MA4720 – Design and Analysis of Experiments

Course Learning Objectives

Upon successful completion of this course, students will be able to:

1. Describe three fundamentals of experimental design and steps to design an experiment using the three fundamentals
2. Perform the randomization for two experimental designs: completely randomized designs with one or more treatment factors and complete block designs
3. Write out mathematical models and describe the assumptions and the meaning of each of components in a model
4. Perform the appropriate statistical analysis of completely randomized designs and complete block designs
    a. Draw appropriate plots to display the data and use it as a guidance to choose an appropriate statistical model
    b. Perform statistical inference including point estimates, confidence intervals, and hypothesis tests of contrasts
    c. Construct the analysis of variance table and perform hypothesis tests of each factor and their interactions
    d. Check the assumptions on the model and perform the appropriate transformation of data
5. Determine sample sizes and power.

Week 0 Learning Objectives

Upon successful completion of this course, students will be able to:

1. State the key points in the syllabus and outline the major assignments and expectations of the course.
2. Describe how and where the course contents are presented (Lessons, Videos, R Files, etc.) and how you communicate and interact with the instructor (Announcements Section, Discussion Board, and Canvas Message System).

3. Know your Student Success Advisors and how to get in touch with them.

Week 1 Learning Objectives

Upon successful completion of this course, students will be able to:

1. Describe common terminologies used in experimental designs and apply three principles of designing experiments including replication, blocking, and randomization.
2. Describe the five steps to design a simple experiment and use R to perform the appropriate randomization for an experiment with one source of variation.
3. Write out the mathematical model, interpret each term in the model, and calculate the least squares estimate of experiments with one source of variation.

Week 2 Learning Objectives

Upon successful completion of this course, students will be able to:

1. Perform the test of equality of treatment effects and interpret the results.
2. Calculate the power for a given sample size or determine the sample size for a given power for designs with one source of variation using Table A.7 or R.
3. Define different types of contrasts and calculate their least squares estimates.

Week 3 Learning Objectives

Upon successful completion of this course, students will be able to:

1. Perform the statistical inference, including point estimation, confidence interval, and hypothesis testing of a single contrast.
2. Choose and apply appropriate method for multiple comparisons.
3. Use appropriate plots and tests to check if the assumptions of ANOVA model are satisfied.

Week 4 Learning Objectives

Upon successful completion of this course, students will be able to:

1. Perform an appropriate data transformation to equalize variances.
2. Construct analysis of variance model with three model include cell-means model, two-way complete model, and two-way main-effects model.
3. Calculate the point estimation, the confidence interval, and hypothesis testing of a single contrast without or with methods of multiple comparisons under different models including cell-means model, two-way complete model, and two-way main-effects model.
4. Use appropriate plots and tests to check if the assumptions of models for experiments with two crossed treatment factors are satisfied.

Week 5 Learning Objectives

Upon successful completion of this course, students will be able to:

1. Construct and evaluate different analysis of variance models for experiments with three or more crossed treatment factors.
2. Calculate point estimation, confidence interval, and perform hypothesis testing of a single contrast without or with methods of multiple comparisons under different models for experiments with three or more crossed treatment factors.
3. Use appropriate plots and tests to check if the assumptions of models for experiments with three or more crossed treatment factors.
4. Obtain type I or Type III sum of squares and use them to test main effect of each factor and interaction effects between them.

Week 6 Learning Objectives

Upon successful completion of this course, students will be able to:

1. Construct orthogonal contrasts and use them to analyze single replicate experiments.
2. Use half-normal probability plot of effect estimates to determine the appropriate model for the analysis of single replicate experiments.
3. Design experiments with randomized complete block deigns and general complete block designs.

4. Use appropriate models for the analysis of randomized complete block deigns and general complete block designs.
5. Calculate sample size and power and check model assumptions for randomized complete block deigns and general complete block designs

Week 7 Learning Objectives

Upon successful completion of this course, students will be able to:

1. Apply the analysis of covariance model to analyze experiments with nuisance factors that cannot be measured or can only be measured until the experiment is conducted.
2. Calculate the least squares estimates and the confidence interval with appropriate methods of multiple comparisons.
3. Calculate sample size and power and check model assumptions

MA4720: Design and Analysis of Experiments

Week 1 At-a-Glance

Title: Introduction of Experimental Design and Statistics

Overview

This week, we will introduce fundamental concepts and three principles used in experimental designs and some basics of statistics. We will describe simple steps to design a simple experiment and demonstrate how to use R to perform the randomization. We will introduce and learn the mathematical model and the least squares estimate of experiments with one source of variation.

Learning Objectives

When you complete this module, you should be able to do the following:

1. Describe common terminologies used in experimental designs and apply three principles of designing experiments including replication, blocking, and randomization.
2. Describe the five steps to design a simple experiment and use R to perform the appropriate randomization for an experiment with one source of variation.
3. Write out the mathematical model, interpret each term in the model, and calculate the least squares estimate of experiments with one source of variation.

Reading (Optional)

1. Sections 1.1 and 1.2 in Chapter 1, Section 10.2 in Chapter 10.
2. Sections 2.1, 2.2, 2.3, and 2.5.2 in Chapter 2.
3. Section 3.1 to Section 3.4 in Chapter 3.

Instruction Content

See other files for details.

Quizzes

Two quizzes. See other files for two quizzes.

Homework

One homework. See other file for details.

MA4720: Design and Analysis of Experiments

Week 1 Instruction Contents

Lesson 1.1 Basic Concepts and Principles of Experimental Designs

Required Reading: Sections 1.1 and 1.2 in Chapter 1, Section 10.2 in Chapter 10.

One of most common questions asked by my collaborators is "How many observations (or samples) to I need to take". This is a kind of questions that cannot be easily answered in a couple of sentences. As a first step towards obtaining an answer, other questions such as "What is the main purpose of running this experiment?" and "What do I hope to be able to show". We can only answer such questions only after we understand the design and analysis of experiments which are the focus of this course.

Let us start with a simple example: an experiment about the battery. This is experiment is referred as the battery experiment throughout this course. One of authors of the textbook was interested in finding which type of non-rechargeable battery was most economical. He considered the batteries with different duties and brands in his experiment. Two different duties including alkaline and heavy duty and two brands including a name brand and a store brand were considered. There are many questions you need to consider when you plan such an experiment including:

1. Sample size: how many batteries are needed?
2. What is objective of this study?
3. How to measure if a battery is more economical?
4. What statistical models should be used in the analysis and what conclusions we can obtain from the statistical analysis?

As I have mentioned, we will defer to answer such questions later. At this stage, we will introduce some basic concepts and principles used in designing such experiments and use the battery experiment as an example.

Basic Concepts and Principles

Response/Dependent variables are variables of interest. They should be tightly related to the objective of the study.

- For the aforementioned battery experiment, the life or the ratio of life and cost (life per unit cost) can be used as a response variable.

- The responsible variable is generally a continuous variable. A **continuous** variable is one that can take any one of an uncountable number of values in an interval. The examples of a continuous variable include the square footage of the house, the life of bulb, etc.
- While a **discrete** variable can assume only accountable number of values. The examples of a discrete variable include the gender, the zip code of a house, the brand of battery.
- In real world, the continuous variables may appear discrete but conceptually take any value in an interval. For example, AGE, generally, we calculate it according to your birthdate not in which minute and hour you were born. Other examples include height, blood pressure.

Factors (Treatment factors): variables of interest that affect the response variable.

- For the aforementioned battery experiment, there are two factors, one treatment factor is the duty of battery and the other is the brand of battery.

Nuisance factors: variables that affect the response variable but not our primary of interest.

- We can classify a nuisance factor as a blocking factor, a covariate, or a noise factor. However, such classification is not always obvious. The decision will often be governed by the goal of the experiment.

Noise factors: Nuisance factors are classified as noise factors if the objective of the experiment is to find settings of the treatment factors whose response is least affected by varying the levels of the nuisance factors.

- Settings of noise factors can usually be controlled during an experiment but are uncontrollable outside the laboratory.
- For the aforementioned battery experiment, the room temperature and humanity can be considered as the noise factors since these factors will affect the lifetime of battery but their effect should be small if the fluctuations in the temperature and humanity are small.

Covariates: Covariates are nuisance factors that cannot be controlled but can be measured prior to, or during, the experiment.

- Sometimes covariates are of interest in their own right, but when they are included in the model as nuisance variables, their effects are used to adjust the responses so that treatments can be compared as though all experimental units were identical.

- For the aforementioned battery experiment, the date of manufacture of the purchased battery can considered as a covariate. This variable could not be controlled in the experiment because the batteries used in the experiment were purchased at different times and in different locations in order to give a wide representation of dates of manufacture. If the dates have been marked on the packets, we will know such information and can include them in the analysis.

Blocking factors: factors are used in a block design

- A block is appropriate when the goal of the experiment is to compare the effects of different treatments averaged over a range of different conditions. The experimental units are grouped into sets in such a way that two experimental units in the same set are similar and can be measured under similar experimental conditions, but two experimental units in different sets are likely to give rise to quite different measurements even when assigned to the same treatment. The sets of similar experimental units are called blocks, and the conditions that vary from block to block form the levels of the blocking factor.
- The intent of blocking is to prevent large differences in the experimental units from masking differences between treatment effects, while at the same time allowing the treatments to be examined under different experimental conditions.
- For the aforementioned battery experiment, if different types of flash light are used then the type of flash light can be considered as a blocking factor.
- As another example, suppose that a researcher would like to know which one of two types of berry trees can give more yield and two fields are available to plant those trees, then the filed should be considered as a blocking factor.

Levels: the specific values of each factor, which can either be discrete or continuous.

- In the battery experiment, there are two levels for duties of battery: alkaline and heavy duty
- Room temperature is continuous so has infinite number of levels
- Treatment and nuisance factors (including noise factors, covariates, and blocking factors) are called **exploratory or independent variables** and both of them.

Experimental Unit: smallest unit that we can apply a treatment combination.

- For the aforementioned battery experiment, the experimental unit can be a single battery, two batteries, or three batteries, depending on how many batteries used in a flashlight.

Observational Unit - is the unit upon which we make the measurement.

- Observational unit may be different from the experimental unit
- For the battery experiment, each battery was tested and its life time was recorded. So the observational unit is same as the experimental unit.

Observational Studies: record data for a period of time. They measure value of response variable without attempting to influence any of the response or explanatory variables.

- Generally, we **cannot draw conclusions about cause and effect**.
- For example, if the alkaline battery is always tested with one flashlight why the heavy duty battery is always tested with another flashlight, the difference of battery lifetime may be just due to the flashlight used not the type of battery.

Designed Experiments: designed experiment involves the application of treatments to experimental units to assess the effects of response variable.

- Need to control all factors (if possible) that affect the response variable
- So we **can draw conclusions about cause and effect** based on a well designed experiment.
- This is the major goal of this course.

There are **three Fundamentals of Experimental Design:**

- **Replication**
- **Blocking**
- **Randomization**

Replication

- **Replication** is repetition of experimental conditions so that the effects of interest can be estimated with greater precision and the associated variability can be estimated.

- In other words, **replication** means that we apply at least one of the treatment combinations to **more than one experimental unit**, which allows us to estimate the experimental error and to perform the formal statistical analysis.
- There is a difference between the replication and repeated measurements. **For example**,
    - Blood pressures from 10 individuals are "replication". The variation reflects the variation of individuals' blood pressure
    - 10 blood pressure measurements from a single person is "repeated measurement". The variation reflects the variation of measuring process.

Blocking:

- Blocking is a technique that can often be used to control and adjust for some of the variation in experimental units.
- The experimental units are signed into groups called blocks in such a way that the experimental units in each block are intended to be relatively similar, so that treatments assigned to experimental units in the same block can be compared under relatively similar experimental conditions.
- If blocking is done well, then comparisons of two or more treatments are made more precisely in the experiment than similar comparisons from an unblocked design.

Example of Blocking Design: in an experiment to compare the effects of two skin ointments for rash,

- **Unblocked design:** two treatments are compared on arms of two different people.
- **Blocking design:** two treatments are compared on two arms of the same person
- More precise for the blocking design since the other conditions of two arms from the same person are similar, so the variation of difference of response variable of two skin ointment is reduced.

Randomization:

- **Randomization**: A random assignment of experimental units to treatments prior to the start of the experiment.
- The use of randomization to:

  - Avoid experimenter bias
  - Avoid systematic bias
  - To ensure valid statistical analysis. Most statistical analysis is based on “simple random sampling”.

- For the aforementioned battery experiment, suppose we have total 20 batteries (5 batteries for each of treatment combinations), the randomization means that these 20 batteries in a random order, not in the order of 5 batteries of name brand and heavy duty, then 4 batteries of name brand and alkaline, etc.
- There are many ways to achieve the randomization. Later we will show how to use a simple function in R to achieve the randomization.

Randomization: lack of randomization can lead to:

- **Experimenter bias,**
  - For example, A doctor can assign his/her preferred drug to the patients most likely to respond well.
- **Systematic bias,**
  - For the aforementioned battery experiment, a systematic bias may be introduced when you test the lifetime by the following order: alkaline/name brand, alkaline/store brand, heavy duty/name brand, heavy duty/store brand. The differences observed may not be due to the battery type and/or battery brand but due to some other factors (e.g., temperature if some tests are performed at morning while some tests are performed at noon).

Lesson 1.2 Steps for Designing an Experiment

Required Reading: Sections 2.1, 2.2, 2.3, and 2.5.2 in Chapter 2.

Step for designing an experiment

- **Step 1**: Identify problem (define objective, determine response variables).
- **Step 2**: Identify all sources of variation, including
    - treatment factors and their levels,
    - experimental units and observation units
    - blocking factors, noise factors, and covariates
- **Step 3**: Determine the sample size and choose a rule to assign experimental units to treatments
- **Step 4**: Conduct experiment
- **Step 5**: Collect data, specify model, analyze data, draw conclusions.

Example 1.1: Steps for Designing an Experiment – Battery Experiment

Step 1 for the battery experiment: Identify problem (define objective, determine response variable).

- **Objective:** find which kind of battery is more economical
- **Response Variable:** the life per unit cost
- **Response Variable can also be:** the life if we do not care the cost or the cost is not important

Step 2 for the battery experiment: Identify all sources of variation, including

- **Treatment factors and their levels**
    - Duties of battery with two levels: alkaline and heavy duty
    - Brands of battery with two levels: name brand and store brand
    - Total of 4 combination from two treatments factors
- **Experimental unit:** each individual battery
- **Blocking factors, noise factors, and covariates**
    - Date of manufacturer of battery and flashlight bulbs are not available

- Running condition: continuous running (used) or on and off running
- Room temperature: at 68F, small fluctuations are not considered as important

Step 3 for the battery experiment: Determine the sample size and choose a rule to assign experimental units to treatments. Suppose that we will test a total of 20 batteries do there are 5 batteries for each treatment combination.

- The following codes are used to represent the treatment combinations

Level	Treatment Combinations	Label
1	Alkaline, name brand (11)	1, 2, 3, 4, 5
2	Alkaline, store brand (12)	6, 7, 8, 9, 10
3	Heavy duty, name brand (21)	11, 12, 13, 14, 15
4	Heavy duty, store brand (22)	16, 17, 18, 19, 20

- 20 batteries: 5 batteries for each treatment combination

Step 3 for the battery experiment: Assign experimental units to treatments

- **Randomization:** no blocking factors, a **completely randomized design** can be used
- Since when each battery was purchased, its type and brand have already been determined. Therefore the randomization here means that the batteries are tested with a random order.
  - There are 20! different orders and all orders are equally, so each order has the same probability of 1/20! being used.
- Another **randomization**: Run 5 times and each time run 4 treatment combinations with random order. [**Question:** how many different orders?]

Randomization: how to? [Please refer to the video about the R program]

- Suppose the batteries are labeled from as 1 to 20 as in the table
- **Complete randomization**:
  - R code: sample(1:20)

- Another **randomization**: run 5 times and each time run 4 treatment combinations with random order
- R code: See R program for more details in canvas in canvas

Steps 4 for the battery experiment: Conduct experiment (Skipped)

Step 5 for the battery experiment: Collect data, specify model, analyze data, draw conclusions.

(**please refer to the corresponding video**)

Table: Data for the battery experiment

Battery Type	Life (Min)	Unit Cost ($)	Life per unit cost	Time order
1	602	0.985	611	1
2	863	0.935	923	2
1	529	0.985	537	3
4	235	0.495	476	4
1	534	0.985	542	5
1	585	0.985	593	6
2	743	0.935	794	7
3	232	0.520	445	8
4	282	0.495	569	9
2	773	0.935	827	10
2	840	0.935	898	11
3	255	0.520	490	12
4	238	0.495	480	13
3	200	0.520	84	14
4	228	0.495	460	15
3	215	0.520	413	16

Step 5: Specify Model – the linear model is used throughout of this course. We will defer details to the next section but give you some brief introduction.

- **Model 1 (Two-way main-effects model):**

  Lifetime/cost = constant + effect of Duty + effect of Brand + error

- **Model 2 (Two-way complete model):**

  Lifetime/cost = constant + effect of Duty + effect of Brand + Interaction + error

- **Model 3 (Cell-means model, which is equivalent to Model 2):**

  Lifetime/cost = constant + effect of treatment combination + error

Example 1.1 Step 3 - Randomization. How to perform the randomization experiment? Assume that the number of treatments is $v = 4$ and the number of replicates for each treatment is $r = 5$.

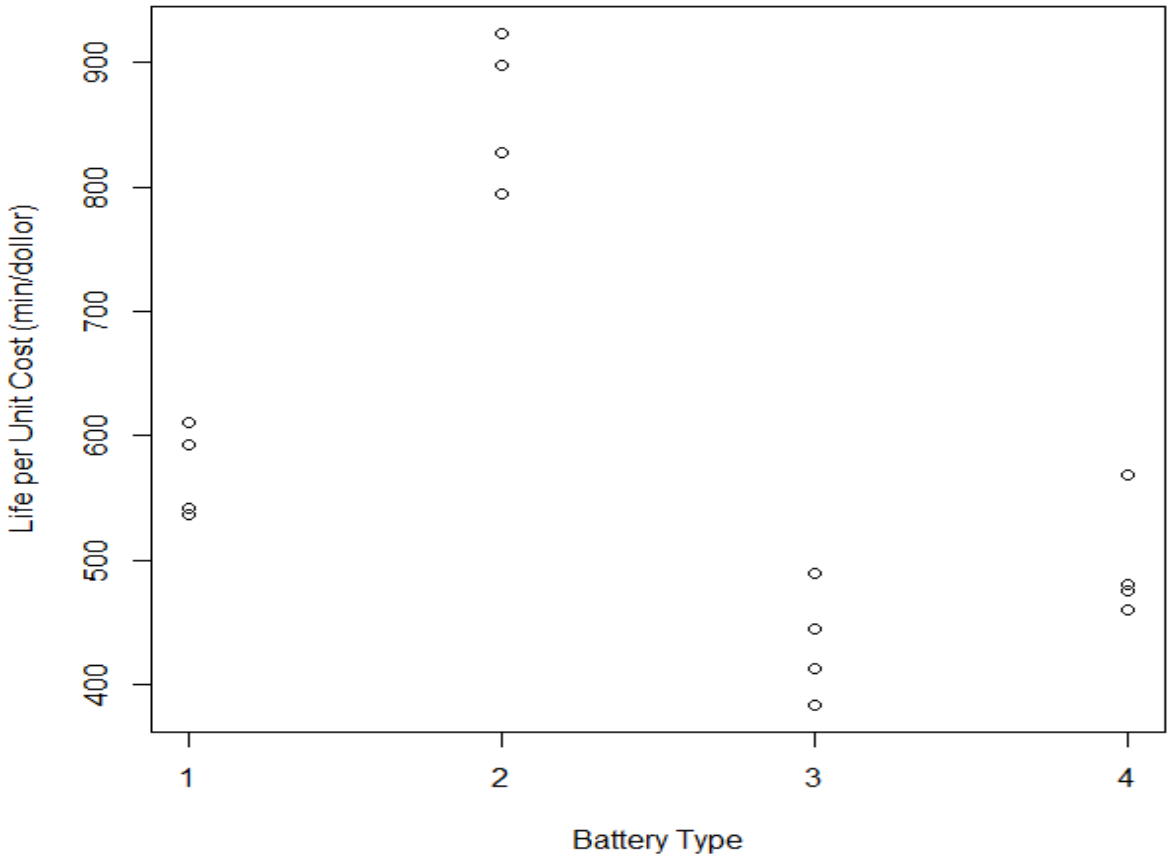
Solution: There are different ways to perform the randomization. One way is to label the batteries from 1 to 20 as in the table and use R function sample. The other way is to run 5 times and each time run 4 treatment combinations with random order. The second way guarantees that

- **Complete randomization**:
  - R code: sample(1:20)

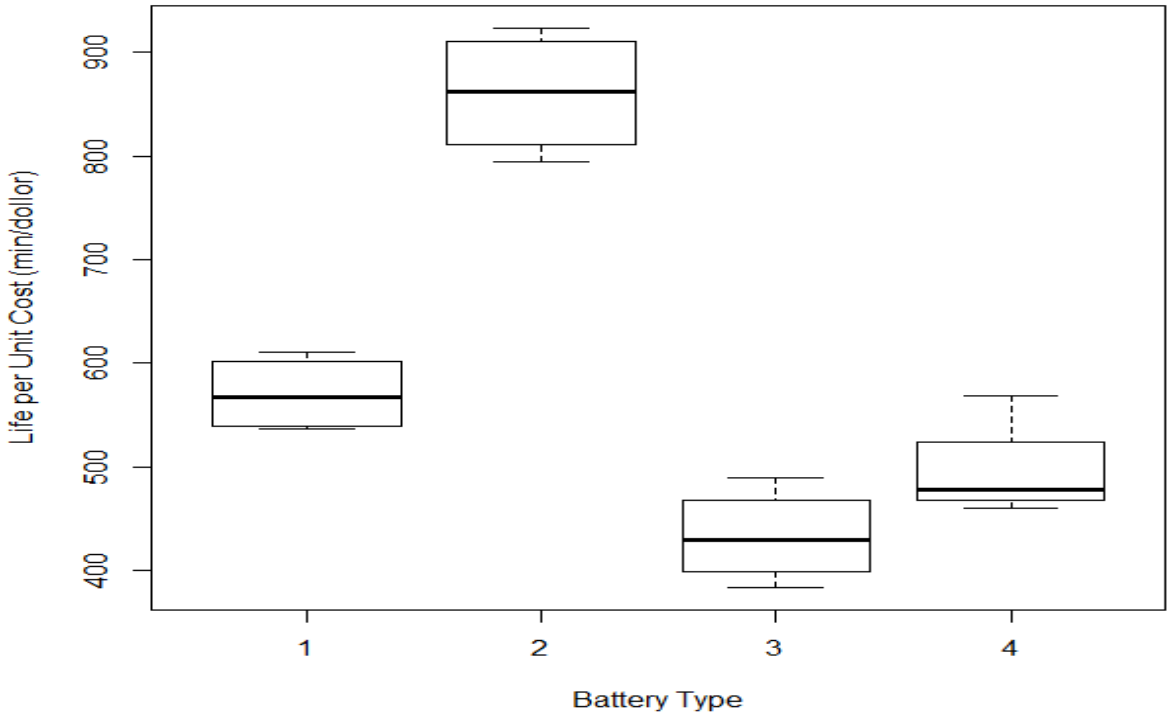
Another **randomization**:

- R code: See R program for more details in canvas in canvas

Example 1.1 Step 5 – Scatter plot and boxplot.



Battery Example: Exploratory Analysis (boxplot)



Video Placeholder: week-1-video-01-example-1-1

Exercise: Humidifier Experiment

Two brands of ultrasonic humidifiers are compared with respect to the rate at which they output moisture. The company chose 8 humidifiers from brand A and 8 from brand B. They measured the maximum outputs (in fluid ounces) per hour in a chamber controlled a temperature of 70F and a relative humidity of 30%.

Consider:

- What is the problem of interest (objective of study)?
- What is the response variable?
- What is the treatment factor? What is the number of levels of the factor? What is the number of treatment combinations?
- What are blocking or noise factors?
- What is the experimental unit? What is the observational unit?
- How would you apply replication, randomization, and blocking?

Lesson 1.3 Introduction of Some Basic Statistics and Designs with One Treatment Factors

Required Reading: Sections 3.1 and 3.2 in Chapter 3

The first part of this section is to help you refresh your knowledge on the statistical inference. Such knowledge will be used in this course.

Statistics deals largely with principles and procedures for collecting, describing, and drawing conclusions from data. There are two type statistics – descriptive statistics and inferential statistics.

Descriptive Statistics – organizing and summarizing data through numerical summaries, tables, and graphics. For example, we can

- Use a figure or a table to show the daily snowfall at Houghton
- Draw the distribution of SAT scores of freshmen at MTU

The descriptive statistics should help you to generate some hypothesizes. Although we can draw some conclusions from the description statistics but these conclusions are not based on the rigorous statistical analysis thus can be susceptible.

An example of descriptive statistics: S&P 500 Index in the stock market within a period of time.

Inferential Statistics – using methods that take results from a ***sample***, extends them to the population, and measures of ***reliability*** of the results. Here are some examples of inferential statistics:

- Predicting political winner from sample polling
- Finding the important factors that influence the elastic strength of yarns
- In the battery example, finding which type of battery is more economical

Here are some basic concepts in statistics.

- **Data -** A set of **data is** a collection of observed values representing one or more characteristics of some objects or units.
- **Population -** A **population** is a data set representing the entire entity of interest. Note that the definition of population depends on the problem of interest. For example, if we are interested in the SAT score distribution of all MTU undergraduate students, then the data set containing the SAT scores of MTU undergraduate students is the population. If we are interested in the SAT score distribution of all undergraduate students in a four-year college in Michigan, then the data set of SAT scores of MTU undergraduate students becomes a sample.
- **Sample -** A **sample** is a data set consisting of a portion of a population. Such sample must be obtained in a way to represent the population. A most commonly used sampling scheme is the random sampling meaning that each individual in the population has the same chance to be selected.
- **Census** – the collection of the data from everyone in a population. Generally, it is cost prohibitive to collect a census data.

Here we use the data from the battery experiment as an illustration:

Example 1.2 Data - Battery Experiment

Objective: find which kind of battery is more economical

Two treatment factors:

Level	Treatment Combinations
1	Alkaline, name brand (11)
2	Alkaline, store brand (12)
3	Heavy duty, name brand (21)
4	Heavy duty, store brand (22)

Data from 16 samples are collected.

Table: Data for the battery experiment

Battery Type	Life (Min)	Unit Cost ($)	Life per unit cost	Time order
1	602	0.985	611	1
2	863	0.935	923	2
1	529	0.985	537	3
4	235	0.495	476	4
1	534	0.985	542	5
1	585	0.985	593	6
2	743	0.935	794	7
3	232	0.520	445	8
4	282	0.495	569	9
2	773	0.935	827	10
2	840	0.935	898	11
3	255	0.520	490	12

4	238	0.495	480	13
3	200	0.520	84	14
4	228	0.495	460	15
3	215	0.520	413	16

This is typical data table, will be the typical format from the file on computers. First row, the name of each column, then followed by each observational unit (battery). Here we use all numerical values (can be others). We have several variables: battery type, life, unit cost, life per unit cost, and the time order.

Questions asked from this example data (practice questions):

1. Is this data a census data or sample data?
    - Sample data
2. What is the population here?
    - All four types of batteries
3. What is the relationship between the population and the sample?
    - Sample is the subset of the population
4. Why we cannot collect census data in this situation and in most of other situations?
    - It is too expensive and/or time consuming, may not be possible

Population and Sample:

Population – its characteristic is described by the **parameter**.

- Usually denoted by Greek letters, such as α, β, μ, etc.
- It is a fixed constant generally unknown.
- It is primary of interest of statistical inference and is estimated from sample.

Sample – its characteristic is described by the **statistics**.

- Usually denoted by alphabetic letters, such as $x,\ y,\ z,\ p$, etc.
- It is unknown before you collect the data but can be calculated from the data and is considered as known once the data is collected.
- It is used to estimate the population parameter.

A statistic and a parameter are very similar. They are both descriptions of a characteristic. For example, "more than 30% of MTU students take MA3710 – Engineering Statistics".

The difference between them is:

Parameter – a numerical value describing some characteristic of a ***population***.

Statistic – a numerical value describing some characteristic of a ***sample***.

Exercise 1.2: Population and Sample – Practice Questions

1. Determine whether the underlined value is a parameter or a statistic: In a survey conducted in the town of Atherton, **25%** of adult respondents reported that they had been involved in at least one car accident in the past ten years. Here is 25% a Statistic or Parameter?
   Answer: Statistic. Since we are interested in the percentage of adults in a town who have been involved in at least one car accident in the past ten years. But 25% is from only adult respondents from a survey.
2. Determine whether the underlined value is a parameter or a statistic. **27.2%** of the mayors of cities in a certain state are from minority groups. Here is 27.2% a Statistic or Parameter?
   Answer: Parameter. Here the population is all the mayors of cities in a certain state. And 27.2% is from all mayors of hat state.
3. From the battery experiment, the average life per unit cost from 8 alkaline batteries is 715.63 based on the data. Is this number a parameter or a statistic?
   Answer: Statistic. Since this number is from a sample.

Statistical Inference

Statistical inference is based on statistics from sample to draw conclusions for population parameters.

A Statistic:

- It unknown before the data is collected, so it is a **random variable**. Thus, we need the probability theory to describe the property of the statistic.
- It is known after the data is collected: it is a fixed value once the data is given.

Example 1.3: Statistics - Battery Experiment

What is the average life per unit cost from 8 alkaline batteries?

- It is unknown before we conduct the experiment and collect the data.
- It can take many possible values, 500, 612, or 437.5, etc.
- How can we describe it? They should be described by the **random variables.**

Once the data is collected:

- 611, 923, 537, 542, 593, 794, 827, 898
- The average is 715.63.
- The average can be another value when another data is collected.

The statistic is described by the random variables. A random variable is a real valued function and is characterized by its cumulative distribution. In general, we use a capital Roman letter to describe a random variable, such as Y, Z, etc.

If Y is a random variable, then the ***cumulative distribution function*** (cdf), denoted by $F(y)$, is given by

$$F(y) = \Pr(Y \leq y) \text{ for all real numbers } y(-\infty < y < \infty).$$

If we know $F(y)$, then we should know everything of Y. Note that the cdf $F(y)$, of a random variable Y, must satisfy several conditions. We will not list these conditions but keep in mind that $0 \leq F(y) \leq 1$ since it is a probability.

In this course, we will focus on the continuous random variable.

Y is **continuous random variable** is it can take on any value in an interval. If Y is a continuous random variable, then $\boldsymbol{F(y)}$ **is a continuous function** and can be written as

$$F(y) = \Pr(Y \le y) = \int_{-\infty}^{y} f(t)dt \text{ for all real numbers } y(-\infty < y < \infty).$$

The function $f(y)$ $(-\infty < y < \infty))$ is called the **probability density function (pdf)** of Y. Also we have the following relationship between $F(y)$ and $f(y)$:

$$F(y) = \Pr(Y \le y) = \int_{-\infty}^{y} f(t)dt \text{ and } \frac{d}{dy}F(y) = f(y)$$

We know one of them we can get another, but in general we often use the probability density function to describe a continuous random variable.

While Y is a **discrete random variable** is it can take on only a countable number of values. If Y is a discreet random variable, then $\boldsymbol{F(y)}$ **is a step function.**

Example of a continuous random variable:

The random variable is the life of the light bulb in years, denoted by Y. Then it is a continuous variable since it can take any value between 0 to $+\infty$.

One commonly used cumulative distribution function to describe the life time of an object is called exponential distribution, For example, we can use $F(y) = \Pr(Y \le y) = 1 - \exp(-\frac{y}{3})$ for $y \ge 0$ and $F(y) = 0$ for $y < 0$. Then the corresponding probability density function of Y is

$$f(y) = \frac{d}{dy}F(y) = \frac{d}{dy}\left[1 - \exp\left(-\frac{y}{3}\right)\right] = \frac{1}{3}\exp(-\frac{y}{3}) \text{ for } y \ge 0$$

$$\text{and } f(y) = 0 \text{ for } y < 0.$$

There are several quantities associated with the random variable Y with the pdf $f(y)$.

The **population mean** of Y, μ, is defined as

$$\mu = E(Y) = \int_{-\infty}^{\infty} yf(y)dy.$$

The **Population variance** of Y, denoted by σ^2, is defined as

$$\sigma^2 = Var(Y) = E[(Y-\mu)^2] = E(Y^2) - \mu^2.$$

Note that the population mean and population variance are parameters of a population, They are generally unknown and must be estimated from the sample.

Exponential Distribution

Example 1.4: Life of a light bulb (in years)

We can model the life of a light bulb by an **exponential distribution**:

$$f(y) = \lambda \exp(-\lambda), y > 0, \lambda > 0$$

Where λ is a positive constant:

- λ is called the **rate** or **inverse scale** parameter.
- $1/\lambda$ is called the **scale** parameter.

The exponential distribution is used to model lifetimes and times between incidents.

Example 1.4: Life of a light bulb (continued)

The probability density function (pdf) of Y is:

$$f(y) = \lambda \exp(-\lambda y), y > 0, \lambda > 0 \text{ (implies that } f(y) = 0 \text{ for } y \leq 0)$$

The cumulative distribution function of (cdf) is (only need to consider the value in the support set of Y):

$$F(y) = \Pr(Y \leq y) = \int_0^y \lambda \exp(-\lambda t)\, dt = 1.0 - \exp(-\lambda y), y > 0$$

The population mean is:

$$\mu = E(Y) = \int_{-\infty}^{\infty} y * f(y) dy = \int_0^{\infty} y * \lambda \exp(-\lambda y)\, dy = \frac{1}{\lambda}.$$

Similarly, you can find

$$E(Y^2) = \int_{-\infty}^{\infty} y^2 * f(y)dy = \int_0^{\infty} y^2 * \lambda \exp(-\lambda y)\, dy = \frac{2}{\lambda^2}.$$

So the population variance is:

$$\sigma^2 = E(Y^2) - [E(Y)]^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

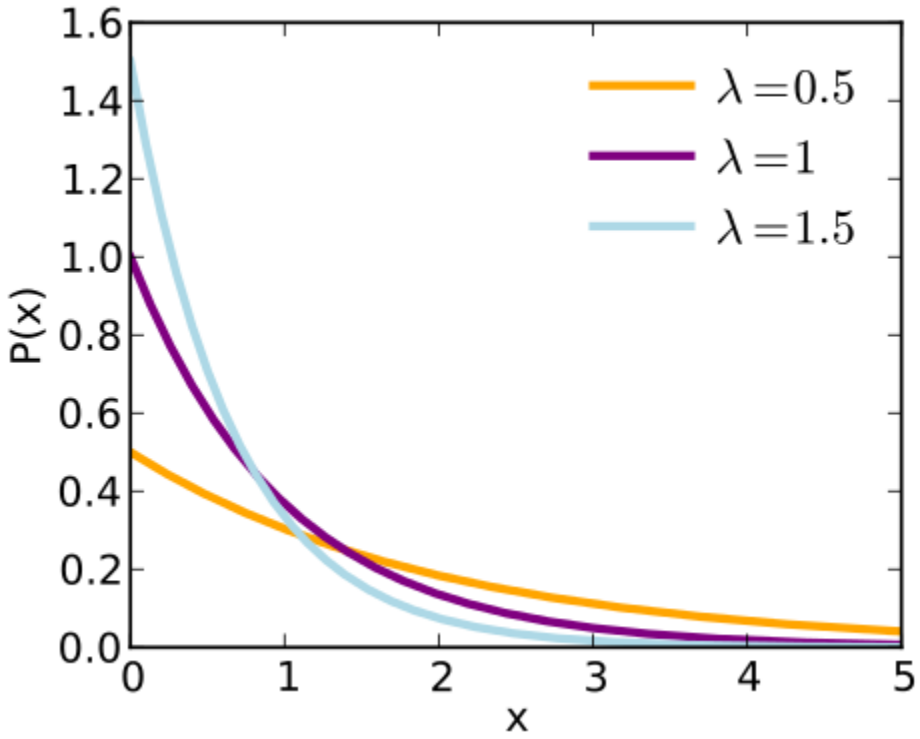
For $\lambda = 0.05$, the cdf is:

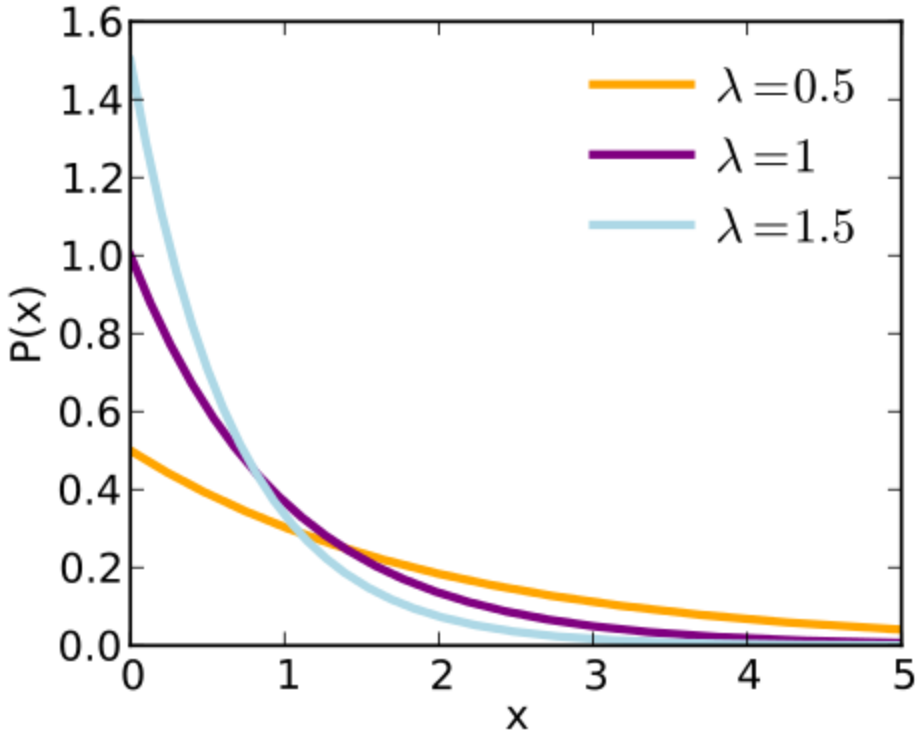
$$F(y) = 1.0 - \exp(-0.05y)$$

The probability that a bulb lasts no more than ten years is:

$$\Pr(Y \leq 10) = 1 - \exp(-0.05 * 10) = 0.393$$

Exponential Distribution: pdf and cdf





Top is the pdf and bottom is the cdf. Both of them are Copied from Wiki: https://en.wikipedia.org/wiki/Exponential_distribution

The probability distribution functions used in our class include the normal distribution, the t-distribution, and the F-distribution.

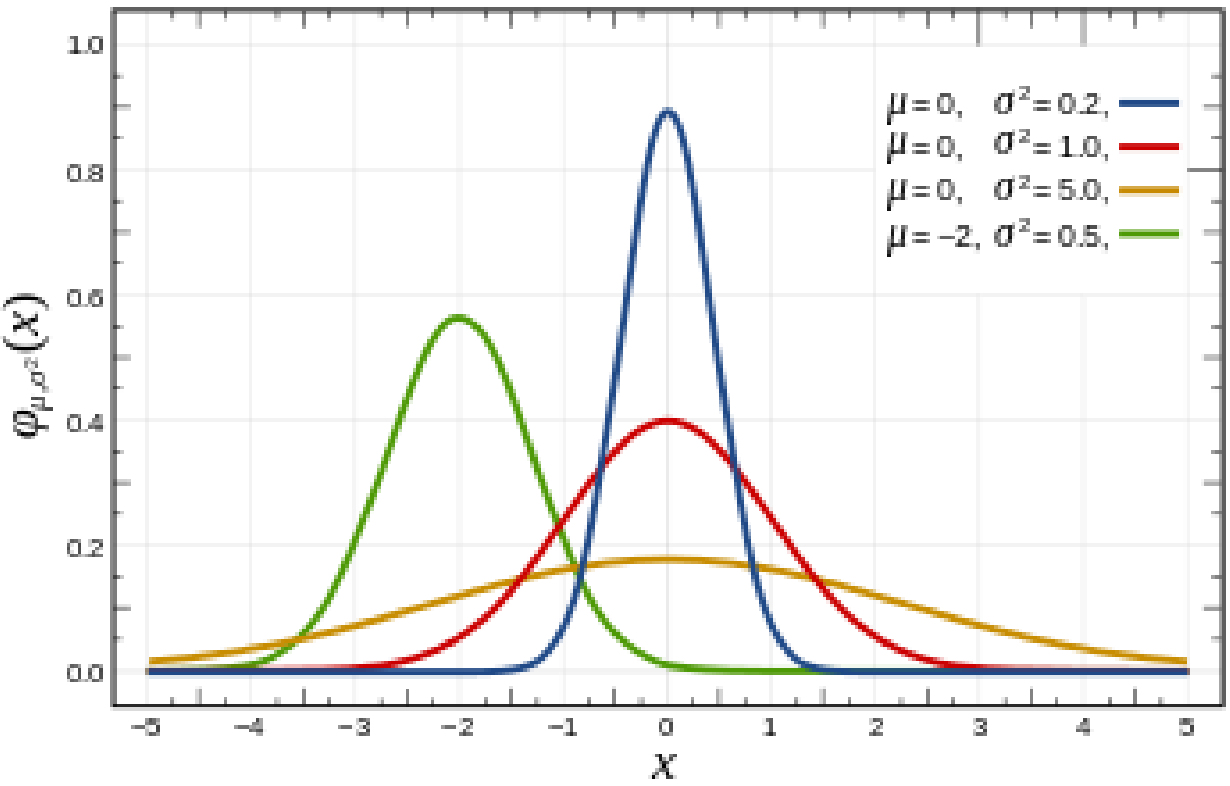
Normal Distribution

The pdf of normal random variable with the location parameter μ and the scale parameter σ^2 is:

$$f(y) = \frac{1}{\sqrt{2*\pi*\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \text{ for } -\infty < y < \infty$$

In our class, we use the simple notation: $Y \sim N(\mu, \sigma^2)$ to denote that the random Y has the normal distribution with the location parameter μ and the scale parameter σ^2.

Normal Distribution – Normal Curve



The properties of the normal distribution include:

(1) The population mean $E(Y)$ and the population variance $Var(Y)$ are the location parameter μ and the scale parameter σ^2, respectively. In other words, If $Y \sim N(\mu, \sigma^2)$, then we have

(2) $\mu = E(Y)$ and $\sigma^2 = Var(Y)$

(3) If $Y \sim N(\mu, \sigma^2)$ and $\mu = 0$ and $\sigma^2 = \sigma = 1$, then Y has a standard normal distribution.

(4) If $Y \sim N(\mu, \sigma^2)$, then for any real numbers a and b,, the random variable $aY + b$ has a normal distribution and $aY + b \sim N(a\mu + b, a^2\sigma^2)$

(5) If $Y \sim N(\mu, \sigma^2)$, then $\frac{Y-\mu}{\sigma} \sim N(0,1)$. This can be directly derived from (4): we have $\frac{Y-\mu}{\sigma} = \frac{1}{\sigma}Y - \frac{\mu}{\sigma} \sim N\left(\frac{1}{\sigma} * \mu + \left(-\frac{\mu}{\sigma}\right), \left(\frac{1}{\sigma}\right)^2 * \sigma^2\right) = N(0,1)$ (here $a = \frac{1}{\sigma}$ and $b = -\frac{\mu}{\sigma}$).

(6) If $Y_1, \cdots, Y_n$ are independent, and $Y_i \sim N(\mu_i, \sigma_i^2)(i = 1, \cdots, n)$), then for any constants $a_1, a_2, \cdots, a_n$ and $b_1, b_2, \cdots, b_n$, we have $\sum_{i=1}^{n}(a_iY_i + b_i) = a_1Y_1 + b_1 + a_2Y_2 + b_2 + \cdots + a_nY_n + b_n \sim N(\sum_{i=1}^{n}(a_i\mu_i + b_i), \sum_{i=1}^{n} a_i^2\sigma_i^2)$

(7) This is a special case of (6). If $Y_1, \cdots, Y_n$ are **identically**, **independently distributed** (i.i.d.) random variables from $N(\mu, \sigma^2)$, then the **sample mean** is

$$\bar{Y} = \frac{Y_1 + \cdots + Y_n}{n} \text{ and } \bar{Y} = \frac{Y_1 + \cdots + Y_n}{n} \sim N(\mu, \frac{\sigma^2}{n})$$

(8) We do not have a closed formula to calculate the cumulative distribution function of normal distribution but we can do so using R very easily.

Note that these properties are very important for our course since we mainly deal with normally distributed data.

Both Chi-square distribution, t-distribution, and F-distribution are frequen used in the statistical inference of normally distributed data. Their probability density functions are quite complicated. However, we can use R to calculate their pdfs and cdfs easily. And for this course, that is what we need to know.

Two theorems related to the distribution of the sample mean:

Theorem: If the sample of size n is a random sample from the normal distribution $N(\mu, \sigma^2)$, then sampling distribution of sample mean is also normal, $N(\mu, \frac{\sigma^2}{n})$.

Theorem (Central Limit Theorem) - If random samples of size n are taken from any distribution with mean μ and variance σ^2, sample mean will have a distribution **approximately** normal with mean μ and variance $\frac{\sigma^2}{n}$ for the large n.

The Central Limit Theorem tells us that even the random samples are not from a normal distribution, we can still treat the sample mean as normally distributed as long as the sample size n is sufficiently large. How large the sample size (n) should be for the Central Limit Theorem:

- $n > 30$ in general.
- Much small n is fine if the data is approximately normal.

Statistical inference is based on statistics from sample to draw conclusions for population parameters. Three types of statistical inference:

- **Point Estimator:** use a statistic as the population parameter (no reliability)
- **Interval Estimator:** usually done in the form of an interval, and reliability is expressed as probability of true value is covered.
- **Hypothesis Testing:** hypothesize that parameters have some specific values or relationships, and make decisions about that. Reliability is the probability that the decision is incorrect.

Example 1.5: Statistical Inference – Battery Experiment

For the battery experiment, suppose that we are interested in the average life per unit cost of alkaline batteries. For this experiment, we only collected the life per unit cost from 8 alkaline batteries.

(1) We use a Greek letter, μ, denote the parameter of interest, the average life per unit cost of alkaline batteries. Note that μ is the parameter from the population. It is a non-negative constant but unknown. We need to conduct the experiment to make the inference on this constant.

(2) The data (the life per unit cost from 8 alkaline batteries) are unknown before we conduct the experiment and collect the data so are described by the random variable. Here, we use the upper-case of Roman letters, $Y_1, \cdots, Y_8$ to describe the **random sample** and can comfortably assume that $Y_1, \cdots, Y_n$ are **identically**, **independently distributed** (i.i.d.) random variables from $N(\mu, \sigma^2)$.

