fit.

- **Sweet Almond Oil** (X_1)
- **Hydrogenated Polyisobutene** (X_2)
- **Octydodecanol** (X_3)

The R code and output are given below.

```
##    ExpNum      X1      X2      X3 Break Soft
## 1       1 1.00000 0.00000 0.00000 325.8 62.0
## 2       2 1.00000 0.00000 0.00000 239.2 53.0
## 3       3 0.00000 1.00000 0.00000 332.9 61.3
## 4       4 0.00000 0.00000 1.00000 242.1 58.8
## 5       5 0.66667 0.33333 0.00000 247.3 53.3
## 6       6 0.33333 0.66667 0.00000 258.9 53.3
## 7       7 0.66667 0.00000 0.33333 342.1 61.5
## 8       8 0.66667 0.00000 0.33333 343.0 57.8
## 9       9 0.33333 0.33333 0.33333 292.5 53.8
## 10     10 0.00000 0.66667 0.33333 299.9 52.8
## 11     11 0.33333 0.00000 0.66667 399.2 61.3
## 12     12 0.00000 0.33333 0.66667 407.6 58.0
## 13     13 0.66667 0.16667 0.16667 258.9 53.3
## 14     14 0.16667 0.66667 0.16667 326.0 55.0
## 15     15 0.16667 0.16667 0.66667 325.1 53.8
## 16     16 0.16667 0.16667 0.66667 436.8 59.0

##
##                coefficients    Std.err    t.value         Prob
## X1                 283.4746   33.32269  8.5069535 0.0001444655
## X2                 331.9844   47.06751  7.0533671 0.0004063754
## X3                 244.3524   47.08335  5.1897831 0.0020349071
## cubic(X1, X2)    -201.1627  393.63036 -0.5110446 0.6275740863
## cubic(X1, X3)    -437.2026  349.65399 -1.2503865 0.2577122053
## cubic(X2, X3)    -667.6002  388.25675 -1.7194813 0.1363220821
## X2:X1            -222.8240  196.31796 -1.1350157 0.2996670546
## X3:X1             464.4747  180.58235  2.5720936 0.0422159307
## X2:X3             314.0409  210.14454  1.4944045 0.1856908476
## X2:X3:X1        -1152.8906 1320.00853 -0.8733963 0.4160293996
##
## Residual standard error:  47.18515  on  6 degrees of freedom
## Corrected Multiple R-squared:  0.7582724

## Analysis of Variance Table
##
## Response: Break
##                Df Sum Sq Mean Sq  F value    Pr(>F)
## X1              1 822487  822487 369.4245 1.283e-06 ***
## X2              1 418251  418251 187.8594 9.375e-06 ***
## X3              1 380430  380430 170.8724 1.236e-05 ***
## cubic(X1, X2)  1    802     802   0.3604   0.57023
## cubic(X1, X3)  1   1624    1624   0.7293   0.42589
## cubic(X2, X3)  1   5300    5300   2.3807   0.17378
## X1:X2           1   7691    7691   3.4546   0.11243
## X1:X3           1  11369   11369   5.1064   0.06457 .
## X2:X3           1   3437    3437   1.5437   0.26042
## X1:X2:X3        1   1698    1698   0.7629   0.41602
## Residuals       6  13358    2226
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Call:
## lm(formula = mixmodnI, data = frame)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -60.722 -15.433   0.876  12.755  50.978
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## X1               283.47      33.32   8.507 0.000144 ***
## X2               331.98      47.07   7.053 0.000406 ***
## X3               244.35      47.08   5.190 0.002035 **
## cubic(X1, X2)  -201.16     393.63  -0.511 0.627570
## cubic(X1, X3)  -437.20     349.65  -1.250 0.257712
## cubic(X2, X3)  -667.60     388.25  -1.719 0.136321
## X1:X2          -222.83     196.32  -1.135 0.299661
## X1:X3           464.47     180.58   2.572 0.042215 *
## X2:X3           314.04     210.14   1.494 0.185689
## X1:X2:X3      -1152.90    1320.00  -0.873 0.416021
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 47.18 on 6 degrees of freedom
## Multiple R-squared:  0.992,  Adjusted R-squared:  0.9786
## F-statistic: 74.25 on 10 and 6 DF,  p-value: 1.76e-05
```