(3) To estimate the average life per unit cost of alkaline batteries, we can use the average life per unit cost of 8 alkaline batteries:

$$\bar{Y} = \frac{1}{8}(Y_1 + \cdots + Y_8) = \frac{1}{8}\sum_{i=1}^{8} Y_i$$

Here, $\bar{Y}$ is called the sample mean. It can be used as a **point estimator** of μ and we denote by $\hat{\mu} = \bar{Y} = \frac{1}{8}\sum_{i=1}^{8} Y_i$. Note that the **point estimator** of μ, $\bar{Y}$ is a random variable

and it has a normal distribution: $N(\mu, \frac{\sigma^2}{8})$. This point estimator and its distribution are used to construct the point estimator and the statistical hypothesis testing for μ.

In general, the sample mean from the sample of size n is defined as:

$$\bar{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n) = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

The distribution of the sample mean can be obtained by the following theorems.

Theorem: If the sample of size n is a random sample from the normal distribution $N(\mu, \sigma^2)$ (meaning that $Y_1, \cdots, Y_n$ are **identically**, **independently distributed** (i.i.d.) random variables from $N(\mu, \sigma^2)$), then distribution of sample mean ($\bar{Y} = \frac{1}{n}(Y_1 + \cdots + Y_n) = \frac{1}{n}\sum_{i=1}^{n} Y_i$) is also normal, $N(\mu, \frac{\sigma^2}{n})$.

Theorem (Central Limit Theorem) - If random samples of size n are taken from any distribution with mean μ and variance σ^2(meaning that $Y_1, \cdots, Y_n$ are **identically**, **independently distributed** (i.i.d.) random variables from a distribution with mean μ and variance σ^2), then sample mean will have a distribution approximately normal with mean μ and variance $\frac{\sigma^2}{n}$.

How large n should be for the Central Limit Theorem:

- $n > 30$ in general.
- Much small n is fine if data is approximately normal.

(4) We can also define the sample variance S^2, as

$$S^2 = \frac{(Y_1 - \bar{Y})^2 + \cdots + (Y_n - \bar{Y})^2}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

Note that S^2 is also a random variable and can be used as a point estimator of the population variance σ^2.

- Once the data is collected, we use the lower-case Roman letter to describe the data. For the battery experiment, we can $y_1, \cdots, y_8$ to describe eight numbers form the battery

experiments: 611, 923, 537, 542, 593, 794, 827, 898. The sample mean (the average of these 8 numbers) is 715.63 and is used as a **point estimate** of the parameter of interest, μ. We still denote it by $\hat{\mu} = \bar{y} = \frac{1}{8}\sum_{i=1}^{8} y_i$.

(5) For the battery experiment, the 95% confidence interval estimates of μ (from your other statistical course) is:

$$\bar{y} \pm \sqrt{s^2/8} * t_{0.025,7} = 715.625 \pm \sqrt{26121.12/8} * 2.365 = (580.49, 850.76)$$

Note that the confidence interval (for this course) can be calculated with the following formula:

$$\text{point estimate} \pm \text{estimated standard error} * \text{critical value}$$

(6) For the hypothesis testing, $H_0: \mu \leq 700$ **versus** $H_1: \mu > 700$**,** we fail to reject the null hypothesis since $\frac{\bar{y}-700}{\sqrt{s^2/8}} = \frac{715.625-700}{\sqrt{26121.12/8}} = 0.2736 > t_{0.05,7} = 1.8932$

Here, we will skip the details of interval estimates and the hypothesis testing which should be covered by other courses already. However, I would emphasize the interpretation of results obtained from such analysis. Here we will use some practice problems to enhance your understanding of such interpretation.

Video Placeholder: week-1-video-02-example-1-5.mp4

Exercise 1.3: Statistical Inference – Practice Problems

Interpretation of CI (Another Example) The blood pressure of 100 patients from a hospital is collected. Based on the data, the 95% confidence interval of mean blood pressure of patients in a hospital is from 95.1 to 143.6, which of the following statements is correct:

1. The true mean blood pressure of patients in that hospital is between 95.1 and 143.6.
2. About 95% of patients in that hospital will have the blood pressure in the interval from 95.1 to 143.6.

3. The probability of true mean blood pressure of patients in that hospital in the interval from 95.1 to 143.6 is 0.95.
4. None of the above is correct.

Answer: 4. It should be stated like this way: **We are 95% confident that the true mean blood pressure of the patients from that hospital is between 95.1 and 143.6.**

Hypothesis testing:

If the null hypothesis is really false, which of these statements characterizes a situation when the value of the test statistic falls in the rejection region?

1. A type I error has been committed.
2. A type II error has been committed.
3. Insufficient information has been given to make a decision.
4. The decision is correct.
5. None of the above is correct.

Answer: 4. When the value of the test statistic falls in the rejection region, we reject the null hypothesis. Since the null hypothesis is really false, so we make a correct decision.

Hypothesis testing:

If the null hypothesis is really false, which of these statements characterizes a situation where the value of the test statistic does not fall in the rejection region?

1. A type I error has been committed.
2. A type II error has been committed.
3. Insufficient information has been given to make a decision.
4. The decision is correct.
5. None of the above is correct.

Answer: 2. When the value of the test statistic does not fall in the rejection region, we fail to reject the null hypothesis. Since the null hypothesis is really false, so we make a Type II error.

Hypothesis testing:

If the value of the test statistic does not fall in the rejection region, the decision is:

1. Reject the null hypothesis.
2. Reject the alternative hypothesis.
3. Fail to reject the null hypothesis.
4. Fail to reject the alternative hypothesis.
5. There is insufficient information to make a decision.

Answer: 3. When the value of the test statistic does not fall in the rejection region, we fail to reject the null hypothesis.

Lesson 1.4 Models for Designs with One Treatment Factors

Required Reading: Sections 3.2 and 3.4 in Chapter 2

Introduction of designs with one source of variation

Here, we consider experiments with one source of variation. For such experiments,

(1) there is only **one treatment factor.**
(2) There are **at least two levels** for the treatment factor.
(3) Blocking factors, noise factors, covariates are not considered.
(4) Observational unit is same as experimental unit.
(5) There are principles: **replication** and **randomization**. Here the replication means that there are at least two observations for each treatment level. We only consider the **completely randomized design.** For **completely randomized design**, the experimental units are assigned to the treatments completely at random, subject to the number of observations to be taken on each treatment.
(6) Experiments with more than one treatment factors can be analyzed as experiments with one treatment factor. In such situation, all treatment combination will be considered as the levels from a single treatment factor. For example, there are two treatment factors and four treatment combinations in the battery experiment, we can consider these four treatment combinations as the four levels from a single treatment factor.

The **completely randomized design:**

- With the help of *R*, it is easy to design a completely randomized experiment.
- Levels of factor are $1, \cdots, v$ and $v \geq 2$.
- The numbers of observations are $r_1, r_2, \cdots, r_v$.
- The total number of observations is $n = r_1 + r_2 + \cdots + r_v$.
- There are many different methods to achieve *Randomization*. Here the randomization means randomly assign n experimental units to v treatment levels and randomly assign the order of v treatment levels

- Suppose that experimental units are labeled from 1 to n. Here we use the following way for randomization.
    - Step 1: Enter one column r_1 of 1's, r_2 of 2's, ⋯, and finally r_v of v's, giving a total of $n = r_1 + r_2 + \cdots + r_v$ entries. These represent the treatment labels.
    - Step 2: Random permute the treatment labels.
    - Step 3: Assign experimental unit t to the treatment whose label is in row t.

Example for randomization: Heart-Lung Pump Experiment (**please refer to the corresponding video**)

Example 1.6: Basic Concepts - Heat-Lung Pump Experiment

It was run by Richard Davis at The Ohio State University in 1987 to determine the effect of the number of revolutions per minute (rpm) of the rotary pump head of an Olson heart–lung pump on the fluid flow rate. The rpm was set directly on the tachometer of the pump console and PVC tubing of size 3⁄8" by 3⁄32" was used.

- The flow rate was measured in liters per minute.
- Five equally spaced levels of the treatment factor "rpm" were selected, namely, 50, 75, 100, 125, and 150 rpm, and these were coded as 1, 2, 3, 4, 5, respectively.
- The experimental design was a completely randomized design with

$$r_1 = r_3 = r_5 = 5, r_2 = 3, r_4 = 2 \text{ so } n = 20$$

- For this example, the randomization is to randomly assign 20 experimental units to five treatment levels and randomly assign the order of 5 treatment levels. This can be done in a single step.
- Randomization procedure:
    - Enter one column 5 of 1's, 5 of 2's, ⋯, and finally 2 of 5's, giving a total of 20 entries. These represent the treatment labels.

- Step 2: Random permute the treatment labels.
- Step 3: Assign experimental unit t to the treatment whose label is in row t.

- See R codes to find these method and **Book Example 3.2.1**.

Video Placeholder: week-video-03-example-1-6.mp4

Statistical Model: Notations

- $Y_{it}(t = 1, \cdots, r_i; i = 1, \cdots, v)$, random variables
  - the response obtained on the tth observation of the ith treatment.
  - These random variables are used to describe the uncertainty during the sampling process: the values are unknown before we take the samples
  - They are used to derive the formulas for the statistical analysis of experiments with one treatment factor.
- $y_{it}(t = 1, \cdots, r_i; i = 1, \cdots, v)$, numbers
  - the value of the tth observation of the ith treatment.
  - These represent the observed values of the response variable once the data is collected.
- $\mu_i = \mu + \tau_i$ $(i = 1, \cdots, v)$: population parameters
  - the "true response" of the ith treatment
  - the response obtained without error
- In our course, we write μ_i as $\mu_i = \mu + \tau_i$ $(i = 1, \cdots, v)$.
  - The parameter μ is a constant
  - The parameter τ_i represents the positive or negative deviation of the response from this constant when the ith treatment is observed.
  - The deviation τ_i, is called the "effect" on the response of the ith treatment.

  - We can see that the examination of differences between the parameters μ_i is equivalent to examination of differences between the parameters τ_i.

- $\epsilon_{it}(t = 1, \cdots, r_i; i = 1, \cdots, v)$, error variables
  - the error for the tth observation of the ith treatment.

Statistical Model:

The modeled used in this chapter is:

$$Y_{it} = \mu + \tau_i + \epsilon_{it}, t = 1, \cdots, r_i; i = 1, \cdots, v$$

This model is also called as the one-way analysis of variance model.

Assumptions:

- μ and τ_i $(i = 1, \cdots, v)$ are unknown population parameters. $\mu + \tau_i$ is the "true response" of the ith treatment
- $\epsilon_{it}(t = 1, \cdots, r_i; i = 1, \cdots, v)$ are independent
- $\epsilon_{it} \sim N(0, \sigma^2)$: normal distribution with mean 0 and variance σ^2
- Therefore, Y_{it} are independent and $Y_{it} \sim N(\mu + \tau_i, \sigma^2)$ (**Why**? Use property of the normal distribution). It is worth noting that here Y_{it} may have different mean according to the treatment level, but they have the same variance of σ^2. This is called the **equal variance assumption**.

What questions should be asked for the following model:

$$Y_{it} = \mu + \tau_i + \epsilon_{it}, t = 1, \cdots, r_i; i = 1, \cdots, v$$

The following questions should be considered:

- How to estimate parameters, including μ, τ_i, and σ^2
- How to estimate the difference of the responses from two levels (one of main interests of experiments with one treatment factor)
- How to estimate functions of parameters such as $2\mu + \tau_1 + \tau_2$

- How to determine if there is a difference across all treatment responses
- How to determine if the responses from two levels are different
- How to check if the assumptions are satisfied

Estimable Functions of Parameters

- A function of the parameters of any linear model is said to be **estimable** if and only if it can be written as the expected value of a linear combination of the response variables.
- Only estimable functions of the parameters have unique linear unbiased estimates.
- In our model, $Y_{it} = \mu + \tau_i + \epsilon_{it}, t = 1, \cdots, r_i; i = 1, \cdots, v$, only the following functions are estimable:

$$b_1 * (\mu + \tau_1) + \cdots + b_v * (\mu + \tau_v) = \sum_{i=1}^{v} b_i * (\mu + \tau_i)$$

where b_i $(i = 1, \cdots, v)$ are constants.

This is because:

$$E\left(\sum_{i=1}^{v}\sum_{t=1}^{r_i} a_{it} * Y_{it}\right) = \sum_{i=1}^{v}\sum_{t=1}^{r_i} E(a_{it} * Y_{it})$$

$$= \sum_{i=1}^{v}\sum_{t=1}^{r_i} a_{it} * E(Y_{it}) = \sum_{i=1}^{v}\sum_{t=1}^{r_i} a_{it} * (\mu + \tau_i)$$

$$= \sum_{i=1}^{v}(\mu + \tau_i)\sum_{t=1}^{r_i} a_{it} = \sum_{i=1}^{v} b_i * (\mu + \tau_i)$$

$$\text{where } b_i = \sum_{t=1}^{r_i} a_{it}\ (i = 1, \cdots, v).$$

Examples: Estimable functions

For the model:

$$Y_{it} = \mu + \tau_i + \epsilon_{it}, t = 1, \cdots, r_i; i = 1, \cdots, v$$

Determine if the following functions are estimable: (**please refer to the corresponding video**)

- $\mu + \tau_1$ (estimable)

- τ_1 (not estimable))
- $\tau_1 - \tau_2$ (estimable)
- $\tau_1 + \tau_2 - 2 * \tau_3$ (estimable)
- $\tau_1 + \tau_2$ (not estimable)

Any function $\sum_{i=1}^{v} c_i * \tau_i$ for which $\sum_{i=1}^{v} c_i = 0$ is called a **contrast**, and all contrasts are estimable.

- Note that if $\sum_{i=1}^{v} c_i = 0$, then

$$\sum_{i=1}^{v} c_i * \tau_i = \sum_{i=1}^{v} c_i * (\mu + \tau_i)$$

So a **contrast** is estimable.

Examples: Contrast

For the model:

$$Y_{it} = \mu + \tau_i + \epsilon_{it}, t = 1, \cdots, r_i; i = 1, \cdots, v$$

Determine if the following functions are contrast:

- $\mu + \tau_1$ (not a contrast, estimable)
- τ_1 (not a contrast, not estimable)
- $2 * \mu + \tau_1 + \tau_2$ (not a contrast, estimable)
- $\tau_1 + \tau_2 - 2 * \tau_3$ (contrast, so estimable too)
- $\tau_1 - 0.5 * \tau_2 - 0.5 * \tau_4$ (contrast, so estimable too)

Estimation of Parameters: Notations

- Sample mean of the ith treatment is defined as:

$$\bar{Y}_{i\cdot} = \frac{1}{r_i}(Y_{i1} + \cdots + Y_{ir_i}) \text{ and } \bar{y}_{i\cdot} = \frac{1}{r_i}(y_{i1} + \cdots + y_{ir_i})$$

- The "dot" notation means "add over all values of the subscript replaced with a dot," and the "bar" means "divide by the number of terms that have been added up."
- So we have

$$Y_{i\cdot} = (Y_{i1} + \cdots + Y_{ir_i}) \text{ and } y_{i\cdot} = (y_{i1} + \cdots + y_{ir_i})$$

$$n = r_{i\cdot} = r_1 + \cdots + r_v \text{ and } \bar{y}_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{v}\sum_{t=1}^{r_i} y_{it} = \frac{1}{n} y_{\cdot\cdot}$$

- Note that if the summation applies to a subscript on two variables, the dot notation cannot be used. For example,

$$\sum_{i=1}^{v} r_i * \tau_i \neq r_{i\cdot} * \tau_{i\cdot} \text{ because } r_{i\cdot} * \tau_{i\cdot} = (\sum_{i=1}^{v} r_i) * (\sum_{i=1}^{v} \tau_i)$$

- Note that when notation involves both a sum and a square, such as $y_{\cdot\cdot}^{\ 2} = (y_{\cdot\cdot})^2$, the sum is taken first and then the sum is squared.
- For a parameter, for example, we use $\hat{\mu}$ to represent the estimate (a number, calculated from samples, y_{it}) and the estimator (a random variable, calculated from Y_{it}).

Least Squares Estimates (please refer to the corresponding video)

- Find parameters that minimize the sum of squared errors:

$$\sum_{i=1}^{v}\sum_{t=1}^{r_i} \epsilon_{it}^2 = \sum_{i=1}^{v}\sum_{t=1}^{r_i} (y_{it} - \mu - \tau_i)^2$$

- The sum of squared errors is differentiated with respect to each of the parameters $\mu, \tau_i, \ldots, \tau_v$ in turn. Then each of the $v + 1$ resulting derivatives is set equal to zero, yielding a set of $v + 1$ equations.
- For example, if we have

$$\frac{d}{d\mu}\sum_{i=1}^{v}\sum_{t=1}^{r_i} (y_{it} - \mu - \tau_i)^2 = \sum_{i=1}^{v}\sum_{t=1}^{r_i} \frac{d}{d\mu}(y_{it} - \mu - \tau_i)^2$$

$$= \sum_{i=1}^{v}\sum_{t=1}^{r_i} -2 * (y_{it} - \mu - \tau_i)$$

$$= -2 * \left(\sum_{i=1}^{v}\sum_{t=1}^{r_i} y_{it} - \sum_{i=1}^{v}\sum_{t=1}^{r_i} \mu - \sum_{i=1}^{v}\sum_{t=1}^{r_i} \tau_i\right)$$

$$= -2 * \left(y_{..} - \mu * \sum_{i=1}^{v} \sum_{t=1}^{r_i} 1 - \sum_{i=1}^{v} \tau_i * \sum_{t=1}^{r_i} 1 \right)$$

$$= -2 * (y_{..} - n * \hat{\mu} - \sum_{i=1}^{v} r_i * \hat{\tau}_i)$$

- Ste the above formula to 0, we have

$$y_{..} - n * \hat{\mu} - \sum_{i=1}^{v} r_i * \hat{\tau}_i = 0$$

- In summary, the estimates must satisfy $v + 1$ **normal equations**:

$$y_{..} - n * \hat{\mu} - \sum_{i=1}^{v} r_i * \hat{\tau}_i = 0$$

$$y_{i.} - r_i * \hat{\mu} - r_i * \hat{\tau}_i = 0 \ (i = 1, \cdots, v)$$

- By solving these $v + 1$ **normal equations,** we obtain the least squares estimates:

$$\hat{\mu} + \hat{\tau}_i = \frac{y_{i.}}{r_i} = \bar{y}_{i.} (i = 1, \cdots, v)$$

- Note that we can estimate $\mu + \tau_i$ but not μ and τ_i seperately.
- But we always have:

$$\hat{\mu} + \hat{\tau}_i = \frac{y_{i.}}{r_i} = \bar{y}_{i.} (i = 1, \cdots, v)$$

Another parameter is σ^2, how can we estimate it? Let us start from the simple case. From other statistical courses, we know how to estimate σ^2 when we have a random sample from $N(\mu, \sigma^2)$. That is, we use the sample variance as a point estimate of the population variance, which is the ratio of squared difference of sample values and sample mean and the degrees of freedom.

Here, for each $i = 1, \cdots, v$, we have a random sample from the normal population. For example, we have $y_{i1}, \cdots, y_{ir_i}$ as a random sample from $N(\mu + \tau_i, \sigma^2)$. So we can use sample variance of data from each treatment level to estimate σ^2. In addition, since responses from all treatment

levels have the same variance σ^2, it is better for us to use all data to estimate σ^2. Therefore, an estimate of σ^2 is given by:

$$\hat{\sigma}^2 = \frac{1}{n-v}\sum_{i=1}^{v}\sum_{t=1}^{r_i}(y_{it} - \bar{y}_{i\cdot})^2$$

- $n-v$: **error degrees of freedom**
- $\sum_{i=1}^{v}\sum_{t=1}^{r_i}(y_{it} - \bar{y}_{i\cdot})^2$: it has many names, such as the **sum of squares of error (ssE)** or the **error sum squares** or **within group sum of squares (ss*W*).** The use of term of **within group sum of squares** is due to that the squared difference of sample values and sample mean from each group (treatment level) is calculated.
- Note that the ssE for the *i*th treatment level, can be written as product of the degrees of freedom and the sample variance of data from the *i*th treatment level. In other words, for $i = 1, \cdots, v$, we have

$$\sum_{t=1}^{r_i}(y_{it} - \bar{y}_{i\cdot})^2 = (r_i - 1)\frac{1}{r_i - 1}\sum_{t=1}^{r_i}(y_{it} - \bar{y}_{i\cdot})^2 = (r_i - 1) * s_i^2$$

- $\frac{1}{n-v}\sum_{i=1}^{v}\sum_{t=1}^{r_i}(y_{it} - \bar{y}_{i\cdot})^2$ *:* the **mean square for error (msE)** or **error mean square**

Video Placeholder: week-1-video-04-statistical-model.mp4

Example 1.7: Heart-Lung Pump Experiment: The data file is **heart.lung.pump.txt.** It was run by Richard Davis at The Ohio State University in 1987 to determine the effect of the number of revolutions per minute (rpm) of the rotary pump head of an Olson heart–lung pump on the fluid flow rate. The rpm was set directly on the tachometer of the pump console and PVC tubing of size 3⁄8” by 3⁄32” was used.

- The flow rate was measured in liters per minute.
- Five equally spaced levels of the treatment factor “rpm” were selected, namely, 50, 75, 100, 125, and 150 rpm, and these were coded as 1, 2, 3, 4, 5, respectively.
- The experimental design was a completely randomized design with

$$r_1 = r_3 = r_5 = 5, r_2 = 3, \text{ and } r_4 = 2$$

Questions: for the one-way analysis of variance model of this experiment,

(1) For $\mu + \tau_2$, state the its meaning, write out its least squares estimator and corresponding distribution, and find the least squares estimate.

(2) For $0.5 * \tau_1 + 0.5 * \tau_2 - \tau_3$, state the its meaning, write out its least squares estimator and corresponding distribution, and find the least squares estimate. and its distribution.

(3) Calculate SSE and find the least squares estimate of σ^2.

Solution:

Step 1:

We can use a calculator or R to calculate the sum, the sample mean, the squares for error, and the sample variance for each treatment level and obtain the following table (see R program for such calculation) (**please refer to the corresponding video**):

Level	**Replicates**	**Sum**	**Sample Mean**	**ssE**	**Sample Variance**
1	$r_1 = 5$	$y_{1\cdot} = 5.676$	$\bar{y}_{1\cdot} = 1.1352$	0.0008208	$s_1^2 = 0.0002052$
2	$r_2 = 3$	$y_{2\cdot} = 5.166$	$\bar{y}_{2\cdot} = 1.7220$	0.0019440	$s_2^2 = 0.0009720$
3	$r_3 = 5$	$y_{3\cdot} = 11.634$	$\bar{y}_{3\cdot} = 2.3268$	0.0011808	$s_3^2 = 0.0002952$
4	$r_4 = 2$	$y_{4\cdot} = 5.850$	$\bar{y}_{4\cdot} = 2.9250$	0.0064980	$s_4^2 = 0.00064980$
5	$r_5 = 5$	$y_{5\cdot} = 17.646$	$\bar{y}_{5\cdot} = 3.5292$	0.0103968	$s_5^2 = 0.0025992$
Overall	$n = 20$	$y_{\cdot\cdot} = 45.972$	$\bar{y}_{\cdot\cdot} = 2.2986$	N/A	N/A

Step 2: If necessary, rewrite the parameter or the estimable function of the parameters as $\sum_{i=1}^{v} b_i * (\mu + \tau_i)$ and use $\sum_{i=1}^{v} b_i * \bar{y}_{i\cdot}$ as its least squares estimate. The key here is to find the coefficients $b_i, i = 1, \cdots, v$.

(1) $\mu + \tau_2 = \sum_{i=1}^{v} b_i * (\mu + \tau_i)$, where only $b_2 = 1$ and all other b_i is 0

$\mu + \tau_2$ is the true (or true average) flow rate of the treatment level 2.

The least squares estimator of $\mu + \tau_2$ is

$$\bar{Y}_{2\cdot}$$

The distribution of $\bar{Y}_{2\cdot}$ is

$$N\left(\mu + \tau_2, \frac{\sigma^2}{r_2}\right) = N(\mu + \tau_2, \frac{\sigma^2}{3})$$

The least squares estimates of $\mu + \tau_2$ is

$$\hat{\mu} + \hat{\tau}_2 = \bar{y}_{2\cdot} = 1.7220$$

(2) Since $0.5 * \tau_1 + 0.5 * \tau_2 - \tau_3$ is equivalent to $0.5 * (\mu + \tau_1) + 0.5 * (\mu + \tau_2) - (\mu + \tau_3)$, $0.5 * \tau_1 + 0.5 * \tau_2 - \tau_3$ represents the difference of the average of flow rates from the treatment levels 1 and 2 and the flow rate from the treatment level 3.

So its least square estimator is:

$$0.5 * \bar{Y}_{1\cdot} + 0.5 * \bar{Y}_{2\cdot} - \bar{Y}_{3\cdot}$$

The distribution of $0.5 * \bar{Y}_{1\cdot} + 0.5 * \bar{Y}_{2\cdot} - \bar{Y}_{3\cdot}$ is:

$$N(0.5 * (\mu + \tau_1) + 0.5 * (\mu + \tau_2) - (\mu + \tau_3), 0.5^2 * \frac{\sigma^2}{r_1} + 0.5^2 * \frac{\sigma^2}{r_2} + 1^2 * \frac{\sigma^2}{r_5})$$

This distribution is:

$$N\left(0.5 * \tau_1 + 0.5 * \tau_2 - \tau_3, \frac{1}{4} * \frac{\sigma^2}{5} + \frac{1}{4} * \frac{\sigma^2}{3} + \frac{\sigma^2}{5}\right) = N\left(0.5 * \tau_1 + 0.5 * \tau_2 - \tau_3, \frac{\sigma^2}{3}\right)$$

So its least square estimate is:

$$0.5 * \bar{y}_{1\cdot} + 0.5 * \bar{y}_{2\cdot} - \bar{y}_{3\cdot} = 0.5 * 1.1352 + 0.5 * 1.7220 - 2.3268 = -0.8982$$

(3) From the above table, we first can calculate the sum of squares for error:

$$SSE = 0.0008208 + 0.0019440 + 0.0011808 + 0.0064980 + 0.0103968 = 0.02084$$

So the least squares estimate of σ^2 is the mean square for error:

$$\hat{\sigma}^2 = MSE = \frac{SSE}{n-v} = \frac{0.02084}{20-5} = 0.00139$$

Video Placeholder: week-1-video-05-example-1-7.mp4

Week 1 Practice Problem Discussion

Consider the following practice problem and participate in the Week 1 Practice Problem Discussion for assistance, or to provide your own suggestions. The data file is **battery.txt**. Still the one-way analysis of variance model will be used. Use R program to create the following table and

(1) For $\tau_2 - \tau_4$, state the its meaning, write out its least squares estimator and corresponding distribution, and find the least squares estimate.
(2) Calculate ssE and find the least squares estimate σ^2.

Level	**Replicates**	**Sum**	**Sample Mean**	**ssE**	**Sample Variance**
1					
2					
3					
4					
Overall					

MA4720: Design and Analysis of Experiments

Week 2 At-a-Glance

Title: Designs with One Source of Variation

Overview

This week, we will continue our lecture on designs with one treatment factor. We will illustrate how to test equality of treatment effects. We will describe how to use Table A.7 and R program to calculate the power and sample size for designs with one source of variation. We will introduce different types of contrasts and their least squares estimates.

Objectives

When you complete this module, you should be able to do the following:

1. Perform the test of equality of treatment effects and interpret the results.
2. Calculate the power for a given sample size or determine the sample size for a given power for designs with one source of variation using Table A.7 or R.
3. Define different types of contrasts and calculate their least squares estimates.

Reading (Optional)

1. Sections 3.5 and 3.6 in Chapter 3.
2. Sections 4.1 and 4.2 in Chapter 4.

Instruction Content

See other files for details.

Quiz

Two quizzes. See other files for details.

Homework

One homework. See other file for details.

MA4720: Design and Analysis of Experiments

Week 2 Instruction Contents

Lesson 2.1 One-Way Analysis of Variance

Required Reading: Sections 3.5 in Chapter 3.

For designs with one source of variance, we use the following one-way analysis of variance model:

$$Y_{it} = \mu + \tau_i + \epsilon_{it}, t = 1, \cdots, r_i; i = 1, \cdots, v$$

With the following assumptions:

- μ and τ_i $(i = 1, \cdots, v)$ are unknown population parameters
- $\epsilon_{it} (t = 1, \cdots, r_i; i = 1, \cdots, v)$ are independent
- $\epsilon_{it} \sim N(0, \sigma^2)$: normal distribution with mean 0 and variance σ^2

In an experiment involving v treatments, an obvious question is whether or not the treatments differ at all in terms of their effects on the response variable. Thus one may wish to test the following null hypothesis:

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_v$$

The alternative hypothesis is:

$$H_1 \text{ or } H_A: \text{at least two of } \tau_i \text{ differ}$$

In other words, the null hypothesis is:

H_0: different levels of treatment factor have the same effect on the response variable

Versus

H_A: at least two levels of treatment factor have the different effect on the response variable

At first glance, the null hypothesis appears to involve nonestimable parameters. However, we can easily rewrite it in terms of $v - 1$ contrasts (thus estimable). For example, we can write H_0 as:

$$H_0: \tau_1 - \tau_2 = 0 \text{ and } \tau_2 - \tau_3 = 0 \text{ and} \cdots \text{and } \tau_{v-1} - \tau_v = 0$$

Or we can use the contrasts $\tau_i - \bar{\tau}.$ where $\bar{\tau}. = \frac{1}{v}(\tau_1 + \tau_2 + \cdots + \tau_v)$ (the mean):

$$H_0 : \tau_1 - \bar{\tau}. = 0 \text{ and } \tau_2 - \bar{\tau}. = 0 \text{ and} \cdots \text{and } \tau_v - \bar{\tau}. = 0$$

Note here we use v contrasts but actually we only need $v - 1$ of them. For example, if we have $\tau_i - \bar{\tau}. = 0$ $(i = 1, \cdots, v - 1)$, then we have

$$\tau_v - \bar{\tau}. = \tau_v - \bar{\tau}. + \tau_1 - \bar{\tau}. + \cdots + \tau_{v-1} - \bar{\tau}.$$

$$= \tau_1 + \cdots + \tau_v - v * \bar{\tau}_\cdot = \tau_1 + \cdots + \tau_v - v * \frac{1}{v}(\tau_1 + \tau_2 + \cdots + \tau_v) = 0$$

Any way that we rewrite H_0 in terms of estimable functions of the parameters, it will always depend on $v - 1$ distinct contrasts. Here the number $v - 1$ is called the **treatment degrees of freedom**.

The basic idea behind an analysis of variance test is that the sum of squares for error measures how well the model fits the data. Consequently, a way of testing H_0 is to compare the sum of squares for error under the original one-way analysis of variance model, known as the ***full model***, with that obtained from the modified model, which assumes that the null hypothesis is true. This modified model is called the ***reduced model***.

Under H_0, the τ_i's are equal to single parameter τ, and the reduced model becomes:

$$Y_{it} = \mu + \tau + +\epsilon_{it}, t = 1, \cdots, r_i; i = 1, \cdots, v$$

We can easily obtain the least squares estimate of $\mu + \tau$:

$\hat{\mu} + \hat{\tau} = \bar{y}_{\cdot\cdot}$

where $\bar{y}_{\cdot\cdot} = \frac{1}{n}\sum_{i=1}^{v}\sum_{t=1}^{r_i} y_{it}$ is the overall mean of the data.

The sum of squares for error in the reduced model is

$$ssE_0 = \sum_{i=1}^{v}\sum_{t=1}^{r_i}(y_{it} - \bar{y}_{\cdot\cdot})^2$$

It is called the **total sum of squares**.

If the null hypothesis is false, and the treatment effects differ, the sum of squares for error *ssE* under the full model is considerably smaller than the sum of squares for error ssE_0 of the reduced model. On the other hand, if the null hypothesis is true, then ssE_0 and *ssE* will be very similar. The analysis of variance test is based on the difference $ssE_0 - ssE$, relative to the size of *ssE*; that is, the test is based on $\frac{ssE_0 - ssE}{ssE}$. We would want to reject the null hypothesis if this quantity is large.

We call $ssT = ssE_0 - ssE$ the *sum of squares for treatments* or the *treatment sum of squares*, since its value depends on the differences between the treatment effects. We can verify that

$$ssT = \sum_{i=1}^{v} r_i * (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{v} r_i * \bar{y}_{i\cdot}{}^2 - n * \bar{y}_{\cdot\cdot}{}^2$$

Put this into the one-way analysis of variance model, we obtain the reduced model

We already know the sum of squares for error, which is

$$ssE = \sum_{i=1}^{v} \sum_{t=1}^{r_i} (y_{it} - \bar{y}_{i\cdot})^2$$

We know that we will reject H_0 if ssT/ssE is large, but we need to study the corresponding random variables of *ssT* and *ssE* to find the rejection region. Let

$$SST = \sum_{i=1}^{v} r_i * (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2 \text{ and } SSE = \sum_{i=1}^{v} \sum_{t=1}^{r_i} (Y_{it} - \bar{Y}_{i\cdot})^2$$

Then when the null hypothesis is true *SST* and *SSE* are independent and

$$\frac{SST/(v-1)}{SSE/(n-v)} \sim F_{v-1,n-v}$$

. We will reject the null hypothesis at the significant level α if

$$\frac{ssT/(v-1)}{ssE/(n-v)} > F_{v-1,n-v,\alpha}$$

where $F_{v-1,n-v,\alpha}$ is the critical value from the *F* distribution with $v-1$ and $n-v$ degrees of freedom with α in the right-hand tail. The probability α is often called the ***significance level*** of the test and is the probability of rejecting H_0 when in fact it is true (a Type I error).

The corresponding p-value is calculated as:

$$p = \Pr\left(F_{v-1,n-v} > \frac{ssT/(v-1)}{ssE/(n-v)}\right) = \Pr\left(F_{v-1,n-v} > \frac{msT}{msE}\right)$$

where $F_{v-1,n-v}$ represents a random variable with *F* distribution with $v-1$ and $n-v$ degrees of freedom. We will reject H_0 if $p < \alpha$.

Also, we call

$v-1$: ***treatment degrees of freedom***

$n-v$: **error degrees of freedom**

$n-1$: ***total degrees of freedom***

$msT = ssT/(v-1)$: **mean square for treatment**

$msE = {ssE}/{(v-1)}$: **mean square for error**

In summary, we can obtain the following one-way analysis of variance (ANOVA) table:

Source of Variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**
Treatment	$v-1$	$ssT = \sum_{i=1}^{v} r_i * (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$ $ssT = \sum_{i=1}^{v} r_i * \bar{y}_{i\cdot}{}^2 - n * \bar{y}_{\cdot\cdot}{}^2$	$msT = \frac{SST}{v-1}$
Error	$n-v$	$ssE = \sum_{i=1}^{v}\sum_{t=1}^{r_i}(y_{it} - \bar{y}_{i\cdot})^2$	$msE = \frac{SSE}{n-v}$
Total	$n-1$	$sstot = \sum_{i=1}^{v}\sum_{t=1}^{r_i}(y_{it} - \bar{y}_{\cdot\cdot})^2$	

The above table can be easily obtained from the data using R. Also, you are required to construct this table based on the sample means and sample variances.

Before we look at an example, I would like to give an intuitive explanation of the F test statistic. This explanation can be considered as an analog of two sample t-test. For the test statistic:

$$F = \frac{msT}{msE} = \frac{{ssT}/{(v-1)}}{{ssE}/{(n-v)}} = \frac{{\sum_{i=1}^{v} r_i * (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}/{(v-1)}}{{\sum_{i=1}^{v}\sum_{t=1}^{r_i}(y_{it} - y_{i\cdot})^2}/{(n-v)}}$$

- Numerator is the sum of squared difference of each treatment response and the average of treatment response.
- Denominator is the estimate of variability of data.

- It is large if the alternative hypothesis is true.
- It is the square of t-test statistics when $v = 2$ (Please refer to the following notes).

This part can be used an additional document.

The one-way ANOVA model is:

$$Y_{it} = \mu + \tau_i + \epsilon_{it}, t = 1, \cdots, r_i; i = 1, \cdots, v$$

Here we will show that the F-test statistic in ANOVA is equivalent to the two sample t-test statistic when the number of treatment levels is 2. We introduce the following notations:

$$v = 2, n = r_1 + r_2$$

r_1, r_2: number of replicates for two treatment levels;

$\bar{y}_{1\cdot}, \bar{y}_{2\cdot}$: sample means for two treatment levels;

s_1^2, s_2^2: sample variances for two treatment levels

s_p^2: pooled sample variance from all data

Note that the overall mean, $\bar{y}_{\cdot\cdot} = (r_1 * \bar{y}_{1\cdot} + r_2 * \bar{y}_2)/(r_1 + r_2)$

For $H_0: \tau_1 = \tau_2$ versus $H_a: \tau_1 \neq \tau_2$, the F-test statistic and t-test statistic are:

$$F = \frac{r_1 * (\bar{y}_{1\cdot} - \bar{y}_{\cdot\cdot})^2 + r_2 * (\bar{y}_{2\cdot} - \bar{y}_{\cdot\cdot})^2}{((r_1 - 1) * s_1^2 + (r_2 - 1) * s_2^2)/(r_1 + r_2 - 2)}$$

$$t = \frac{\bar{y}_{1\cdot} - \bar{y}_{2\cdot}}{\sqrt{s_p^2(\frac{1}{r_1} + \frac{1}{r_2})}} = \frac{\bar{y}_{1\cdot} - \bar{y}_{2} \cdot}{\sqrt{\frac{(r_1 - 1) * s_1^2 + (r_2 - 1) * s_2^2}{(r_1 + r_2 - 2)} * \frac{r_1 + r_2}{r_1 * r_2}}}$$

So we have:

$$t^2 = \frac{(\bar{y}_{1\cdot} - \bar{y}_{2\cdot})^2 * \frac{r_1 * r_2}{r_1 + r_2}}{((r_1 - 1) * s_1^2 + (r_2 - 1) * s_2^2)/(r_1 + r_2 - 2)}$$

To show that $F = t^2$, we just need to show the denominators of F-test statistic and t-test statistic are same, that is:

$$r_1 * (\bar{y}_{1\cdot} - \bar{y}_{\cdot\cdot})^2 + r_2 * (\bar{y}_{2\cdot} - \bar{y}_{\cdot\cdot})^2 = (\bar{y}_{1\cdot} - \bar{y}_{2\cdot})^2 * \frac{r_1 * r_2}{r_1 + r_2}$$

Note that $\bar{y}_{\cdot\cdot} = (r_1 * \bar{y}_{1\cdot} + r_2 * \bar{y}_2)/(r_1 + r_2)$, so

$$r_1 * (\bar{y}_{1\cdot} - \bar{y}_{\cdot\cdot})^2 + n_2 * (\bar{y}_{2\cdot} - \bar{y}_{\cdot\cdot})^2$$

$$= r_1 * \left(\bar{y}_{1\cdot} - \frac{r_1 * \bar{y}_{1\cdot} + r_2 * \bar{y}_2}{r_1 + r_2}\right)^2 + r_2 * \left(\bar{y}_{2\cdot} - \frac{r_1 * \bar{y}_{1\cdot} + r_2 * \bar{y}_2}{r_1 + r_2}\right)^2$$

$$= r_1 * \left(\frac{r_2 * (\bar{y}_{1\cdot} - \bar{y}_{2\cdot})}{r_1 + r_2}\right)^2 + r_2 * \left(\frac{r_1 * (\bar{y}_{2\cdot} - \bar{y}_{1\cdot})}{r_1 + r_2}\right)^2$$

$$= \frac{r_1 * (r_2)^2 + r_2 * (r_1)^2}{(r_1 + r_2)^2} * (\bar{y}_{1\cdot} - \bar{y}_{2\cdot})^2 = (\bar{y}_{1\cdot} - \bar{y}_{2\cdot})^2 * \frac{r_1 * r_2}{r_1 + r_2}$$

Example 2.1: On-Way Analysis of Variance - Heart-Lung Pump Experiment

Show how to obtain the following one-way ANOVA table (please refer to the corresponding video):

Source of Variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**Statistics**	***p*-value**
Treatment	4	16.126	4.031	2902	$<2 * 10^{-16}$
Error	15	0.021	0.00139		
Total	19	16.147			

Method 1: Use R function aov (simplest way)

Method 2: Use R and corresponding R functions such as mean and var

Method 3: Use the corresponding sample means and sample variances. This is useful when you only know the summary statistics. This is can help you to understand the formulas of ssT, ssE, and sstot and their relationships.

Level	**Replicates**	**Sample Mean**	**Sample Variance**
1	$r_1 = 5$	$\bar{y}_{1\cdot} = 1.135$	$s_1^2 = 0.000205$
2	$r_2 = 3$	$\bar{y}_{2\cdot} = 1.722$	$s_2^2 = 0.000972$
3	$r_3 = 5$	$\bar{y}_{3\cdot} = 2.327$	$s_3^2 = 0.000295$
4	$r_4 = 2$	$\bar{y}_{4\cdot} = 2.925$	$s_4^2 = 0.00650$
5	$r_5 = 5$	$\bar{y}_{5\cdot} = 3.529$	$s_5^2 = 0.00260$
Overall	$n = 20$	$\bar{y}_{\cdot\cdot} = 2.299$	$s^2 = 0.8498$

Since $v = 5$, the treatment degrees of freedom is $5 - 1 = 4$.

Since $n = 20$, the total degrees of freedom is $20 - 1 = 19$

The error degrees of freedom is $n - v = (n - 1) - (v - 1) = 15$

From this table, the easiest way to calculate ssT is to calculate ssE and $sstot$ first.

First, calculate ssE:

$$ssE = \sum_{i=1}^{v} \sum_{t=1}^{r_i} (y_{it} - \bar{y}_{i\cdot})^2 = \sum_{i=1}^{v} (r_i - 1) * s_i^2$$

$$= (5 - 1) * 0.000205 + (3 - 1) * 0.000972 + (5 - 1) * 0.000295 +$$

$$(2 - 1) * 0.00650 + (5 - 1) * 0.00260 == 0.020844$$

Then calculate $sstot$:

$$sstot = \sum_{i=1}^{v} \sum_{t=1}^{r_i} (y_{it} - \bar{y}_{\cdot\cdot})^2 = (n - 1) * s^2 = 19 * 0.8498 = 16.146$$

Therefore,

$$ssT = sstot - ssE = 16.146 - 0.020844 = 16.12516$$

There are other but more tedious ways to calculate ssT. For example.

$$ssT = \sum_{i=1}^{v} r_i * \bar{y}_{i\cdot}{}^2 - n * \bar{y}_{\cdot\cdot}{}^2$$

$$= 5 * 1.135^2 + 3 * 1.722^2 + 5 * 2.327^2 + 2 * 2.925^2 + 5 * 3.529^2 - 20 * 2.299^2$$

$$= 16.08406$$

Also, we can have

$$ssT = \sum_{i=1}^{v} r_i \, (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$$

$$= 5 * (1.135 - 2.299)^2 + 3 * (1.722 - 2.299)^2 + 5 * (2.327 - 2.299)^2 +$$

$$2 * (2.925 - -2.299)^2 + 5 * (3.529 - -2.299)^2 = 16.12544$$

Note: There are some slight inconsistencies for results obtained from different methods. This is mainly due to the round up errors since we only use 3 decimal digits in the calculations. Such inconsistencies should disappear if more decimal digits are used.\.

Exercise 2.1: One-Way Analysis of ANOVA - Battery Experiment

Construct the following one-way ANOVA table:

Source of	**Degrees**	**Sum of Squares**	**Mean Square**	**Statistics**	***p*-value**

Variation	**of Freedom**				
Treatment	3	427915	142638	60.24	$1.66 * 10^{-7}$
Error	12	28412	2368		
Total	15	456327			

Method 1: Use R function aov (simplest way)

Method 2: Use R and corresponding R functions such as mean and var

Method 3: Use the corresponding sample means and sample variances from the following table.

Level	**Replicates**	**Sample Mean**	**Sample Variance**
1	$r_1 = 4$	$\bar{y}_{1\cdot} = 570.75$	$s_1^2 = 1360.250$
2	$r_2 = 4$	$\bar{y}_{2\cdot} = 860.50$	$s_2^2 = 3619.000$
3	$r_3 = 4$	$\bar{y}_{3\cdot} = 433.00$	$s_3^2 = 2064.667$
4	$r_4 = 4$	$\bar{y}_{4\cdot} = 496.25$	$s_4^2 = 2426.917$
Overall	$n = 16$	$\bar{y}_{\cdot\cdot} = 590.125$	$s^2 = 30421.85$

Solution: Please refer to R programs.

Lesson 2.2 Sample Sizes and Power

Required Reading: Sections 3.6 in Chapter 3.

Introduction

Before an experiment can be run, the experimenter must determine the number of observations that should be taken on each treatment. This can be achieved by different methods. One method, which involves specifying the desired length of confidence intervals, will be presented later. The other method, which involves specifying the power required of the analysis of variance, is the topic of this lesson.

There are several situations that requires the power and sample size calculation. For a fixed budget, a rough calculation can be made of the maximum number, n, of observations that can be afforded. After this step, the power and sample calculation may show that it is unnecessary to take as many as n observations, in which case valuable resources can be saved. Alternatively, and unfortunately the more likely, it may be found that more than n observations are needed in order to fulfill all the experimenter's requirements of the experiment. In this case, the experimenter needs to go back and review the decisions made so far in order to try to relax some of the requirements. Otherwise, an increase in budget needs to be obtained. There is little point in running the experiment with smaller sample sizes than those required without finding out what effect this will have on the analysis.

The power and sample size calculation also occur at the grant application stage. The experimenter often conducts the power and sample size calculation first to find the appropriate sample size that can achieve higher power (for example, the power is greater than 80%). Then the experimenter can make the budget according to the corresponding sample size and seek the funding.

Basic Concepts and Notations

Before we conduct the sample size and power calculation. Let us review some basic concept in hypothesis testing. In hypothesis testing, we have the null hypothesis H_0, the alternative hypothesis H_A, the test statistic T, the rejection region R, and the significance level α. Note that the test statistic T has different distributions depending on which hypothesis is true. The rejection

region, R, is constructed in such a way so that the type I error rate, which is the probability that the null hypothesis is true but is rejected (the probability that the null hypothesis is true but the test statistic falls in the rejection region), does not exceed the pre-determined significance level, α. In other words, we have $P(T \in R) \leq \alpha$ when H_0 is true. In many simple cases including the case in this lesson, we can construct a rejection region such that $P(T \in R) = \alpha$ when H_0 is true. Then the probability that the alternative hypothesis is true but is rejected (the probability that the alternative hypothesis is true but the test statistic falls in the rejection region). In other words, the power is $P(T \in R)$ when H_A is true.

In this lesson, for the power and sample size calculation from experiments with one source of variation, we assume that there are v levels and each level has r samples (in other words, $r_1 = \cdots = r_v = r$), so the sample size will be $n = v * r$. For the one-way analysis of variance model,

$$Y_{it} = \mu + \tau_i + \epsilon_{it}, t = 1, \cdots, r; i = 1, \cdots, v$$

we have the following:

The null hypothesis: $H_0: \tau_1 = \tau_2 = \cdots = \tau_v$

The alternative hypothesis: H_A: at least two of τ_i differ

The test statistic: $T = \frac{MST}{MSE} = \frac{SST/(v-1)}{SSE/(n-v)} = \frac{\sum_{i=1}^{v} r*(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2/(v-1)}{\sum_{i=1}^{v}\sum_{t=1}^{r}(Y_{it}-\bar{Y}_{i\cdot})^2/(n-v)}$

For a given significance level α, the rejection region: reject H_0 if $\frac{msT}{msE} \geq F_{v-1,n-v,\alpha}$. Here $F_{v-1,n-v,\alpha}$ is the critical value from F distribution with $v-1$ and $n-v$ degrees of freedom with α in the right-hand tail.

Note that the test statistic T has a F distribution with $v-1$ and $n-v$ degrees of freedom when the null hypothesis is true.

The power of this test becomes: $P(\frac{MST}{MSE} \geq F_{v-1,n-v,\alpha})$. Note that the distribution of the test statistic T is **NOT** a F distribution with $v-1$ and $n-v$ degrees of freedom when the alternative hypothesis is true. When the alternative hypothesis is true, $\frac{MST}{MSE}$ has a related distribution called a noncentral F distrbution. The null central F distribution is denoted by $F_{v-1,n-v,\delta^2}$, where δ^2 is called noncentrality parameter and is a function of $v, \tau_1 \cdots, \tau_v$, and σ^2:

$$\delta^2 = (v-1)Q(\tau_1 \cdots, \tau_v)/\sigma^2$$

So in theory, if we know $\tau_1 \cdots, \tau_v$, and σ^2, we can easily calculate the power. However, it is generally difficult to know each of $\tau_1 \cdots, \tau_v$, so the power is generally expressed as a function of Δ, denoted by $\pi(\Delta)$. The power at Δ, $\pi(\Delta)$ is the probability of rejecting H_0 when the effects of at least two of the treatments differ by Δ.

In summary, for the power and sample size calculation, we have the following notations:

- α: significance level
- v: number of treatment levels
- $r = r_1 = \cdots = r_v$: number of replicates for each level
- π: power
- σ^2: variance in the model
- Δ: a threshold for the differences in the effects of the treatments

We have the following calculations:

1. Find power (π), for the given $\alpha,\ v,\ r,\ \sigma^2,\ \Delta$

   or

2. Find sample size (r) for the given α, v, σ^2, Δ, π

We can use either Table A.7 or R program to do the calculations.

Power calculation with Table A.7

Here we will show how to calculate the power π for the given $\alpha,\ v,\ r,\ \sigma^2,\ \Delta$. In Table A.7, the power π is tabulated by two degrees of freedom and the parameter ϕ and

$$v_1 = v - 1, v_2 = n - v = v * r - v, \phi = \sqrt{\frac{r*\Delta^2}{2*v*\sigma^2}}$$

So the procedures for the power calculation are: **[You may use a flow chart for this]**

Step 1: Find the corresponding table according to $v_1 = v - 1$ and α.

Step 2: Find the corresponding row according to $v_2 = v * r - v$. Note if v_2 is not in the table, choose one that is closest to v_2.

Step 3: Find the corresponding column according to $\phi = \sqrt{\frac{r*\Delta^2}{2*v*\sigma^2}}$. Note if ϕ is not in the table, choose one that is closest to ϕ.

Step 4: Report the number according to the row and column from Step 2 and 3. This number is the power (approximate).

Example 2.2: Power Calculation - Heart-Lung Pump Experiment

What is the power of current study for $\alpha = 0.05, v = 5, r = 4, \sigma^2 = 0.0014$

If $\Delta = 0.10$:

Solution:

Step 1: Find the corresponding table according to $v_1 = v - 1 = 4$ and $\alpha = 0.05$

Step 2: Find the corresponding row according to $v_2 = v * r - v = 15$. Note that 15 is in the table, so 15 is used.

Step 3: Find the corresponding column according to $\phi = \sqrt{\frac{r*\Delta^2}{2*v*\sigma^2}} = \sqrt{\frac{4*0.10^2}{2*5*0.0014}} = 1.69$.

Since 1.69 is not in the table, 1.75 is chosen since it is closest to 1.69 in the table.

Step 4: The corresponding number is 0.771 which is the approximate power.

Exercise: if we set $\Delta = 0.05$, what is the power.

Solution:

Step 1: Find the corresponding table according to $v_1 = v - 1 = 4$ and $\alpha = 0.05$

Step 2: Find the corresponding row according to $v_2 = v * r - v = 15$. Note that 15 is not in the table, so 15 is used.

Step 3: Find the corresponding column according to $\phi = \sqrt{\frac{r*\Delta^2}{2*v*\sigma^2}} = \sqrt{\frac{4*0.05^2}{2*5*0.0014}} = 0.84$.

Since 0.84 is not in the table, 1.00 is chosen since it is closest to 0.84 in the table.

Step 4: The corresponding number is 0.298 which is the approximate power.

Sample Size Calculation with Table A.7

Here we will show how to calculate the sample size r for the given $\alpha,\ v,\ \sigma^2,\ \Delta, \pi$. Comparing with the power calculation, there are two differences here. First, both $v_2 = n - v = v * r - v$

and $\phi = \sqrt{\frac{r*\Delta^2}{2*v*\sigma^2}}$ depend on the unknow sample size r. Second, for the sample size calculation, we need to make sure that the calculated sample size that can achieve the desired power. For these reasons, we can use the following procedures: **[You may use a flow chart for this]**

Step 0: Start with $r = 2$

Step 1: Find the corresponding table according to $v_1 = v - 1$ and α.

Step 2: Find the corresponding row according to $v_2 = v * r - v$. Note if v_2 is not in the table, choose one that is **smaller** than v_2.

Step 3: Find the corresponding column according to $\phi = \sqrt{\frac{r*\Delta^2}{2*v*\sigma^2}}$. Note if ϕ is not in the table, choose one that is **smaller** to ϕ.

Step 4: Find the corresponding power. If the power $\geq \pi$, report the corresponding r and $n = v * r$ as the sample size. Otherwise increase r and repeat steps 1-4.

Note:

1. For Steps 1 and 2, we choose the smaller numbers. Note that the power increases with the increased v_2 and ϕ, so the obtained power from Step 4 will be lower than the actual power. Therefore, we can guarantee that the power from the calculated sample size is not less than the given power.
2. Our procedures are different from the book. Our procedures may need more calculations but are more straightforward.

Example 2.3 (Also with Exercise): Sample Size Calculation - Heart-Lung Pump Experiment

What is the sample size required for current study with $\alpha = 0.05, v_1 = 4, \sigma^2 = 0.0014, \Delta = 0.10, \pi = 0.90$

Solution:

Start with $r = 2$:

Step 1: $v_1 = v - 1 = 4$ and $\alpha = 0.05$

Step 2: $v_2 = v * r - v = 5 * 2 - 5 = 5$. Since 5 is in the table, so 5 is used.

Step 3: Find the corresponding column according to $\phi = \sqrt{\frac{r*\Delta^2}{2*v*\sigma^2}} = \sqrt{\frac{2*0.10^2}{2*5*0.0014}} = 1.19$.

Since 1.19 is not in the table, 1.00 is chosen.

Step 4: The corresponding power is 0.19. Since $0.19 < 0.90$, we repeat Steps 2 to 4 for $r = 3$.

For $r = 3$:

Step 1: $v_1 = v - 1 = 4$ and $\alpha = 0.05$

Step 2: $v_2 = v * r - v = 5 * 3 - 5 = 10$. Since 10 is in the table, so 10 is used.

Step 3: Find the corresponding column according to $\phi = \sqrt{\frac{r*\Delta^2}{2*v*\sigma^2}} = \sqrt{\frac{3*0.10^2}{2*5*0.0014}} = 1.46$.

Since 1.46 is not in the table, 1.25 is chosen. (Note here 1.25 instead of 1.50 is used although 1.50 is closer to 1.46 than 1.25 to 1.46).

Step 4: The corresponding power is 0.398. Since $0.398 < 0.90$, we repeat Steps 2 to 4 for $r = 4$.

For $r = 4$: (**As an exercise and the solution is attached**)

Step 1: $v_1 = v - 1 = 4$ and $\alpha = 0.05$

Step 2: $v_2 = v * r - v = 5 * 4 - 5 = 15$. Since 15 is in the table, so 15 is used.

Step 3: Find the corresponding column according to $\phi = \sqrt{\frac{r*\Delta^2}{2*v*\sigma^2}} = \sqrt{\frac{4*0.10^2}{2*5*0.0014}} = 1.69$.

Since 1.69 is not in the table, 1.50 is chosen.

Step 4: The corresponding power is 0.622. Since $0.622 < 0.90$, we repeat Steps 2 to 4 for $r = 5$.

For $r = 5$: (**As an exercise and the solution is attached**)

Step 1: $v_1 = v - 1 = 4$ and $\alpha = 0.05$

Step 2: $v_2 = v * r - v = 5 * 6 - 5 = 20$. Since 20 is in the table, so 20 is used.

Step 3: Find the corresponding column according to $\phi = \sqrt{\frac{r*\Delta^2}{2*v*\sigma^2}} = \sqrt{\frac{5*0.10^2}{2*5*0.0014}} = 1.89$.

Since 1.89 is not in the table, 1.75 is chosen.

Step 4: The corresponding power is 0.806. Since $0.806 < 0.90$, we repeat Steps 2 to 4 for $r = 6$.

For $r = 6$: (**As the part of Example**)

Step 1: $v_1 = v - 1 = 4$ and $\alpha = 0.05$

Step 2: $v_2 = v * r - v = 5 * 6 - 5 = 25$. Since 25 is in the table, so 20 is used.

Step 3: Find the corresponding column according to $\phi = \sqrt{\frac{r*\Delta^2}{2*v*\sigma^2}} = \sqrt{\frac{6*0.10^2}{2*5*0.0014}} = 2.07$. Since 2.07 is not in the table, 2.00 is chosen.

Step 4: The corresponding power is 0.907. Since $0.907 > 0.90$, so the sample size is $r = 6$.

Final Answer:

The required sample size is $r = 6, n = 5 * 6 = 30$. The power for this sample size is at least 0.907. (Note that we use $v_2 = 20$ and $\phi = 2.00$ while the actual $v_2 = 25$ and $\phi = 2.07$ are larger, so the actual power for $r = 6$ is higher than 0.907.

Questions about Power: Determine if the power increases or decreases if

(1) the sample size increases, all others are unchanged (**Answer: the power will increase**)

(2) Δ increases, all others are unchanged (**Answer: the power will increase**)

(3) σ^2 increases, all others are unchanged (**Answer: the power will decrease**)

(4) the significance level α increases, all others are unchanged (**Answer: the power will decrease**)

Power and Sample Size Calculation Using R

One disadvantage of using Table A.7 is that the results are approximate because not all values of v_2 and ϕ are listed in Table A.7. While the use of R does not have this: the results are accurate for each value of v_2 and ϕ.

The R function used for the power and sample size calculation is **pwr.anova.test** in the package **pwr.** The following parameters and their meaning should be specified:

k: number of levels, corresponding to v here

n: number of samples for each treatment level, corresponding to r

sig.level: significance level, corresponding to α

power: the power of the test

f: effect size, is calculated by $f = \sqrt{\frac{\Delta^2}{2*v*\sigma^2}}$. You can see that $\phi = \sqrt{r} * f$.

Example 2.4 (also with Exercise): Power and Sample Size Calculation with R - Heart-Lung Pump Experiment:

Example: Use R function **pwr.anova.test** for the following power and sample size calculation.

(1) Calculate the power for current study with $\alpha = 0.05, v = 5, r = 4, \sigma^2 = 0.0014$, and $\Delta = 0.10$.

(2) Find the sample size for $\alpha = 0.05, v = 5, \sigma^2 = 0.0014$, $\Delta = 0.10$, and $\pi = 0.90$.

Solution:

(1) The power is: 0.739

The R program: pwr.anova.test()

pwr.anova.test(k = 5, n = 4, sig.level = 0.05, f = sqrt(0.10 * 0.10/(2 * 5 * 0.0014)))

(2) The sample size is: $r = 6$ ($r = 5.36$ from R)

The R program: pwr.anova.test()

pwr.anova.test(k = 5, sig.level = 0.03, f = sqrt(0.05 * 0.05/(2 * 5 * 0.0014)), power = 0.80)

Exercise: Use R function **pwr.anova.test** for power and sample size calculation**.**

(1) Calculate the power for current study with $\alpha = 0.05, v = 5, r = 4, \sigma^2 = 0.0014$, and $\Delta = 0.05$.

(2) Find the sample size for $\alpha = 0.03, v = 5, \sigma^2 = 0.0014$, $\Delta = 0.05$, and $\pi = 0.80$.

Solution:

(1) The power is: 0.219

The R program: pwr.anova.test()

pwr.anova.test(k = 5, n = 4, sig.level = 0.05, f = sqrt(0.05 * 0.05/(2 * 5 * 0.0014)))

(2) The sample size is: $r = 17$ ($r = 16.24$ from R)

The R program: pwr.anova.test()

pwr.anova.test(k = 5, sig.level = 0.03, f = sqrt(0.05 * 0.05/(2 * 5 * 0.0014)), power = 0.80)

Additional Thinking questions:

(1) Compare the results of (1) to (3) with the results from A.7. Can you explain why there are some differences?

Power and Sample Size Calculation: A Case Study

Example 2.5: Power and Sample Size Calculation - A Case Study

The project is from Prof. William Cooke in Department of Kinesiology and Integrative Physiology at MTU. Prof. Cooke would like to submit a grant proposal to NIH for funding.

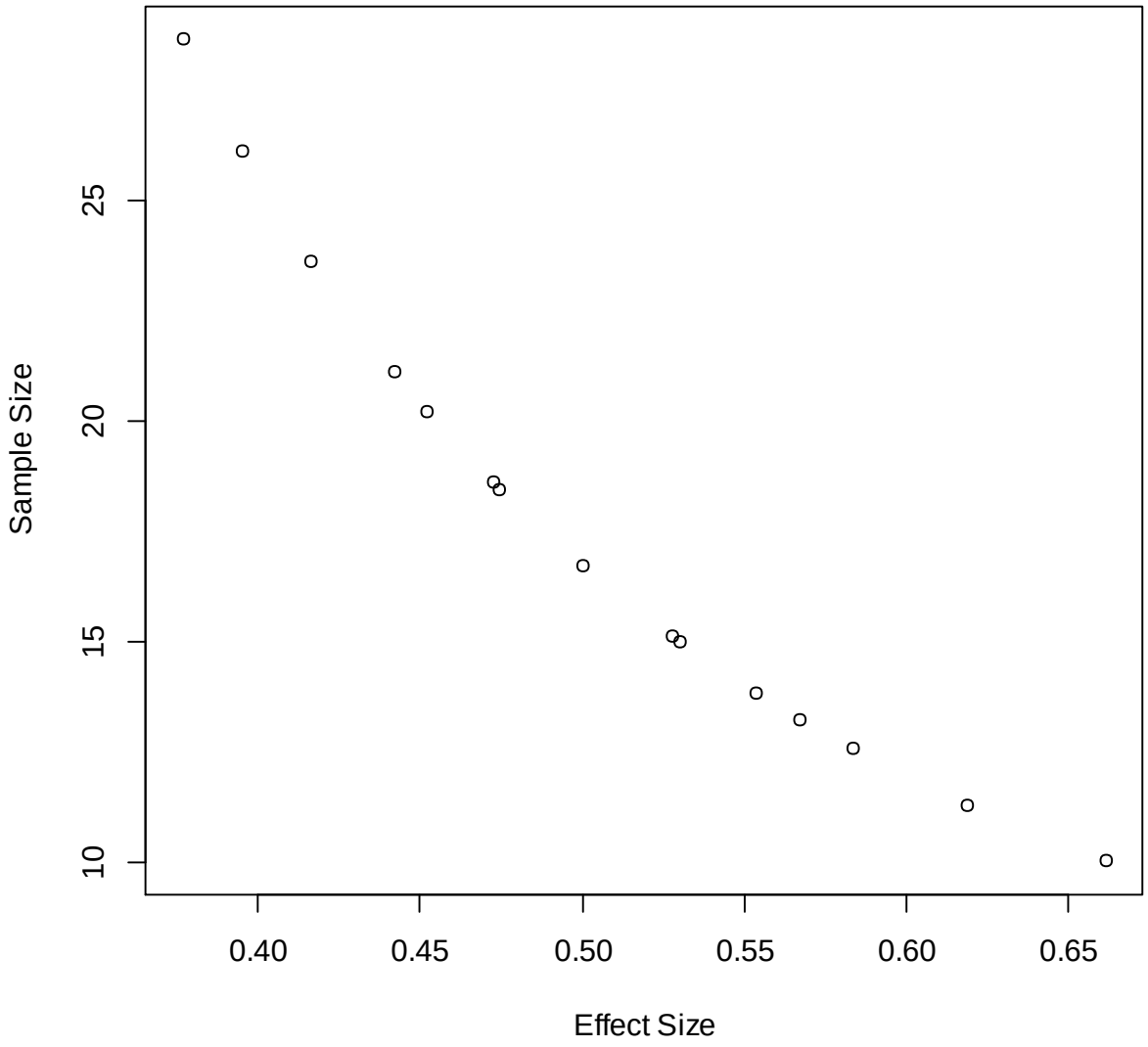
- **Study design**: Samples from two groups - smokers who shift from cigarettes to electronic cigarettes containing no nicotine (placebo) and smokers who shift from cigarettes to electronic cigarettes containing nicotine (treatment group), will be collected.
- **Response variable**: muscle sympathetic nerve activity.
- **Preliminary results**: From a preliminary study of 8 samples, the estimated mean and standard deviation of difference in muscle sympathetic nerve activity before and after using an electronic cigarette are **3** bursts per minute and **3**.
- **Question:** How many samples are needed to achieve 80% power with a significance level of 0.05?
- **Solution:** The sample size can be calculated from a two sample t-test, which is equivalent to ANOVA with two levels.
- **Parameters:**

$$\text{Power} = 0.80,\ \alpha = 0.05,\ v = 2,\ \sigma^2 = 9,\ \Delta = 3.$$

- **Additional considerations:** since difference and standard deviation are estimated from the preliminary study, we need to take the variability into consideration. The solution is to used several numbers around the estimate to see how the sample size are effect by them and then choose an appropriate sample size.
- So we want to see how the sample size changes according to σ^2 and Δ and used different σ^2 and Δ:

$$\sigma^2 = 7.00, 8.00,\ 9.00, 10.00, 11.00 \text{ and } \Delta = 2.5,\ 3,\ 3.5$$

Solution: See R program for the details. The plot is here:



The x-axis is the effect size and y-axis is the corresponding sample size required for achieving 80% of power. Since the power increases with the increased effect size the sample size required to achieve a desired power decreases with the increased effect size. From the plot, you can see that for the sample size of 20 ($r = 20$), the power is at least 80% in most situations considered here. So we proposed used 20 samples per group (a total of 40 samples for two groups) in our grant application.

Lesson 2.3 Introduction of Contrasts

Required Reading: Sections 4.1 and 4.2 in Chapter 4.

Introduction

For experiments with one source of variation, the following analysis variance model is used:

$$y_{it} = \mu + \tau_i + \epsilon_{it}, t = 1, \cdots, r_i; i = 1, \cdots, v$$

For such model, the first step is to test if all τ_i are same. In other words, the first step is to test if all treatment effects are same. We have learned that the use of ANOVA table to perform such test. If the null hypothesis of that all treatment effects are same is rejected, what is the next step? In general, the objective of an experiment is often much more specific than determining whether or not all of the treatments have the same effect. For example, let us look the following experiment first.

Example 2.6: Contrasts - Shrimp Data Experiment

An experiment to determine the effect of various diets on the weight of a certain type of shrimp larvae involved the following seven diets. Five 1-liter containers with 100 shrimp larvae each were fed one of the seven diets in a random assignment. After a period of time the containers were drained and the dry weight of the 100 larvae determined.

- The number of levels: $v = 7$
- The number of replicates: $r_1 = \cdots = r_7 = 5$

Four Experimental diets contained a basal compound diet and three standard diets are used.

- Four experimental diets contained a basal compound diet:

    1. corn and fish oil in a 1:1 ratio,

    2. corn and linseed oil in a 1:1 ratio,

    3. fish and sunflower oil in a 1:1 ratio, and

    4. fish and linseed oil in a 1:1 ratio.

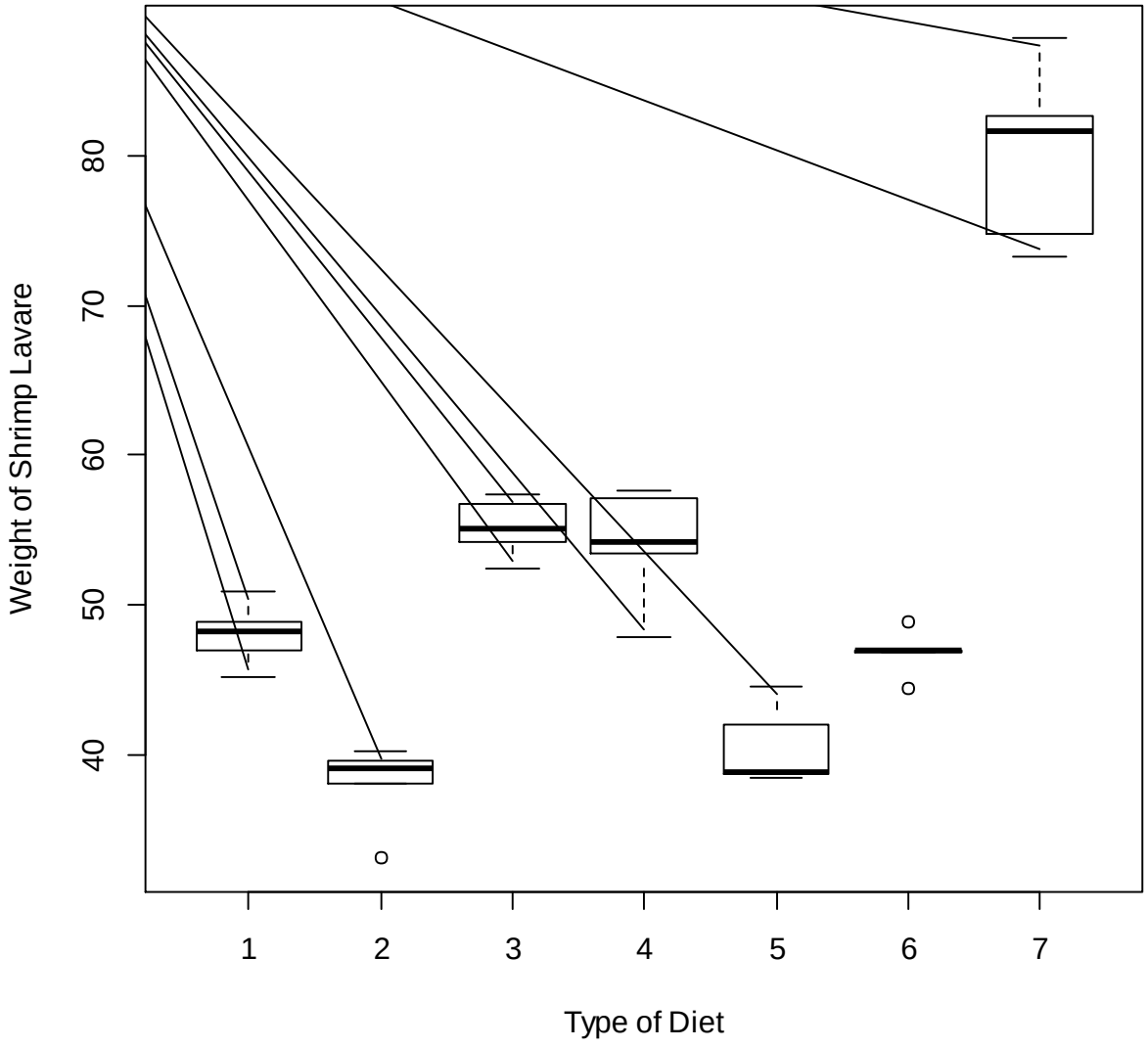
- Three standard diets are used.

5. basal compound diet (a standard diet),

6. live micro algae (a standard diet), and

7. live micro algae and *Artemia* nauplii.

The data can be found in XXX. The box plot of the data is shown below:



Based on the boxplot, we would hypothesize that the effects of seven diets are different. And the results from the ANOVA table confirms our initial thought (The construction of ANOVA table

will be used a quiz problem). So what is our next step? Actually, there are a lot of interesting questions can be answered. For example, we may be interested in:

- Which pair of diets has different effects?
- Is the effect of experimental diets different from the effect of standard diets?
- Is the effect of diets containing corn different from the effect of diets without corn?
- Is the effect of diets containing linseed oil different from the effect of diets without linseed oil?
- What is the CI of the difference between the effect of diets containing Artemia and the effect of diets without Artemia?

These questions can be answered by **studying contrasts**. The purpose of this lesson and some lessons in Week 3 is to provide confidence intervals and hypothesis tests about treatment comparisons by studying contrasts.

Contrasts

Note that if a function of parameters is estimable and/or a contrast depending on the model. Here we use the one-way analysis of variance model:

$$Y_{it} = \mu + \tau_i + \epsilon_{it}, t = 1, \cdots, r_i; i = 1, \cdots, v$$

From Lesson 1.4, we know that all estimable functions can be written as:

$$\sum_{i=1}^{v} b_i * (\mu + \tau_i)$$

And if $\sum_{i=1}^{v} b_i = 0$, then $\sum_{i=1}^{v} b_i * (\mu + \tau_i) = \sum_{i=1}^{v} b_i * \tau_i$ is a contrast. To differentiate the contrast and a general estimable function of parameters, we use c_i instead of b_i for the contrast. Therefore, we can define the contrast as:

Any function $\sum_{i=1}^{v} c_i * \tau_i$ for which $\sum_{i=1}^{v} c_i = 0$ is called a **contrast**, and all contrasts are estimable.

It is convenient to represent a contrast by listing only the coefficients of the parameters $\tau_1, \tau_2, \cdots \tau_v$. Thus the contrast $\sum_{i=1}^{v} c_i * \tau_i = c_1\tau_1 + \cdots + c_v\tau_v$ would be represented by the list of ***contrast coefficients***:

$$[c_1, c_2, \cdots, c_v]$$

Since $\sum_{i=1}^{v} c_i * \tau_i = \sum_{i=1}^{v} c_i * (\mu + \tau_i)$ if $\sum_{i=1}^{v} c_i = 0$, so the least squares estimator of a contrast $\sum_{i=1}^{v} c_i * \tau_i$ is:

$$\sum_{i=1}^{v} c_i * \hat{\tau}_i = \sum_{i=1}^{v} c_i * (\hat{\mu} + \hat{\tau}_i) = \sum_{i=1}^{v} c_i * \bar{Y}_{i\cdot}$$

We also know that $\bar{Y}_{i\cdot} \sim N((\mu + \tau_i, \frac{\sigma^2}{r_i})(i = 1, \cdots, v)$ and $\bar{Y}_{i\cdot}(i = 1, \cdots, v)$ are independent, so the least squares estimator $\sum_{i=1}^{v} c_i * \bar{Y}_{i\cdot}$ also has normal distribution with the mean of $\sum_{i=1}^{v} c_i * (\mu + \tau_i) = \sum_{i=1}^{v} c_i * \tau_i$ and the variance of $\sum_{i=1}^{v} c_i^2 * \frac{\sigma^2}{r_i}$. Then the estimate variance of $\sum_{i=1}^{v} c_i * \bar{Y}_{i\cdot}$ is:

$$\widehat{Var}(\sum_{i=1}^{v} c_i * \bar{Y}_{i\cdot}) = \sum_{i=1}^{v} c_i^2 * \frac{\hat{\sigma}^2}{r_i} = \hat{\sigma}^2 * \sum_{i=1}^{v} \frac{c_i^2}{r_i}$$

where $\hat{\sigma}^2$ is an estimate of σ^2, which is msE (mean square for error).

The estimated standard deviation (also called standard error of an estimator) is: $\sqrt{\hat{\sigma}^2 * \sum_{i=1}^{v} \frac{c_i^2}{r_i}}$.

The distribution of $\sum_{i=1}^{v} c_i * \bar{Y}_{i\cdot}$ and the estimated standard deviation of $\sum_{i=1}^{v} c_i * \bar{Y}_{i\cdot}$ will be used to construct the confidence interval of the contrast $\sum_{i=1}^{v} c_i * \tau_i$.

Four types of contrasts will be studied in our course: **contrasts for pairwise comparisons**, **contrasts for treatment versus controls**, **contrasts for difference of averages**, and **contrasts for trend**. Here, we will use the Battery Experiment or the Heart-Lung Pump Experiment as an example. Note that for the **Battery Experiment**, we have $\hat{\sigma}^2 = 2367.71$ and the following table:

Level	**Replicates**	**Sum**	**Mean**

1	$r_1 = 4$	$y_{1\cdot} = 2283$	$\bar{y}_{1\cdot} = 570.75$
2	$r_2 = 4$	$y_{2\cdot} = 3442$	$\bar{y}_{2\cdot} = 860.50$
3	$r_3 = 4$	$y_{3\cdot} = 1732$	$\bar{y}_{3\cdot} = 433.00$
4	$r_4 = 4$	$y_{4\cdot} = 1985$	$\bar{y}_{4\cdot} = 496.25$
Overall	n =16	$y_{\cdot\cdot} = 9442$	$\bar{y}_{\cdot\cdot} = 590.13$

For the Heart-Lung Pump Experiment, we have $\hat{\sigma}^2 = 0.0014$ and the following table:

Level	**Replicates**	**Sum**	**Mean**
1	$r_1 = 5$	$y_{1\cdot} = 5.676$	$\bar{y}_{1\cdot} = 1.1352$
2	$r_2 = 3$	$y_{2\cdot} = 5.166$	$\bar{y}_{2\cdot} = 1.7220$
3	$r_3 = 5$	$y_{3\cdot} = 11.634$	$\bar{y}_{3\cdot} = 2.3268$
4	$r_4 = 2$	$y_{4\cdot} = 5.850$	$\bar{y}_{4\cdot} = 2.9250$
5	$r_5 = 5$	$y_{5\cdot} = 17.646$	$\bar{y}_{5\cdot} = 3.5292$
Overall	$n = 20$	$y_{\cdot\cdot} = 45.972$	$\bar{y}_{\cdot\cdot} = 2.2986$

Contrasts for Pairwise Comparisons

As the name suggests, pairwise comparisons are simple differences $\tau_u - \tau_s$ of pairs of parameters τ_u and τ_s ($u \neq s$). These are of interest when the experimenter wishes to compare each treatment with every other treatment. The list of contrast coefficients for the pairwise difference $\tau_u - \tau_s$ is

$$[0,0,1,0,\cdots,0,-1,0,0]$$

where the 1 and −1 are in positions u and s, respectively. For $\tau_u - \tau_s$:

- The least squares estimate: $\hat{\tau}_u - \hat{\tau}_s = \bar{y}_{u\cdot} - \bar{y}_{s\cdot}$
- The least squares estimator: $\hat{\tau}_u - \hat{\tau}_s = \bar{Y}_{u\cdot} - \bar{Y}_{s\cdot}$
- The distribution of $\bar{Y}_{u\cdot} - \bar{Y}_{s\cdot}$ is $N(\tau_u - \tau_s, \sigma^2 * (\frac{1}{r_u} + \frac{1}{r_s}))$
- The estimated standard deviation of $\hat{\tau}_u - \hat{\tau}_s = \bar{Y}_{u\cdot} - \bar{Y}_{s\cdot}$ (the standard deviation of an estimator is also called the standard error) is:

$$\sqrt{\hat{\sigma}^2 * \left(\frac{1}{r_u} + \frac{1}{r_s}\right)} = \sqrt{msE * (\frac{1}{r_u} + \frac{1}{r_s})}$$

Example 2.7: Contrasts for Pairwise Comparisons - Battery Experiment

For this experiment, The average lives per unit cost (in minutes/dollar) for the four batteries, calculated from the data are:

$$\bar{y}_{1\cdot} = 570.75, \bar{y}_{2\cdot} = 860.50, \bar{y}_{3\cdot} = 433.00, \bar{y}_{4\cdot} = 496.25$$

The least squares estimates of the pairwise differences are, therefore,

$$\hat{\tau}_1 - \hat{\tau}_2 = \bar{y}_{1\cdot} - \bar{y}_{2\cdot} = 570.25 - 860.50 = -289.75\,,$$

$$\hat{\tau}_1 - \hat{\tau}_3 = 137.75\,, \hat{\tau}_1 - \hat{\tau}_4 = 74.50\,,$$

$$\hat{\tau}_2 - \hat{\tau}_3 = 427.50\,, \hat{\tau}_2 - \hat{\tau}_4 = 364.25\,, \hat{\tau}_3 - 4 = -63.25$$

From these estimated pairwise differences, we can suggest, for example, that battery type 2 (alkaline, store brand) is vastly superior to the other three battery types in terms of the mean life per unit cost. However, we need to investigate whether or not these perceived differences might be due only to random fluctuations in the data. This will be achieved by calculating the confidence intervals of these pairwise differences. At this stage, we can calculate the estimated standard error for each pairwise comparison. Because the mean square for error is 2367.71, so $\hat{\sigma}^2 = 2367.71$ and the estimated standard error for $\hat{\tau}_1 - \hat{\tau}_2$ is:

$$\sqrt{\hat{\sigma}^2 * \left(\frac{1}{r_1} + \frac{1}{r_2}\right)} = \sqrt{2367.71 * \left(\frac{1}{4} + \frac{1}{4}\right)} = 34.41$$

Contrasts Treatment Versus Control

Sometimes, we would like to compare the effects of one special treatment with the effects of each of the other treatment. The special treatment is called the **control.** These contrasts are called **treatment versus control contrasts**. For example, a pharmaceutical experiment might involve one or more experimental drugs together with a standard drug that has been on the market for some years. Frequently, the objective of such an experiment is to compare the effect of each experimental drug with that of the standard drug but not necessarily with the effects of any of the other experimental drugs. The standard drug is then the control. Note there are total of $v - 1$ contrasts and each of them is a contrast for pairwise comparison.

Contrasts for Difference of Averages

Sometimes the levels of the treatment factors divide naturally into two or more groups, and the experimenter is interested in the **difference of averages contrast** that compares the average effect of one group with the average effect of the other group(s).

Example 2.8: Contrasts for Difference of Averages - Battery Experiment

For this experiment, we may be interested in comparing the life per unit cost for name brand and store brand batteries. Since the name brand batteries are in the first and third group while the store brand batteries are in the second and fourth group, the contrast for this comparison is

$$\frac{1}{2} * (\tau_1 + \tau_3) - \frac{1}{2} * (\tau_2 + \tau_4)$$

which is the difference of the average life per unit cost of name brand and store brand batteries. For this contrast,

- The corresponding coefficients are: $\left[\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right] = [0.5, -0.5, 0.5, -0.5]$
- The least squares estimate is

$$\frac{1}{2} * (\hat{\tau}_1 + \hat{\tau}_3) - \frac{1}{2} * (\hat{\tau}_2 + \hat{\tau}_4) = \frac{1}{2} * (\bar{y}_{1\cdot} + \bar{y}_{3\cdot}) - \frac{1}{2} * (\bar{y}_{2\cdot} + \bar{y}_{4\cdot})$$

$$= \frac{1}{2} * (570.25 + 433.00) - \frac{1}{2} * (860.50 + 496.25)$$

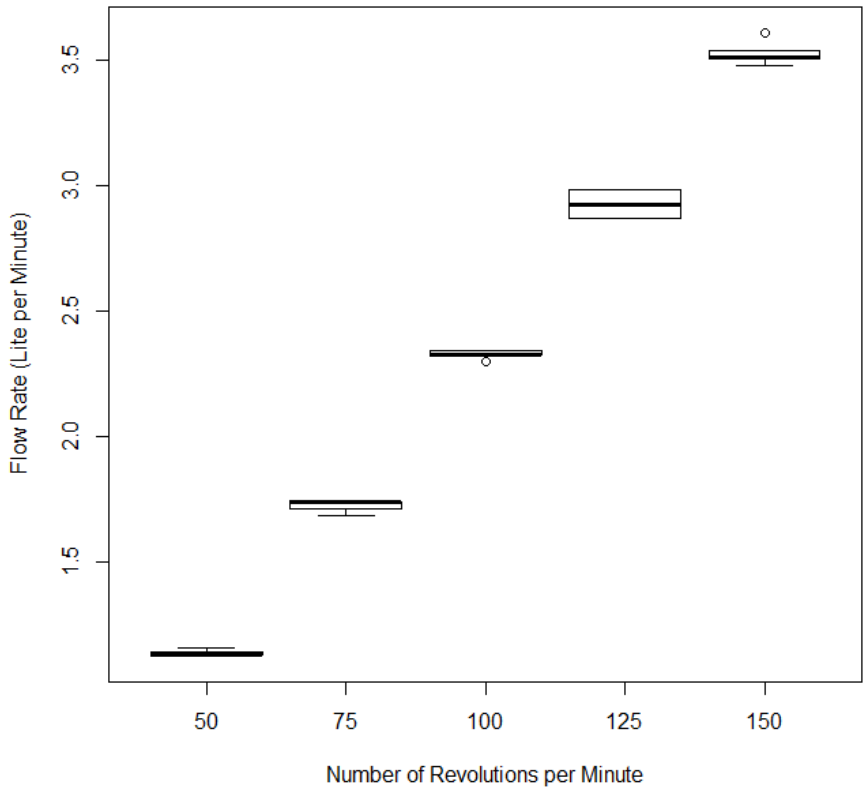
$$= -176.75$$

- The estimated standard error is:

$$\sqrt{\hat{\sigma}^2 * \sum_{i=1}^{v} \frac{c_i^2}{r_i}} = \sqrt{2367.71 * (\frac{0.5^2}{4} + \frac{(-0.5)^2}{4} + \frac{0.5^2}{4} + \frac{(-0.5)^2}{4})} = 24.33$$

Contrasts for Trend

Trend contrasts may be of interest when the levels of the treatment factor are quantitative and have a natural ordering. For example, in Heart-Lung Pump Experiment, the treatment levels, the number of revolutions per minute (rpm) of the rotary pump, are 50, 75, 100, 125, and 150 (although they are coded as 1, 2, 3, 4, 5). **We may want to know the fluid flow rate increases or decreases as the number of revolutions increases and, if so, whether the rate of change remains constant.** If we look at the corresponding box plot below, we can see the rate of change seems to remain constant. This can be done by estimating linear trends in the response. In this course, we will focus on the linear trend.



The trend contrast coefficients for v **equally spaced levels** of a treatment factor and **equal sample sizes** are listed in Table A.2 for values of v between 3 and 7. Table A.2 does not contain the contrast coefficients for unequally spaced levels or for unequal sample sizes. The general method of obtaining the coefficients of the trend contrasts involves fitting a regression model to the noncoded levels of the treatment factor. Here we introduce how to calculate the contrast coefficients for the linear trend.

Let x_i be the ith un-coded level of treatment, then the contrast coefficients for the linear trend is:

$$c_i = r_i * (x_i - \bar{x}.)(i = 1,2,\cdots,v)$$

where $\bar{x}. = \frac{r_1 * x_1 + \cdots + r_v * x_v}{n} = \frac{r_1 * x_1 + \cdots + r_v * x_v}{r_1 + \cdots + r_v}$ is the weighted average of x_i. Note that $\bar{x}. \neq \frac{x_1 + \cdots + x_v}{v}$ unless $r_1 = r_2 = \cdots = r_v = r$.

Example 2.9: Contrasts for Trend - Heart-Lung Pump Experiment

The treatment levels, the number of revolutions per minute (rpm) of the rotary pump, are 50, 75, 100, 125, and 150. The treatment levels are equally spaced since $25 = 150 - 125 = 125 - 100 = 100 - 75 = 75 - 50$. The sample sizes are not equal therefore we cannot use the contrast coefficients from Table A.2. To calculate the contrast coefficients for the linear trend, we can construct and use the following table:

Such calculation can be easily tabulated by the following way:

	r_i	x_i	$r_i * x_i$	$r_i * \bar{x}_.$	$c_i = r_i * (x_i - \bar{x}_.)$ $= r_i * x_i - r_i * \bar{x}_.$
	$r_1 = 5$	$x_1 = 50$	$5 * 50 = 250$	$5 * 98.75 = 493.75$	$c_1 = 250 - 493.75 = -243.75$
	$r_2 = 3$	$x_2 = 75$	$3 * 75 = 225$	$3 * 98.75 = 296.25$	$c_2 = 225 - 296.25 = -71.25$
	$r_3 = 5$	$x_3 = 100$	$5 * 100 = 500$	$5 * 97.75 = 493.75$	$c_3 = 500 - 493.75 = 6.25$
	$r_4 = 2$	$x_4 = 125$	$2 * 125 = 250$	$2 * 98.75 = 197.50$	$c_4 = 250 - 197.50 = 52.50$
	$r_5 = 5$	$x_5 = 150$	$5 * 150 = 750$	$5 * 98.75 = 493.75$	$5 = 750 - 493.75 = 256.25$
Total	$n = 20$	N/A	$\bar{x}$ =1975 $\bar{x}_. = \frac{1975}{20} = 98.75$	N/A	N/A

Note:

- Column for r_i: the number of samples for each treatment level. Take the sum to obtain the total number of samples, n
- Column for x_i: the un-coded levels
- Column for $r_i * x_i$: calculated correspondingly. Take the sum to obtain $\bar{x}$, then calculate $\bar{x}_.$.
- Columns for $r_i * \bar{x}_.$ and c_i: calculated correspondingly

Therefore, the contrast coefficients for the linear trend is:

$$[-243.75, -71.25, 6.25, 52.50, 256.25]$$

The least squares estimate is:

$$\sum_{i=1}^{5} r_i * \bar{y}_{i\cdot} = -243.75 * 1.1352 + \cdots + 256.25 * 3.5292 = 673.065$$

The estimated standard deviation for the contrast of the linear trend is:

$$\sqrt{\hat{\sigma}^2 * \sum_{i=1}^{v} \frac{c_i^2}{r_i}} = \sqrt{0.0014} \sqrt{\frac{(-243.75)^2}{5} + \frac{(-71.27)^2}{3} + \cdots + \frac{256.25^2}{5}} = 6.2476$$

Exercise 2.2: Contrasts - Shrimp Experiment

There are seven levels for the diet – the treatment factor:

- Four experimental diets contained a basal compound diet:

  1. corn and fish oil in a 1:1 ratio,

  2. corn and linseed oil in a 1:1 ratio,

  3. fish and sunflower oil in a 1:1 ratio, and

  4. fish and linseed oil in a 1:1 ratio.

- Three standard diets are used.

  5. basal compound diet (a standard diet),

  6. live micro algae (a standard diet), and

  7. live micro algae and *Artemia* nauplii.

From the data, we can obtain:

$$msE = \hat{\sigma}^2 = 11.064$$

$$\bar{y}_{1\cdot} = 48.04, \bar{y}_{2\cdot} = 38.04, \bar{y}_{3\cdot} = 55.20, \bar{y}_{4\cdot} = 54.06, \bar{y}_{5\cdot} = 40.54, \bar{y}_{6\cdot} = 46.84, \bar{y}_{7\cdot} = 80.06$$

$$r = r_1 = \cdots = r_7 = 5$$

Find the contrast coefficients, the least squares estimates, and the estimated standard errors for the following comparisons:

(1) The pairwise comparison between two diets containing fish.

(2) The compassion between the experimental diets and the standard diets.

Solution:

(1) Contrast coefficients: $[0, 0, 1, -1, 0, 0, 0]$ since diets 3 and 4 contain the fish.

Least squares estimate: $55.20 - 54.06 = 1.14$

Estimated standard error: $\sqrt{11.064 * (\frac{1}{5} + \frac{1}{5})} = 2.104$

(2) Contrast coefficients: $[\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}]$ since first four diets are the experimental diet and the last three diets are the standard diet. To get the average effect of the experimental diet, we need to take the average of effects from four experimental diets. Similar, we can get the average effect of three standard diets.

Least squares estimate: $\frac{1}{4} * 48.04 + \cdots + \left(-\frac{1}{3}\right) * 80.06 = -6.978$

Estimated standard error: $\sqrt{11.064 * (\frac{(1/4)^2}{5} + \cdots + \frac{(-1/3)^2}{5})} = 1.136$

Week 2 Practice Problem Discussion

A manufacturer of concrete bridge supports is interested in determining the effect of varying the sand content of concrete on the strength of the supports. Five supports are made for each of four different amounts of sand in the concrete mix and each support tested for compression resistance. The four level of percent of sand are: 15%, 20%, 25%, and 30%.

(1) Are the treatment levels equally spaced?

(2) Are the sample sizes equal?

(3) Find the contrast coefficients for the liner trend from Table A.2.

(4) Manually calculate the contrast coefficients for the linear trend.

(5) What is the relationship between the contrast coefficients from (2) and (3)?

Solution:

(1) The treatment levels are equally spaced.

(2) The sample sizes are equal.

(3) From Table A.2, the contrast coefficients for the liner trend are: $[-3,-1,1,3]$

(4) From the following table, the contrast coefficients for the liner trend are: $[-37.5,-12.5,12.5,37.5]$

	r_i	x_i	$r_i * x_i$	$r_i * \bar{x}.$	$c_i = r_i * (x_i - \bar{x}.)$ $= r_i * x_i - r_i * \bar{x}.$
	5	15	$5 * 15 = 75$	$5 * 22.5 = 112.5$	$c_1 = 75 - 112.5 = -37.5$
	5	20	$5 * 20 = 100$	$5 * 22.5 = 112.5$	$c_2 = 100 - 112.5 = -12.5$
	5	25	$5 * 25 = 125$	$5 * 22.5 = 112.5$	$c_3 = 125 - 112.5 = 12.5$
	5	30	$5 * 30 = 150$	$5 * 22.5 = 112.5$	$c_4 = 150 - 112.5 = 37.5$
Total	$n = 20$	N/A	$\bar{x}$ =450 $\bar{x}. = \frac{450}{20} = 22.5$	N/A	N/A

(5) We have $12.5 * [-3,-1,1,3] = [-37.5,-12.5,12.5,37.5]$. So the conclusions for testing the linear trend based on these two contrasts will be the same. Actually, for a formal comparison of least squares estimates of two different contrasts, we should normalized them (Please refers to Page 70 of the textbook) so they are at the same scale. But such normalization will be introduced later when we need it. After the normalization, the contrasts from Table A.2 and the contrast from our manual calculation would be same.