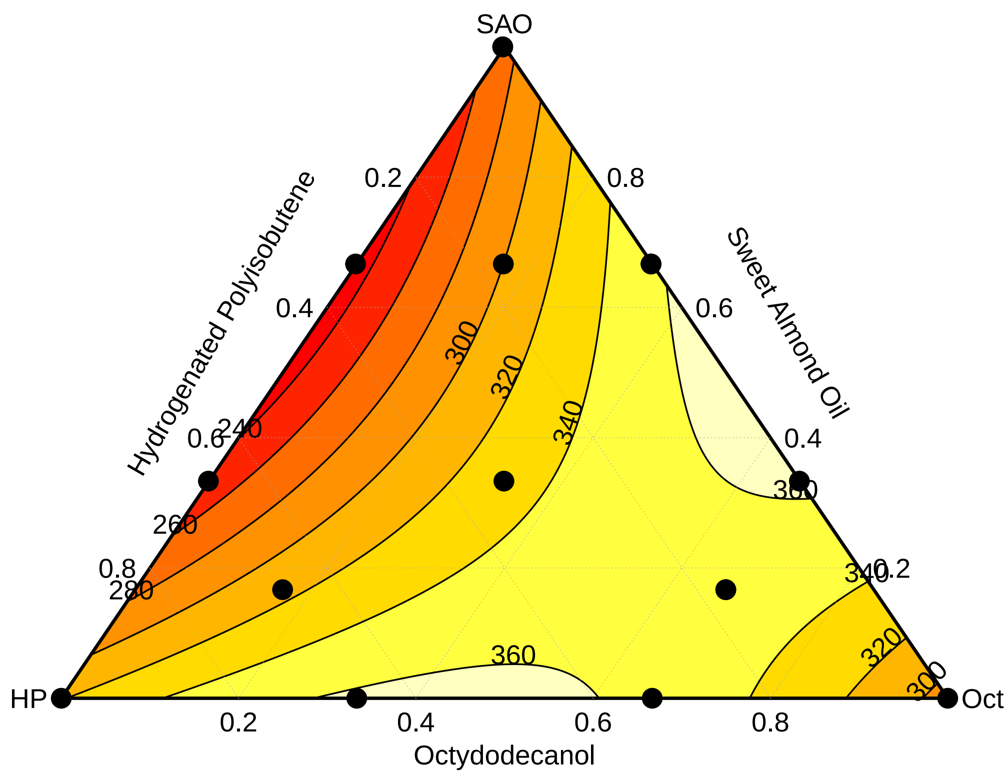


Figure 13.3: Mixture Plot for lipstick study

```
##         X1      X2      X3 Break
## 1  1.00000 0.00000 0.00000 325.8 283.4746
## 2  1.00000 0.00000 0.00000 239.2 283.4746
## 3  0.00000 1.00000 0.00000 332.9 331.9841
## 4  0.00000 0.00000 1.00000 242.1 244.3521
## 5  0.66667 0.33333 0.00000 247.3 235.2266
## 6  0.33333 0.66667 0.00000 258.9 281.1992
## 7  0.66667 0.00000 0.33333 342.1 341.2641
```

```
## 8  0.66667 0.00000 0.33333 343.0 341.2641
## 9  0.33333 0.33333 0.33333 292.5 305.6438
## 10 0.00000 0.66667 0.33333 299.9 323.1077
## 11 0.33333 0.00000 0.66667 399.2 392.9941
## 12 0.00000 0.33333 0.66667 407.6 392.8011
## 13 0.66667 0.16667 0.16667 258.9 263.7996
## 14 0.16667 0.66667 0.16667 326.0 285.0699
## 15 0.16667 0.16667 0.66667 325.1 385.8225
## 16 0.16667 0.16667 0.66667 436.8 385.8225
```

The highest breaking strengths (> 360) appear at blends of approximately 50/50% between Sweet Almond Oil and Octydodecanol, based on the plot of the simplex.

∇

Bibliography

M.I. Aksu, H. Imik, and M. Karoglu. Influence of dietary sorghum (sorghum vulgare) and corn supplemented with methionine on cut-up pieces weights of broiler carcass and quality properties of breast and drumsticks meat. *Food Science and Technology International*, 13(5):361–367, 2007.

C. Beata, J. Cordel, and N. Marlois. Effect of alpha-casozepine (zylkene) on cats with anxiety. *Journal of Veterinary Behavior: Clinical Applications and Research*, 2(2):40–46, 2007.

D. Bishop, C.E. Frazier, L. Lanza-Kaduce, and L. Winner. The transfer of juveniles to criminal court: Does it make a difference? *Crime & Delinquency*, 42:171–191, 1996.

F. Bronet, M-C. Nogales, E. Martinez, M. Ariza, C. Rubio, J-A. Garcia-Velasco, and M. Meseguer. Is there a relationship between time-lapse parameters and embryo sex? *fertility and Sterility*, 103(2):396–401, 2015.

D.J. Carr, C. Lankester, A. Peare, N. Fabri, and N. Gridley. Does quilting improve the fragment protective performance of body armour? *Textile Research Journal*, 82(9):883–888, 2012.