MA4720: Design and Analysis of Experiments

Week 3 At-a-Glance

Title: Inference on Contrasts and Checking Model Assumptions

Overview

This week, we will cover the statistical inference including point estimation, confidence interval, and hypothesis testing of a single contrast. We will discuss several methods for adjust of multiple comparisons when several contrasts are considered simultaneously. We will introduce methods for checking if the assumptions of one-way analysis of variance model are satisfied.

Objectives

When you complete this module, you should be able to do the following:

1. Perform the statistical inference, including point estimation, confidence interval, and hypothesis testing of a single contrast.
2. Choose and apply appropriate method for multiple comparisons.
3. Use appropriate plots and tests to check if the assumptions of ANOVA model are satisfied.

Reading (Optional)

1. Section 4.3 in Chapter 4.
2. Sections 4.4 and 4.5 in Chapter 4.
3. Sections 5.1, 5.2, 5.3, 5.4, 5.5., and 5.6 in Chapter 5.

Instruction Content

See other files for details.

Quiz

Two quizzes. See other files for details.

Homework

One homework. See other files for details.

MA4720: Design and Analysis of Experiments

Week 3 Instruction Contents

Lesson 3.1 Statistical Inference for Individual Contrasts

Required Reading: Sections 4.3 in Chapter 4.

For designs with one source of variance, we use the following one-way analysis of variance model:

$$Y_{it} = \mu + \tau_i + \epsilon_{it}, t = 1, \cdots, r_i; i = 1, \cdots, v$$

with the following assumptions:

- μ and τ_i $(i = 1, \cdots, v)$ are unknown population parameters
- $\epsilon_{it}(t = 1, \cdots, r_i; i = 1, \cdots, v)$ are independent
- $\epsilon_{it} \sim N(0, \sigma^2)$: normal distribution with mean 0 and variance σ^2

Under this model, we are interested in the statistical inference, including the point estimation, the confidence interval, and the hypothesis testing of an individual contrast $\sum_{i=1}^{v} c_i \tau_i$ where $\sum_{i=1}^{v} c_i = 0$.

Point Estimation for an Individual Contrast

We have learned this from Lesson 2.3. The least squares estimate of the contrast $\sum_{i=1}^{v} c_i \tau_i$ is:

$$\sum_{i=1}^{v} c_i \bar{y}_{i\cdot}$$

where $\bar{y}_{i\cdot}(i = 1, \cdots, v)$ is the sample mean of response variable from the ith treatment level.

Confidence Interval for an Individual Contrast

In this subsection, we obtain a formula for a confidence interval for an individual contrast. We give the formula first, and explain the intuition behind this formula. The detailed deviations which can be found in the textbook will not be presented here.

A $100(1-\alpha)\%$ confidence interval for the contrast $\sum_{i=1}^{v} c_i \tau_i$ is:

$$\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} \pm t_{n-v, \frac{\alpha}{2}} * \sqrt{\sum_{i=1}^{v} c_i^2 * \frac{msE}{r_i}}$$

At the first glance, this formula seems to be very complicated. However, this formula is very similar with the confidence interval of population mean from a set of random samples from a normal population. For a set of random samples of size n, $x_1, x_2, \cdots, x_n$, from a normal population with mean μ, a $100(1-\alpha)\%$ confidence interval of μ is:

$$\bar{x}_{\cdot} \pm t_{n-1,\alpha/2} * \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x}_{\cdot})^2 / n-1}{n}}$$

Actually, this $100(1-\alpha)\%$ confidence interval of μ can written as this way:

$$\text{point estimate of } \mu \ \pm t_{degrees\ of\ freedom,\alpha/2} * \sqrt{estimated\ varaince\ of\ point\ estimator}$$

Therefore, if we consider $\sum_{i=1}^{v} c_i\tau_i$ as a single parameter, because its least squares estimator is

$$\sum_{i=1}^{v} c_i * \bar{Y}_{i\cdot} \text{ and } \sum_{i=1}^{v} c_i * \bar{Y}_{i\cdot} \sim N(\sum_{i=1}^{v} c_i\tau_i, \sqrt{\sum_{i=1}^{v} c_i^2 * \frac{\sigma^2}{r_i}})$$

We have:

the point estimate (least squares estimate) of $\sum_{i=1}^{v} c_i\tau_i$ is: $\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot}$

the estimated variance of least squares estimator of $\sum_{i=1}^{v} c_i\tau_i$ is:

$$\sqrt{\sum_{i=1}^{v} c_i^2 * \frac{\hat{\sigma}^2}{r_i}} = \sqrt{\sum_{i=1}^{v} c_i^2 * \frac{msE}{r_i}} = \sqrt{msE * \sum_{i=1}^{v} \frac{c_i^2}{r_i}}$$

The degrees of freedom for the error is: $n - v$, which will be used in the calculation of critical value from the t-distribution.

Hypothesis Testing for an Individual Contrast

First, the outcome of a hypothesis test can be deduced from the corresponding confidence interval in the following way.

For the following hypothesis testing problem:

$$H_0: \sum_{i=1}^{v} c_i * \tau_i = h \text{ versus } H_A: \sum_{i=1}^{v} c_i * \tau_i \neq h$$

we can first construct a $100(1-\alpha)\%$ confidence interval for $\sum_{i=1}^{v} c_i\tau_i$. If such confidence interval contains h, then the null hypothesis $H_0: \sum_{i=1}^{v} c_i * \tau_i = h$ will be rejected at significance

level α. If such confidence interval contains h. Then we fail to reject the null hypothesis $H_0: \sum_{i=1}^{v} c_i * \tau_i = h$ at significance level α.

We can make this more explicit, as follows. Suppose we wish to test the hypothesis

$$H_0: \sum_{i=1}^{v} c_i * \tau_i = h \text{ against } H_A: \sum_{i=1}^{v} c_i * \tau_i \neq h$$

Using the same arguments for obtaining the confidence interval of $\sum_{i=1}^{v} c_i * \tau_i$, we can easily show that we will reject the null hypothesis at significance level of α if:

$$\left| \frac{\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} - h}{\sqrt{msE * \sum_{i=1}^{v} \frac{c_i^2}{r_i}}} \right| > t_{n-v,\frac{\alpha}{2}}$$

One drawback of the use of confidence interval or the rejection region for the hypothesis testing is that we need to construct the confidence interval or rejection region according to the significance level. If the significance level is changed, then the confidence level or the rejection region need to be calculated again accordingly. For this reason, the p-value is appealing since it does not depend on the significance level α and the researchers can make the decision based on the p-value and the significance level they choose. Similarly, the p-value can be calculated as:

$$2 * \Pr(T_{n-v} > \left| \frac{\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} - h}{\sqrt{msE * \sum_{i=1}^{v} \frac{c_i^2}{r_i}}} \right|)$$

where T_{n-v} is a random variable with t-distribution with $n - v$ degrees of freedom.

The test statistic can be squared, so the rejection region will reject the null hypothesis at significance level of α if:

$$\left(\frac{\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} - h}{\sqrt{msE * \sum_{i=1}^{v} \frac{c_i^2}{r_i}}}\right)^2 = \frac{\left(\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} - h\right)^2}{msE * \sum_{i=1}^{v} \frac{c_i^2}{r_i}} = \frac{ssc}{msE} \geq t^2_{n-v,\frac{\alpha}{2}} = F_{1,n-v,\alpha}$$

Here we call the quality $\frac{(\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} - h)^2}{\sum_{i=1}^{v} \frac{c_i^2}{r_i}}$ as the **sum of squares for contrasts** or **contrast sum of squares** (even though it is the "sum" of only one squared term).

$$\Pr(F_{1,n-v} > \frac{(\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} - h)^2}{msE * \sum_{i=1}^{v} \frac{c_i^2}{r_i}} = \frac{ssc}{msE})$$

where $F_{1,n-v}$ is a random variable with F-distribution with 1 and $n - v$ degrees of freedom.

The above test is a two-tailed test, since the null hypothesis will be rejected for both large and small values of the contrast. Both t-test statistic and F-test statistic can be used for two-tailed tests and will results in identical conclusions and p-values. However, only t-test statistic instead of F-test statistic can be used for one-tailed tests.

For the test of

$$H_0: \sum_{i=1}^{v} c_i * \tau_i = h \text{ against } H_A: \sum_{i=1}^{v} c_i * \tau_i > h$$

or

$$H_0: \sum_{i=1}^{v} c_i * \tau_i \leq h \text{ against } H_A: \sum_{i=1}^{v} c_i * \tau_i > h$$

We reject the null hypothesis when the t-test statistic is large, so we will reject the null hypothesis if

$$\frac{\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} - h}{\sqrt{msE * \sum_{i=1}^{v} \frac{c_i^2}{r_i}}} > t_{n-v,\alpha}$$

Equivalently, we will reject the null hypothesis if h does not fall in the right-tailed $100(1-\alpha)\%$ confidence interval:

$$(\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} - t_{n-v,\alpha} * \sqrt{msE * \sum_{i=1}^{v} \frac{c_i^2}{r_i}}, \infty)$$

The p-value is:

$$\Pr(T_{n-v} > \frac{\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} - h}{\sqrt{msE * \sum_{i=1}^{v} \frac{c_i^2}{r_i}}})$$

For the test of

$$H_0: \sum_{i=1}^{v} c_i * \tau_i = h \text{ against } H_A: \sum_{i=1}^{v} c_i * \tau_i < h$$

or

$$H_0: \sum_{i=1}^{v} c_i * \tau_i \geq h \text{ against } H_A: \sum_{i=1}^{v} c_i * \tau_i < h$$

We reject the null hypothesis when the t-test statistic is small, so we will reject the null hypothesis if

$$\frac{\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} - h}{\sqrt{msE * \sum_{i=1}^{v} \frac{c_i^2}{r_i}}} < -t_{n-v,\alpha}$$

Equivalently, we will reject the null hypothesis if h does not fall in the left-tailed $100(1-\alpha)\%$ confidence interval:

$$(-\infty, \sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} + t_{n-v,\alpha} * \sqrt{msE * \sum_{i=1}^{v} \frac{c_i^2}{r_i}}, \infty)$$

The p-value is:

$$\Pr(T_{n-v} < \frac{\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} - h}{\sqrt{msE * \sum_{i=1}^{v} \frac{c_i^2}{r_i}}})$$

Example 3.1: Statistical Inference of Contrasts – Shrimp Experiment

Here we illustrate how to make statistical inference about a single contrast $\frac{1}{2}(\tau_1 + \tau_2) - \frac{1}{2}(\tau_3 + \tau_4)$ and how to interpret the results obtained using the data from shrimp experiment. The questions of interest include:

(1) What is the point estimate of this contrast?

(2) What is a 95% confidence interval of this contrast?

(3) At the significance level 0.05, perform the following hypothesis testing:

$$H_0: \frac{1}{2}(\tau_1 + \tau_2) - \frac{1}{2}(\tau_3 + \tau_4) = 0 \text{ versus } H_A: \frac{1}{2}(\tau_1 + \tau_2) - \frac{1}{2}(\tau_3 + \tau_4) \neq 0$$

using the confidence interval approach, the rejection region approach, and the p-value approach. What is your conclusion here?

(4) At the significance level 0.05, what is your conclusion for the hypothesis testing:

$$H_0: \frac{1}{2}(\tau_1+\tau_2)-\frac{1}{2}(\tau_3+\tau_4) \geq -8.0 \text{ versus } H_A: \frac{1}{2}(\tau_1+\tau_2)-\frac{1}{2}(\tau_3+\tau_4) < -8.0$$

Note that the contrast $\frac{1}{2}(\tau_1+\tau_2)-\frac{1}{2}(\tau_3+\tau_4)$ is a contrast for the difference of averages. Actually, it represents the difference of effects between two experimental diets with corn and two experiment diets without corn, Before we answer these questions, we first list the information needed for our calculation here:

$$n=35, v=7, r=r_1=\cdots=r_7=5, msE=\hat{\sigma}^2=11.064$$

$$\bar{y}_{1\cdot}=48.04, \bar{y}_{2\cdot}=38.04, \bar{y}_{3\cdot}=55.20, \bar{y}_{4\cdot}=54.06, \bar{y}_{5\cdot}=40.54, \bar{y}_{6\cdot}=46.84, \bar{y}_{7\cdot}=80.06$$

Solution: Note that the contrast coefficients for $\frac{1}{2}(\tau_1+\tau_2)-\frac{1}{2}(\tau_3+\tau_4)$ is

$$[0.5, 0.5, -0.5, -0.5, 0, 0, 0]$$

(1) The point estimate of $\frac{1}{2}(\tau_1+\tau_2)-\frac{1}{2}(\tau_3+\tau_4)$ is:

$$\frac{1}{2}(\bar{y}_{1\cdot}+\bar{y}_{2\cdot})-\frac{1}{2}(\bar{y}_{3\cdot}+\bar{y}_{4\cdot})=\frac{1}{2}*(48.04+38.04)-\frac{1}{2}*(55.20-54.06)=-11.59$$

(2) The point estimate is: -11.59 (from (1)

The critical value is: $t_{n-v,\frac{\alpha}{2}}=t_{35-7,\frac{0.05}{2}}=2.0484$

The estimated standard error is:

$$\sqrt{mse*\sum_{i=1}^{v}\frac{c_i^2}{r_i}}=\sqrt{11.064*\frac{0.5^2+0.5^2+(-0.5)^2+(-0.5)^2}{5}}=1.4875$$

(3) Therefore, a 95% confidence interval $\frac{1}{2}(\tau_1+\tau_2)-\frac{1}{2}(\tau_3+\tau_4)$ is:

$$-11.59 \pm 2.0484*1.4875=(-16.64,-8.54)$$

(4) One hypothesis test of interest is to see the difference of effects is different from zero. This can be expressed as the following:

$$H_0: \frac{1}{2}(\tau_1+\tau_2)-\frac{1}{2}(\tau_3+\tau_4)=0 \text{ versus } H_A: \frac{1}{2}(\tau_1+\tau_2)-\frac{1}{2}(\tau_3+\tau_4)\neq 0$$

With the significance level of 0.05, since 0 is not in the 95% confidence interval of this contrast, the null hypothesis is rejected.

The test statistic is:

$$\frac{\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} - 0}{\sqrt{msE * \sum_{i=1}^{v} \frac{c_i^2}{r_i}}} == \frac{-11.59}{1.4875} = -7.79$$

Since $|-7.79| > t_{n-v,\frac{\alpha}{2}} = t_{35-7,\frac{0.05}{2}} = 2.0484$, we reject the null hypothesis.

The corresponding p-value is:

$$p - \text{value} = 2 * \Pr\left(T_{n-v} > \left| \frac{\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} - 0}{\sqrt{msE * \sum_{i=1}^{v} \frac{c_i^2}{r_i}}} \right| \right) = 2 * \Pr\left(T_{35-7} > \left| \frac{-11.59}{1.4875} \right| \right) = 1.74 * 10^{-8}$$

Again we reject the null hypothesis since the p-value is less than the significance level of 0.05.

But how we interpret that the null hypothesis is rejected here? We need to state that we have sufficient evidence to show that the difference of effects between two experimental diets with corn and two experiment diets without corn is statistically significant. Alternatively, we can conclude that there is significant difference of effects between two experimental diets with corn and two experiment diets without corn is statistically significant.

(4) In addition to test if a contrast is different from 0, in many situations to have practical significance, we are also interested in if the contrast is greater or less than a threshold. For example, if we look at the difference of effects between two experimental diets with corn and two experiment diets without corn, it may be useful if such difference is greater than a threshold. In such situations, the one-tailed test will be appropriate. Hypothetical, suppose here we are interest in testing of:

$$H_0: \frac{1}{2}(\tau_1 + \tau_2) - \frac{1}{2}(\tau_3 + \tau_4) \geq -8.0 \text{ versus } H_A: \frac{1}{2}(\tau_1 + \tau_2) - \frac{1}{2}(\tau_3 + \tau_4) < -8.0$$

Here we use the p-value approach and the p-value is:

$$p-\text{value} = \Pr\left(T_{n-v} < \frac{\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} - h}{\sqrt{msE * \sum_{i=1}^{v} \frac{c_i^2}{r_i}}}\right) = \Pr\left(T_{35-7} < \frac{-11.59-(-8.0)}{1.4865}\right) = 0.0112$$

With the significance level of 0.01, we fail to reject the null hypothesis since the *p*-value is greater than 0.01.

Keep in mind that we need to state we fail to reject the null hypothesis instead of acceptance of the null hypothesis. **The conclusion here is that we do not have sufficient evidence to state that the difference of effects between two experimental diets with corn and two experiment diets without corn is greater than -8.0.**

We can also use the confidence interval approach. First construct 99% left-tailed confidence interval:

$$\left(-\infty, \sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} + t_{n-v,\alpha} * \sqrt{msE * \sum_{i=1}^{v} \frac{c_i^2}{r_i}}, \infty\right)$$

$$= (-\infty, -11.59 + 2.4671 * 1.4865) = (-\infty, -7.922)$$

Since $h = -8.0$ belongs to the 99% left-tailed confidence interval, we fail to reject the null hypothesis.

Statistical Inference of Individual Contrasts Using R [A video may be added here]

R provides functions that can be used to calculate point estimate, confidence interval, and hypothesis testing for any individual contrast. Please refer to the R program for details.

Exercise Problem 3.1 (Shrimp Experiment) Statistical Inference of Contrasts. Suppose we are interest in the difference of effect between the first experiment diet and two experimental diets without corn.

(1) Write out the corresponding contrast (contrast coefficients).

(2) What is the point estimate of this contrast?

(3) What is a 95% confidence interval of this contrast?

(4) At the significance level 0.05, perform the following hypothesis testing:

$$H_0: \text{contrast} = 0 \text{ versus } H_A: \text{contrast} \neq 0$$

Solution: please refer to R program.

Lesson 3.2 Methods of Multiple Comparisons

Required Reading: Sections 4.4 and 4.5 in Chapter 4.

Multiple Confidence Intervals

Experimenters are generally interested in statistical inference of multiple contrasts. For example, experimenters may be interest in confidence intervals of all pairwise contrasts. The confidence level for a single confidence interval is based on the probability, that the random interval will be "correct" (meaning that the random interval will contain the true value of the contrast or function). It is shown below that when several confidence intervals are calculated, the probability that they are all simultaneously correct can be alarmingly small. Similarly, when several hypotheses are to be tested, the probability that at least one hypothesis is incorrectly rejected can be uncomfortably high. Much research has been done over the years to find ways around these problems. The resulting techniques are known as ***methods of multiple comparison***, the intervals are called ***simultaneous confidence intervals***, and the tests are called ***simultaneous hypothesis tests***.

Here we illustrated the problem with the confidence interval. For simplicity, suppose that an experimenter would like to calculate *m* confidence intervals, each having a $100(1 - \alpha^*)\%$ confidence level. Let S_j be the event that the jth confidence interval contains the true parameter estimated and $\bar{S}_j$ be the complement of S_j (the event that the jth confidence interval contains the true parameter estimated), $(j = 1, \cdots, m)$. $P(S_1 \cap \cdots \cap S_m)$, the probability that the m intervals will simultaneously be correct is called the ***overall confidence level***. The probability that at least one of m intervals is not correct, $1 - P(S_1 \cap \cdots \cap S_m)$ is called ***overall significance level*** or ***experiment-wise*** *error rate*. Although we can use the standard rules for probabilities of unions and intersections of events to show that the overall significance level has lower bound:

$$P(S_1 \cap \cdots \cap S_m) \geq 1 - m * \alpha^*$$

the exact calculation of $P(S_1 \cap \cdots \cap S_m)$ can be difficult. For a large m, the lower bound of overall confidence level can be too small to provide any useful information for a given α^*. For example, if we use a typical value of 0.05 as α^*, for $m = 10$, we only know that the overall confidence level is at least $50\% = 1 - 10 * 0.05$ which is not very informative. However, based

on this inequality, we can use different α^* according to m such that the overall significance level greater than the desired value.

Similar comments apply to the hypothesis testing situation. If hypotheses for m different contrasts are to be tested, each at significance levelα^* , then the probability that at least one hypothesis is incorrectly rejected is at most $m\alpha^*$.

Various methods have been developed to ensure that the overall confidence level is not too small and the overall significance level is not too high. Some methods are completely general, that is, they can be used for any set of estimable functions, while others have been developed for very specialized purposes such as comparing each treatment with a control. Which method is best depends on which contrasts are of interest and the number of contrasts to be investigated.

Summary of Multiple Comparison Methods

In this section, four methods are discussed that control the overall confidence level and overall significance level. We need to understand the terms ***preplanned contrasts*** and ***contrasts after data snooping*** first.

Preplanned contrasts: are those contrasts determined by the experimenters before the experiment is conducted and the data is collected. For example,

Data snooping: After the data have been collected, the experimenters usually look carefully at the data to see whether unplanned contrasts may turn out to be the most interesting. Allowing the data to suggest additional interesting contrasts is called data snooping.

Contrasts after data snooping: are those additional contrasts identified by the experimenters after data snooping.

Here, we will summarize four methods for multiple comparisons and give some details and examples later.

Bonferroni method for preplanned comparisons

- Applies to **any** m **preplanned contrasts** or estimable functions of the parameters.
- Gives shorter confidence intervals than the other methods listed if m is small.
- The confidence interval of a specific contrast changes according to m.
- **Can be used for any design**.
- **Cannot be used for data snooping**.

Scheffé method for all comparisons

- Applies to **any** m **contrasts** or estimable functions of the parameters.
- Gives shorter intervals than Bonferroni's method if m is large.
- The confidence interval of a specific contrast does not change according to m.
- **Allows data snooping**.
- **Can be used for any design**.

Tukey method for all pairwise comparisons

- Applies to **any** m pairwise **contrasts**.
- Best if **all pairwise comparisons** are considered.
- The confidence interval of a specific contrast does not change according to m.
- Can be used for completely randomized designs, randomized block designs, and balanced incomplete block designs.
- **Allows data snooping if only pairwise contrasts are considered.**

Dunnett method for treatment-versus-control comparisons

- Best for **all treatment-versus-control contrasts**.
- Can be used for completely randomized designs, randomized block designs, and balanced incomplete block designs.
- **Allows data snooping if only treatment-versus-control contrasts are considered.**

Example 3.2: Selection of Methods for Multiple Comparison - Shrimp Experiment

Using Shrimp data, list multiple comparison methods that can be used and discuss which method may be better than the others.

(1) Before the experiment, we decide to do all pairwise comparisons.

(2) Before the experiment, we decide to do all pairwise comparisons, and a comparison between the standard diets and the experiments diets.

(3) Before the experiment, we decide to do the comparisons between 7^{th} diet and other diets since many other experiments using 7^{th} level as a control.

(4) After looking at data, we find there is a difference between 7^{th} diet and 4 experimental diets, so we decide to do these 4 pairwise comparisons.

(5) In addition to all pairwise comparisons, we decide to compare the effect of 7^{th} diet and the average effect of 4 experimental diets after we find there is a difference between 7^{th} diet and 4 experimental diets.

Solution: Since this is a completely randomized experiment, the design is not a concern here.

(1) These are preplanned pairwise contrasts. So Bonferroni method, Scheffé method, and Tukey method can be used. Tukey method should be best because all pairwise contrasts are considered. The number of pairwise contrasts is $m = \frac{7*6}{2} = 21$. This is a relatively large m, so Scheffé method could be better here.

(2) These are preplanned contrasts. The comparison between the standard diets and the experiments diets is a contrast for averages. The total number of contrasts is $m = 22 = 21 + 1$. Both Bonferroni method and Scheffé method can be used. Scheffé method could be better here for this relatively large m.

(3) These are preplanned contrasts. The contrasts are for treatment-versus-control comparisons. So Bonferroni method, Scheffé method, Tukey method, and Dunnet method can be used. Dunnet method would be best. Tukey method should better than Scheffé method since all contrasts are for pairwise comparisons. Since $m = 6$ is small, Bonferroni method should be better than Scheffé method. It is unclear if Bonferroni method is better than Tukey method.

(4) These contrasts are not preplanned contrasts. All contrasts are for pairwise comparisons. So only Scheffé method and Tukey method can be used and Tukey method should be better.

(5) These contrasts are not preplanned contrasts. One contrast is a contrast for averages. So only Scheffé method can be used here.

Example 3. 3: Methods for Multiple Comparison - Shrimp Experiment

Suppose that before the experiment, we decide to do the comparisons between 7th diet and other diets since many other experiments using 7th level as a control. As we have discussed in the previous example, all four methods for multiple comparisons, including Bonferroni method, Scheffé method, Tukey method, and Dunnet method can be used. Here we use the pairwise contrast $\tau_7 - \tau_3$ to illustrate how its 95% simultaneous confidence interval can be calculated. For your convenience, we list the information needed for our calculation here:

$$n = 35, v = 7, r = r_1 = \cdots = r_7 = 5, msE = \hat{\sigma}^2 = 11.064$$

$$\bar{y}_{1\cdot} = 48.04, \bar{y}_{2\cdot} = 38.04, \bar{y}_{3\cdot} = 55.20, \bar{y}_{4\cdot} = 54.06, \bar{y}_{5\cdot} = 40.54, \bar{y}_{6\cdot} = 46.84, \bar{y}_{7\cdot} = 80.06$$

Also, we have:

The point estimate of $\tau_7 - \tau_3 = \bar{y}_{7\cdot} - \bar{y}_{3\cdot} = 80.06 - 55.20 = 24.86$

The estimated standard error is:

$$\sqrt{mse * \sum_{i=1}^{v} \frac{c_i^2}{r_i}} = \sqrt{11.064 * \frac{1^2 + (-1)^2}{5}} = 2.1037$$

Confidence Interval without Methods for Multiple Comparisons

First let us first refresh our memory about the confidence interval for an individual contrast. For any contrast, $\sum_{i=1}^{v} c_i * \tau_i$, the $100 * (1 - \alpha)\%$ confidence interval of $\sum_{i=1}^{v} c_i * \tau_i$ is given by:

$$\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} \pm t_{n-v,\frac{\alpha}{2}} * \sqrt{msE \sum_{i=1}^{v} \frac{c_i^2}{r_i}}$$

As we have illustrated, if $\sum_{i=1}^{v} c_i * \tau_i$ is one of m contrasts considered, the overall confidence level of the confidence intervals of m contrasts has a lower bound of $1 - m * \alpha$. To guarantee

that the overall significance level is at least $1-\alpha$, we need to increase the critical value. Actually, all four methods, Bonferroni method, Scheffé method, Tukey method, and Dunnet method, for multiple comparisons, have the same formula to construct the $100*(1-\alpha)\%$ confidence interval of $\sum_{i=1}^{v} c_i * \tau_i$:

$$\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} \pm w * \sqrt{msE \sum_{i=1}^{v} \frac{c_i^2}{r_i}}$$

The only difference among them is how to determine the critical value w which is greater than $t_{n-v,\frac{\alpha}{2}}$.

Example 3. 3 Continued –95% Confidence Interval of the contrast $\tau_7 - \tau_3$ without methods for multiple comparisons

Solution: The critical value w is

$$w = t_{n-v,\frac{\alpha}{2}} = t_{35-7,\frac{0.05}{2}} = 2.0484$$

So the 95% confidence interval of $\tau_7 - \tau_3$ is:

$$24.86 \pm 2.0484 * 2.1037 = (20.551, 29.169)$$

Bonferroni Method for Preplanned Comparisons

The Bonferroni method is a simple and popular method for simultaneous confidence intervals. To ensure that the overall confidence level of m contrasts is at least $100(1-\alpha)\%$, the Bonferroni method uses $\alpha^* = \frac{\alpha}{m}$ to construct each confidence interval with the confidence level of $100(1-\alpha^*)\%$. According to the aforementioned inequality, the overall confidence level is greater than or equal to $100(1-m*\alpha^*)\% = 100\left(1-m*\frac{\alpha}{m}\right)\% = 100(1-\alpha)\%$.

In summary, to obtain a set of simultaneous $100(1-\alpha)\%$ confidence intervals for m preplanned contrasts, we use the following formula to obtain the confidence interval of $\sum_{i=1}^{v} c_i * \tau_i$:

$$\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} \pm t_{n-v,\frac{\alpha}{2m}} * \sqrt{msE \sum_{i=1}^{v} \frac{c_i^2}{r_i}}$$

Where the critical value, $w_B = t_{n-v,\frac{\alpha}{2m}}$.

Bonferroni Method for p-values of Multiple Comparisons

Suppose that $\sum_{i=1}^{v} c_i * \tau_i$ is one of m preplanned contrasts and the unadjusted p-value is obtained by testing:

$$H_0: \sum_{i=1}^{v} c_i * \tau_i = 0 \text{ versus } H_A: \sum_{i=1}^{v} c_i * \tau_i \neq 0$$

with the following formula:

$$p - \text{value} = 2 * \Pr\left(T_{n-v} > \left| \frac{\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} - 0}{\sqrt{msE * \sum_{i=1}^{v} \frac{c_i^2}{r_i}}} \right| \right)$$

The Bonferroni adjusted p-value is the unadjusted p-value times the number of contrasts, m. For the significance level α, the null hypothesis is rejected only if the unadjusted p-value is less than $\frac{\alpha}{m}$, or equivalently, the null hypothesis is rejected only if the Bonferroni adjusted p-value is less than α.

Example 3. 3 Continued – Simultaneous 95% Confidence Interval of the contrast $\tau_7 - \tau_3$ with Bonferroni method

Solution: Note that we have $m = 6$. So the critical value w_B is

$$w_B = t_{n-v,\frac{\alpha}{2*m}} = t_{35-7,\frac{0.05}{2*6}} = 2.8389$$

So the 95% confidence interval of $\tau_7 - \tau_3$ is:

$$24.86 \pm 2.8389 * 2.1037 = (18.89, 30.83)$$

The unadjusted p-value is:

$$2 * \Pr\left(T_{35-7} > \left|\frac{24.86 - 0}{2.1037}\right|\right) = 2.138 * 10^{-12}$$

The Bonferroni adjusted p-value is:

$$6 * 2.138 * 10^{-12} = 1.283 * 10^{-11}$$

Scheffé Method of Multiple Comparisons

From the formula, we can see that simultaneous confidence intervals from the Bonferroni method depends on the number of preplanned contrasts, m. If m is large, the confidence intervals from Bonferroni method can become very wide. Scheffé's method, on the other hand, provides a set of simultaneous $100(1-\alpha)\%$ confidence intervals whose widths are determined only by the number of treatments and the number of observations in the experiment, no matter how many contrasts are of interest.

Scheffé's method is based on the fact that every possible contrast $\sum_{i=1}^{v} c_i * \tau_i$ can be written as a linear combination of the set of $v-1$ treatment-versus-control contrasts, $\tau_2 - \tau_1$, $\tau_3 - \tau_1$, $\cdots$, and $\tau_v - \tau_1$. Actually, you can easily verify that $\sum_{i=1}^{v} c_i * \tau_i = \sum_{i=2}^{v} c_i * (\tau_i - \tau_1)$ since $\sum_{i=1}^{v} c_i = 0$ for a contrast. Once the experimental data have been collected, it is possible to find a set of simultaneous $100(1-\alpha)\%$ confidence intervals for these $v-1$ treatment-versus-control contrasts. The confidence interval not only determines confidence bounds for each treatment-versus-control contrast, it determines bounds for ***every possible*** contrast $\sum_{i=1}^{v} c_i * \tau_i$, and in fact, for *any number* of contrasts, while the overall confidence level remains fixed.

For a contrast, $\sum_{i=1}^{v} c_i * \tau_i$ among m contrasts, let α be the overall significance level, then for Scheffé's method, the simultaneous $100 * (1-\alpha)\%$ confidence interval of $\sum_{i=1}^{v} c_i * \tau_i$ from one-way analysis of variance model of a completely randomized deign is given by:

$$\sum_{i=1}^{v} c_i * \bar{y}_{i\cdot} \pm \sqrt{(v-1)F_{v-1,n-v,\alpha}} * \sqrt{msE \sum_{i=1}^{v} \frac{c_i^2}{r_i}}$$

Where the critical value, $w_S = \sqrt{(v-1)F_{v-1,n-v,\alpha}}$. Note that the F-distribution here is same as the F-distribution used in the one-way analysis of variance table. The two degrees of freedoms are the degrees of freedom for the treatment $(v-1)$ and the degrees of freedom for the error $(n-v)$, respectively.

Example 3. 3 Continued – Simultaneous 95% Confidence Interval of the contrast $\tau_7 - \tau_3$ with Scheffé method

Solution: The critical value w_S is

$$w_S = \sqrt{(v-1)F_{v-1,n-v,\alpha}} = \sqrt{(7-1)F_{7-1,35-7,0.05}} = 3.8303$$

So the 95% confidence interval of $\tau_7 - \tau_3$ is:

$$24.86 \pm 3.8303 * 2.1037 = (16.80, 32.92)$$

Here we can see that $w_S = 3.8303 > w_B = 2.8389$. So the simultaneous 95% confidence interval from Scheffé method is wider than it from Bonferroni method. This is due to the smaller number of contrasts.

Tukey Method for All Pairwise Comparisons

In some experiments, confidence intervals may be required only for pairwise difference contrasts. Tukey, in 1953, proposed a method that is specially tailored to handle this situation and that gives shorter intervals for pairwise differences than do the Bonferroni and Scheffé methods.

For a pairwise contrast, $\sum_{i=1}^{v} c_i * \tau_i = \tau_u - \tau_s$ among m contrasts, let α be the overall significance level, then for Tukey method, the simultaneous $100 * (1-\alpha)\%$ confidence interval of $\tau_u - \tau_s$ from one-way analysis of variance model of a completely randomized deign is given by:

$$\bar{y}_{u\cdot} - \bar{y}_{s\cdot} \pm \frac{q_{v,n-v,\alpha}}{\sqrt{2}} * \sqrt{msE(\frac{1}{r_u} + \frac{1}{r_s})}$$

Where the critical value, $w_T = \frac{q_{v,n-v,\alpha}}{\sqrt{2}}$. Here, $q_{v,n-v,\alpha}$ corresponds to a probability of α in the right-hand tail of the Studentized range distribution. Some values of $q_{v,n-v,\alpha}$ can be found in Table A.8 but we will use R to find this critical value. Note that the first degrees of freedom is v instead of $v-1$ while the second degrees of freedom is still the degrees of freedom for the error $(n-v)$.

The deviation of Tukey method is based on Studentized range distribution. Tukey method is also called **Tukey HSD (Honest Significant Difference)** Test. When the sample sizes are equal ($r = r_1 = \cdots = r_v$), the overall confidence level from Tukey method is ***exactly* 100*(1 − α)*%.** When the sample sizes are unequal, the confidence level is ***at least*** $100(1 - \alpha)\%$.

Example 3. 3 Continued – Simultaneous 95% Confidence Interval of the contrast $\tau_7 - \tau_3$ with Tukey method

Solution: The critical value w_T is

$$w_T = \frac{q_{v,n-v,\alpha}}{\sqrt{2}} = \frac{q_{7,35-7,0.05}}{\sqrt{2}} = \frac{4.4861}{\sqrt{2}} = 3.1722$$

So the 95% confidence interval of $\tau_7 - \tau_3$ is:

$$24.86 \pm 3.1722 * 2.1037 = (18.19, 31.53)$$

Here we can see the simultaneous 95% confidence interval of $\tau_7 - \tau_3$ from Tukey method is shorter than it from Scheffé method but wider than it from Bonferroni method.

Dunnett Method for Treatment-Versus-Control Comparisons

In 1955, Dunnett developed a method of multiple comparisons that is specially designed to provide a set of simultaneous confidence intervals for preplanned treatment-versus-control contrasts. The intervals obtained from Dunnett method are shorter than those given by the Scheffé, Tukey, and Bonferroni methods, but the method should not be used for any other type of contrasts.

Without loss of generality, we assume that level 1 is used as the control and the $v - 1$ treatment-versus-control contrasts are $\tau_2 - \tau_1, \cdots, \tau_v - \tau_1$. The formulae for the simultaneous confidence intervals are based on the joint distribution of the estimators $\bar{Y}_{i\cdot} - \bar{Y}_{1\cdot}$ $(i = 2, \ldots, v)$. This distribution is a special case of the multivariate t-distribution and depends on the correlation between $\bar{Y}_{i\cdot} - \bar{Y}_{1\cdot}$

For a pairwise contrast, $\sum_{i=1}^{v} c_i * \tau_i = \tau_u - \tau_s$ among $v-1$ preplanned treatment-versus-control contrasts, let α be the overall significance level, then for Dunnett method, the simultaneous $100 * (1-\alpha)\%$ confidence interval of $\tau_u - \tau_s$ from one-way analysis of variance model of a completely randomized deign **with the equal sample size** is given by:

$$\bar{y}_{u\cdot} - \bar{y}_{s\cdot} \pm |t|^{(0.5)}_{v-1,n-v,\alpha} * \sqrt{msE(\frac{1}{r_u} + \frac{1}{r_s})}$$

Where the critical value, $w_{D2} = |t|^{(0.5)}_{v-1,n-v,\alpha}$. Here, $|t|^{(0.5)}_{v-1,n-v,\alpha}$is the upper critical value for the maximum of the absolute values of a multivariate t-distribution with correlation 0.5 and $n-v$ error degrees of freedom, corresponding to the chosen value of α in the right-tail. Some values of $|t|^{(0.5)}_{v-1,n-v,\alpha}$can be found in Table A.10 but we will use R to find this critical value.

The deviation of Tukey method is based on Studentized range distribution. Tukey method is also called **Tukey HSD (Honest Significant Difference)** Test. When the sample sizes are equal ($r = r_1 = \cdots = r_v$), the overall confidence level from Tukey method is ***exactly* 100*(1 − α)*%**. When the sample sizes are unequal, the confidence level is ***at least*** $100(1-\alpha)\%$.

Example 3. 3 Continued – Simultaneous 95% Confidence Interval of the contrast $\tau_7 - \tau_3$ with Dunnett method

Solution: he critical value w_T is

$$w_{D2} = |t|^{(0.5)}_{v-1,n-v,\alpha} = |t|^{(0.5)}_{7-1,35-7,0.05} = 2.7314$$

So the 95% confidence interval of $\tau_7 - \tau_3$ is:

$$24.86 \pm 2.7314 * 2.1037 = (19.11, 30.61)$$

Here we can see the simultaneous 95% confidence interval of $\tau_7 - \tau_3$ from Dunnett method is shortest.

Example 3.4: Methods for Multiple Comparison – Shrimp Experiment

Suppose that before the experiment, we decide to do all pairwise comparisons, and a comparison between the standard diets and the experiments diets. Discuss what methods of multiple comparisons can used and use them to construct the simultaneous 95% confidence interval of the

contrast for the comparison between the standard diets and the experiments diets. Also calculate its 95% confidence interval without any method of multiple comparisons.

Solution: As we have discussed before, only Bonferroni method and Scheffé method can be used. The contrast coefficients of the contrast for the comparison between the standard diets and the experiments diets are:

$$\left[-\frac{1}{4},-\frac{1}{4},-\frac{1}{4},-\frac{1}{4},\frac{1}{3},\frac{1}{3},\frac{1}{3}\right]=\frac{1}{12}[-3,-3,-3,-3,4,4,4]$$

With the following summary statistics:

$$n=35, v=7, r=r_1=\cdots=r_7=5, msE=\hat{\sigma}^2=11.064$$

$$\bar{y}_{1\cdot}=48.04, \bar{y}_{2\cdot}=38.04, \bar{y}_{3\cdot}=55.20, \bar{y}_{4\cdot}=54.06, \bar{y}_{5\cdot}=40.54, \bar{y}_{6\cdot}=46.84, \bar{y}_{7\cdot}=80.06$$

We have the point estimate of:

$$\frac{1}{12}(-3\tau_1-3\tau_2-3\tau_3-3\tau_4+4\tau_5+4\tau_6+4\tau_7)$$

$$=\frac{1}{12}(-3\bar{y}_{1\cdot}-3\bar{y}_{2\cdot}-3\bar{y}_{3\cdot}-3\bar{y}_{4\cdot}+4\bar{y}_{5\cdot}+4\bar{y}_{6\cdot}+4\bar{y}_{7\cdot})$$

$$=\frac{1}{12}(-3*48.04-3*38.04-3*55.20-3*54.06+4*40.54+4*46.84+4*80.06)$$

$$=6.978$$

The estimated standard error is:

$$\sqrt{mse*\sum_{i=1}^{v}\frac{c_i^2}{r_i}}=\sqrt{11.064*\frac{4*(-3)^2+3*4^2}{5*12^2}}=1.1361$$

95% confidence interval without methods of multiple comparisons:

The critical value w is

$$w=t_{n-v,\frac{\alpha}{2}}=t_{35-7,\frac{0.05}{2}}=2.0484$$

So the 95% confidence interval is:

$$6.978\pm 2.0484*1.1361=(4.651,9.306)$$

Simultaneous 95% confidence interval with Bonferroni method:

Note that we have $m=\frac{7*6}{2}+1=22$. So the critical value w_B is

$$w_B = t_{n-v,\frac{\alpha}{2*m}} = t_{35-7,\frac{0.05}{2*22}} = 3.3585$$

So the 95% confidence interval is:

$$6.978 \pm 3.3585 * 1.1361 = (3.1626, 10.794)$$

The unadjusted p-value is:

$$2 * \Pr\left(T_{35-7} > \left|\frac{6.978 - 0}{1.1361}\right|\right) = 1.250 * 10^{-6}$$

The Bonferroni adjusted p-value is:

$$22 * 1.250 * 10^{-6} = 2.750 * 10^{-05}$$

Simultaneous 95% confidence interval with Scheffé method:

The critical value w_S is

$$w_S = \sqrt{(v-1)F_{v-1,n-v,\alpha}} = \sqrt{(7-1)F_{7-1,35-7,0.05}} = 3.8303$$

So the 95% confidence interval of $\tau_7 - \tau_3$ is:

$$6.978 \pm 3.8303 * 1.1361 = (2.627, 11.330)$$

Conclusion:

The confidence interval from Scheffé method is still wider that it from Bonferroni method. Comparing with our previous example, the critical value of Bonferroni method is changed since the number of contrasts is changed. While the critical value of Scheffé method remain unchanged.

Methods of Multiple Comparisons Using R [A video may be added here]

R provides functions that can be used to calculate point estimate, confidence interval, and hypothesis testing of multiple contrast with multiple contrasts with several methods of multiple comparisons including Bonferroni method, Scheffé method, Tukey method, and Dunnett method. Please refer to the R program for details.

A Special Note for Using Scheffé Method and Tukey Method with R Function lsmeans

I would like to bring up some issues about how to obtain correct simultaneous confidence intervals of a set of contrasts with Scheffé method and Tukey method with R function lsmeans. Due to the

way of the implementation of these methods in R, we need to pay additional attentions when we use Scheffé method or Tukey method with R function lsmeans. Let us start with the following example.

Example 3.5: Pairwise Comparison – Shrimp Experiment

Suppose that before the experiment, we decide to do all pairwise comparisons between an experimental diet and diet 7. The first four diets are the experimental diets and the last three diets are the standard diets. We can use Bonferroni method, Scheffe method, and Tukey method to obtain 95% simultaneous confidence intervals for four contrasts of interest. Here, we focus on the CIs from Scheffe method and Tukey method. From **Example 3.4**, we can get the critical value from Scheffe method is:

$$w_s = \sqrt{(7-1)F_{7-1,28,0.05}} = 3.8303$$

The critical value from Tukey method is:

$$w_s = \frac{q_{7,28,0.05}}{\sqrt{2}} = 3.1722$$

The 95% simultaneous confidence intervals for four contrasts of interest can be obtained correspondingly:

Contrast	Least Squares Estimate	Standard Error	95% CI - Scheffe	95% CI – Tukey
$\tau_1 - \tau_7$	-32.02	2.10484	(-40.078, -23.962)	(-38.693, -25.347)
$\tau_2 - \tau_7$	-42.02	2.10484	(-50.078, -33.962)	(-48.693, -35.347)
$\tau_3 - \tau_7$	-24.86	2.10484	(-32.918, -16.802)	(-31.533, -18.187)
$\tau_4 - \tau_7$	-26.00	2.10484	(-34.058, -17.942)	(-32.673, -19.327)

If we would like to use R functions, aov, lsmeans, contrast, and summary to obtain the same results, we need to be careful due to its "**smart**" implementation. First, we can get the data, fit the model, and use lsmeans function:

```
#------------------------------------------------------------
# use shrimp data
```

```
shrimp <- read.table(file = "shrimp.txt", header = TRUE, stringsAsFactors = FALSE)
# analysis of variance model
sh.lm <- aov(weight ~ factor(diet), data = shrimp)
sh.lm,diet <- lsmeans(sh.lm, ~ diet)
```

Since all four contrasts of interest are the pairwise comparison, the correct way is to use method = "pairwise" in summary function:

```
summary(contrast(sh.lm.diet, method = "pairwise", adjust = "scheffe"), infer = c(T, T))
summary(contrast(sh.lm.diet, method = "pairwise", adjust = "tukey"), infer = c(T, T))
```

You can verify that the results are same as the results from the above table. Also, we have the following output from R for Scheffe method:

```
Conf-level adjustment: scheffe method with rank 6
P value adjustment: scheffe method with rank 6
```

Since there are 7 treatment levels, so the rank (dimensionality) of Scheffe method should be $7 - 1 = 6$. The output from R indicates that the adjustment from Scheffe method is correct.

We also have the following output from R for Tukey method:

```
Conf-level adjustment: tukey method for comparing a family of 7 estimates
P value adjustment: tukey method for comparing a family of 7 estimates
```

Again, this output indicates that the adjustment from Tukey method is correct.

Since we are only interested in four contrasts, you may just want to obtain the 95% simultaneous CI for those four contrasts of interest and use the following R code:

```
# statistical inference for contrasts with Scheffe method
my.con <- list(c17 = c(1, 0, 0, 0, 0, 0, -1), c27 = c(0, 1, 0, 0, 0, 0, -1),
                    c37 = c(0, 0, 1, 0, 0, 0, -1), c47 = c(0, 0, 0, 1, 0, 0, -1))
summary(contrast(sh.lm.diet, method = my.con, adjust = "scheffe"), infer = c(T, T))
summary(contrast(sh.lm.diet, method = my.con, adjust = "tukey"), infer = c(T, T))
```

However, the different results are obtained this time. For example, the 95% simultaneous CI for $\tau_1 - \tau_7$ from Scheffe method is (-39.0, -25.1) which is different from (-40.1, -24.0). This is because that it uses $w_s = \sqrt{4F_{4,28,0.05}}$ instead of $w_s = \sqrt{(7-1)F_{7-1,28,0.05}}$. From the output, we

can see that the **rank 4** is used. According to our course, we should use the **rank $6 = 7 - 1$**. Note that this is from the newest version of lsmeans package while the rank 6 is still used in older version of lsmeans packages. I have contacted the developer of the package about this problem, who stated: "Your old lsmeans must be VERY old, because I changed this some time ago, from a **"textbook" calculation of rank (dimensionality)** to **an actual rank determination**. In its underlying theory, Scheffe method is based on the maximum possible t-test statistic among all linear combinations of contrasts under test, and its rank is the rank of the linear space containing all those underlying tests. If you were testing all possible contrasts among the 7 means, that rank is indeed $7 - 1 = 6$. But you specified only 4 contrasts, and so the rank cannot exceed 4. If these 4 contrasts are the only ones you are interested in, then the Scheffe correction with rank 6 is way, way conservative and the results you have are preferable. If you really do want to view these 4 contrasts as just 4 among a vast number of others, then **put more contrasts in the mix so that the dimensionality comes out 6**; and then just ignore the results for all but those four. "

Basically, **the "textbook" calculation of rank is $v - 1$ (in our example, $v = 7$), while the actual rank is based on the set of contrasts. In our example here, the actual rank is 4 and that is why 4 is used**. I only partially agree with his statement. For our course, we should stick to the **"textbook" calculation of rank which is always $v - 1$**.

For Tukey method, the results are also different from the results from our manual calculation. The 95% simultaneous confidence interval of $\tau_1 - \tau_7$ is $(-37.6, -26.4)$, which is different from our manual calculation: $(-38.7, -25.3)$. This is because that "**sidak method for 4 estimates**" is used. Sidak method is a method that is similar to Bonferroni.

In summary, if we would like to use R functions such as lsmeans and contrast to calculate the simultaneous confidence intervals of a set of contrasts with Scheffe method or Tukey method, we should:

- For Tukey method, always **use method = "pairwise"** option in R function contrast.
- For Scheffe method with only pairwise comparisons, use **method = "pairwise"** option in R function summary.
- For Scheffe method with other types of contrasts, put **more contrasts in the mix so that the actual rank is $v - 1$**. The simplest way is use emmeans:::consec.emmc(1:v) to add a set of pairwise contrasts with the rand $v - 1$ to your contrasts. Note that

emmeans:::consec.emmc(1:v) will add $v-1$ contrasts $\tau_2-\tau_1, \tau_3-\tau_2, \cdots, \tau_v-\tau_{v-1}$ and this set of $v-1$ contrasts has he rank $v-1$. For example, we can use the following R code:

```
my.con <- list(c17 = c(1, 0, 0, 0, 0, 0, -1), c27 = c(0, 1, 0, 0, 0, 0, -1),
        c37 = c(0, 0, 1, 0, 0, 0, -1), c37 = c(0, 0, 0, 1, 0, 0, -1))
my.con.mix <- c(my.con, emmeans:::consec.emmc(1:7))

summary(contrast(sh.lm.diet, method = my.con.mix, adjust = "scheffe"), infer = c(T, T))
```

- For Scheffe method with other types of contrasts, the best and simplest way is to use sceffe.rank implemented in R function summary to explicitly specify the correct rank. For example, we can use the following R code:

```
my.con <- list(c17 = c(1, 0, 0, 0, 0, 0, -1), c27 = c(0, 1, 0, 0, 0, 0, -1),

        c37 = c(0, 0, 1, 0, 0, 0, -1), c37 = c(0, 0, 0, 1, 0, 0, -1))

summary(contrast(sh.lm.diet, method = my.con, adjust = "scheffe"), scheffe.rank = 6, infer = c(T, T))
```

Sample Size Calculation for Desired Length of Confidence Interval

Before an experiment can be run, it is necessary to determine the number of observations that should be taken on each level of each treatment factor. In Lesson 2.2, we presented a method to calculate the power/or sample size of the test of

$$H_0: \tau_1 = \tau_2 = \cdots = \tau_v$$

from experiments of one treatment factor with v levels. In this subsection, we show how to determine the sample sizes to achieve confidence intervals of specified lengths.

The lengths of confidence intervals decrease as sample sizes increase. Consequently, if the length of an interval is specified, it should be possible to calculate the required sample sizes, especially when these are equal. As we have mentioned before, there is a problem. Since the experimental data have not yet been collected, the value of the mean squared error is not known. However, if the value of the mean squared error can be reasonably well be guessed at, either from previous experience or from a pilot study, then a trial and error approach to the problem can Conf-level adjustment: tukey method for comparing a family of 7 estimatesbe followed. For

illustration, we assume that the sample sizes are same ($r = r_1 = \cdots = r_v$) and only consider the pairwise contrasts here. For unequal size sample sizes and other types of contrasts, similar procedure can be used.

Recall that the formula for (simultaneous) confidence interval of a pairwise comparison

$$\bar{y}_{u\cdot} - \bar{y}_{s\cdot} \pm w * \sqrt{msE * \frac{2}{r}}$$

where w is the critical value and depends on the method of multiple comparisons used.

Note the length of confidence interval is $2 * w * \sqrt{msE * \frac{1}{r}}$ and is same for all pairwise contrasts if the sample sizes are equal. Therefore, the following steps are suggested to calculate the sample size needed for a given threshold of the length of confidence interval l:

Step 0: Set $r = 2$.

Step 1: Calculate corresponding w and the length of confidence interval

$$2 * w * \sqrt{msE * \frac{2}{r}}$$

Step 2: If the length of confidence interval is less than l. Record r and this is the sample size required. Otherwise increase r by 1 and repeat Steps 1 and 2.

These procedures can be easily implemented in R. Since the calculation is quite intensive, we suggest to use R to do the calculation. Here we use the following example for the illustration.

Example 3.5: Sample Size Calculation for Desired Length of Confidence Interval - Bean-soaking Experiment

Suppose we were to plan an experiment to compare the effects of $v = 5$ different soaking times on the growth rate of mung bean seeds. The response variable will be the length of a shoot of a mung bean seed 48 hours after soaking. Suppose that a pilot experiment has indicated that the mean square for error is likely to be not more than 10 mm^2, and suppose that we would like a set of 95% simultaneous confidence intervals for pairwise differences of the soaking times, with

each interval no wider than 6mm (that is, the half width or minimum significant difference should be no greater than 3mm). Find the sample sizes required.

Solution: The required sample sizes also depend on the method of multiple comparisons used. Since we plan to perform all pairwise comparison, we will use the best method of multiple comparison, Tukey method. With $v = 5$, the estimated msE of 10, and the sample size of r for each treatment level, the simultaneous 95% confidence interval of a pairwise contrast $\tau_u - \tau_s$ is:

$$\bar{y}_{u\cdot} - \bar{y}_{s\cdot} \pm \frac{q_{v,n-v,\alpha}}{\sqrt{2}} * \sqrt{msE * \left(\frac{1}{r_u} + \frac{1}{r_s}\right)} = \bar{y}_{u\cdot} - \bar{y}_{s\cdot} \pm \frac{q_{5,5*r-5,0.95}}{\sqrt{2}} * \sqrt{10 * \frac{2}{r}}$$

The half length of such confidence interval is: $\frac{q_{5,5*r-5,0.95}}{\sqrt{2}} * \sqrt{10 * \frac{2}{r}}$

We need to find r such that

$$\frac{q_{5,5*r-5,0.95}}{\sqrt{2}} * \sqrt{10 * \frac{2}{r}} \leq 3$$

Or equivalently,

$$q^2_{5,5*r-5,0.95} \leq 0.9 * r$$

r	$5 * r - 5$	$\frac{q_{5,5*r-5,0.95}}{\sqrt{2}} * \sqrt{10 * \frac{2}{r}}$	$q^2_{5,5*r-5,0.95}$	$0.9 * r$
15	70	3.233	15.682	13.5
16	75	3.124	15.627	14.4
17	**80**	**3.027**	**15.579**	**15.3**
18	**85**	**2.938**	**15.537**	**16.2**
19	90	2.856	15.500	17.1
20	95	2.781	15.466	18.0

The following table shows values of $\frac{q_{5,5*r-5,0.95}}{\sqrt{2}} * \sqrt{10 * \frac{2}{r}}$ for some r calculated using R. From the table, we can see if $r = 17$ observations are taken on each of the five soaking times, and if the mean square for error is approximately 10 mm^2 in the main experiment, then the half length of 95% Tukey simultaneous confidence intervals for pairwise comparisons is which is little over 3mm. If $r = 18$ observations are taken, the half length is litter shorter than the 3mm required.

If the cost of the experiment is high, then $r = 17$ can be selected; otherwise, $r = 18$ can be selected.

Exercise Problem 3.2: Sample Size Calculation for Desired Length of Confidence Interval - Bean-Soaking Experiment

For Example 3.5, find the required sample size with Bonferroni method.

Lesson 3.3 Checking Model Assumptions

Required Reading: Sections 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, and 5.7 in Chapter 5.

Introduction

For deigns of one source of variance, we use the following one-way analysis of variance model:

$$Y_{it} = \mu + \tau_i + \epsilon_{it}, t = 1, \cdots, r_i; i = 1, \cdots, v$$

with the following assumptions:

- μ and τ_i $(i = 1, \cdots, v)$ are unknown population parameters
- $\epsilon_{it}(t = 1, \cdots, r_i; i = 1, \cdots, v)$ are independent
- $\epsilon_{it} \sim N(0, \sigma^2)$: normal distribution with mean 0 and variance σ^2

Once the data is collected, we need to check if these assumptions are satisfied. We first present a general strategy, including the order in which model assumptions should be checked. For checking model assumptions, we rely heavily on residual plots. We do so because while examination of residual plots is more subjective than would be testing for model lack-of-fit, the plots are often more informative about the nature of the problem, the consequences, and the corrective action.

Strategy for Checking Model Assumptions

A good strategy for checking the assumptions about the model is to use the following sequence of checks.

- Check the form of the model - are the mean responses for the treatments adequately described by $\mathrm{E}(Y_{it}) = \mu + \tau_i, t = 1, \cdots, r_i; i = 1, \cdots, v$
- Check for outliers - are there any unusual observations (outliers)?
- Check for independence - do the error variables ϵ_{it} appear to be independent?
- Check for constant variance - do the error variables ϵ_{it} have similar variances for each treatment?
- Check for normality - do the error variables ϵ_{it} appear to be a random sample from a normal distribution?

Residuals and Standardized Residuals

We rely on residuals for checking assumptions. From the model, we can have the following terms:

(1) $\epsilon_{it} = Y_{it} - (\mu + \tau_i)$

(2) $\hat{e}_{it} = Y_{it} - (\hat{\mu} + \hat{\tau}_i) = Y_{it} - \bar{Y}_{i\cdot}$

(3) $\hat{e}_{it} = y_{it} - (\hat{\mu} + \hat{\tau}_i) = y_{it} - \bar{y}_{i\cdot}$

The first term, which is already introduced, is called the error term. They are independent normal random variables with the mean 0 and variance σ^2.

The second term, is called the residual. It is an estimator of the error. Thus hey are normal random variables with mean 0. However, The residuals with each treatment level are not independent. With some calculation, we can verify that the variance of $\hat{e}_{it} = Y_{it} - (\hat{\mu} + \hat{\tau}_i) = Y_{it} - \bar{Y}_{i\cdot}$ is $\frac{r_i - 1}{r_i}$. Therefore, if the sample sizes are equal, the residuals do not have the same variance. The theory of the use of residuals for checking model assumptions is based on the distributions of the residuals.

The third term, is also called the residual. It is an estimate of the error. These are the values calculated from a given data and are used for checking model assumptions.

The residuals, $\hat{e}_{it} = y_{it} - \bar{y}_{i\cdot}$, which are the estimates of the errors, has the following properties:

(1) The sample mean of $\hat{e}_{it}$ is 0.

To show that the sample mean of $\hat{e}_{it}$ is 0, we just need to show that the sum of $\hat{e}_{it}$ is 0. Actually, we can show that the sum of residuals from each treatment level is 0. For the treatment level $i (i = 1,2, \cdots, v)$, we have

$$\sum_{t=1}^{r_i} \hat{e}_{it} = \sum_{t=1}^{r_i} (y_{it} - \bar{y}_{i\cdot}) = \sum_{t=1}^{r_i} y_{it} - \sum_{t=1}^{r_i} \bar{y}_{i\cdot}$$

$$= r_i * \frac{\sum_{t=1}^{r_i} y_{it}}{r_i} - \bar{y}_{i\cdot} \sum_{t=1}^{r_i} 1 = r_i * \bar{y}_{i\cdot} - r_i * \bar{y}_{i\cdot} = 0$$

(2) The sample variance of $\hat{e}_{it}$ is $\frac{ssE}{n-1}$. Since the sample mean of $\hat{e}_{it}$ is 0, the sample variance of $\hat{e}_{it}$ is

$$\frac{\sum_{i=1}^{v}\sum_{t=1}^{r_i}(\hat{e}_{it}-0)^2}{n-1}=\frac{\sum_{i=1}^{v}\sum_{t=1}^{r_i}(y_{it}-\bar{y}_{i\cdot})^2}{n-1}=\frac{ssE}{n-1}$$

Therefore, we can use the standardized residuals which is achieved by dividing the residuals by their standard deviation:

$$z_{it}=\frac{\hat{e}_{it}-0}{\sqrt{ssE/(n-1)}}=\frac{y_{it}-\hat{y}_{i\cdot}}{\sqrt{ssE/(n-1)}}$$

Therefore, the standardized residuals have the sample mean of 0 and sample variance of 1.

Note that the standardized residuals are different from the **Studentized residuals. Studentized residuals** are more preferred in many situations. However, there is little distinction between approaches using the standardized residuals and approached using studentized residuals for analysis of variance models for data that is balanced or nearly so. Therefore, we focus on the sue of the standardized residuals throughout of this course.

Residual Plots

A ***residual plot*** is a scatter plot of the standardized residuals z_{it}(on the y-axis) versus another variable (on the x-axis), the choice of which depends on the assumption being checked. The following figure shows a plot of the standardized residuals versus the levels of the treatment factor for the **trout experiment**. Plots like this are useful for evaluating the assumption of constant error variance as well as the adequacy of the model.

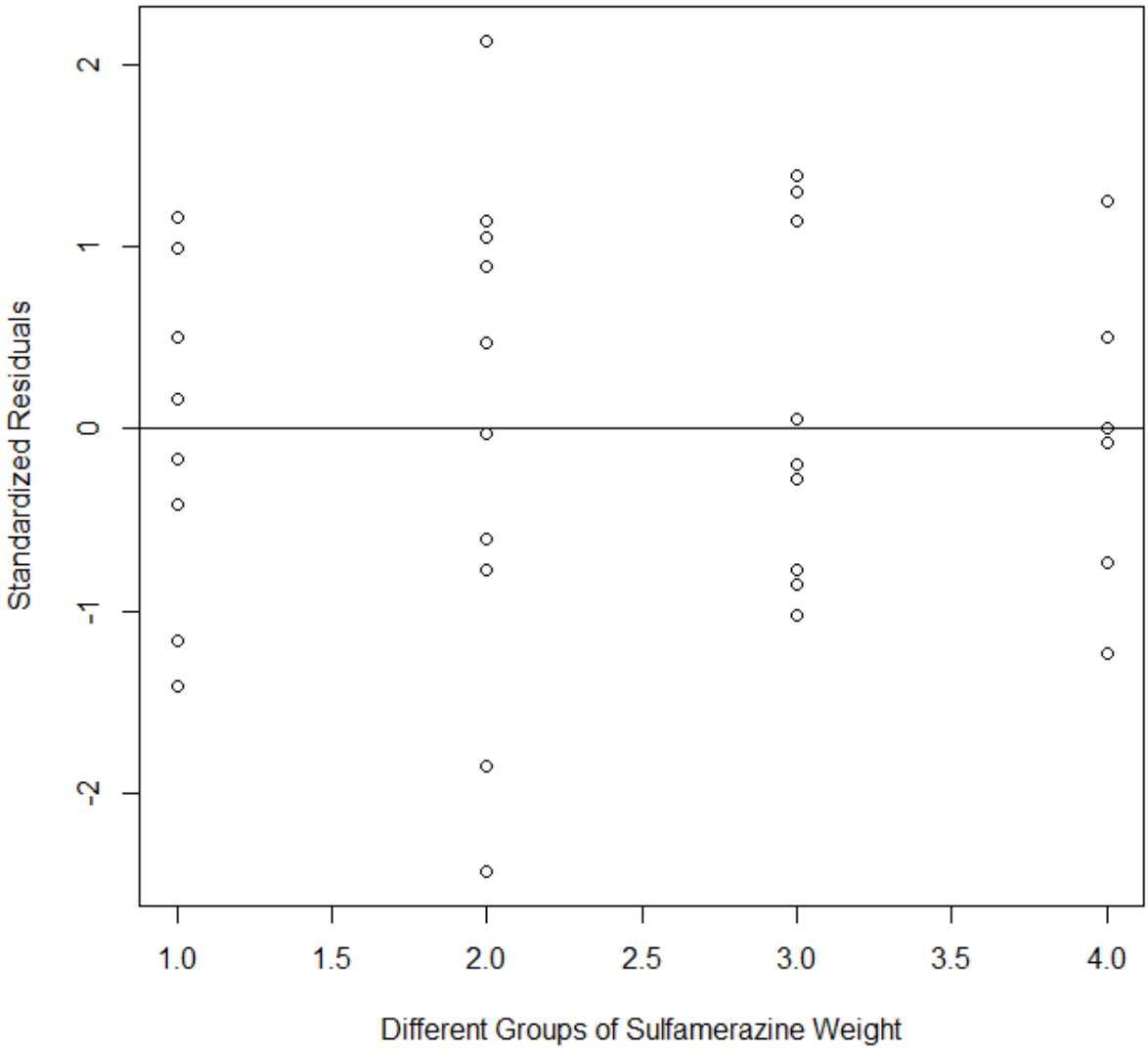
Example 3.6: Residual Plot - Trout Experiment

The data show the measurements of hemoglobin (grams per 100 ml) in the blood of brown trout. The trout were placed at random in four different troughs. The fish food added to the troughs contained, respectively, 0, 5, 10, and 15 g of sulfamerazine per 100 pounds of fish (coded 1, 2, 3, 4). The measurements were made on ten randomly selected fish from each trough after 35 days.

First, we can get $ssE = 56.471$ from the analysis of variance model. To calculate a specific residual or standardized residual, for example, $\hat{e}_{23}$ and z_{23},we have:

$$\hat{e}_{23} = y_{23} - \hat{y}_{2\cdot} = 10.4 - 9.33 = 1.07$$

$$z_{23} = \frac{y_{23} - \hat{y}_{2\cdot}}{\sqrt{ssE/(n-1)}} = \frac{1.07}{\sqrt{56.471/39}} = 0.889$$



From this plot, we can see that the standardized residuals seem fairly well scattered around zero, although the spread of the residuals for treatment 2 seems a little larger than the spread for the other three treatments. This could be interpreted as a sign of unequal variances of the error

variables or that the data values having standardized residuals are outliers, or it could be attributed to chance variation.

Exercise Problem 3.3 (Battery Experiment) Residual Plot. For Battery experiment,

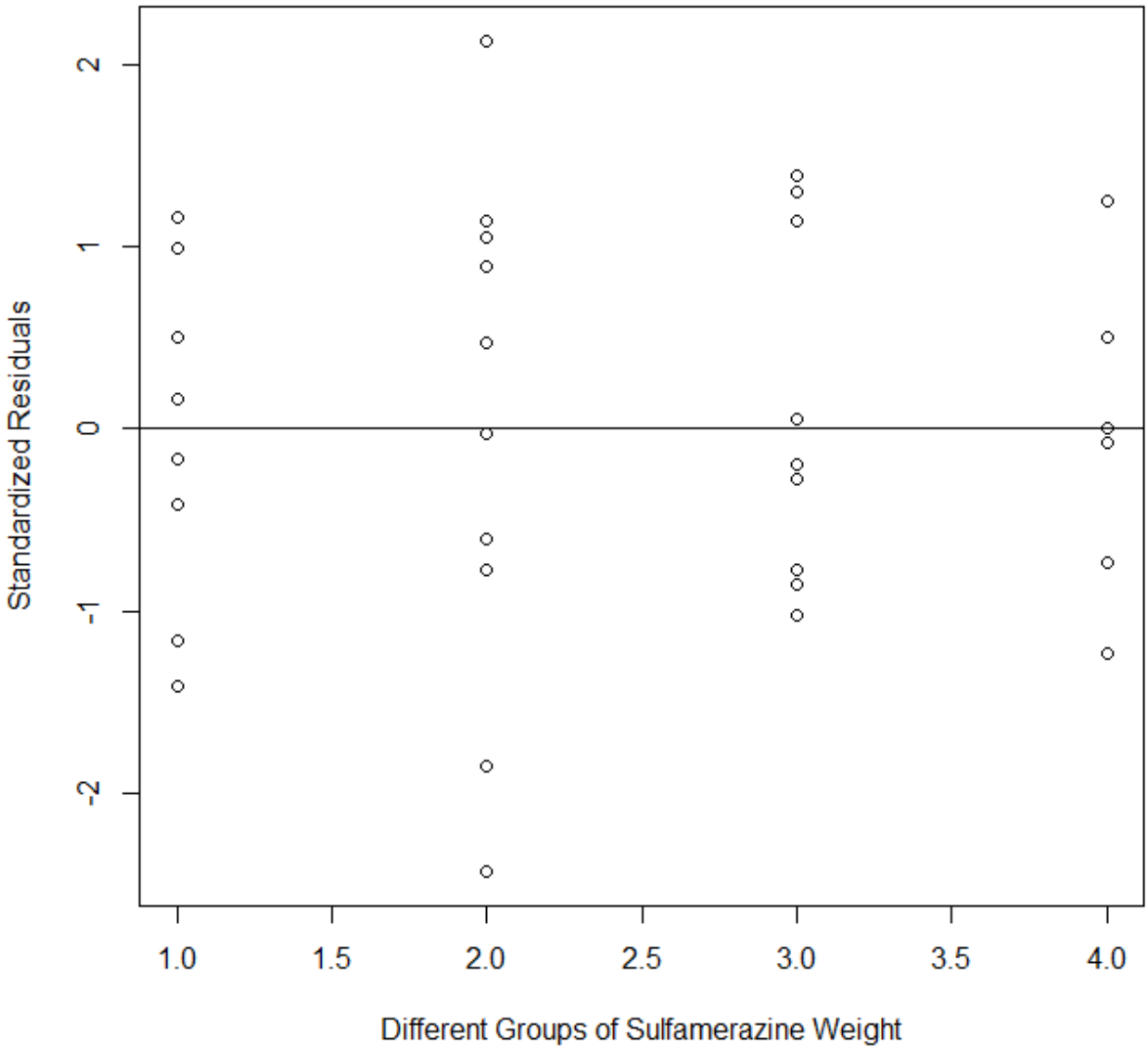
(1) Manually calculate all residuals from treatment level 2. Show that the sum of these residuals is 0.
(2) Draw a residual plot and use the standardized residuals as the y-axis and the treatment levels as the x-axis.

Checking the Fit of the Model

The first assumption to be checked is the assumption that the model $E(Y_{it}) = \mu + \tau_i$ for the mean response is correct. One purpose of running a pilot experiment is to choose a model that is a reasonable description of the data. If this is done, the model assumption checks for the main experiment should show no problems. If the model for mean response does not adequately fit the data, then there is said to be **model lack of fit**.

Checking the fit of the model for experiments with one treatment factor can be done by plotting the standardized residuals versus the treatment levels. Lack of fit is indicated if the residuals exhibit a nonrandom pattern about zero in any such plot, being too often positive for some levels of the independent variable and too often negative for others. Since the sum of residuals of each treatment level is zero, lack of fit would only be detected if there were a number of unusually large or small observations.

Example 3.7: Residual Plot for Checking Model Fit - Trout Experiment
For the following residual plot, we would see the one-way analysis of variance model is adequate to describe the data. For each treatment level, we did not observe an unusually large or small residuals.



Checking for Outliers

An ***outlier*** is an observation that is much larger or much smaller than expected. This is indicated by a residual that has an unusually large positive or negative value. Outliers are fairly easy to detect from a plot of the standardized residuals versus the levels of the treatment factors. Any outlier should be investigated. Sometimes such investigation will reveal an error in recording the data, and this can be corrected. Otherwise, outliers may be due to the error variables not being normally distributed, or having different variances, or an incorrect specification of the model.

Checking outliers again can be done by plotting the standardized residuals versus the treatment levels and looking at $|z_{it}|$, the absolute value of standardized residuals. We need to pay attention on the observations whose absolute standardized residuals are large. An empirical rule is:

Susceptible outliers: if $2 < |z_{it}| < 3$

Outliers: if $|z_{it}| \geq 3$

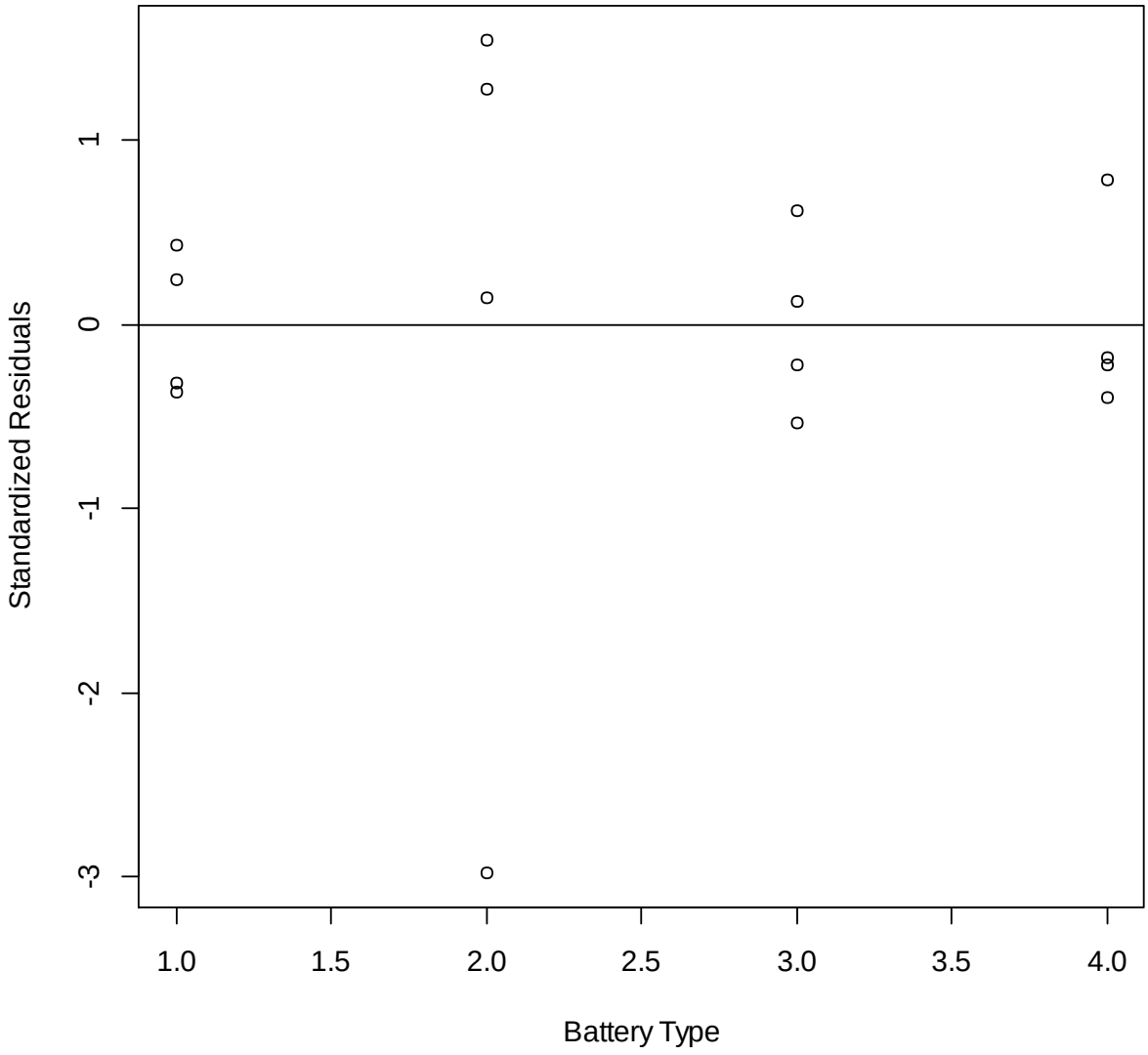
This is because that z_{it} approximately follows a standard normal distribution when all the assumptions of one-way analysis of variance model hold. **Therefore, approximately 68% of the standardized residuals should be between -1 and +1, approximately 95% between -2 and +2, and approximately 99.7% between -3 and +3.**

Note that the outliers cannot be automatically removed from the analysis. The flow charts describe how the outliers should be dealt with.

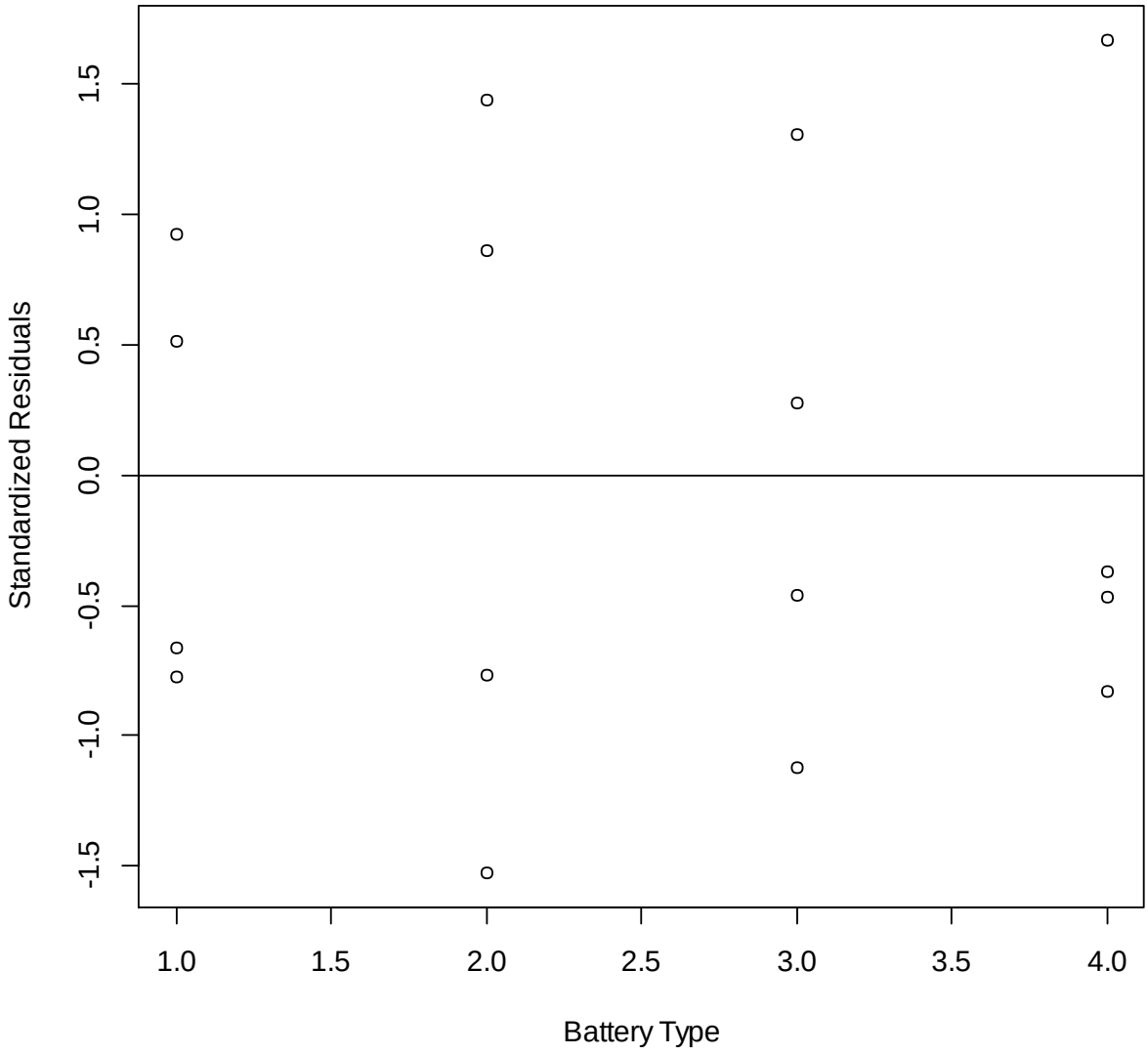
[See file week-3-outlier-flow-char.pptx for the flow chart]

Example 3.8: Residual Plot for Checking Outlies - Battery Experiment

In the battery experiment, four observations on battery life per unit cost were collected for each of four battery types. The data as originally entered into the computer for analysis. The original data file is battery-error.txt. From the scatter plot of the standardized residuals versus the treatment levels based on the data from file battery-error.txt shows two related anomalies. First, there is one apparent outlier for battery type 2, the residual value being -2.98. Second, *all* of the standardized residuals for the other three battery types are less than one in magnitude. This is many more than the 68% expected.



An investigation of the outlier revealed a data entry error. For the observation with the standardized residual of -2.98, the value of 473 was typed, but the recording sheet for the experiment showed the correct value to be 773 minutes. After correcting the error, the model was fitted again and the standardized residuals were replotted. From the plot we can see that two problems from the data with error entry disappear.

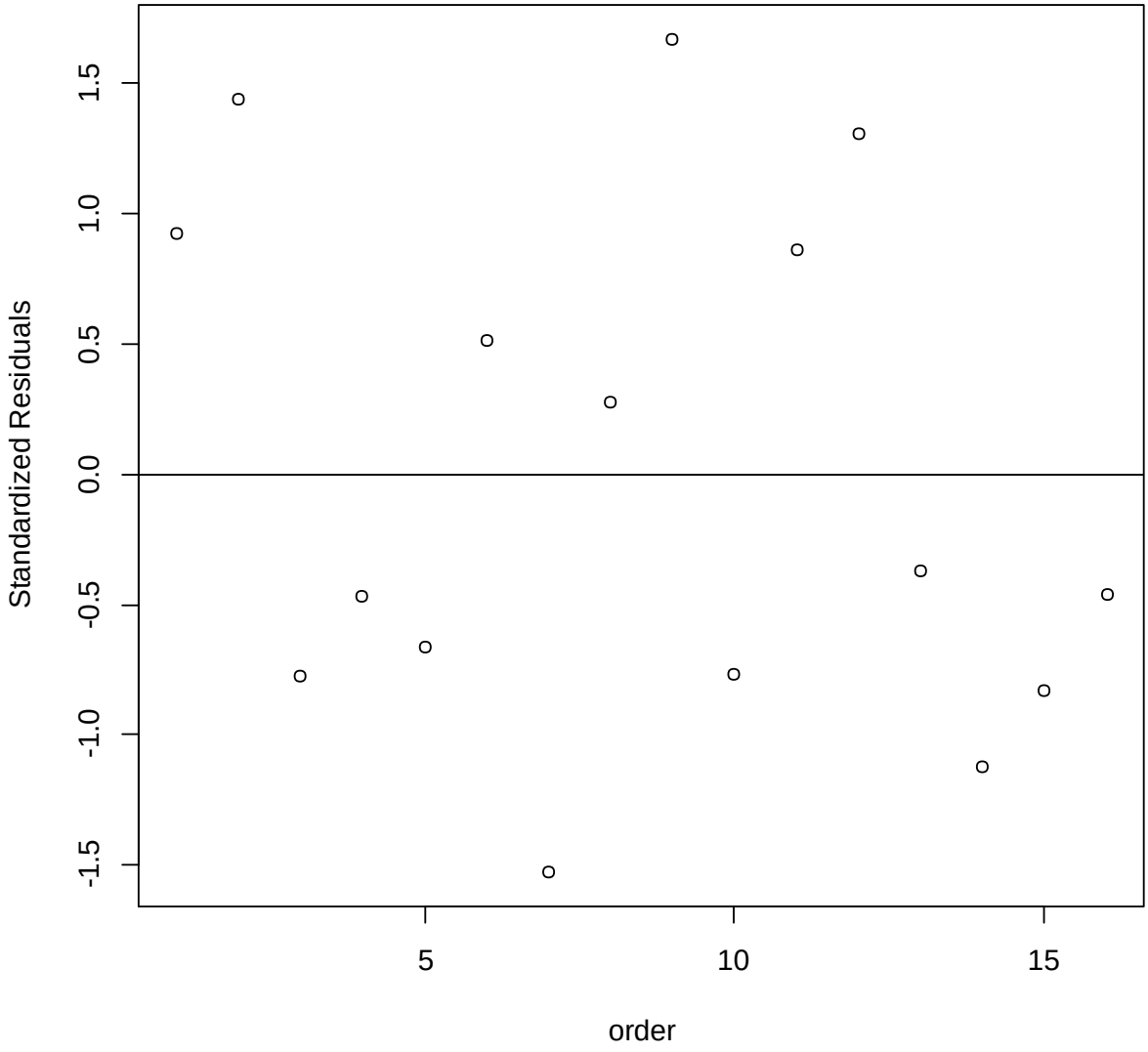


Checking Independence of Error

Checking independence of error is done for plotting the standardized residuals against the order in which the corresponding observations were collected and against any spatial arrangement of the corresponding experimental units. If the independence assumption is satisfied, the residuals should be randomly scattered around zero with no discernible pattern. If the plot were to exhibit a strong pattern, then this would indicate a serious violation of the independence assumption, as illustrated in the following example.

Example 3.9: Residual Plot for Checking Independence of Error - Battery Experiment

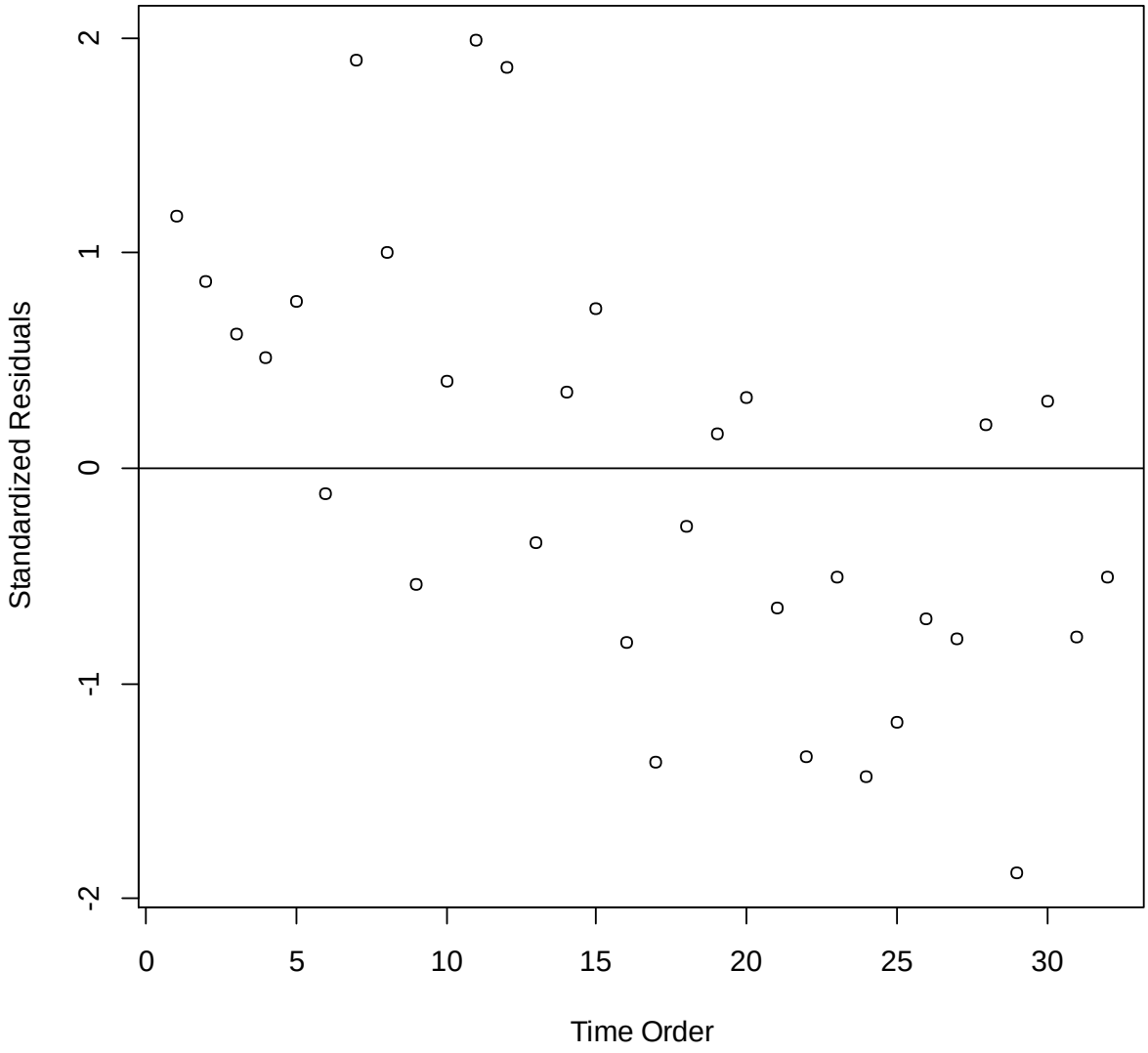
The following plot is the scatter plot of the standardized residuals against the order in which the corresponding observations were collected. It can be seen that the residuals are randomly scattered around zero with no discernible pattern. So we can conclude that the assumption of independence of error holds.



Example 3.10: Residual Plot for Checking Independence of Error - Balloon Experiment

Prior to 1985, the experimenter (Meily Lin) had observed that some colors of birthday balloons seem to be harder to inflate than others. She ran this experiment to determine whether balloons of different colors are similar in terms of the time taken for inflation to a diameter of 7 inches. Four colors were selected from a single manufacturer. An assistant blew up the balloons and the experimenter recorded the times (to the nearest 1/10 second) with a stop watch. The data, in the order collected, are given in the file **balloon.txt**, where the codes 1, 2, 3, 4 denote the colors pink, yellow, orange, blue, respectively. The scatter plot of the standardized residuals against the order in which the corresponding observations were collected is given below. It can be seen that there appears to be a strong downward drift in the residuals as time progresses.

So the observations are clearly dependent. The reason that the lack of independence is due to that the experimenter had used a single subject to blow up all the balloons in the experiment, and the subject had become an expert balloon blower before the experiment was finished!



If the there is a lack of independence, a more complicated (e.g., a model with a random effect) that can account such dependency can be used. However, this is not covered by this course.

Checking the Equal Variance Assumption

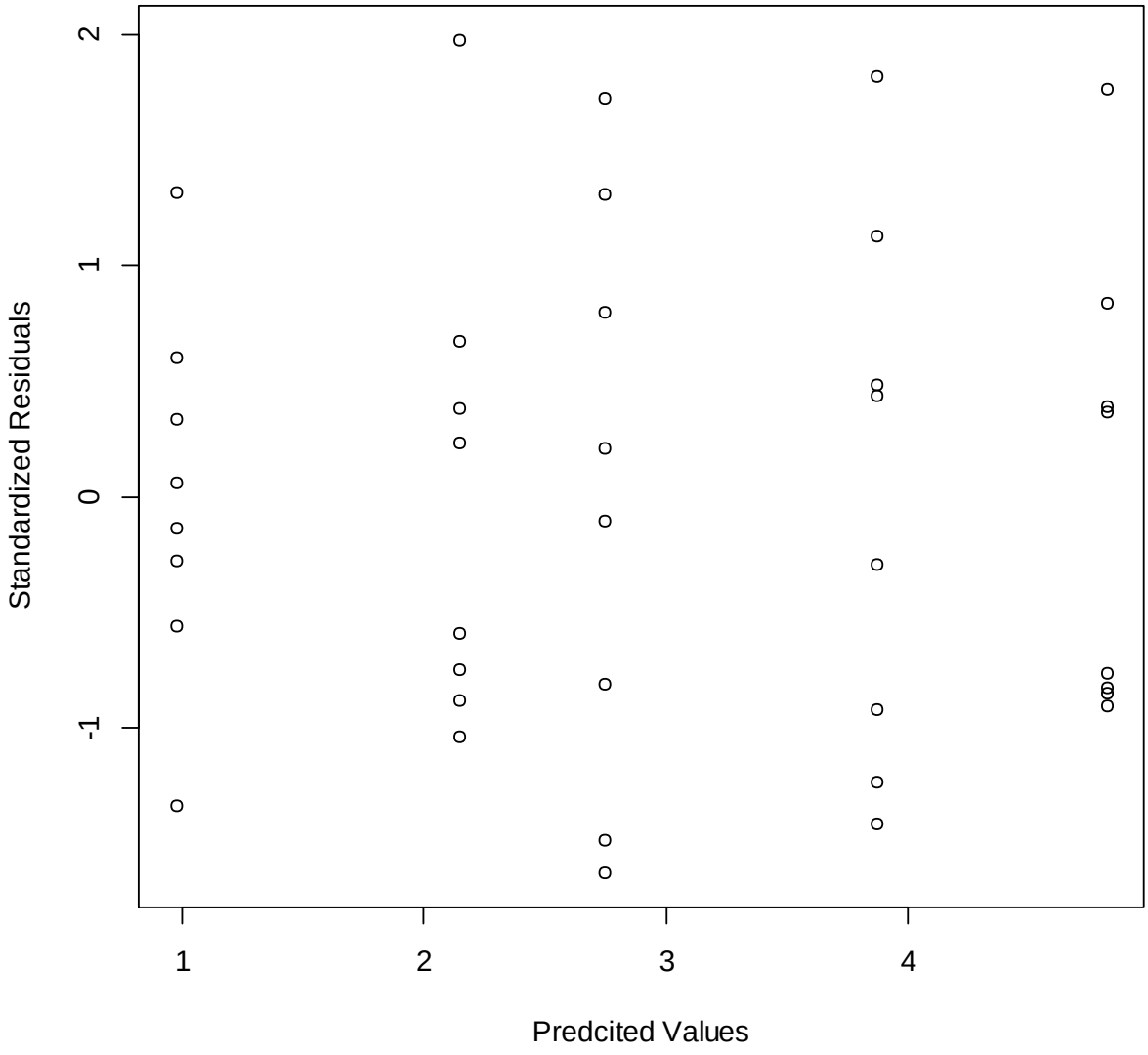
In this section, we will introduce three methods to check the equal variance assumption: (1) a residual plot; (2) an empirical procedure that compares the sample variances between groups; (3) Levene's test of equal variance assumption.