A. Carraro, C.A. Elliott, and E. Gobbi. Perceived treadmill function is correlated with enjoyment of use in trained runners: A user-centered approach. *Applied Ergonomics*, 74:37–40, 2019.

P.-C. Chang, S.-Y. Chou, and K.-K. Shieh. Reading performance and visual fatigue when using electronic displays in long-duration reading tasks under various lighting conditions. *Displays*, 34:208–214, 2013.

H. Chaudhary, K. Kohli, S. Amin, P. Rathee, and V. Kumar. Optimization and formulation design of gels of diclfenac and curcumin for transdermal drug delivery by box-behnken statistical design. *Journal of Pharmaceutical Sciences*, 100(2):580–593, 2011.

S.R. Cook and G. Proulx. Mechanical evaluation and performance improvement of the rotating jaw conibear 120 trap. *Journal of Testing and Evaluation*, 17(3):190–195, 1989.

S. Deverajan, R.H. McQueen, and S. Wen. Can common finishing treatments used in chef jacket fabrics improve protection against scald injury? *Fashion & Textiles*, 4(19):1:15, 2017. doi: 10.1186/s40691-017-0103-3.

M. Fidaleo, M. Moresi, A. Cammaroto, M. Ladrange, and R. Nardi. Soy sauce desalting by electrodialysis. *Journal of Food Engineering*, 110:175–181, 2012.

M.L. Garcia and J. Diaz. Combability measurements on human hair. *Journal of the Society of Cosmetic Chemists*, 27:379–398, 1976.

G. Genova, P. Iacopini, M. Baldi, A. Ranieri, P. Storchi, and L. Sebastiani. Temperature and storage effects on antioxidant activity of juice from red and white grapes. *International Journal of Food Science & Technology*, 47: 13–23, 2012.

A.N. Grand and L.N. Bell. Caffeine content of fountain and private label store brand carbonated beverages. *Journal of the American Dietetic Association*, 97(2):179–182, 1997.

R.G. Gullberg. Employing components-of-variance to evaluate forensic breath test instruments. *Science and Justice*, 48:2–7, 2008.

T. Hanihara, Y. Dodo, O. Kondo, T. Nara, N. Doi, and N. Sensui. Intra- and inter-observer errors in facial flatness measurements. *Anthropological Science*, 107(1):25–39, 1999.

M. Hinneh, E.E. Abotsi, D. Van de Walle, D.A. Tzompa-Sosa, A. De Winne, J. Simonis, J. Messens, K. anf Van Durme, E.O. Afoakwa, L. De Cooman, and K. Dewettink. Pod storage with roasting: A tool to diversifying the flavor profiles of dark chocolates produced from ???bulk??? cocoa beans? (part ii: Quality and sensory profiling of chocolates). *Food Research International*, 132(109116), 2020.

S.-H. Hsu and S.-P. Wu. An investigation for determining the optimal length of chopsticks. *Applied Ergonomics*, 22(6):395–400, 1991.

W.N. Hui, S. Tamburic, and P. Grant-Poss. Lip-smacking results. *Cosmetics & Toiletries*, 132(3):48–68, 2017.

K. Hyllegard, J. Ogle, and R.-N. Yan. The impact of advertising message strategy - fair labour v sex appeal - upon gen y consumers' intent to patronize an apparel retailer. *Journal of Fashion Marketing and Management*, 13(1): 109–127, 2009.

S. Ismail, S. Siddiqui, F. Shafiq, M. Ishaq, and S. Khan. Blood transfusion in patients having caesarean section: a prospective multicentre observational study of practice in three pakistan hospitals. *International Journal of Obstetric Anesthesia*, 23:253–259, 2014.

C.R. Jensen. Variance components calculations: Common methods and misapplications in the semiconductor industry. *Quality Engineering*, 14(4):647–657, 2002.

J. Jung and Y.J. Ahn. Effects of interface on procedural skill transfer in virtual training: Lifeboat launching operation study. *Computer Animation and Virtual Worlds*, 29:e1812, 2018. doi: 10.1002/cav.1812. URL https://doi.org/10.1002/cav.1812.

Z. Keshtkaran, F. Sharif, and M. Rambod. Students' readiness for and perception of inter-professional learning: A cross-sectional study. *Nurse Education Today*, 34:991–998, 2014.

S. Kowarit, K. Sathpornprasath, and S.N. Jansri. Application of hot air-derived rsm conditions and shading for solar drying of avocado pulp and its properties. *Solar Energy*, 278:1–12, 2024. doi: 10.1016/j.solener.2024.112768. URL https://doi.org/10.1016/j.solener.2024.112768.