Checking the equal variance assumption can be done by plotting the standardized residuals against the predicted values. For the one-way analysis of variance model, the predicted value of y_{it} is just $\bar{y}_{i\cdot}$. In other words, we have

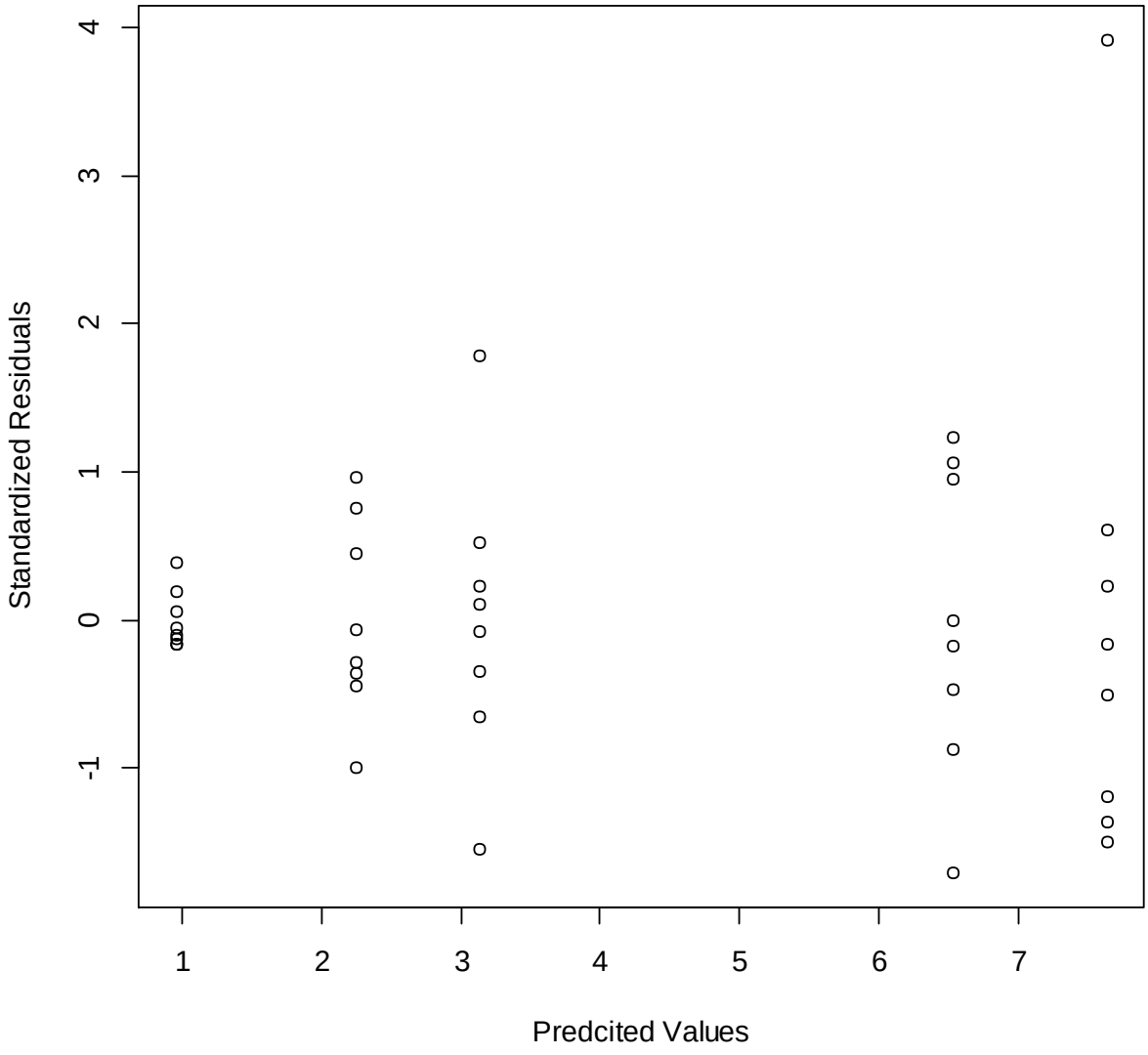
$$\hat{y}_{it} = \bar{y}_{i\cdot}$$

If the residuals are randomly scattered around zero and the variances of residuals are similar across different predicted values, the plot indicates that the equal variance assumption holds. If the plot shows some patterns such as megaphone (right fan), left fan, bulge in middle, then it indicates that the equal variance is violated.

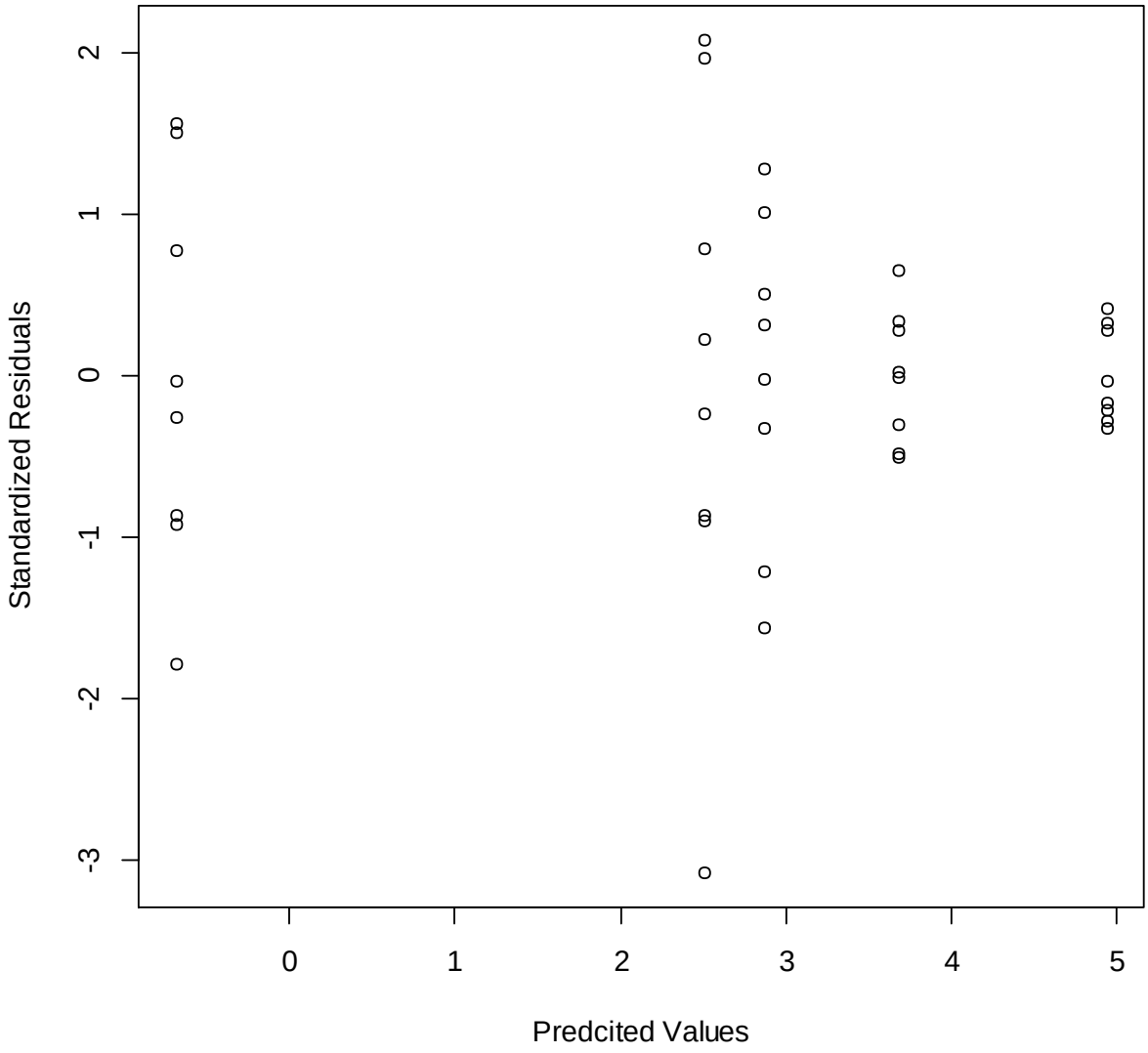
The following plots from simulated data show different patterns that validate or violate the equal variance assumption.



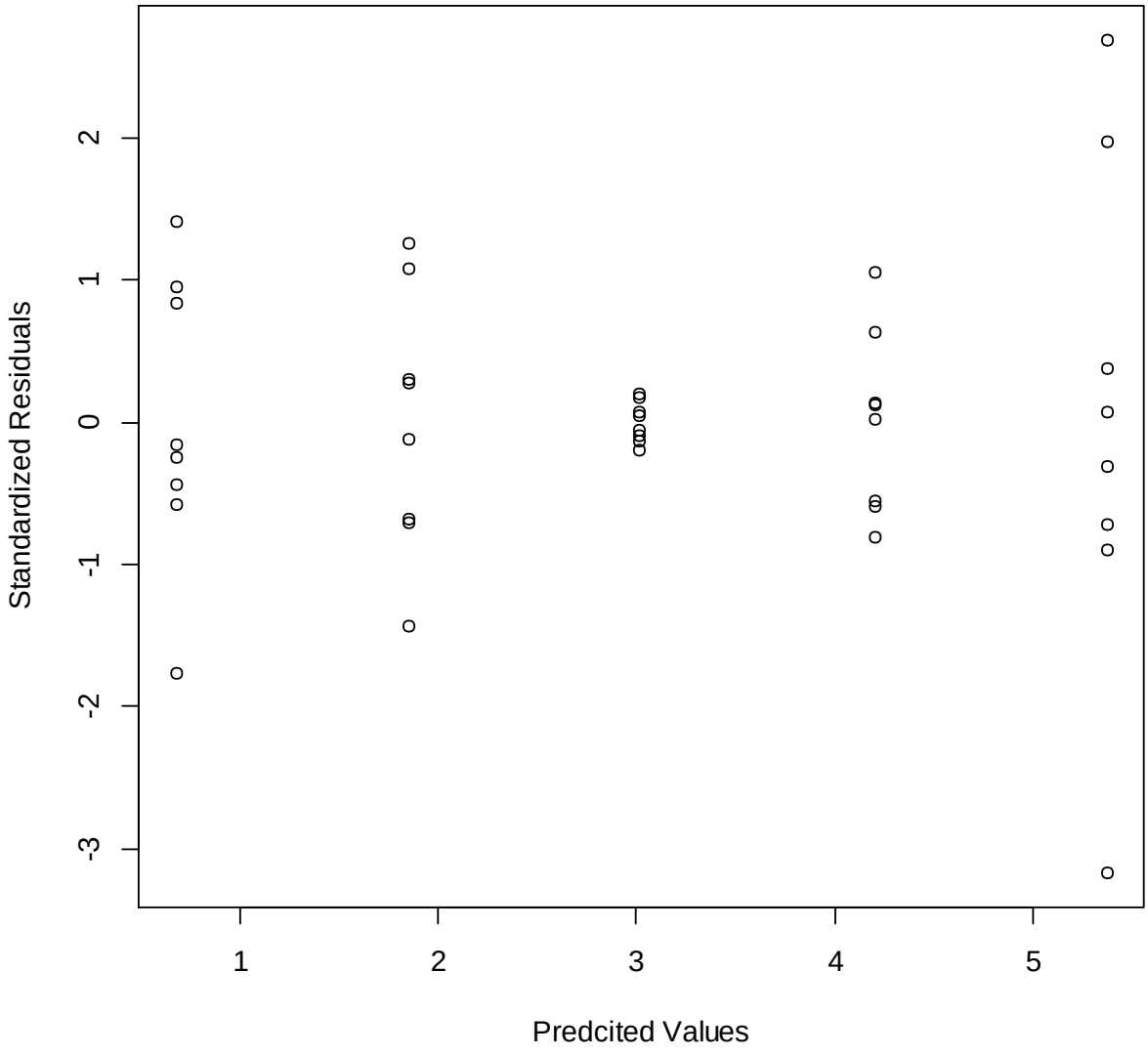
This Plot: the equal variance assumption is fine.



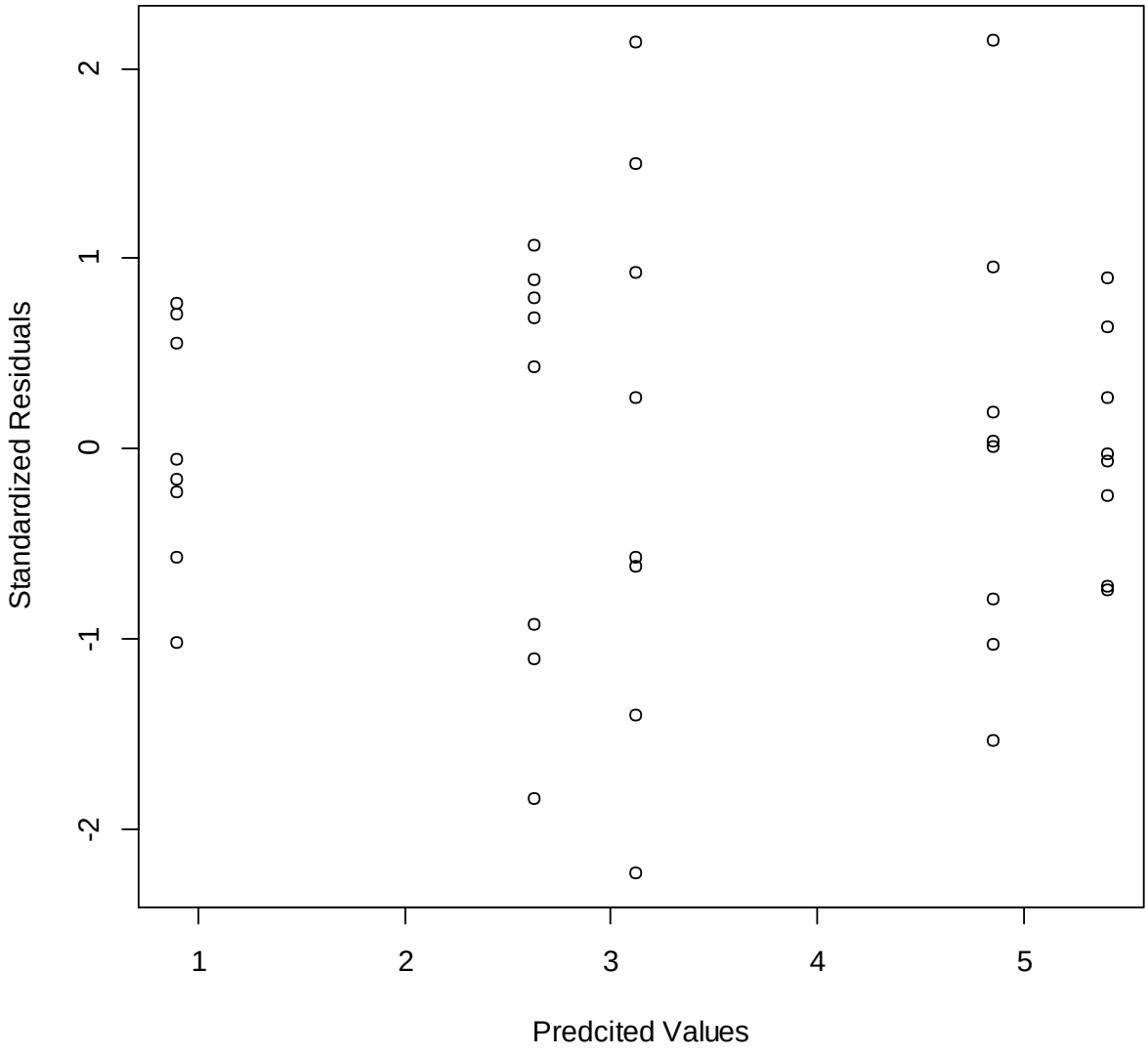
This plot: a megaphone (right fan) pattern that indicates the error variance increases as the mean response increases.



This plot: a left fan pattern that indicates the error variance decreases as the mean response increases.



This plot: a bulge in middle pattern that indicates the error variance first decreases then increases the mean response increases.

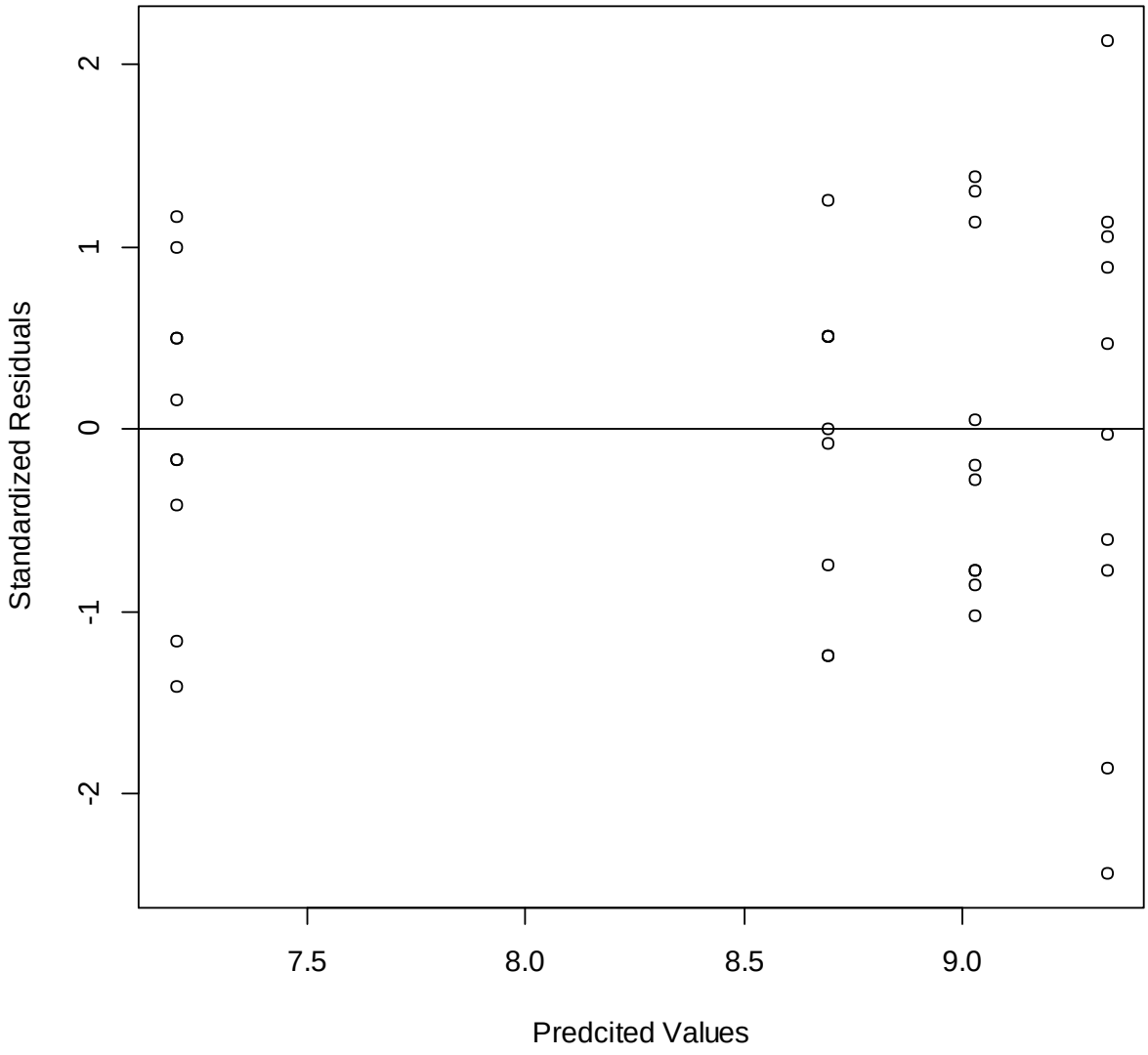


This plot: a bulge in middle pattern that indicates the error variance first increases then decreases as the mean response increases.

Example 3.11: Residual Plot for Checking Equal Variance Assumption - Trout Experiment

From the scatter plot of the standard residuals against the predicted values from Trout Experiment, we can see the error variance from the level 4 is greater than the error variances from the other levels. One explanation is that the equal variance assumption may be violated. Another explanation

is that we have a small sample size ($r = 10$ for each treatment level) so such fluctuation of error variances just occurred by chance.



The second method is based on the sample variances of treatment levels. The sample variance of ith treatment level,

$$s_i^2 = \frac{1}{r_i - 1}\sum_{t=1}^{r_i}(y_{it} - \bar{y}_{i\cdot})^2$$

Is unbiased estimator of error variance for the ith treatment level. If the equal variance assumption holds, then $s_i^2 (i = 1, \cdots, v)$ should be similar. Therefore, we can look at the ratio of of the maximum of $s_i^2 (i = 1, \cdots, v)$ and the minimum of $s_i^2 (i = 1, \cdots, v)$. A rule of thumb to declare that the equal variance assumption is violated is the ratio is greater than 3. However, this method tends to be conservation, meaning that it is possible, and perhaps even likely, for the ratio of $\frac{s_{max}^2}{s_{min}^2}$ to exceed three, even when the equal variance assumption holds.

Example 3.12: Testing Equal Variance Assumption - Trout Experiment

We can use R to calculate the sample mean sample variance for each treatment level. From the following table:

i	1	2	3	4
s_i^2	1.04	2.95	1.29	1.00

We have $\frac{s_{max}^2}{s_{min}^2} = \frac{2.95}{1.00} = 2.95$, which satisfies our rule of thumb, but only just.

A formal test called Levene's test can be used to test the equal variance assumption. The testing here is:

H_0: the error variances are equal versus H_A: the error variances are not equal

The test statistic is basically a F-test statistic from the following one-way analysis of variance model:

$$d_{it} = \mu + \tau_i + \epsilon_{it}, t = 1, \cdots, r_i; i = 1, \cdots, v$$

where $d_{it} = |y_{it} - \bar{y}_{it}| = |\hat{e}_{it}|$ is the absolute value of residuals.

Example 3.13: Testing Equal Variance Assumption - Trout Experiment

We can use R to calculate the absolute value of residuals and perform the ANOVA analysis. The ANOVA table is presented below:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	Statistics	*p*-value
Treatment	3	2.217	0.7389	1.635	0.198
Error	36	16.268	0.4519		
Total	39	18.485			

With the p-value of 0.198, we fail to reject the null hypothesis with the significance level of 0.05. In other words, we do not have sufficient evidence to suggest the violation of equal variance assumption.

Note Levene's test tend to have low power unless there are large numbers of observations on each treatment factor level. Therefore, the failure to reject the null hypothesis of equal variance assumption may just due to the low power of Levene's test.

Checking the Normality Assumption

Again, checking the normality assumption can be down by plotting the standardized residuals against their **normal scores**. Such plot is called normal probability plot or quantile-quantile (q-q) plot. ***Normal scores*** are percentiles of the standard normal distribution. If all assumptions of one-way ANOVA model hold, then the standardized residuals would look similar to n independent observations from the standard normal distribution. In particular, the qth smallest standardized residual would be approximately equal to the $100[q/(n+1)]$th percentile of the standard normal distribution. Consequently, when the model assumptions hold, a plot of the q th smallest standardized residual against the $100[q/(n+1)]$th percentile of the standard normal distribution for each $q = 1, 2, \ldots, n$ would show points roughly on a straight line through the origin with slope

equal to 1.0. However, if any of the model assumptions fail, and in particular if the normality assumption fails, then the normal probability plot shows a nonlinear pattern.

Many normal scores, which are only slightly different between them, are proposed. Here, we introduce the normal score recommended by Blom, in 1958. The corresponding normal score if the qh smallest standardized residual is the value ζ_p for which:

$$P(Z \leq \zeta_q) = \frac{q - 0.375}{n + 0.25} = 1 - (1 - \frac{q - 0.375}{n + 0.25})$$

where Z is a random variable with the standard normal distribution. Note the we have used z_α as:

$$P(Z > z_\alpha) = \alpha \text{ or equivalently, } P(Z \leq z_\alpha) = 1 - \alpha$$

Thus, we can re-write ζ_p as:

$$\zeta_q = z_{1-\frac{q-0.375}{n+0.25}}$$

In summary, we can draw a normal probability plot by the following procedure:

Step 1. Order the standardized residuals from the smallest to the largest.

Step 2. For the qth smallest standardized residual, calculate the corresponding normal score $\zeta_q = z_{1-\frac{q-0.375}{n+0.25}}$.

Step 3: Draw a scatter plot of the standardized residuals against normal scores.

If the normality assumption holds, the points are roughly on a straight line through the origin with slope equal to 1.0. However, if any of the model assumptions fail, and in particular if the normality assumption fails, then the normal probability plot shows a nonlinear pattern.

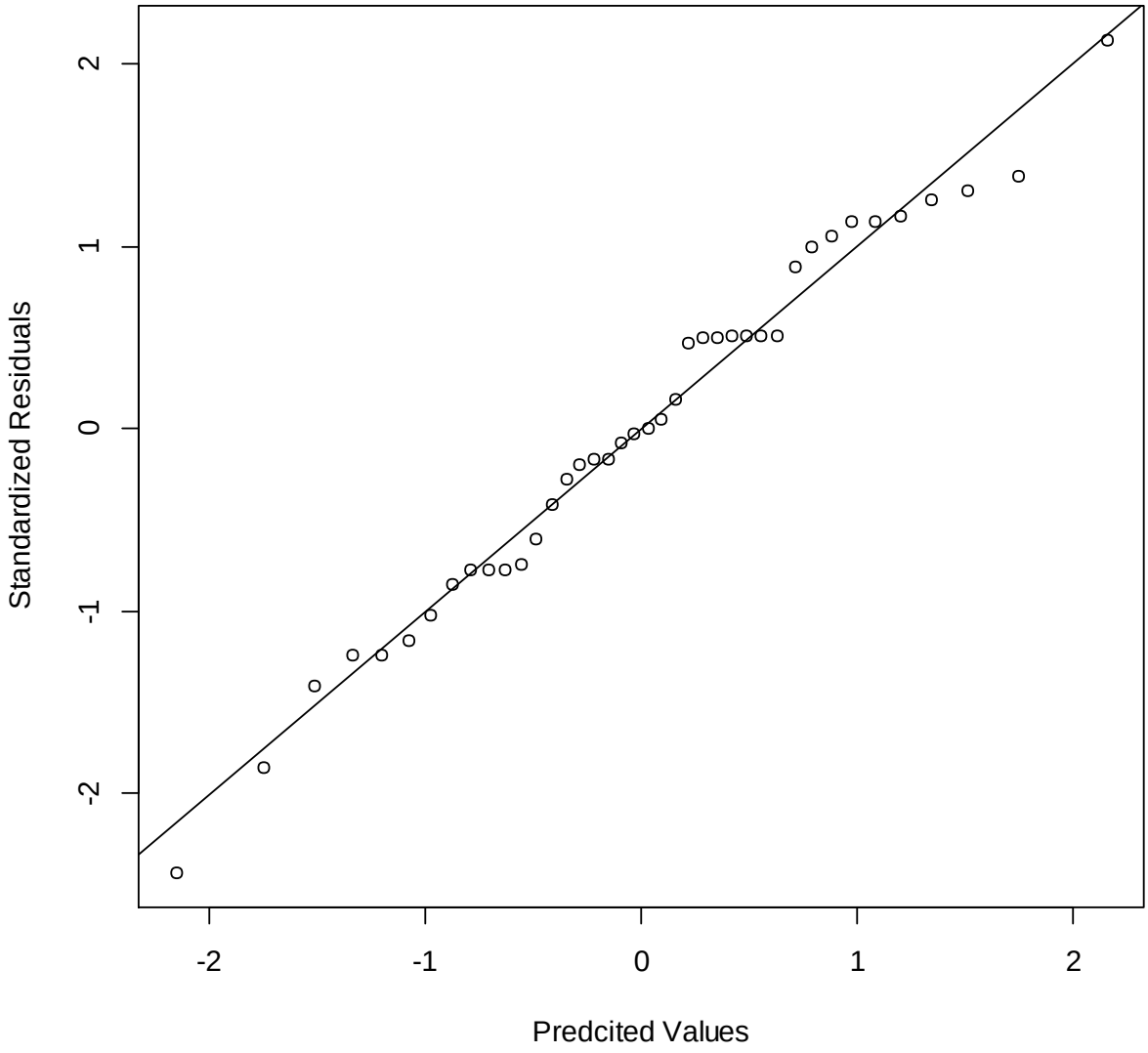
Example 3.14: Checking Normality Assumption - Trout Experiment

The following table show several standardized residuals and their normal scores. Note that the standardized residuals are ordered from the smallest to the largest.

z_{it}	q	$\frac{q - 0.375}{n + 0.25}$	$\zeta_q = z_{1-\frac{q-0.375}{n+0.25}}$

-2.4340	1	0.01552	-2.1564
…	…	…	…
-0.7729	11	0.2640	-0.6311
-0.7396	12	0.2888	-0.5568
…	…	…	…
2.1358	40	0.9845	2.1564

From the following normal probability (or Q-Q) plot, we can see that all points are close tor on the straight line with a slope of 1. Therefore, we conclude that the normality assumption holds.



Example 3.15: Checking Assumptions of One-Way ANOVA Model - Mug Bean Experiment

An experiment was run in 1993 by K.H. Chen, Y.F. Kuo, R. Sengupta, J. Xu, and L.L. Yu to compare watering schedules and growing mediums for mung bean seeds. There were 6 treatment levels. Forty-eight beans of approximately equal weights were randomly selected for the experiment. These were all soaked in water in a single container for two hours. After this time, the beans were placed in separate containers and randomly assigned to a treatment (water/medium) combination in such a way that eight containers were assigned to each treatment combination. The 48 containers were placed on a table in a random order. The shoot lengths of the beans were

measured (in mm) after one week. Use appropriate plots and tests to checking assumptions of one-way ANOVA model.

Solution: Please refer to R program to find how we performed such analysis.

Checking the Fit of the Model:

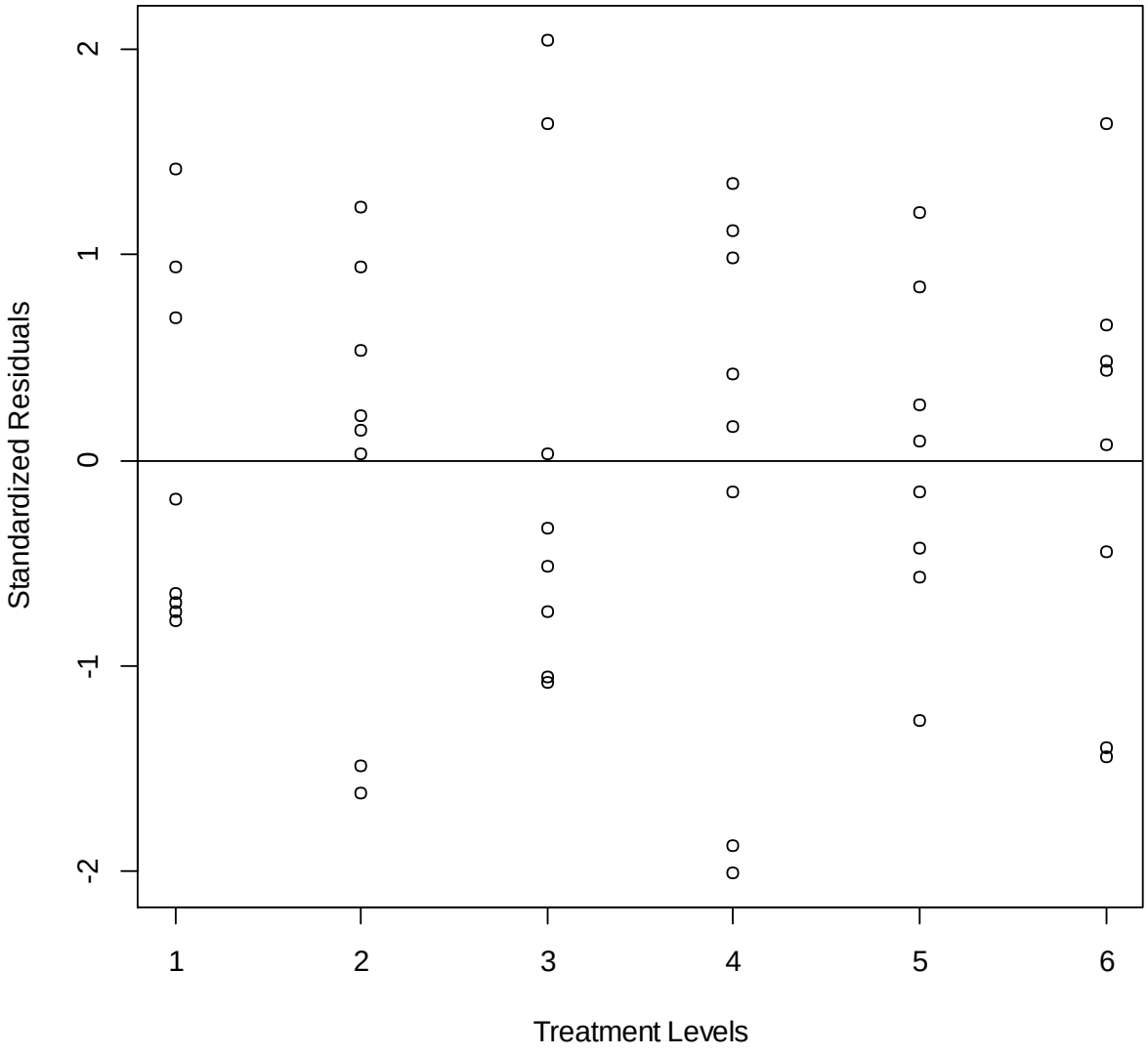
Plot: scatter plot of the standardized residuals against the treatment level

Conclusion: one-way ANOVA model is adequate to describe the data since the residuals are scattered around zero for each treatment level. There is no unusual patterns.

Checking outliers

Plot: scatter plot of the standardized residuals against the treatment level

Conclusion: most of observations have the standardized residuals between -2 and 2. Observation 34 has a standardized residual of -2.01 and observation 46 has a standardized residual of 2.05. These two observations should be our major concern but we may want to check the data to make sure there is no data entry errors.

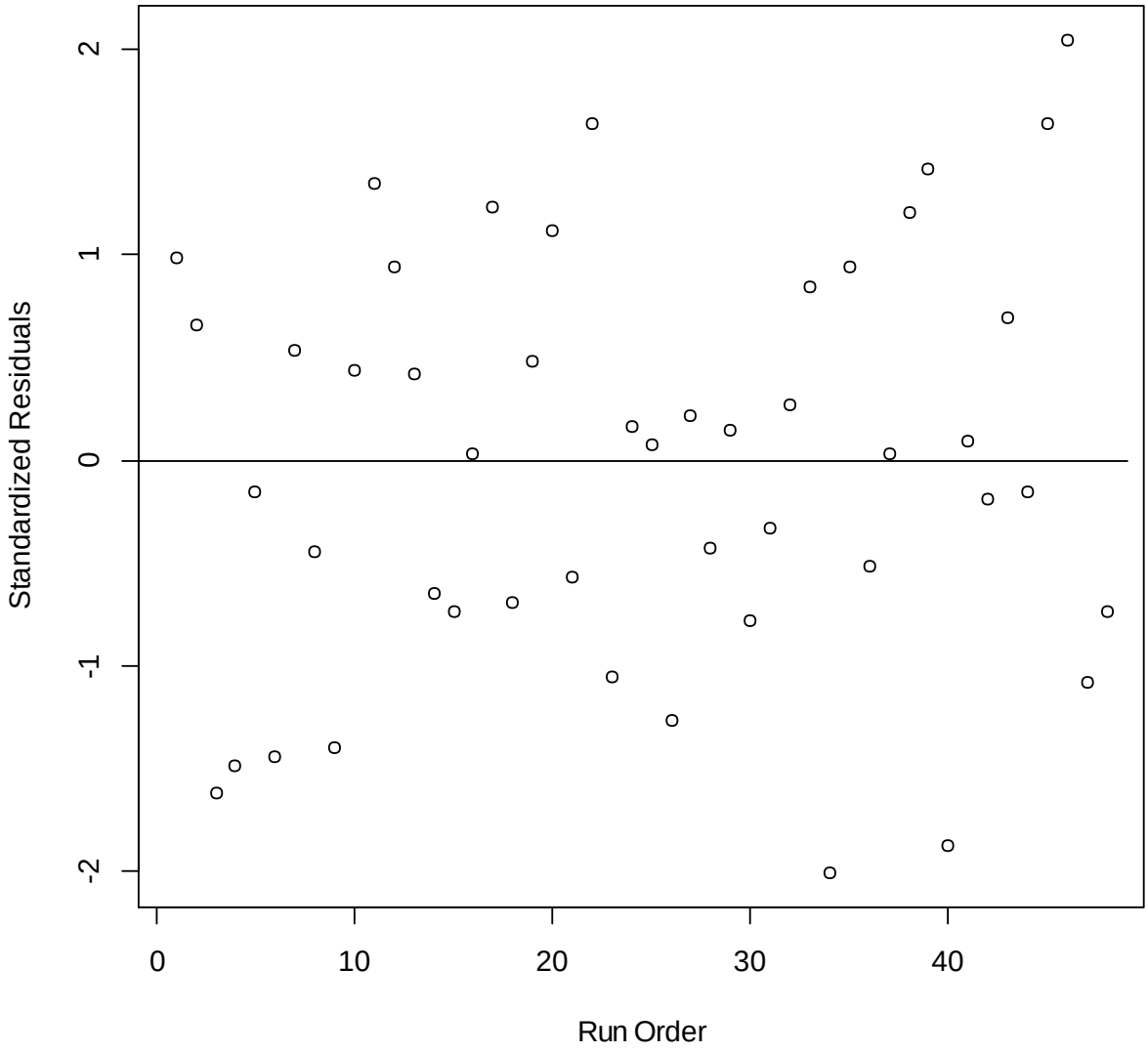


.

Checking outliers

Plot: scatter plot of the standardized residuals against the order in which the corresponding observations were collected

Conclusion: The independence assumption holds since the residuals are randomly scattered around zero with no discernible pattern.



Checking the Equal Variance Assumption

Plot: scatter plot of the standardized residuals against the predicted values

Conclusion: The equal variance assumption holds since the residuals are randomly scattered around zero with no discernible pattern.

Checking the Equal Variance Assumption

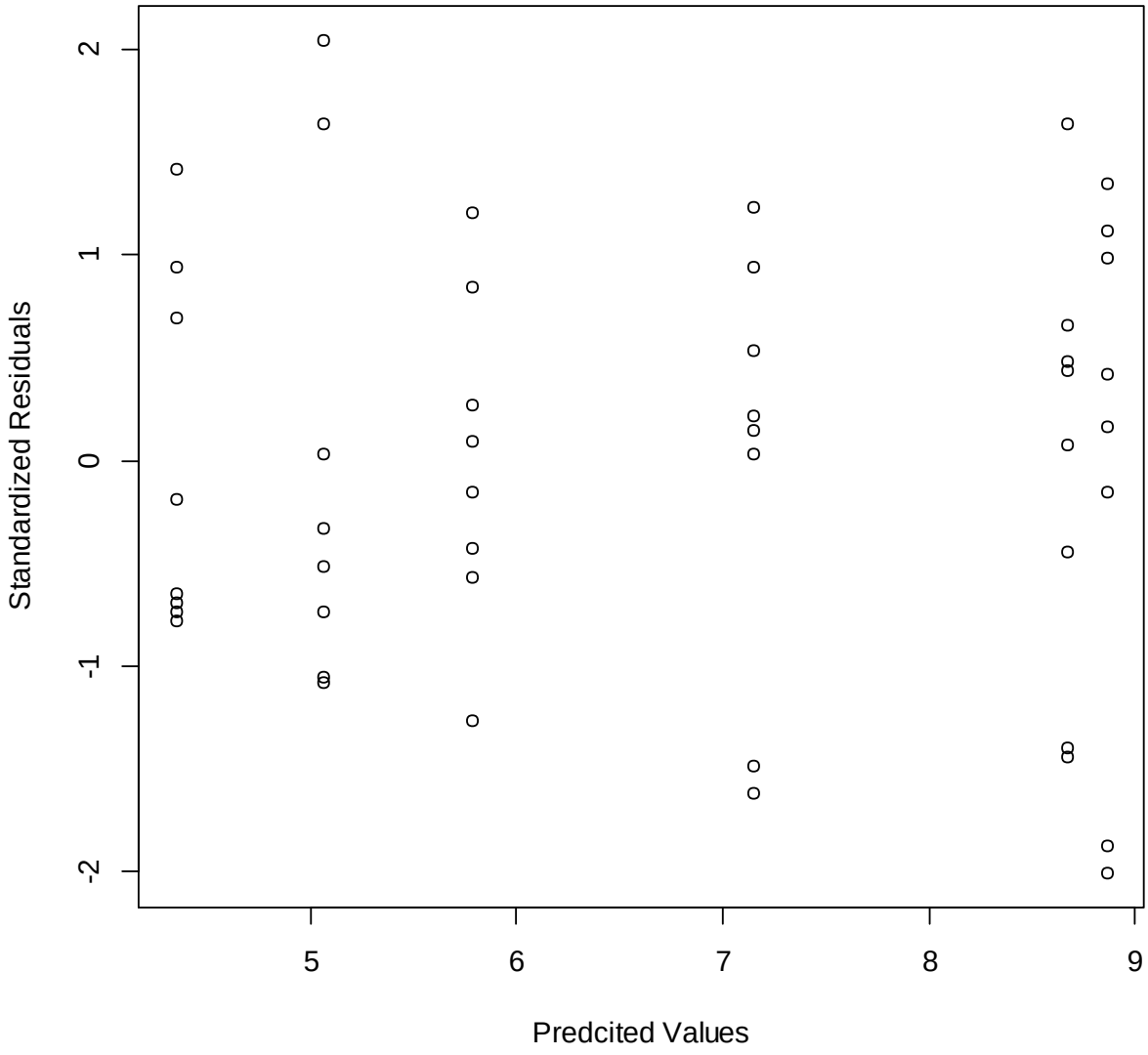
$\frac{s_{max}^2}{s_{min}^2} = 2.04 < 3$

Conclusion: The equal variance assumption holds since the residuals are randomly scattered around zero with no discernible pattern.

Checking the Equal Variance Assumption

Levene's Test: p-value $= 0.797$

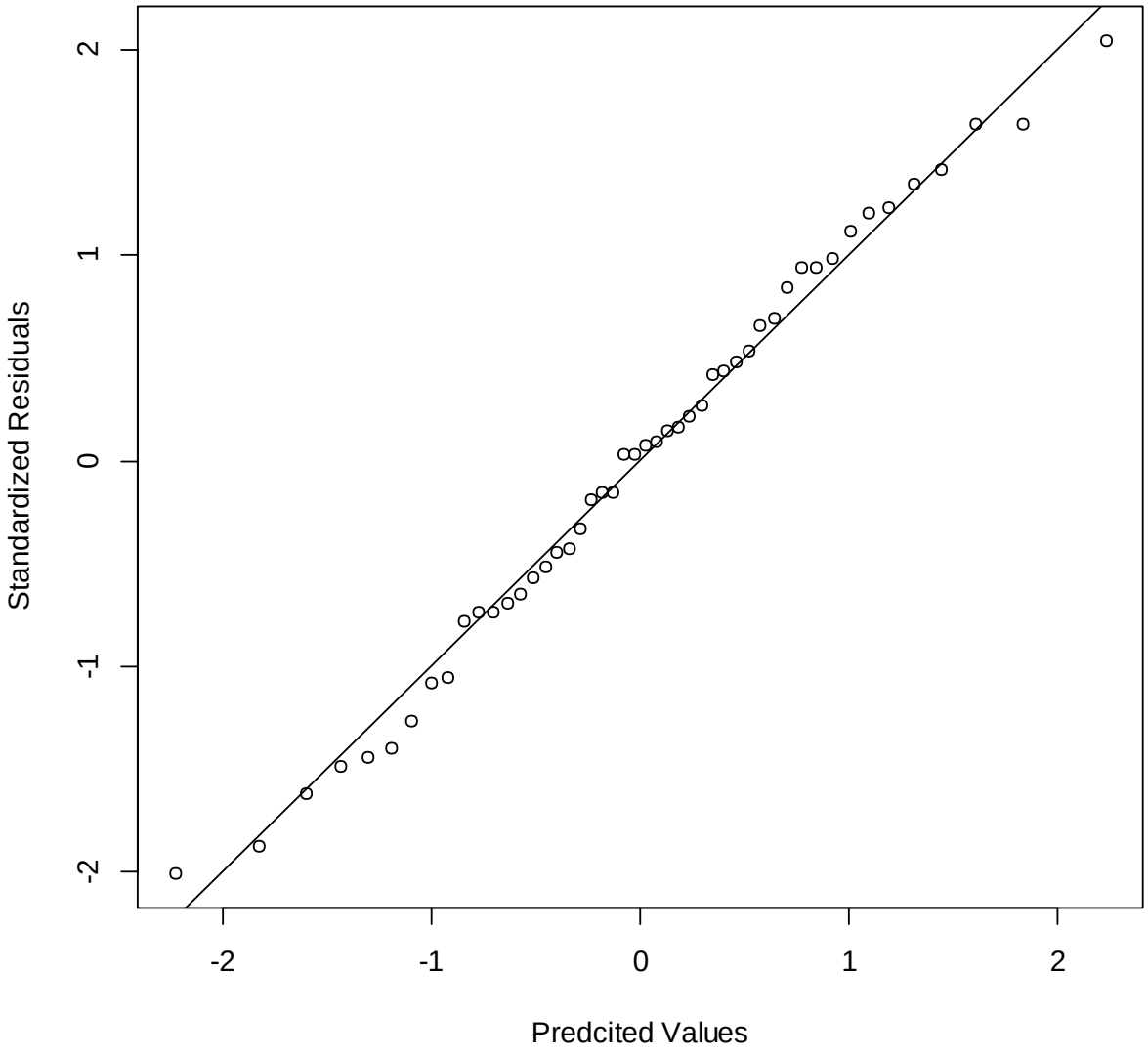
Conclusion: Fail to reject the null hypothesis of the equal variance assumption at the significance level of 0.05.



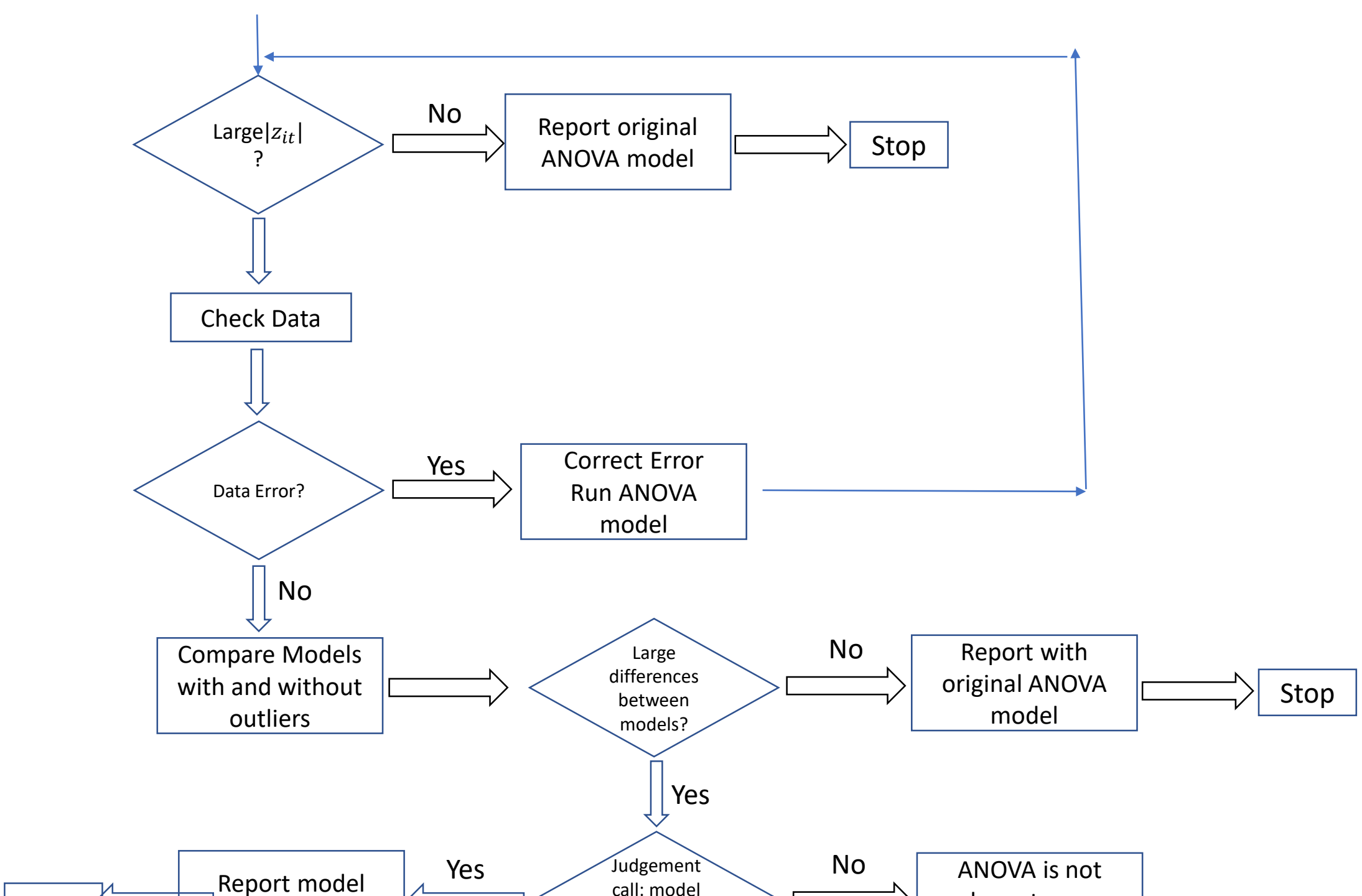
Checking the Normality Assumption

Plot: normal probability (or Q-Q) plot

Conclusion: The normality assumption holds since all points are close tor on the straight line with a slope of 1.



Exercise Problem 3.4 (Shrimp Experiment) Checking Assumptions of On-Way ANOVA Model. Perform the one-way ANOVA analysis using the data from Shrimp Experiment then check the assumptions using appropriate plots and tests.


Large$|z_{it}|$?
No
Report original ANOVA model
Stop
Check Data
Data Error?
Yes
Correct Error Run ANOVA model
No
Compare Models with and without outliers
Large differences between models?
No
Report with original ANOVA model
Stop
Yes
Judgement call: model
No
ANOVA is not
Yes
Report model

MA4720: Design and Analysis of Experiments

Week 4 At-a-Glance

Title: Data Transformations and Experiments with Two Crossed Treatment Factors

Overview

This week, we will describe a data transformation approach to equalize variances. We will present three models that can be used to analyze experiments with two crossed treatment factors. We will describe the methods that detect the effects of each treatment factor and the interactions between two treatment factors. We will cover the statistical inference including point estimation, confidence interval, and hypothesis testing of a single contrast without or with methods of multiple comparisons under different models.

Objectives

When you complete this module, you should be able to do the following:

1. Perform an appropriate data transformation to equalize variances.
2. Construct analysis of variance model with three model include cell-means model, two-way complete model, and two-way main-effects model.
3. Calculate the point estimation, the confidence interval, and hypothesis testing of a single contrast without or with methods of multiple comparisons under different models including cell-means model, two-way complete model, and two-way main-effects model.
4. Use appropriate plots and tests to check if the assumptions of models for experiments with two crossed treatment factors are satisfied.

Reading (Optional)

1. Section 5.6.2 in Chapter 5.
2. Sections 6.1, 6.2, 6.3, 6.4, and 6.5 in Chapter 6.

Instruction Content

See other files for details.

Quiz

Two quizzes. See other files for details.

Homework

One homework. See other files for details.

MA4720: Design and Analysis of Experiments

Week 4 Instruction Contents

Lesson 4.1: Data Transformation to Equalize Variances

Required Reading: Sections 5.6.2 in Chapter 5.

Introduction

In Lessen 3.3, we presented methods for checking the assumptions of one-way analysis of variance model including checking the equal variance assumption. two types of approaches can be applied if the equal variance assumption does not hold. One approach is to use a method of data analysis that is designed for nonconstant variances. However, such method is approximate and tends to be less powerful than the one-way analysis of variance methods of transformed data. Therefore, here we will only describe the approaches of data transformation.

Data Transformations to Equalize Variances

Finding a transformation of the data to equalize the variances of the error variables involves finding some function $h(y_{it})$ of the data and using $h(y_{it})$ instead of y_{it} as the response variance so that the one-analysis of variance model

$$h(Y_{it}) = \mu^* + \tau_i^* + \epsilon_{it}^*, t = 1, \cdots, r_i; i = 1, \cdots, v$$

holds and $\epsilon_{it}^* \sim N(0, \sigma^2)$ and the ϵ_{it}^*'s are mutually independent for all $t = 1, \cdots, r_i; i = 1, \cdots, v$.

An appropriate transformation can generally be found if there is a clear relationship between the error variance $\sigma_i^2 = Var(\epsilon_{it})$ and the mean response $E[Y_{it}] = \mu + \tau_i$ for $i = 1, \ldots, v$. For example, if the variance and the mean increase together, or if one increases as the other decreases, then the relationship between σ_i^2 and $\mu + \tau_i$ is often of the form

$$\sigma_i^2 = k * (\mu + \tau_i)^q$$

where k and q are constants. In this case, the function $h(y_{it})$ should be chosen to be

$$h(y_{it}) = \begin{cases} (y_{it})^{1-q/2} & q \neq 2 \\ \ln(y_{it}) & q = 2 \text{ and } y_{it}'s \text{ are nonzero} \\ \ln(y_{it}+1) & q = 2 \text{ and some } y_{it}'s \text{ are zero} \end{cases}$$

Here "ln" denotes the natural logarithm, which is the logarithm to the base e.

There are different ways to obtain the unknown parameter q in the transformation. First, the value of q is sometimes suggested by theoretical considerations if the response Y_{it} can be approximated by a commonly used distributions. For example, if Y_{it} follows the Poisson distribution, then $q = 1$ and the square-root transformation $h(y_{it}) = \sqrt{y_{it}}$ would be appropriate. If y_{it} follows the binomial distribution with $E(Y_{it}) = m * p$ and $Var(Y_{it}) = m * p * (1-p)$, then the transformation is $h(y_{it}) = \arcsin(\sqrt{y_{it}/m})$.

Another empirical way to obtain q is to find the slop from the simple linear regression with $\ln(s_i^2)$ as the response variable and $\ln(\bar{y}_{i.})$ as the regressor. The rationale behind this is that we assume that the relationship between σ_i^2 and $\mu + \tau_i$ has the following:

$$\sigma_i^2 = k * (\mu + \tau_i)^q$$

Then we have

$$\ln(\sigma_i^2) = \ln(k) + q * \ln(\mu + \tau_i)$$

Substitute σ_i^2 and $\mu + \tau_i$ by their estimates, we have

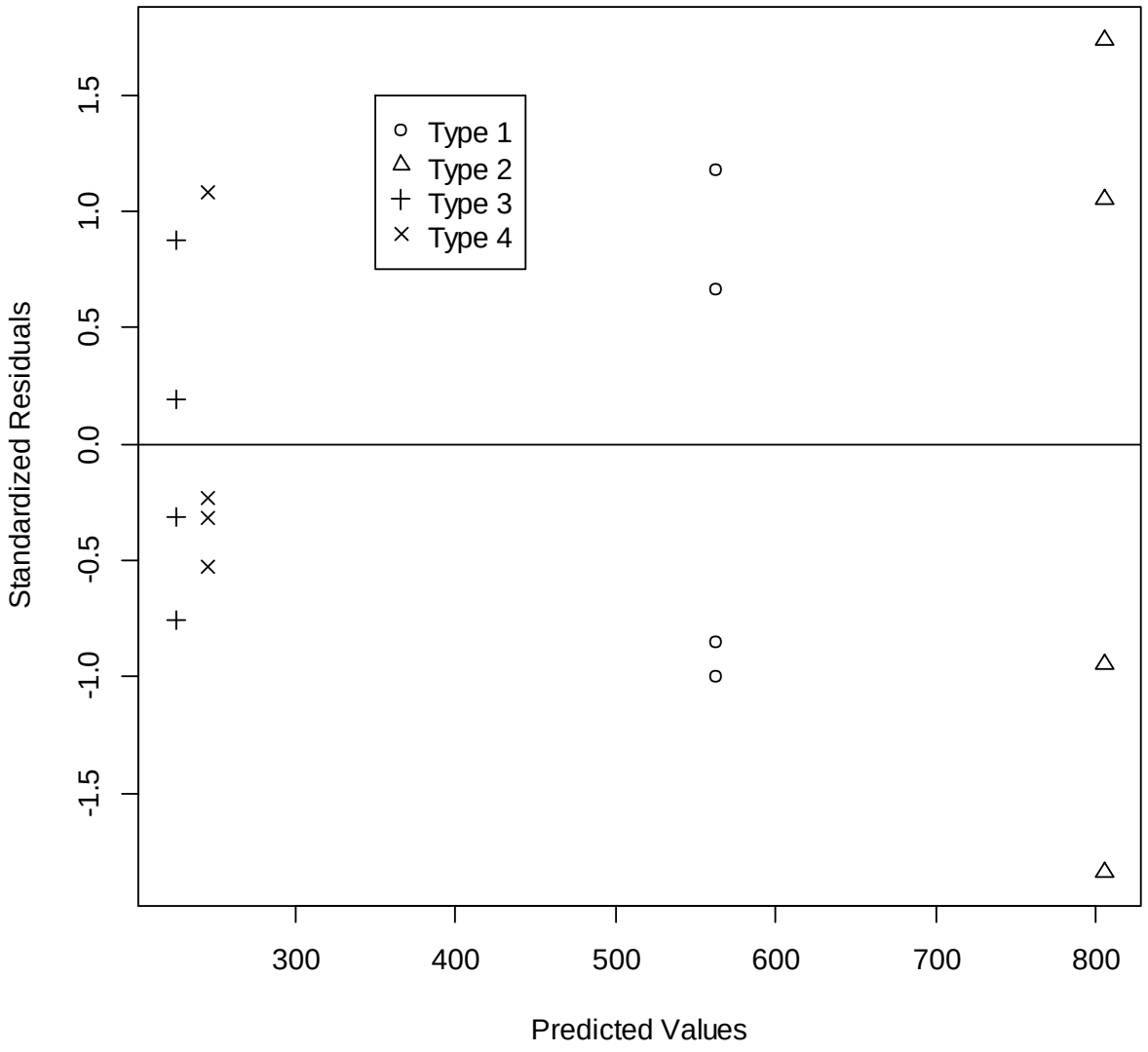
$$\ln(s_i^2) = \ln(k) + q * \ln(\bar{y}_{i.})$$

Example 4.1: Transformation - Battery Experiment

The following table contains the sample means and variances of each treatment level of battery life (not battery life per unit) from Battery Experiment. We can see that the sample variances increase as the sample means increase. Such observation is also consistent with the scatter plot of the standardized residuals against the predict values from the one-way analysis of variance

model: the error variances increase as the predicted values increase. In addition, the p-value from Levene's test is 0.016 suggesting that the equal variance assumption does not hold.

i	$\bar{y}_{i\cdot}$	$\ln(\bar{y}_{i\cdot})$	s_i^2	$\ln(s_i^2)$
1	562.50	6.3324	1333.67	7.1957
2	804.75	6.6905	3152.25	8.0559
3	225.50	5.4183	557.67	6.3238
4	245.75	5.5043	601.67	6.3996



Such relationship between the error variances and the predicted values suggest that we can use the following transformation:

$$h(y_{it}) = \begin{cases} (y_{it})^{1-q/2} & q \neq 2 \\ \ln(y_{it}) & q = 2 \text{ and } y_{it}'s \text{ are nonzero} \\ \ln(y_{it}+1) & q = 2 \text{ and some } y_{it}'s \text{ are nonzero} \end{cases}$$

To determine q, we perform a linear regression of $\ln(s_i^2)$ on $\ln(\bar{y}_{i\cdot})$ and find that the corresponding slope is 1.268. We take $q = 1.25$ and transformation is

$$h(y_{it}) = (y_{it})^{1-1.25/2} = (y_{it})^{0.375}$$

Since $(y_{it})^{0.375}$ is close to $(y_{it})^{0.5}$, and since the square root of the data values is perhaps more meaningful than $(y_{it})^{0.375}$, we will take the square root transformation in our analysis. It can be shown that the equal variance assumption holds for the transformed data. And simultaneous 99% confidence intervals from Tukey's method indicate that all pairwise differences are significantly different from zero except for the comparison of battery types 3 and 4. Furthermore, it is reasonable to conclude that type 2 (alkaline, store brand) is best, followed by type 1 (alkaline, name brand). However, more detailed interpretation of the results is muddled by use of the transformation, since the comparisons use mean values of $\sqrt{\text{life}}$.

	Contrast	**Point Estimate**	**99% CI**
Brand 1 versus Brand 2	$\tau_1 - \tau_2$	-4.647	(-6.94, -2.36)
Brand 1 versus Brand 3	$\tau_1 - \tau_3$	8.706	(6.42, 11.00)
Brand 1 versus Brand 4	$\tau_1 - \tau_4$	8.045	(5.75, 10.34)
Brand 2 versus Brand 3	$\tau_2 - \tau_3$	13.354	(11.06, 15.64)
Brand 2 versus Brand 4	$\tau_2 - \tau_4$	12.693	(10.40, 14.98)
Brand 3 versus Brand 4	$\tau_3 - \tau_4$	-0.661	(-2.95, 1.63)

Exercise 4.1: Transformation - Battery Experiment

For Battery Experiment, check the equal variance assumption if the life per unit is used as the response variable. Even the equal variance assumption is not violated, as a practice, please follow the steps in **Example 4.1** to find a transformation and obtain simultaneous 95% confidence intervals of all pairwise comparisons with Tukey method based on the transformed data.

Lesson 4.2: Introduction of Experiments with Two Crossed Treatment Factors

Required Reading: Section 6.1 and 6.2 in Chapter 6

Crossed and Nested Treatment Factors

In this lesson, we consider the use of completed randomized design for experiments with two **crossed** treatment factors. We use the upper case of Roman letters to represent the treatment and the lower case of Roman letters to represent the number of levels:

First treatment factor A with a levels and are coded as: $1,2,\cdots,a$

Second treatment factor B with b levels and are coded as: $1,2,\cdots,b$

Factors are crossed if every combination of levels are observed. In other words, there is at least one observation in every combination of levels for two factors.

For two crossed treatment factors, the total number of combinations is $v = a * b$ and they are coded as $11, 12, \cdots, ab, 21, 22, \cdots, 2b, \cdots, a1, a2, \cdots, ab$. In previous lessons, we recoded the treatment combinations as $1,2,\cdots,v$ with the **lexical order** and used the one-way analysis of variance model. In this lesson, we need to investigate the contributions of that each of the factors make individually to the response, so it is more convenient to retain the 2-digit code ij for a treatment combination in which factor A is at level i and factor B is at level j.

Also note that the lexical (or lexicographical order) is a generalization of the way to define a total order of sequences. For two crossed factors, the combinations of levels are first ordered by the level from the first treatmen factor then ordered by the level from the second factor.

Example 4.2: Two Crossed Factors - Battery Experiment

We have two treatment factors: duty and brand. These two treatment factors are crossed and they are coded as the following:

Factor A (duty): $a = 2$ levels, alkaline is coded as 1, and heavy duty is coded as 2

Factor 2 (brand): $b = 2$ levels, name brand is coded as 1, and store brand is coded as 2

Therefore, there are $v = 2 * 2 = 4$ treatment combinations, and these four treatment combinations are coded as:

1-digit code	1	2	3	4
2-digite code	11	12	21	22
Treatment combination	Alkaline Name brand	Alkaline Store Brand	Heavy duty Name Brand	Heavy Duty Store Brand

A factor is **nested** with another factor when each level of the first factor co-occurs within only one level of the other factor. In other words, an observation has to be within one level of factor 2 in order to have a specific level of factor 1. Not all treatment combinations are represented.

Example 4.3: An Example of Nested Factors

To test the packing system of a manufacturer, the researchers take:

Factor 1: two factories

Factor 2: three batches from each factory

There are two factors and the factor 2 is nested with factor 1. Note that there are six levels for factor 2 and three levels are within one factory while the other three levels are within another factory. For these two factors, the total number of treatment combinations is $v = 6$ instead of $v = 2 * 6 = 12$.

The Meaning of Interaction [Need to add a video here]

For experiments with two crossed factors, not only the **main effects**, the contributions of that each of the factors make individually to the response, but also the **interaction effects** need to be considered. For the contributions of that each of the factors to the response, an **interaction effect** is change in the main effect of one factor over levels of the other factor. Such effects can be easily seen in the **interaction plot**. In the interaction plot, the y-axis represents $\bar{y}_{ij.}$, the mean response from the level i of factor A and the level j of factor B, the x-axis represents the levels of one factor while the different symbols represent the levels of another factor. Here we use a hypothetical example to illustrate the use of interaction plot to understand the main effects and the interaction effects of two treatment factors.

Example 4.4: Interaction – A Hypothetical Experiment

Suppose that a hypothetical statistics department wishes to know to what extent student performance in an introductory online course is affected by the primary presentation format (textbook reading assignments, videotaped lectures, or interactive software) and course structure (structured, with regular deadlines throughout the term; or unstructured, with only a deadline to finish by the end of the term).

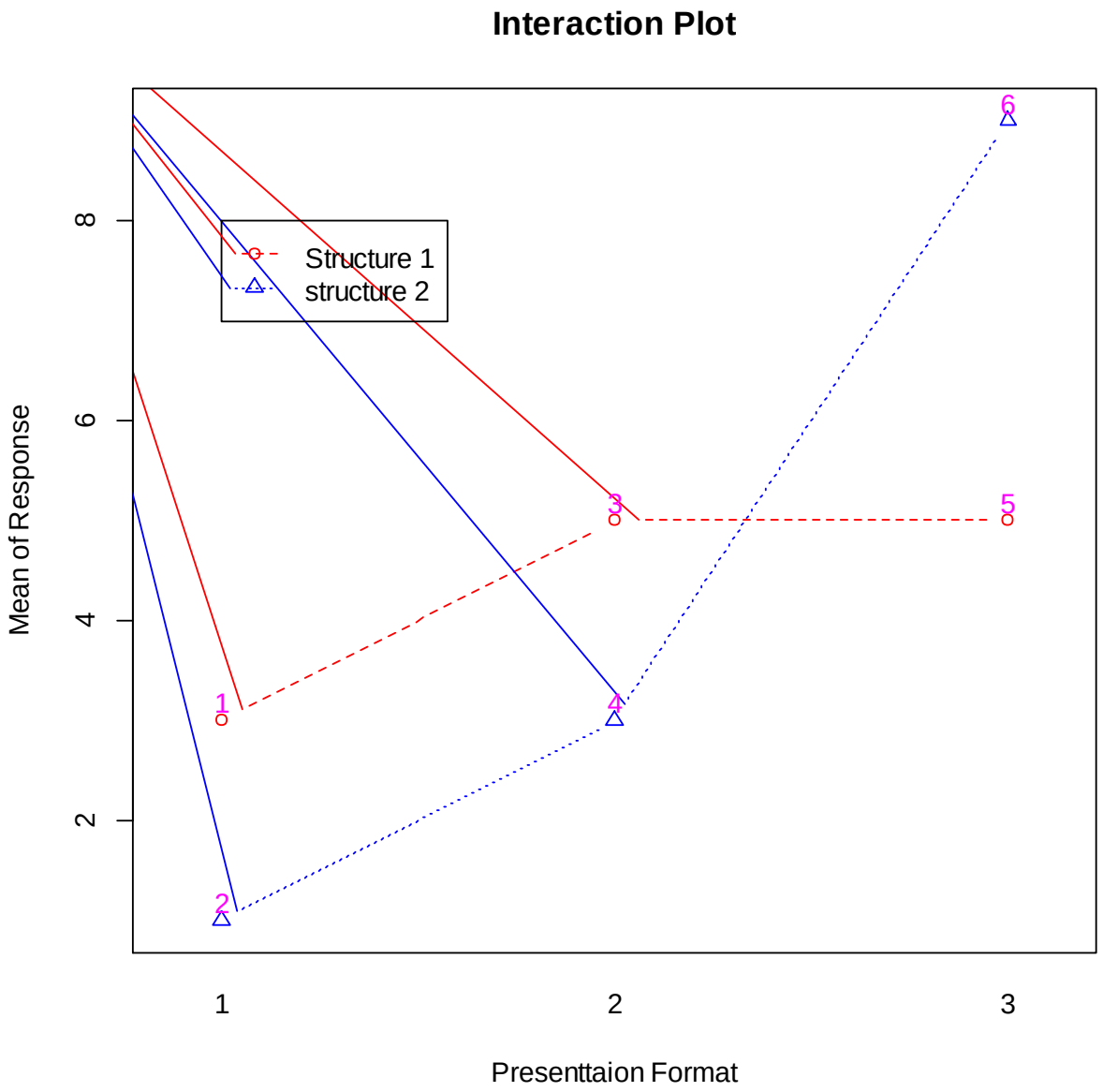
There are two treatment factors of interest, namely "presentation format," which has three levels, coded 1, 2, and 3, and "course structure," which has two levels, coded 1 and 2. The students who enroll in the introductory course are the experimental units and are allocated at random to one of the six treatment combinations in such a way that approximately equal numbers of students are assigned to each combination of presentation format and course structure. Student performance is to be measured by means of a computer-graded multiple-choice examination, and an average exam score $\bar{y}_{ij.}$ for each treatment combination will be obtained, averaging over students for each treatment combination. There are eight different types of situations that could occur, and these can be depicted in the interaction plots.

The following is an example of interaction plot and other situations are presented in the video. From this plot, we can see there are six points, representing the mean responses of 6 treatment

combinations. The 1-digit codes of 6 treatment combinations are also added at the top of points. To look at the main effect of each treatment factor, we need to look at if the average responses of each level of that factor over all levels of the other factor and see if they are roughly same. To look at the interaction effects, we need to look at the lines are paralleled. Let us use the plot as an illustration:

- **Main effect of presentation format:** this factor has three levels thus three averages (the average of points 1 and 2, the average of points 3 and 4, and the average of points 5 and 6, in terms of $\bar{y}_{ij.}$) are not same, so we conclude that there are main effects of presentation format, meaning the responses differ according to the different presentation format.
- **Main effect of structure:** this factor has two levels thus two averages (the average of points 1, 3, and 5 and the average of points 2, 4, and 6, in terms of $\bar{y}_{ij.}$) are quite similar. However, it is actually misleading to state there are no main effects of structure. We must be careful here because there are interaction effects between two treatment factors. In general, **if there is an interaction between two treatment factors, then it may not be sensible to examine either of the main effects separately**. Instead, it will often be preferable to compare the effects of the treatment combinations themselves.
- **Interaction effects:** A simplest form of interaction is pairwise interaction that refers to the changes in the main effect of two levels of one factor over two levels of the other factor. The different changes indicate that there is an interaction effect and the lines on the interaction plot are not parallel. The similar changes indicate that there is no interaction effect and the lines on the interaction plot are close to parallel. For the following interaction plot, the change in the main effect of presentation formats 1 and 2 over structure 1 (the difference of points 1 and 3 in terms of $\bar{y}_{ij.}$) is same as the change in the main effect of presentation formats 1 and 2 (the difference of points 2 and 4 in terms of $\bar{y}_{ij.}$) . These two lines are parallel, indicating there is no or little pairwise interaction between formats 1 and 2 and structures 1 and 2. However, the change in the main effect of presentation formats 2and 3 over structure 1 (the difference of points 3 and 5 in terms of $\bar{y}_{ij._}$) are not same as the change in the main effect of presentation formats 1 and 2 over structure 2 (the difference of points 4 and 6 in terms of $\bar{y}_{ij.}$). So the two lines are not

parallel, indicating there is a pairwise interaction between formats 2 and 3 and structures 1 and 2. In summary, the two lines in the plot are not parallel indicating that there are interactions between the presentation format and structure.



[Add a video here]

Models for Two Treatment Factors

Three models will be introduced here: the **cell-means model**, the **two-way complete model**, and the **two-way main-effects model**.

Cell-Means Model

From previous lessons, we have learned how to use the one-way analysis of variance model to analyze experiments with two crossed treatments factors. In that model, the combinations of levels from two treatment factors are considered as the levels of a new treatment factor. The cell-means model is same as the one-way analysis model. However, for convenience, the treatment combinations are coded with two-digit codes and the **cell-means model** is:

$$Y_{ijt} = \mu + \tau_{ij} + \epsilon_{ijt}, t = 1, \cdots, r_{ij}; i = 1, \cdots, a; j = 1, \cdots, b$$

with the following assumptions:

- μ and τ_{ij} $(i = 1, \cdots, a; j = 1, \cdots, b)$ are unknown population parameters
- $\epsilon_{ijt} (t = 1, \cdots, r_{ij}; i = 1, \cdots, a; j = 1, \cdots, b)$ are independent
- $\epsilon_{ijt} \sim N(0, \sigma^2)$: normal distribution with mean 0 and variance σ^2

Here, i and j are the levels of factors A and B, respectively. The "cell" here refers to the cell of table whose row represent the levels of A and whose columns represent the levels of B.

Two-way Complete Model

It is often useful to model the effect on the response of treatment combinations ij to be the sum of the individual effect of the two factors, together with their interaction; that is:

$$\tau_{ij} = \alpha_i + \beta_j + (\alpha\beta)_{ij}$$

Here, α_i is the effect (positive on negative) on the response due to the fact that the ith level of factor A is observed, β_j is the effect (positive on negative) on the response due to the fact that the jth level of factor B is observed, and $(\alpha\beta)_{ij}$ is the extra effect (negative or positive) on the

response of observing levels $i$and j of factors A and B together. Thus, the corresponding model, which is called **two-way complete model** or **two-way analysis of variance model**, is as follows:

$$Y_{ijt} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijt}, t = 1, \cdots, r_{ij}; i = 1, \cdots, a; j = 1, \cdots, b$$

with the following assumptions:

- μ, α_i, β_j, and $(\alpha\beta)_{ij}$ $(i = 1, \cdots, a; j = 1, \cdots, b)$ are unknown population parameters
- $\epsilon_{ijt}(t = 1, \cdots, r_{ij}; i = 1, \cdots, a; j = 1, \cdots, b)$ are independent
- $\epsilon_{ijt} \sim N(0, \sigma^2)$: normal distribution with mean 0 and variance σ^2

Here, the phrase ”two-way” refers to the fact that there are two primary sources of variance, namely, the two treatment factors. We can see that the cell-means model and the two-way complete model just use different parameters but are actually equivalent.

Two-way Main-effects Model

If the interaction effect are not observed from the interaction plot, then it can be removed from the two-way complete model and the model becomes:

$$Y_{ijt} = \mu + \alpha_i + \beta_j + \epsilon_{ijt}, t = 1, \cdots, r_{ij}; i = 1, \cdots, a; j = 1, \cdots, b$$

with the following assumptions:

- μ, α_i, β_j, $(i = 1, \cdots, a; j = 1, \cdots, b)$ are unknown population parameters
- $\epsilon_{ijt}(t = 1, \cdots, r_{ij}; i = 1, \cdots, a; j = 1, \cdots, b)$ are independent
- $\epsilon_{ijt} \sim N(0, \sigma^2)$: normal distribution with mean 0 and variance σ^2

This model is a “submodel” or “reduced model” of the two-way complete model and is called a **two-way main-effect model**, or **two-way additive model**, since the effect on the response of treatment combination ij is modeled as the sum of the individual effects of the two factors. If an additive model is used when the factors really do interact, then inferences on main effects can be very misleading.

Checking Assumptions on the Model

The assumptions implicit in three models are that the error terms have equal variance, are mutually independent, and are normally distributed. The strategy and the methods for checking the model assumptions are the same as those in previous lessons. The standardized residuals are calculated as

$$z_{ijt} = \frac{y_{ijt} - \hat{y}_{ijt}}{\sqrt{SSE/(n-1)}}$$

where $\hat{y}_{ijt}$ is the estimated response, and depending on the model, we have

$$\hat{y}_{ijt} = \hat{\mu} + \hat{\tau}_{ij} \text{ or } \hat{y}_{ijt} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \widehat{(\alpha\beta)}_{ij} \text{ or } \hat{y}_{ijt} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j.$$

The residuals are plotted against

(1) the order of observation to check the independence
(2) the levels of each factors or treatment combination to check the model fit
(3) the levels of each factors or treatment combination to check for the outlies
(4) $\hat{y}_{ijt}$ to check for equality of variances
(5) the normal scores to check the normality assumption

Statistical Inferences

For the aforementioned models (the cell-means model, the two-way complete model, and the two-way main-effects model), we will perform the statistical inference in the following order:

(1) Find the least squares estimates (or estimators) of the model parameters, such as $\tau_{ij}, \alpha_i, \beta_j, (\alpha\beta)_{ij}$, and σ^2.
(2) Perform the overall analysis to test if there is any effect on the response from two treatment factors.

(3) Perform the analysis to see if there is any effect on the response from any one of two treatment factors.

(4) Perform the statistical inference for individual contrasts including the point estimation, the confidence interval and the hypothesis testing with appropriate methods for multiple comparisons.

Lesson 4.3: Analysis of Cell-means Model and Two-way Complete Model: Contrasts

Required Reading: Section 6.3 from Chapter 6

In this lesson, we will learn how to perform the analysis of experiments with two crossed treatment factors using the cell-means model and the two-way complete model since these two models are equivalent with different parametrization. In this lesson, we will use the data from the meat-cooking experiment as an example.

Meat Cooking Experiment (Balanced)

(L. Alvarez, M. Burke, R. Chow, S. Lopez, and C. Shirk, 1998)

An experiment was run to investigate the amount of weight lost (in grams) by ground beef hamburgers after grilling or frying, and how much the weight loss is affected by the percentage fat in the beef before cooking. The experiment involved two factors: cooking method (factor *A*, with two levels frying and grilling, coded 1, 2), and fat content (factor *B*, with three levels 10, 15, and 20%, coded 1, 2, 3). Thus there were six treatment combinations 11, 12, 13, 21, 22, 23, relabeled as treatment levels 1, 2, …, 6, respectively. Hamburger patties weighing 110 g each were prepared from meat with the required fat content. There were 30 "cooking time slots" which were randomly assigned to the treatments in such a way that each treatment was observed five times ($r = 5$).

Response variable: weight loss after cooking (in grams)

Treatment Levels: cooking method, with two levels frying and grilling, coded 1, 2 and fat content, with three levels 10, 15, and 20%, coded 1, 2, 3

1 – frying, 10% fat (11) 4 – grilling, 10% fat (21)

2 – frying, 15% fat (12) 5 – grilling, 15% fat (22)

3 – frying, 20% fat (13) 6 – grilling, 20% fat (23)

Summary statistics:

Level	Treatment	Replicates	Sample Mean	Sample Variance

	Combinations			
1	11	$r_1 = 5$	$\bar{y}_{1\cdot} = 84.4$	$s_1^2 = 6.3$
2	12	$r_2 = 5$	$\bar{y}_{2\cdot} = 81.8$	$s_2^2 = 4.2$
3	13	$r_3 = 5$	$\bar{y}_{3\cdot} = 76.0$	$s_3^2 = 18.5$
4	21	$r_4 = 5$	$\bar{y}_{4\cdot} = 83.4$	$s_4^2 = 3.8$
5	22	$r_5 = 5$	$\bar{y}_{5\cdot} = 86.0$	$s_5^2 = 4.5$
6	23	$r_6 = 5$	$\bar{y}_{6\cdot} = 78.4$	$s_6^2 = 6.3$
Overall		$n = 30$	$\bar{y}_{\cdot\cdot} = 81.667$	$s^2 = 18.437$

Balance Experiment and Unbalanced Experiment: An experiment is **balanced** when all factor levels (or treatment groups) have the same number of experimental units (or items receiving a treatment). Unbalanced experiments add complexity to the analysis of the data. To illustrate the analysis of unbalanced experiments, we create an unbalanced data from the meat-cooking experiment by removing the last observation. From the Following table, we can see there are some smaller differences among summary statistics.

Summary statistics from Unbalanced Data:

Level	**Treatment Combinations**	**Replicates**	**Sample Mean**	**Sample Variance**
1	11	$r_1 = 5$	$\bar{y}_{1\cdot} = 84.4$	$s_1^2 = 6.3$
2	12	$r_2 = 5$	$\bar{y}_{2\cdot} = 81.8$	$s_2^2 = 4.2$
3	13	$r_3 = 5$	$\bar{y}_{3\cdot} = 76.0$	$s_3^2 = 18.5$
4	21	$r_4 = 5$	$\bar{y}_{4\cdot} = 83.4$	$s_4^2 = 3.8$
5	22	$r_5 = 5$	$\bar{y}_{5\cdot} = 86.0$	$s_5^2 = 4.5$
6	23	$r_6 = 4$	$\bar{y}_{6\cdot} = 77.5$	$s_6^2 = 3.0$
Overall		$n = 29$	$\bar{y}_{\cdot\cdot} = 81.655$	$s^2 = 19.091$

Balance is nonessential but desirable if equal accuracy, power, or confidence interval width for treatment comparisons is important. Severe imbalance can induce factor confounding (correlated factors or non-independent treatment levels).

Estimable Functions and Contrasts

Use the same arguments from the previous lesson, an estimable function of the parameters must have the following form:

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(\mu + d_{ij}\tau_{ij}) = \sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij}(\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij})$$

where $d_{ij}(i = 1, \cdots, a; j = 1, \cdots, b)$ are constants.

A contrast of the parameters must have the following form:

$$\sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij}\tau_{ij} = \sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij}(\alpha_i + \beta_j + (\alpha\beta)_{ij})$$

where $d_{ij}(i = 1, \cdots, a; j = 1, \cdots, b)$ are constants and the sum of them is 0 (in other words, $\sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij} = 0$).

Among all possible contrasts, we are particularly interested in the following four types of contrasts:

- **Pairwise Difference between Two Treatment Combinations ij and sh:**

$$\tau_{ij} - \tau_{sh} = (\alpha_i + \beta_j + (\alpha\beta)_{ij}) - (\alpha_s + \beta_h + (\alpha\beta)_{sh})$$

- **Interaction between Two Levels s and u of Factor A and Two Levels h and q of Factor B:**

$$(\tau_{sh} - \tau_{uh}) - (\tau_{sq} - \tau_{uq})$$

We can verify that the above contrast is indeed an interaction because it only involves the term of $(\alpha\beta)_{ij}$:

$$(\tau_{sh} - \tau_{uh}) - (\tau_{sq} - \tau_{uq}) = ((\alpha\beta)_{sh} - (\alpha\beta)_{uh}) - ((\alpha\beta)_{sq} - (\alpha\beta)_{uq})$$

Actually, the above contrast is the difference of two pairwise differences between two treatment combinations. Note the subscript for the level of factor B (here h) is fixed in the first pairwise difference, the subscript for the level of factor B (here q) is fixed in the second pairwise difference, and the orders of subscripts of factor A (here s and u) are same for two pairwise differences. Based on these rules, the following contrasts are can be used for the interaction between two Levels s and u of factor A and two Levels h and q of factor B:

$$(\tau_{sh} - \tau_{sq}) - (\tau_{uh} - \tau_{uq})$$

$$(\tau_{uh} - \tau_{sh}) - (\tau_{uq} - \tau_{sq})$$

While the following contrast is not an interaction contrast because the orders of subscripts of factor A (here s and u) are not same for two pairwise differences:

$$(\tau_{sh} - \tau_{uh}) - (\tau_{uq} - \tau_{sq})$$

- **Average Effects of Different Levels of Factor *A* or *B***

The contrast for the average effects of different levels of factor A is:

$$\sum_{i=1}^{a} c_i \bar{\tau}_{i\cdot} \text{ with } \sum_{i=1}^{a} c_i = 0 \; \tau_{i\cdot} = \sum_{j=1}^{b} \frac{\tau_{ij}}{b} = \sum_{j=1}^{b} \frac{\alpha_i + \beta_j + (\alpha\beta)_{ij}}{b}$$

Re-write the formula, we can simplify

$$\sum_{i=1}^{a} c_i \bar{\tau}_{i\cdot} = \sum_{i=1}^{a} \sum_{j=1}^{b} c_i * \frac{\alpha_i + \beta_j + (\alpha\beta)_{ij}}{b} =$$

and get

$$\sum_{i=1}^{a} c_i \bar{\tau}_{i\cdot} = \sum_{i=1}^{a} c_i (\alpha_i + (\overline{\alpha\beta})_{i\cdot})$$

where $(\overline{\alpha\beta})_{i\cdot} = \frac{1}{b}\sum_{j=1}^{b}(\alpha\beta)_{ij}$. From this formula, we can see clearly that we have averaged over any interaction effect that might present. We often write

$$\alpha_i^* = \alpha_i + (\overline{\alpha\beta})_{i\cdot}$$

and the corresponding contrast as:

$$\sum_{i=1}^{a} c_i \bar{\tau}_{i\cdot} = \sum_{i=1}^{a} c_i \alpha_i^*$$

Similarly, the contract for the average effects of different levels of factor B is:

$$\sum_{j=1}^{b} k_j \bar{\tau}_{\cdot j} = \sum_{j=1}^{b} k_j \beta_j^*$$

where $\sum_{j=1}^{b} k_j = 0$, $\bar{\tau}_{\cdot j} = \frac{1}{a}\sum_{i=1}^{a} \tau_{ij}$, $\beta_j^* = \beta_j + (\overline{\alpha\beta})_{\cdot j}$, and $(\overline{\alpha\beta})_{\cdot j} = \frac{1}{a}\sum_{i=1}^{a}(\alpha\beta)_{ij}$.

Writing Contrasts as Coefficient Lists

Instead of writing out a contrast explicitly, it is sometimes sufficient, and more convenient, to list the contrast coefficients only. **More importantly, in many situations, the contrast coefficients are used in R function lsmeans and contrast for the statistical inference of a contrast.** We can refer to contrasts as either a list of coefficients of the parameters α_i^*, β_j^*, and $(\alpha\beta)_{ij}$ or as a list of coefficients of the τ_{ij}. In addition, for $(\alpha\beta)_{ij}$ or τ_{ij}, they are arranged by the lexical order. For example, τ_{ij} are arranged by the following order:

$$\tau_{11}, \tau_{12}, \cdots, \tau_{1b}, \tau_{21}, \tau_{22}, \cdots, \tau_{2b}, \cdots, \tau_{a1}, \tau_{a2}, \cdots, \tau_{ab}$$

This is illustrated in the following example.

Example 4.5: Writing Contrasts as Coefficient Lists - Meat-Cooking Experiment

Note that factor A, cooking method, has two levels: frying and grilling while factor B, fat content, has three levels of 10%, 15%, and 20%. The factor B has three equally spaced quantitative levels. Find the contrast coefficients for the following contrasts:

(1) Pairwise difference between treatment combination 11 and 21 in terms of τ_{ij}.
(2) Interaction between levels of 1 and 2 from factor A and levels 1 and 3 from factor B in terms of τ_{ij} or $(\alpha\beta)_{ij}$.
(3) The difference of response of level 3 and the average response of levels 1 and 2 from factor B in terms of τ_{ij} or β_i^*.

(4) The quadratic trend of factor B from the balanced experiment in terms of τ_{ij} or β_i^*.

(5) The linear trend of factor B from the unbalanced experiment in terms of τ_{ij} or β_i^*.

Solution:

(1) The contrast for the difference between treatment combination 11 and 21 is:

$$\tau_{11} - \tau_{21}$$

So the corresponding contrast coefficients in terms of τ_{ij} are:

$$[1, 0, 0, -1, 0, 0]$$

(2) One contrast for the Interaction between levels of 1 and 2 from factor A and levels 1 and 3 from factor B is:

$$(\tau_{11} - \tau_{21}) - (\tau_{13} - \tau_{23}) = ((\alpha\beta)_{11} - (\alpha\beta)_{21}) - ((\alpha\beta)_{13} - (\alpha\beta)_{23})$$

So the corresponding contrast coefficients in terms of τ_{ij} are:

$$[1, 0, -1, -1, 0, 1]$$

So the corresponding contrast coefficients in terms of $(\alpha\beta)_{ij}$ are also:

$$[1, 0, -1, -1, 0, 1]$$

(3) The contrast for the difference of response of level 3 and the average response of levels 1 and 2 from factor B is:

$$\beta_3^* - \frac{1}{2}(\beta_1^* + \beta_2^*) = \bar{\tau}_{\cdot 3} - \frac{1}{2}(\bar{\tau}_{\cdot 1} + \bar{\tau}_{\cdot 2})$$

$$= \frac{1}{2}(\tau_{13} + \tau_{23}) - \frac{1}{2}\left(\frac{1}{2}(\tau_{11} + \tau_{21}) + \frac{1}{2}(\tau_{12} + \tau_{22})\right)$$

So the corresponding contrast coefficients in terms of β_i^* are:

$$[-0.5, -0.5, 1]$$

So the corresponding contrast coefficients in terms of τ_{ij} are:

$$[-0.25, -0.25, 0.5, -0.25, -0.25, 0.5]$$

(4) The factor B has three equally spaced levels, so the contrast coefficients for the quadratic trend in terms of β_i^*, can be obtained from Table A.2:

$$[1, -2, 1].$$

Therefore, the contrast is:

$$\beta_1^* - 2\beta_2^* + \beta_3^* = \bar{\tau}_{\cdot 1} - 2\bar{\tau}_{\cdot 2} + \bar{\tau}_{\cdot 3} = \frac{1}{2}(\tau_{11} + \tau_{21}) - 2 * \frac{1}{2}(\tau_{12} + \tau_{22}) + \frac{1}{2}(\tau_{13} + \tau_{23})$$

So the corresponding contrast coefficients in terms of τ_{ij} are:

$$[0.5, -1, 0.5, 0.5, -1, 0.5]$$

(5) For unbalanced experiment, we need to calculate the contrast coefficients of the linear trend for the factor B according to the previous lesson. Here we have

$$x_1 = 10, x_2 = 15, x_3 = 20$$

$$r_1 = 10, r_2 = 10, r_3 = 9$$

$$\bar{x}_{\cdot} = \frac{r_1 x_1 + r_2 x_2 + r_3 x_3}{r_1 + r_2 + r_3} = \frac{10 * 10 + 10 * 15 + 9 * 20}{10 + 10 + 9} = \frac{430}{29}$$

$$c_1 = r_1 * (x_1 - \bar{x}_{\cdot}) = 10 * \left(10 - \frac{430}{29}\right) = -48.28$$

$$c_2 = r_2 * (x_2 - \bar{x}_{\cdot}) = 10 * \left(15 - \frac{430}{29}\right) = 1.72$$

$$c_3 = r_3 * (x_3 - \bar{x}_{\cdot}) = 9 * \left(20 - \frac{430}{29}\right) = 46.56$$

Therefore, the corresponding contrast coefficients for the linear trend of factor B in terms of β_i^* are:

$$[-48.28, 1.72, 46.56]$$

Since

$$-48.28 * \beta_1^* + 1.72 * \beta_2^* + 46.56 * \beta_3^* = -48.28 * \bar{\tau}_{\cdot 1} + 1.72 * \bar{\tau}_{\cdot 2} + 46.56 * \bar{\tau}_{\cdot 3}$$

$$= -48.28 * \frac{1}{2}(\tau_{11} + \tau_{21}) + 1.72 * \frac{1}{2}(\tau_{12} + \tau_{22}) + 46.56 * \frac{1}{2}(\tau_{13} + \tau_{23})$$

The corresponding contrast coefficients for the linear trend of factor B in terms of τ_{ij} are:

$$[-24.14, 0.86, 23.28, -24.14, 0.86, 23.28]$$

Week 4 Practice Problem Discussion: Writing Contrasts as Coefficient Lists for the Meat-Cooking Experiment

Note that factor A, cooking method, has two levels: frying and grilling while factor B, fat content, has three levels of 10%, 15%, and 20%. The factor B has three equally spaced quantitative levels. Find the contrast coefficients for the following contrasts:

(1) Pairwise difference between treatment combination 21 and 23 in terms of τ_{ij}.

(2) Interaction between levels of 1 and 2 from factor A and levels 2 and 3 from factor B in terms of τ_{ij} or $(\alpha\beta)_{ij}$.

(3) The difference of response of levels 1 and 2 from factor A in terms of τ_{ij} or α_i^*.

(4) The linear of factor B from the balanced experiment in terms of τ_{ij} or β_i^*.

Solution:

(1) The contrast for the difference between treatment combination 21 and 23 is:

$$\tau_{21} - \tau_{23}$$

So the corresponding contrast coefficients in terms of τ_{ij} are:

$$[0, 0, 0, 1, 0, -1]$$

(2) One contrast for the interaction between levels of 1 and 2 from factor A and levels 2 and 3 from factor B is:

$$(\tau_{12} - \tau_{22}) - (\tau_{13} - \tau_{23}) = ((\alpha\beta)_{12} - (\alpha\beta)_{22}) - ((\alpha\beta)_{13} - (\alpha\beta)_{23})$$

So the corresponding contrast coefficients in terms of τ_{ij} are:

$$[0,1, -1 ,0, -1, 1]$$

So the corresponding contrast coefficients in terms of $(\alpha\beta)_{ij}$ are also:

$$[0,1, -1 ,0, -1, 1]$$

(3) The contrast for the difference of response of levels 1 and from factor A is:

$$\alpha_1^* - \alpha_1^* = \bar{\tau}_{1\cdot} - \bar{\tau}_{2\cdot}$$

$$= \frac{1}{3}(\tau_{11} + \tau_{12} + \tau_{13}) - \frac{1}{3}(\tau_{21} + \tau_{22} + \tau_{23})$$

So the corresponding contrast coefficients in terms of α_i^* are:

$$[1, -1]$$

So the corresponding contrast coefficients in terms of τ_{ij} are:

$$[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}]$$

(4) The factor B has three equally spaced levels, so the contrast coefficients for the linear trend in terms of β_i^*, can be obtained from Table A.2:

$$[-1,0,1].$$

Therefore, the contrast is:

$$-\beta_1^* + \beta_3^* = -\bar{\tau}_{\cdot 1} + \bar{\tau}_{\cdot 3} = -\frac{1}{2}(\tau_{11} + \tau_{21}) + \frac{1}{2}(\tau_{13} + \tau_{23})$$

So the corresponding contrast coefficients in terms of τ_{ij} are:

$$[-0.5, 0, 0.5, -0.5, 0, 0.5]$$

Least Squares Estimators and Estimates

Since the two-complete model is equivalent to the cell-means model and the cell-means model is actually a one-analysis of variance model, we can easily use the results from the one-way analysis of variance model to obtain the least squares estimators of $\mu + \tau_{ij}$ from the cell-means model and $\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$ from the two-way complete model and σ^2.