D. Landy and H. Sigall. Beauty is talent: Task evaluation as a function of the performer's physical attraction. *Journal of Personality and Social Psychology*, 29(3):299–304, 1974.

Z. Ma'Or, S. Yehuda, and W. Voss. Skin smoothing effects of dead sea minerals: Comparative profilometric evaluation of skin surface. *International Journal of Cosmetic Science*, 19:105–110, 1997.

D. Marchiori, O. Cornielle, and O. Klein. Container size influences snack food intake independently of portion size. *Appetite*, 58:814–817, 2012.

M. Meilgaard. Hop analysis, cohumulone factor and the bitterness of beer: Review and critical evaluation. *Journal of the Institute of Brewing*, 66(1):35–50, 1960.

D. O'Hare and N. Stenhouse. Under the weather: An evaluation of different modes of presenting meteorological information for pilots. *Applied Ergonomics*, 40:688–693, 2009.

C. O'Keeffe and R. Wiseman. Testing alleged mediumship: Methods and results. *British Journal of Psychology*, 96:165–179, 2005.

A.S. Oliveira, F.M. Dalla Nora, R.O. Mello, P.A. Mello, B. Tischer, A.B. Costa, and J.S. Barin. One-shot, reagent-free determination of the alcohol content of distilled beverages by thermal infrared enthalpimetry. *Talanta*, 171: 335–340, 2017. doi: 10.1016/j.talanta.2017.05.011. URL http://dx.doi.org/10.1016/j.talanta.2017.05.011.

W. Ouedrhiria, M. Balouiri, S. Bouhdid, M., S. Moja, F.O. Chahdie, M. Taleb, and H. Greche. Mixture designs of origanum compactum, origanum majorana and thymus serpyllum essential oils: Optimization of their antibacterial effect. *Industrial Crops and Products*, 89:1–9, 2016.

A. Ranthinam, J.R. Rao, and B.U. Nair. Adsorption of phenol onto activated carbon from seaweed: Determination of the optimal experimental parameters using factorial design. *Journal of the Taiwan Institute of Chemical Engineers*, 42:952–956, 2011.

A.M. Schultz and H.H. Biswell. Competition between grasses reseeded on burned brushlands in california. *Journal of Range Management*, 5(5):338–345, 1952.

D.K. Simonson. Creative life cycles in literature: Poets vs novelists or conceptualists versus experimentalists? *Psychology of Aesthetics, Creativity, and the Arts*, 1(3):133–139, 2007.

K.V. Smith. Stock price and economic indexes for generating efficient portfolios. *The Journal of Business*, 42(3): 332–336, 1969.

P.K. Smith, T. Niiler, and P.W. McCullough. Evaluating makiwara punching board performance. *Journal of Asian Martial Arts*, 19(2):34–45, 2010.

X-j. Song, M. Zhang, and A.S. Mujumdar. Optimization of vacuum microwave predrying and vacuum frying conditions to produce fried potato chips. *Drying Technology*, 25:2027–2034, 2007.

G. Suaria and S. Aliani. Floating debris in the mediterranean sea. *Marine Pollution Bulletin*, 86:494–504, 2014.

K.S. Tan, S.V. Wong, R.S. Radin Umar, A.M.S. Hamouda, and N.K. Gupta. Impact behavior modeling of motorcycle front wheel-tire assembly. *International Journal of Automotive Technology*, 10(3):329–339, 2009.

W.M. To, T.W. Yu, T.M. Lai, and S.P. Li. Characterization of commercial clothes dryers based on energy-efficiency analysis. *International Journal of Clothing Science and Technology*, 19(5):270–290, 2007.

K.J. Wohlgemuth, A. Jesko, V. Frost, M.J. Conner, and J.A. Mota. Characteristics of slow and fast performers on a firefighter air consumption test. *Applied Ergonomics*, 118, 2024. doi: 10.1016/j.apergo.2024.104262. URL https://doi.org/10.1016/j.apergo.2024.104262.

F-G. Wu, H. Lin, and M. You. The enhanced navigator for the touch screen: A comparative study on navigational techniques of web maps. *Displays*, 32:284–295, 2011.

R. Zeng, J. Lin, S. Wu, L. Chen, S. Chen, H. Gao, Y. Zheng, and H. Ma. A randomized controlled trial: Preoperative home-based combined tai chi and strength training (tcst) to improve balance and aerobic capacity in patients with total hip arthroplasty (tha). *Archives of Gerontology and Geriatrics*, 60:265–271, 2015.

I. Zouid, R. Siret, F. Jourjon, E. Mehinagic, and L. Rolle. Impact of grapes heterogeneity according to sugar level on both physical and mechanical berries properties and their anthocyanins extractability at harvest. *Journal of Texture Studies*, 44:95–103, 2013.