The least squares estimator of $\mu + \tau_{ij}$ and $\mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$ is just the sample mean of response for treatment combination ij, which is:

$$\hat{\mu} + \hat{\tau}_{ij} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \widehat{(\alpha\beta)}_{ij} = \bar{Y}_{ij\cdot}$$

The unbiased estimate of σ^2 is:

$$\hat{\sigma}^2 = msE = \frac{ssE}{n - a * b}$$

The sum of squares for error (ssE) is:

$$ssE = \sum_{i=1}^{a} \sum_{j=1}^{b} \sum_{t=1}^{r_{ij}} (y_{ijt} - \bar{y}_{ij\cdot})^2$$

Since $\bar{Y}_{ij\cdot}$ has a normal distribution with mean of $\mu + \tau_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij}$ and variance of $\frac{\sigma^2}{r_{ij}}$, we can easily derive the point estimate, the confidence interval, and the hypothesis testing of a contrast $\sum_{i=1}^{a} \sum_{j=1}^{b} d_{ij}\tau_{ij}$=$\sum_{i=1}^{a} \sum_{j=1}^{b} d_{ij} (\alpha_i + \beta_j + (\alpha\beta)_{ij})$ where $\sum_{i=1}^{a} \sum_{j=1}^{b} d_{ij} = 0$:

(1) The point estimate of the contrast $\sum_{i=1}^{a} \sum_{j=1}^{b} d_{ij}\tau_{ij} = \sum_{i=1}^{a} \sum_{j=1}^{b} d_{ij} (\alpha_i + \beta_j + (\alpha\beta)_{ij})$ is: $\sum_{i=1}^{a} \sum_{j=1}^{b} d_{ij}\tau_{ij} = \sum_{i=1}^{a} \sum_{j=1}^{b} d_{ij} \left(\alpha_i + \beta_j + (\alpha\beta)_{ij}\right) = \sum_{i=1}^{a} \sum_{j=1}^{b} d_{ij}\bar{y}_{ij\cdot}$

(2) The $(1 - \alpha)100\%$ confidence interval of the contrast $\sum_{i=1}^{a} \sum_{j=1}^{b} d_{ij}\tau_{ij} = \sum_{i=1}^{a} \sum_{j=1}^{b} d_{ij} (\alpha_i + \beta_j + (\alpha\beta)_{ij})$ is:

$$\sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij}\bar{y}_{ij\cdot} \pm t_{n-ab,\alpha/2}\sqrt{\sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij}^2 \frac{\hat{\sigma}^2}{r_{ij}}}$$

where $\hat{\sigma}^2 = msE = \frac{ssE}{n-ab}$ is the estimate of σ^2.

(3) The $p-$value for the two-sided test of

$$H_0: \sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij}\tau_{ij} = \sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij}\left(\alpha_i + \beta_j + (\alpha\beta)_{ij}\right) = 0$$

can be calculated by the following formula:

$$p - value = 2 * \Pr(T_{n-ab} > \frac{|\sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij}\bar{y}_{ij\cdot}|}{\sqrt{\sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij}^2 \frac{\hat{\sigma}^2}{r_{ij}}}})$$

For the unbalance data, we do not have the simple formulas of the least squares estimates of $\mu + \alpha_i$, $\mu + \beta_j$, or $\mu + (\alpha\beta)_{ij}$. So when we perform the statistical inference for contrast, we should rely on τ_{ij} and the contrast coefficients in terms of τ_{ij}.

Methods for Multiple Comparisons

When planning the experiment, the experimenter should specify which treatment contrasts are of interest, together with overall error rates for hypothesis tests and overall confidence levels for confidence intervals. If the two-way complete model has been selected, the following four parts of comparison may be of interest: (1) treatment combinations, (2) comparison of main effects of A, (3) comparison of main effects of B, and (4) interaction contrasts. If α is the overall significance, then the experimenter needs to set the significance level of ith part, $\alpha_i (i = 1,2,3,4)$ and $\alpha = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$. According to the Bonferroni method, the experiment-wise error rate will be not greater than $\alpha = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$. For the ith part, the experimenter can select the appropriate methods for multiple comparisons to control the overall error rate is not greater than α_i. For this sub-section, we outline how to perform the methods for multiple comparisons for a given part.

- **Comparing Treatment Combinations**

  When comparison of treatment combinations is of most interest, the cell-means model should be used. The formulae for the Bonferroni, Scheffé, Tukey, and Dunnett methods can all be used in the same way as was done in Lesson 3, but with $v = ab$. In other words, for a contrast for comparing treatment combinations, $\sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij}\tau_{ij}$, its simultaneous $100 * (1-\alpha)\%$ confidence interval is:

$$\sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij}\bar{y}_{ij\cdot} \pm w\sqrt{\sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij}^2\frac{\hat{\sigma}^2}{r_{ij}}}$$

  where the critical coefficient w for each of the four methods is, respectively,

  Bonferroni Method: $w_B = t_{n-ab,\alpha/(2*m)}$, m is the number pre-specified contrasts

  Scheffe Method: $w_s = \sqrt{(ab-1)F_{ab-1,n-ab,\alpha}}$

  Tukey Method: $w_T = \frac{q_{ab,n-ab,\alpha}}{\sqrt{2}}$ (best and only for pair-wise comparison)

  Dunnett Method: $w_{D2} = |t|^{(0.5)}_{ab-1,n-ab,\alpha}$ (best and only for treatment-versus-control comparison)

- **Interaction Contrasts**

  Again, the cell-means model should be used. Note that the interaction contrasts are not contrasts for pairwise comparisons nor for treatment-versus-control comparisons, so only the Bonferroni method and Scheffé method may be applied in this situation. The formulas of simultaneous $100 * (1-\alpha)\%$ confidence intervals are same as the formulas for calculating simultaneous $100 * (1-\alpha)\%$ confidence intervals of contrasts for comparing treatment combinations.

- **Comparing Main Effects**

  Main-effect contrasts compare the effects of the levels of one factor averaging over the levels of the other factor and may not be of interest if the two factors interact. If main-effect contrasts are to be examined, then the Bonferroni, Scheffé, Tukey, and Dunnett methods can be used for each factor separately.

  A contrast for comparing main effects of factor A of has the following formula:

$$\sum_{i=1}^{a} c_i \overline{\tau_i}_{\cdot} = \sum_{i=1}^{a} c_i \alpha_i^* = \sum_{i=1}^{a} \sum_{j=1}^{b} d_{ij} \tau_{ij}$$

where $\sum_{i=1}^{a} c_i = \sum_{i=1}^{a} \sum_{j=1}^{b} d_{ij} \tau_{ij} = 0$. Therefore, the simultaneous $(1-\alpha)100\%$ confidence interval is:

$$\sum_{i=1}^{a} \sum_{j=1}^{b} d_{ij} \bar{y}_{ij\cdot} \pm w \sqrt{\sum_{i=1}^{a} \sum_{j=1}^{b} d_{ij}^2 \frac{\hat{\sigma}^2}{r_{ij}}}$$

where the critical coefficient w for each of the four methods is, respectively,

Bonferroni Method: $w_B = t_{n-ab,\alpha/(2*m)}$, m is the number pre-specified contrasts

Scheffe Method: $w_s = \sqrt{(a-1)F_{a-1,n-ab,\alpha}}$

Tukey Method: $w_T = \frac{q_{a,n-ab,\alpha}}{\sqrt{2}}$ (best and only for pair-wise comparison)

Dunnett Method: $w_{D2} = |t|^{(0.5)}_{a-1,n-ab,\alpha}$ (best and only for treatment-versus-control comparison)

Note that the degrees of freedom for error is still $n-ab$ but the degrees of freedom for groups is a since the factor A has a levels. Similarly, the simultaneous $(1-\alpha)100\%$ confidence intervals of contrasts for comparing main effects of factor B can obtained.

Example 4.6: Statistical Inference of Contrasts - Meat-Cooking Experiment

Note that factor A, cooking method, has two levels: frying and grilling while factor B, fat content, has three levels of 10%, 15%, and 20%. Suppose the experimenter are interested in the following groups of treatment contrasts:

(1) All pairwise comparisons of treatment combinations.
(2) All pairwise comparisons of main effects of factor B.

Construct the simultaneous 95% confidence intervals of contrasts based on the balanced data.

Solution: Based on the Bonferroni method, we can select the select a significance level, α_1, for the first group of contrast and a significance level α_2, to achieve an experiment-wise error rate of 0.05 as long as $\alpha_1 + \alpha_2 = 0.05$. For illustration, we use $\alpha_1 = \alpha_2 = 0.025$.

To perform the statistical inference, we first calculate the sample mean of the response of treatment combinations:

$$\bar{y}_{11\cdot} = 84.4, \bar{y}_{12\cdot} = 81.8, \bar{y}_{13\cdot} = 76.0, \bar{y}_{21\cdot} = 83.4, \bar{y}_{22\cdot} = 86.0, \bar{y}_{23\cdot} = 78.4$$

For the balance data, we have

$$r_{ij} = 5; i = 1,2; j = 1,2,3$$

Then the mean square for error (msE) is also obtained:

$$msE = \hat{\sigma}^2 = \frac{ssE}{30-6} = 7.26667$$

(1) For this group of contrasts, we use a contrast for comparing the response between treatment combinations 12 (frying with 10% fat) and 22 (grilling with 10% fat) as an illustration.

The contrast is:

$$\tau_{12} - \tau_{22}$$

n terms of τ_{ij}, the contrast coefficients are:

$$[0,1,0,0,-1,0]$$

The point estimate is:

$$\bar{y}_{12\cdot} - \bar{y}_{22\cdot} = 81.8 - 86.0 = -4.2$$

The standard error is:

$$\sqrt{msE * \sum_{i=1}^{a}\sum_{j=1}^{b}\frac{d_{ij}^2}{r_{ij}}} = \sqrt{7.2667 * (\frac{1^2}{5} + \frac{(-1)^2}{5})} = 1.7049$$

The simultaneous 97.5% confidence interval can be written as:

$$-4.2 \pm w * 1.7049$$

The 97.5% confidence interval without method for multiple comparisons:

$$w = t_{n-ab,\alpha/2} = t_{30-2*3,0.025/2} = 2.3909$$

$$-4.2 \pm 2.3909 * 1.7049 = (-8.2763, -0.1237)$$

The simultaneous 97.5% confidence interval with the Bonferroni method:

$$m = 15, w = w_B = t_{n-ab,\alpha/(2*m)} = t_{30-2*3,0.025/(2*15)} = 3.5405$$

$$-4.2 \pm 3.5405 * 1.7049 = (-10.2361, 1.8361)$$

The simultaneous 97.5% confidence interval with Scheffe method is:

$$w = w_s = \sqrt{(ab-1)F_{ab-1,n-ab,\alpha}} = \sqrt{(2*3-1)F_{2*3-1,30-2*3,0.025}} = 3.9717$$

$$-4.2 \pm 3.9717 * 1.7049 = (-10.9713, 2.5712)$$

The simultaneous 97.5% confidence interval with Tukey method is:

$$w = w_T = \frac{q_{a*b,n-ab,\alpha}}{\sqrt{2}} = \frac{q_{2*3,30-2*3,0.025}}{\sqrt{2}} = 3.4051$$

$$-4.2 \pm 3.4051 * 1.7049 = (-10.0054, 1.6054)$$

Dunnett method is not applicable here. The confidence interval without multiple comparison adjustment is just for illustration here and is not appropriate either. The confidence interval from Tukey is shortest. The confidence intervals obtained from the Bonferroni method, Scheffe method, and Tukey method contain 0, meaning that we do have sufficient evidence to support that the treatment combination of frying with 10% fat and the treatment combination of grilling with 10% fat have the different effect on the weight loss.

(2) For this group of contrasts, we use a contrast for comparing the response between 15% fat content and 20% fat content as an illustration.

The contrast is

$$\beta_2^* - \beta_3^* = \bar{\tau}_{\cdot 2} - \bar{\tau}_{\cdot 3}$$
$$= \frac{1}{2}(\tau_{12} + \tau_{22}) - \frac{1}{2}(\tau_{13} + \tau_{23})$$

So the corresponding contrast coefficients in terms of β_i^* are:

$$[0, 1, -1]$$

So the corresponding contrast coefficients in terms of τ_{ij} are:

$$[0, 0.5, -0.5, 0, 0.5, -0.5]$$

The point estimate is:

$$0.5 * (\bar{y}_{12\cdot} + \bar{y}_{22\cdot}) - 0.5 * (\bar{y}_{13\cdot} + \bar{y}_{23\cdot})$$

$$= 0.5 * (81.8 + 86.0) - 0.5 * (76.0 + 78.4) = 6.7$$

The standard error is:

$$\sqrt{msE * \sum_{i=1}^{a} \sum_{j=1}^{b} \frac{d_{ij}^2}{r_{ij}}} = \sqrt{7.2667 * \left(\frac{0.5^2}{5} + \frac{(-0.5)^2}{5} + \frac{0.5^2}{5} + \frac{(-0.5)^2}{5}\right)} = 1.2055$$

The simultaneous 97.5% confidence interval can be written as:

$$6.7 \pm w * 1.2055$$

The 97.5% confidence interval without method for multiple comparisons:

$$w = t_{n-ab,\alpha/2} = t_{30-2*3,0.025/2} = 2.3909$$

$$6.7 \pm 2.3909 * 1.2055 = (3.8176, 9.5824)$$

The simultaneous 97.5% confidence interval with the Bonferroni method:

$$m = 3, w = w_B = t_{n-ab,\alpha/(2*m)} = t_{30-2*3,0.025/(2*3)} = 2.8751$$

$$6.7 \pm 2.8751 * 1.2055 = (3.2340, 10.1661)$$

The simultaneous 97.5% confidence interval with Scheffe method is:

$$w = w_s = \sqrt{(a * b - 1)F_{b-1,n-ab,\alpha}} = \sqrt{(3-1)F_{3-1,30-2*3,0.025}} = 2.9390$$

$$6.7 \pm 2.9390 * 1.2055 = (3.1569, 10.2430)$$

The simultaneous 97.5% confidence interval with Tukey method is:

$$w = w_T = \frac{q_{a*b,n-ab,\alpha}}{\sqrt{2}} = \frac{q_{2*3,30-2*3,0.025}}{\sqrt{2}} = 2.8161$$

$$6.7 \pm 2.8161 * 1.2055 = (3.3051, 10.0949)$$

Dunnett method is not applicable here. The confidence interval without multiple comparison adjustment is just for illustration here and is not appropriate either. The confidence interval from Tukey is shortest. The confidence intervals obtained from the Bonferroni method, Scheffe method, and Tukey method do not contain 0, meaning that we can state 15% fat content and 20% fat content have different effects on the weight loss, averaged over the two cooking methods. With experiment-wise confidence level at

least 95%, the weight loss with 15% fat content is between 3.31 and 10.09 greater that that with 20% fat content from Tukey method.

Exercise Problem – Meat-Cooking Experiment.

Repeat the above analysis with unbalanced data.

Lesson 4.4 Analysis of Cell-means Model and Two-way Complete Model: Analysis of Variance

Required Reading: Section 6.4 from Chapter 6

Analysis of Variance

The first hypothesis for the cell-means model or the two-way complete model that is usually examined is if the treatment combinations have the different effects on the response; that is:

$$H_0: \tau_{11} = \tau_{12} = \cdots = \tau_{ab} \text{ (in other words, here } H_0 \text{ is: all } \tau_{ij} \text{ are same)}$$

versus

$$H_1: at\ least\ two\ of\ \tau_{ij} \text{ are different}$$

Or equivalently, for the two-way complete model, the null hypothesis is:

$$H_0: \alpha_1 + \beta_1 + (\alpha\beta)_{11} = \alpha_1 + \beta_2 + (\alpha\beta)_{12} = \cdots = \alpha_a + \beta_b + (\alpha\beta)_{ab}$$

The null hypothesis is actually based on $ab - 1$ estimable contrasts. According to the previous lesson, the F-test statistic can be used:

$$F = \frac{SST/(ab-1)}{SSE/(n-ab)}$$

Such test statistic has a F-distribution with $ab - 1$ and $n - ab$ degrees of freedom, and

$$SSE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{t=1}^{r_{ij}}\left(Y_{ijt} - \bar{Y}_{ij\cdot}\right)^2$$

$$SSTot = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{t=1}^{r_{ij}}\left(Y_{ijt} - \bar{Y}_{\cdots}\right)^2$$

$$SST = SSTot - SSE$$

As we have mentioned, such F-test statistic can be derived by comparing two models: the full model which is the model used and the reduced model which is the model under the null hypothesis. For this null hypothesis, the reduced model is:

$$Y_{ijt} = \mu + \epsilon_{ijt}, t = 1, \cdots, r_{ij}; i = 1, \cdots, a; j = 1, \cdots, b$$

The full model is:

$$Y_{ijt} = \mu + \tau_{ij} + \epsilon_{ijt} = \mu + \ \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijt}, t = 1, \cdots, r_{ij}; i = 1, \cdots, a; j = 1, \cdots, b$$

The F-test statistic is:

$$F = \frac{(SSE_{Reduced} - SSE_{Full})/(df\ of\ SSE\ in\ Reduced\ Model - df\ of\ SSE\ in\ Full\ Model)}{SSE_{Full}/(df\ of\ SSE\ in\ Full\ Model)}$$

Note the reduced model and the full model has the same total sum of squares, we can re-write the test statistic as:

$$F = \frac{(SST_{Full} - SST_{Reduced})/(d.f\ of\ SST\ in\ Full\ Model - d.f\ of\ SST\ in\ Reduced\ Model)}{SSE_{Full}/(d.f\ of\ SSE\ in\ Full\ Model)}$$

It is easy to verify that for this reduced model, the sum of squares for error is same as the total sum of squares, the sum of squares for treatments is 0 and the degrees freedom for treatment is also 0, so this F-test statistic is same as we used before:

$$F = \frac{SST_{Full}/(ab-1)}{SSE_{Full}/(n-ab)}$$

Once this hypothesis is rejected, there are three standard hypotheses that are usually examined when the two-way complete model is used. The first hypothesis is that the interaction between treatment factors A and B is negligible; that is,

$$H_0^{AB}: \{(\alpha\beta)_{ij} - (\alpha\beta)_{iq} - (\alpha\beta)_{sj} + (\alpha\beta)_{sq} = 0\ for\ all\ i \neq s, j \neq q\}$$

Although this null hypothesis contains $\frac{a*(a-1)}{2} * \frac{b*(b-1)}{2}$equaitons, it is actually based on $(a-1)(b-1)$ estimable contrasts, and the test is based on $(a-1)(b-1)$ degrees of freedom. To test hypotheses, the F-test statistic can be written as:

$$F = \frac{(SSE_{Reduced} - SSE_{Full})/(d.f\ of\ SSE\ in\ Reduced\ Model - d.f\ of\ SST\ in\ Full\ Model)}{SSE_{Full}/(d.f\ of\ SSE\ in\ Full\ Model)}$$

Here, the full model is:

$$Y_{ijt} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijt}, t = 1, \cdots, r_{ij}; i = 1, \cdots, a; j = 1, \cdots, b$$

And the reduced model, which is the model when H_0^{AB} is true, is actually the two-way main-effects model:

$$Y_{ijt} = \mu + \alpha_i + \beta_j + \epsilon_{ijt}, t = 1, \cdots, r_{ij}; i = 1, \cdots, a; j = 1, \cdots, b$$

For simplicity, we always use SSE to represent the sum of squares for error in the full model.

The degrees of freedom in the reduced model and full model are $(a - 1 + b - 1)$ and $ab - 1$, respectively. Therefore, the F-test statistic can be written as:

$$F = \frac{(SSE_{Reduced} - SSE)/(ab - 1 - (a - 1 + b - 1))}{SSE/(n - ab)}$$

$$= \frac{SSAB/((a-1)(b-1))}{SSE/(n-ab)} = \frac{MSAB}{MSE}$$

In the above formula, $SSAB$ and $MSAB$ are the sum of squares for interaction, and the mean square for interaction, respectively. The F-test statistic has a F-distribution with $(a-1)(b-1)$ and $n - ab$ degrees of freedom when the null hypothesis (H_0^{AB}) is true. Thus, H_0^{AB} is rejected if

$$F = \frac{ssAB/((a-1)(b-1))}{ssE/(n-ab)} = \frac{msAB}{msE} > F_{(a-1)(b-1),n-ab,\alpha}$$

For a balanced experiment, the relatively simple formulas for the calculation of $ssAB$ can be derived. For an unbalanced experiment, there are no simple formulas for the calculation of $ssAB$ can be derived. Therefore, we mainly rely on R to calculate $ssAB$ for the balanced or unbalanced experiment.

The other two standard hypothesis are the main-effect hypotheses:

$$H_0^A: \alpha_1^* = \alpha_2^* = \cdots = \alpha_a^*$$

$$H_0^B: \beta_1^* = \beta_2^* = \cdots = \beta_b^*$$

where $\alpha_i^* = \alpha_i + (\overline{\alpha\beta})_{i\cdot}$ and $\beta_j^* = \beta_j + (\overline{\alpha\beta})_{\cdot j}$. However, these main-effect hypotheses may not be of interest if there is a sizable interaction. Each of the main-effect hypotheses can be rephrased in terms of estimable contrasts in the parameters, and so can be tested. As in the previous lessons, the tests will be based on $(a - 1)$ and $(b - 1)$ degrees of freedom, respectively. Correspondingly, we reject H_0^A if:

$$F = \frac{ssA/(a-1)}{SSE/(n-ab)} = \frac{msA}{msE} > F_{a-1,n-ab,\alpha}$$

We reject H_0^B if:

$$F = \frac{ssB/(b-1)}{SSE/(n-ab)} = \frac{msB}{msE} > F_{b-1,n-ab,\alpha}$$

Again, we rely on R to calculate ssA and ssB.

In summary, the following analysis of variance table can be obtained for the two-way complete model:

Source of Variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**F-test Statistic**
Factor A	$a-1$	ssA	$msA = \frac{ssA}{a-1}$	$\frac{msA}{msE}$
Factor B	$b-1$	ssB	$msB = \frac{ssB}{b-1}$	$\frac{msB}{msE}$
AB	$(a-1)(b-1)$	$ssAB$	$msAB = \frac{ssAB}{(a-1)(b-1)}$	$\frac{msAB}{msE}$
Error	$n-ab$	ssE	$msE = \frac{ssE}{n-ab}$	

Total	$n-1$	$sstot$		

Type I and Type III Sum of Squares

There are different ways to calculate ssA, ssB, and $ssAB$. Here, we introduce two types of sum of squares: Type I and Type III sum of squares.

Type I sum of squares is also called sequential sum of squares. Type I sum of squares of each term is adjusted for only term that precedes it in the model. For a two-way complete model,

$$y_{ijt} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijt}$$

We assume that the factor A, factor B, and the interaction of AB are added to the model one by one, then we can have the following four models:

$$M0: y_{ijt} = \mu + \epsilon_{ijt}$$

$$M1: y_{ijt} = \mu + \alpha_i + \epsilon_{ijt}$$

$$M2: y_{ijt} = \mu + \alpha_i + \beta_j + \epsilon_{ijt}$$

$$M3: y_{ijt} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijt}$$

Then the type I sum of squares for factor $A, B,$ and their interaction can be calculated by the following way:

$$ssA = ssE\ of\ model\ M0 - \ ssE\ of\ model\ M1$$

$$ssB = ssE\ of\ model\ M1 - \ ssE\ of\ model\ M2$$

$$ssAB = ssE\ of\ model\ M2 - \ ssE\ of\ model\ M3$$

For example, ssA is the difference of the sum of squares for error between $M0$ and $M1$ since the factor A is first term in the model, while ssB is the difference of the sum of squares for error between $M1$ and $M2$ since the factor B is second term after the factor A in the model. Therefore, the order of terms in the model affect the type I sum of squares. In addition, we have the following equation for the type I sum of squares:

$$sstot = ssA + ssB + ssAB + ssE$$

And

$$ssT = ssA + ssB + ssAB$$

Type III sum of squares for each term is adjusted for any other terms that don't contain the effect and orthogonal to any effects that contain the effect. For example, for a two-way complete model, ssA is the difference of the sum of squares for error between the two-way complete model and a reduced model. Since the factor B does not contain the effects of factor A and the interaction effects contain the effects of factor A, here the reduce model contain the effects of factor B and the effects from the interaction that are orthogonal to the interaction effects that contain the effects of factor A. Therefore, the order of terms in the model does not affect the type III sum of squares. To test the hypotheses H_0^{AB}, H_0^A, or H_0^B, the type III sum of squares should be used.

For balanced experiments, the type I sum of squares of a term is same its type III sum of squares; while for unbalance experiments, the type I sum of squares of a term is generally different from its type III sum of squares.

For the type III sum of squares to be correct, the way R codes factors has to be changed from its default with the options(contrasts =…) function before R function aov is used. Changing this will not affect the type I sum of squares.

- Default: options(contrasts = c("contr.treatment", "contr.poly"))
- For correct type III sum of squares: options(contrasts = c("contr.sum", "contr.poly"))

Example 4.7: Type I and Type III Sum of Squares from Meat-Cooking Experiment, Balanced

The following model is used in the example:

$$y_{ijt} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijt}$$

In R program, we use two orders of terms. order of term

Note that in this model, the order of terms are factor A (cooking method), factor B (fat content), and the interaction of factors A and B.

The analysis of variance table for the type I sum of squares:

Source of Variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**F-test Statistic**	***p*-value**
Cooking Method	1	26.133	26.133	3.596	0.070
Fat Content	2	299.267	149.633	20.592	$6.21 * 10^{-6}$
Interaction	2	34.867	17.433	2.399	0.112
Error	24	174.400	7.267		
Total	29	534.667			

For the null hypothesis of

$$H_0: \alpha_1 + \beta_1 + (\alpha\beta)_{11} = \alpha_1 + \beta_2 + (\alpha\beta)_{12} = \cdots = \alpha_a + \beta_b + (\alpha\beta)_{ab}$$

We can calculate F-test statistic from the type I sum of squares:

$$F = \frac{msT}{msE} = \frac{(26.133 + 299.267 + 34.867)\Big/(1 + 2 + 2)}{7.267} = 9.916$$

The corresponding p-value is:

$$p - value = \Pr\left(F_{5,24} \geq 9.916\right) = 3.08 * 10^{-5}$$

The F-test statistic and p-value can be directly obtained by either using a cell-means model or the comparison of the two-way complete model and the reduced model under H_0. Our conclusion here is that the different treatment combinations have different effects on the weight loss of meat.

For this balanced experiment, the type III sum of squares is same as the type I sum of squares. For the above table, the p-values for the hypotheses of H_0^A, H_0^B, and H_0^{AB} are $0.070, 6.21 * 10^{-6}$, and 0.112, respectively. So we have the following conclusions:

- At the significance level of 0.05, we fail to reject H_0^{AB}so we do not have sufficient evidence to support that there are non-negligible interaction effects between two treatment factors: the cooking method and the meat content.
- We fail to reject H_0^A at the significance level of 0.05 and conclude that there is no sufficient evidence for a difference in weight loss of two cooking methods averaged over three fat contents.
- We reject H_0^B at the significance level of 0.05 and conclude that there is a significant difference in weight loss of three fat contents averaged over two cooking methods.

Example 4.8: Type I and Type III Sum of Squares from Meat-Cooking Experiment, Unbalanced

The following model is used in the example:

$$y_{ijt} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijt}$$

In R program, the first order of terms is factor A (cooking method), factor B (fat content), and the interaction of factors A and B while the second order of terms is factor B (fat content), factor A (cooking method), and the interaction of factors A and B

The analysis of variance table for the type I sum of squares is (the order of terms: Cooking Method, Fat Content, and interaction):

Source of Variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**F-test Statistic**	***p*-value**
Cooking Method	1	26.404	26.404	3.839	0.062

Fat Content	2	316.132	158.066	22.981	$3.28 * 10^{-6}$
Interaction	2	33.815	16.980	2.458	0.108
Error	23	158.200	6.878		
Total	28	534.551			

The analysis of variance table for the type I sum of squares is (the order of terms: Fat Content, Cooking Method, and interaction):

Source of Variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**F-test Statistic**	***p*-value**
Fat Content	2	324.75	162.376	23.607	$2.68 * 10^{-6}$
Cooking Method	1	17.78	17.785	2.586	0.122
Interaction	2	33.815	16.980	2.458	0.108
Error	23	158.200	6.878		
Total	28	534.551			

Note that ssA and ssB are different while $ssAB$ are same since the interaction is the last term in both models.

For the null hypothesis of

$$H_0: \alpha_1 + \beta_1 + (\alpha\beta)_{11} = \alpha_1 + \beta_2 + (\alpha\beta)_{12} = \cdots = \alpha_a + \beta_b + (\alpha\beta)_{ab}$$

We can calculate F-test statistic from the type I sum of squares of either tables:

$$F = \frac{msT}{msE} = \frac{(324.75 + 17.78 + 33.815)\Big/(1 + 2 + 2)}{6.878} = 10.943$$

The corresponding p-value is:

$$p - value = \Pr\left(F_{5,23} \geq 10.943\right) = 1.76 * 10^{-5}$$

For this unbalanced experiment, the type III sum of squares is different from the Type I sum of squares:

Source of Variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**F-test Statistic**	***p*-value**
Cooking Method	1	17.672	17.672	2.569	0.123
Fat Content	2	314.600	157.300	22.869	$3.41 * 10^{-6}$
Interaction	2	33.815	16.980	2.458	0.108
Error	23	158.200	6.878		
Total	28	534.551			

Note the type III sum of squares of ssA and ssB are different from those type I sum of squares. All conclusions from the unbalanced experiment are same as those from the balanced experiment.

Exercise 4.3: Type I and Type III Sum of Squares – Mung Bean Experiment

An experiment was run in 1993 by K.H. Chen, Y.F. Kuo, R. Sengupta, J. Xu, and L.L. Yu to compare watering schedules and growing mediums for mung bean seeds. There were 6 treatment levels. Forty-eight beans of approximately equal weights were randomly selected for the experiment. These were all soaked in water in a single container for two hours. After this time, the beans were placed in separate containers and randomly assigned to a treatment (water/medium) combination in such a way that eight containers were assigned to each treatment combination. The

48 containers were placed on a table in a random order. The shoot lengths of the beans were measured (in mm) after one week. There are two treatment factors: "amount of water" with three levels (1, 2, and 3 teaspoons of water per day) and "growing medium" having two levels (tissue and paper towel, coded 1 and 2).

1 – water level 1, tissue towel (11)	2 – water level 1, paper towel (12)
3 – water level 2, tissue towel (21)	4 – water level 2, paper towel (22)
5 – water level 3, tissue towel (31)	6 – water level 3, paper towel (32)

(1) This is a balanced experiment. Following the steps of the above example, construct the analysis of variance table for type I and type III sum of squares from the two-way complete model, perform the hypothesis testing for H_0, H_0^A, H_0^B, and H_0^{AB}, and draw your conclusions.

(2) Remove the last observation from your data and repeat the analysis of (1).

R Program for Lesson 4.4

Please refer to the R program and videos for more details.

Lesson 4.5 Analysis of Two-way Main-Effects Model

Require Reading: Section 6.5 from Chapter 6

If the interaction is negligible, the following two-way main-effects model can be used:

$$Y_{ijt} = \mu + \alpha_i + \beta_j + \epsilon_{ijt}, t = 1, \cdots, r_{ij}; i = 1, \cdots, a; j = 1, \cdots, b$$

with the following assumptions:

- μ, α_i, β_j, and $(\alpha\beta)_{ij}$ $(i = 1, \cdots, a; j = 1, \cdots, b)$ are unknown population parameters
- $\epsilon_{ijt}(t = 1, \cdots, r_i; i = 1, \cdots, a; j = 1, \cdots, b)$ are independent
- $\epsilon_{ijt} \sim N(0, \sigma^2)$: normal distribution with mean 0 and variance σ^2

The analysis of this model is similar to the analysis of two-way complete model.

Contrasts and Estimable Functions

An estimable function from the two-way main-effects model has the following representation:

$$\sum_{i=1}^{a} \sum_{j=1}^{b} d_{ij}(\ \mu + \alpha_i + \beta_j)$$

If $\sum_{i=1}^{a} \sum_{j=1}^{b} d_{ij} = 0$, then it is also a contrast.

Example 4.9: Contrasts of Two-Way Main Effects Model

Determine if the following functions of parameters from the two-way main-effects model are estimable and/or a contrast.

(1) $\mu + \alpha_1$

(2) $\mu + \frac{\alpha_1 + \alpha_2}{2} + \beta_1$

(3) $\alpha_2 - \alpha_1$

Solution:

(1) Not estimable.

(2) Estimable but not a contrast. Since we can let $d_{11} = d_{21} = \frac{1}{2}$ and all other $d_{ij} = 0$, then

$$\sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij}(\ \mu + \alpha_i + \beta_j) = \frac{1}{2}(\mu + \alpha_1 + \beta_1) + \frac{1}{2}(\mu + \alpha_2 + \beta_1)$$

$$= \mu + \frac{\alpha_1 + \alpha_2}{2} + \beta_1$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij} = \frac{1}{2} + \frac{1}{2} = 1$$

(3) Contrast. Since we can let $d_{11} = -1$ and $d_{21} = 1$ and all other $d_{ij} = 0$, then

$$\sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij}(\ \mu + \alpha_i + \beta_j) = -(\mu + \alpha_1 + \beta_1) + (\mu + \alpha_2 + \beta_1) = \alpha_2 - \alpha_1$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b} d_{ij} = -1 + 1 = 0$$

Since there is no interaction in the two-way main-effects model, the following contrasts are our primary of interest:

$$\textstyle\sum_{i=1}^{a} c_i\alpha_i \text{ where } \sum_{i=1}^{a} c_i = 0$$

And

$$\textstyle\sum_{j=1}^{b} k_j\beta_j \text{ where } \sum_{j=1}^{b} k_j = 0$$

Least Squares Estimators of Parameters for Balanced Experiments

The least squares estimator for $\mu + \alpha_i + \beta_j$ has a nice formula for balance experiments. For balanced experiments (all $r_{ij} = r$), the least squares estimator for $\mu + \alpha_i + \beta_j$ is:

$$\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j = \bar{Y}_{i\cdot\cdot} + \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot}$$

Then the least squares estimator for the contrast $\sum_{i=1}^{a} c_i\alpha_i$ is:

$$\sum_{i=1}^{a} c_i\hat{\alpha}_i = \sum_{i=1}^{a} c_i\left(\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j\right) = \sum_{i=1}^{a} c_i\left(\bar{Y}_{i\cdot\cdot} + \bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot}\right)$$

$$= \sum_{i=1}^{a} c_i\,\bar{Y}_{i\cdot\cdot} \sim N\left(\sum_{i=1}^{a} c_i\alpha_i\,, \frac{\sigma^2}{br}\sum_{i=1}^{a} c_i^2\right)$$

Similarly, the least squares estimator for the contrast $\sum_{j=1}^{b} k_j\beta_j$ is:

$$\sum_{j=1}^{b} k_j\beta_j = \sum_{j=1}^{b} k_j\,\bar{Y}_{.j.} \sim N\left(\sum_{j=1}^{b} k_j\beta_j, \frac{\sigma^2}{ar}\sum_{j=1}^{b} k_j^2\right)$$

The least squares estimator of σ^2 is:

$$\hat{\sigma}^2 = MSE = \frac{SSE}{n-a-b+1} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{t=1}^{r_{ij}}\left(Y_{ijt} - \hat{Y}_{ijt}\right)^2}{n-a-b+1}$$

$$= \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{t=1}^{r_{ij}}\left(Y_{ijt} - (\bar{Y}_{i..} + \bar{Y}_{.j.} - \bar{Y}_{...})\right)^2}{n-a-b+1}$$

The degrees of freedom for error in the two-way main-effects model is

$$(n-1) - (a-1+b-1)$$

Since any contrasts can be written as a function of $a-1$ contrasts in terms of α_i and $b-1$ contrasts in terms of β_j.

Based on these least squares estimators, we can perform the statistical inference of an contrast (for example, $\sum_{i=1}^{a} c_i\alpha_i$ or $\sum_{j=1}^{b} k_j\beta_j$) including the point estimate, the confidence interval, and the hypothesis testing. Specifically, for a contrast $\sum_{i=1}^{a} c_i\alpha_i$:

- The least squares estimate is: $\sum_{i=1}^{a} c_i\,\bar{y}_{i..}$
- The $(1-\alpha)*100\%$ confidence interval is:

$$\sum_{i=1}^{a} c_i\,\bar{y}_{i..} \pm t_{n-a-b+1,\alpha/2}\sqrt{\frac{msE}{br}\sum_{i=1}^{a} c_i^2}$$

- The p-value for testing $H_0: \sum_{i=1}^{a} c_i\alpha_i = 0$ is:

$$p-value = 2 * \Pr\left(T_{n-a-b+1} \geq \frac{\left|\sum_{i=1}^{a} c_i\,\bar{y}_{i..}\right|}{\sqrt{\frac{msE}{br}\sum_{i=1}^{a} c_i^2}}\right)$$

- The simultaneous $(1-\alpha)*100\%$ confidence interval is:

$$\sum_{i=1}^{a} c_i\,\bar{y}_{i..} \pm w * \sqrt{\frac{msE}{br}\sum_{i=1}^{a} c_i^2}$$

Bonferroni Method: $w_B = t_{n-a-b+1,\alpha/(2*m)}$, m is the number pre-specified contrasts

Scheffe Method: $w_s = \sqrt{(a-1)F_{a-1,n-a-b+1,\alpha}}$

Tukey Method: $w_T = \frac{q_{a,n-a-b+1,\alpha}}{\sqrt{2}}$ (best and only for pair-wise comparison)

Dunnett Method: $w_{D2} = |t|^{(0.5)}_{a-1,n-a-b+1,\alpha}$ (best and only for treatment-versus-c control comparison)

Note: Although the above formulas look very complicated, they are same as the formulas used in a one-way analysis of variance model in which the treatment factor is A with a levels and br replicates for each level, except that the degrees of freedom for errors $n-a-b+1$ from the two-way main-effects model instead of the degrees of freedom for errors $n-a$ from the one-way analysis of variance model is used.

Similarly, the statistical inference for a contrast $\sum_{j=1}^{b} k_j\beta_j$ can be derived.

Example 4.10: Two-Main Main-Effects Model - Meat-Cooking Experiment, Balanced

Suppose that we are interested in the pairwise differences in the effects of three fat contents, the corresponding contrasts are:

$$\beta_1-\beta_2, \beta_1-\beta_3, \beta_2-\beta_3$$

From the data, we can get:

$$\bar{y}_{\cdot 1\cdot} = 83.9, \bar{y}_{\cdot 2\cdot} = 83.9, \bar{y}_{\cdot 3\cdot} = 77.2, \hat{\sigma}^2 = msE = 8.049, a = 2, b = 3, r = 5, n = 30$$

Note that for balance experiments, we can easily verify that

$$\bar{y}_{i\cdot\cdot} = \frac{\bar{y}_{i1\cdot} + \bar{y}_{i2\cdot} + \cdots + \bar{y}_{ib\cdot}}{b}$$

$$\bar{y}_{\cdot j\cdot} = \frac{\bar{y}_{1j\cdot} + \bar{y}_{2j\cdot} + \cdots + \bar{y}_{aj\cdot}}{a}$$

Therefore, we can also obtain $\bar{y}_{i\cdot\cdot}$ and $\bar{y}_{\cdot j\cdot}$ through $\bar{y}_{ij\cdot}$ for the balanced experiments.

Here we use $\beta_1 - \beta_3$ as an illustration:

- The estimated standard error is: $\sqrt{\frac{8.049}{2*5} * (1^2 + (-1)^2)} = 1.2688$
- The point estimation: $\hat{\beta}_1 - \hat{\beta}_3 = \bar{y}_{\cdot 1\cdot} - \bar{y}_{\cdot 3\cdot} = 83.9 - 77.2 = 6.7$
- 95% confidence interval:

  Without adjustment for multiple comparisons:

$$w = t_{n-a-b+1,\alpha/2} = t_{30-2-3+1,0.05/2} = 2.0555$$

$$6.7 \pm 2.0555 * 1.2688 = (4.09, 9.31)$$

Bonferroni Method:

$$w_B = t_{n-a-b+1,\alpha/(2*m)} = t_{30-2-3+1,0.05/(2*3)} = 2.5589$$

$$6.7 \pm 2.5589 * 1.2688 = (3.45, 9.95)$$

Scheffe Method:

$$w_s = \sqrt{(b-1)F_{b-1,n-a-b+1,\alpha}} = \sqrt{(3-1)F_{3-1,30-2-3+1,0.05}} = 2.5958$$

$$6.7 \pm 2.5958 * 1.2688 = (3.41, 9.99)$$

Tukey Method:

$$w_T = \frac{q_{b,n-a-b+1,\alpha}}{\sqrt{2}} = \frac{q_{3,30-2-3+1,0.05}}{\sqrt{2}} = 2.4849$$

$$6.7 \pm 2.4849 * 1.2688 = (3.55, 9.85)$$

Tukey method gives the shortest simultaneous 95% confidence interval for $\beta_1 - \beta_3$. Since 0 is not in the confidence interval, we have sufficient evidence to say that the 10% fat content and 20% fat content have the different effects on the weight loss.

Least Squares Estimators of Parameters for Unbalanced Experiments

However, for unbalanced experiments, the algebra expressions of least squares estimators are general complicated, and we will leave the statistical inference of a contrast for balanced and unbalanced experiments to R.

Analysis of Variance

For a two-way main-effect model, the two standard hypotheses are the main-effect hypotheses:

$$H_0^A: \alpha_1 = \alpha_2 = \cdots = \alpha_a$$

$$H_0^B: \beta_1 = \beta_2 = \cdots = \beta_b$$

Correspondingly, we reject H_0^A if:

$$F = \frac{ssA/(a-1)}{SSE/(n-a-b+1)} = \frac{msA}{msE} > F_{a-1,n-a-b+1,\alpha}$$

We reject H_0^B if:

$$F = \frac{ssB/(b-1)}{SSE/(n-a-b+1)} = \frac{msB}{msE} > F_{b-1,n-a-b+1,\alpha}$$

Similarly, ssA and ssB can be calculated by comparing the sum of squares of the full model and reduced model. Again, we rely on R to calculate ssA and ssB.

For the two-way main-effects model with the factor A as the first term and the factor B as the second term, the type I ssA is calculated based on the following two model:

Reduced Model: $Y_{ijt} = \mu + \epsilon_{ijt}$

Full Model: $Y_{ijt} = \mu + \alpha_i + \epsilon_{ijt}$ (one-way analysis of variance model)

The type I ssB is calculated based on the following two model:

Reduced Model: $Y_{ijt} = \mu + \alpha_i + \epsilon_{ijt}$

Full Model: $Y_{ijt} = \mu + \alpha_i + \beta_j + \epsilon_{ijt}$

While the type III ssA is calculated based on the following two model:

Reduced Model: $Y_{ijt} = \mu + \beta_j + \epsilon_{ijt}$

Full Model: $Y_{ijt} = \mu + \alpha_i + \beta_j + \epsilon_{ijt}$

The type III ssB is calculated based on the following two model:

Reduced Model: $Y_{ijt} = \mu + \alpha_i + \epsilon_{ijt}$

Full Model: $Y_{ijt} = \mu + \alpha_i + \beta_j + \epsilon_{ijt}$

In summary, the following analysis of variance table can be obtained for the two-way main - effects model:

Source of Variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**F-test Statistic**
Factor A	$a-1$	ssA	$msA = \frac{ssA}{a-1}$	$\frac{msA}{msE}$
Factor B	$b-1$	ssB	$msB = \frac{ssB}{b-1}$	$\frac{msB}{msE}$
Error	$n-a-b+1$	ssE	$msE = \frac{ssE}{n-ab}$	
Total	$n-1$	$sstot$		

Example 4.11: Analysis of Variance - Meat-Cooking Experiment, Unbalanced

From the previous example, we can fail to reject the null hypothesis that the interaction of the cooking method and the fat content is negligible. Here we will use the two-way main-effects model to analyze this experiment again. The following model is used in the example:

$$y_{ijt} = \mu + \alpha_i + \beta_j + \epsilon_{ijt}$$

In R program, the factor A (cooking method) is the first term and the factor B (fat content) is the second term.

The analysis of variance table for the type I sum of squares is:

Source of Variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**F-test Statistic**	***p*-value**
Cooking Method	1	26.404	26.404	3.438	0.076
Fat Content	2	316.132	158.066	20.580	$5.21 * 10^{-6}$
Error	25	192.015	7.681		
Total	28	534.551			

For the null hypothesis of

$$H_0: \alpha_1 = \cdots = \alpha_a \; and \; \beta_1 = \cdots = \beta_b$$

We can calculate F-test statistic from the type I sum of squares:

$$F = \frac{msT}{msE} = \frac{(26.404 + 316.132)/(1+2)}{7.681} = 14.866$$

The corresponding p-value is:

$$p - value = 2 * \Pr(F_{3,25} \geq 14.866) = 9.28 * 10^{-6}$$

For this unbalanced experiment, the type III sum of squares is different from the Type I sum of squares:

Source of Variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**F-test Statistic**	***p*-value**
Cooking	1	17.785	17.785	2.316	0.141

Method					
Fat Content	2	316.132	158.066	20.580	$5.21 * 10^{-6}$
Error	25	192.015	7.681		
Total	28	534.551			

Note that the type III sum of squares of ssA and ssB are different from those type I sum of squares. The conclusions on the factors A and B from the two-way main-effects model are same as those from the two-way complete model. This is expected since we fail to reject that the interaction is negligible.

Exercise 4.4: Analysis of Variance - Meat-Cooking Experiment, Balanced

Find the type I and type III ssA and ssB using the two-way main-effects model. Are these sums of squares same as those from the two-way complete model? Why?

Exercise 4.5: Two-Way Main-Effects Model - Mung Bean Experiment

1. This is a balanced experiment. Suppose we are interested in the pairwise difference of the effects of 3 levels from factor B. Following the steps of the above example, construct **98%** confidence intervals without multiple comparisons, and simultaneous **98%** confidence intervals with Bonferroni method, Scheffe method, and Tukey method.
2. Remove the last observation from your data and repeat the analysis of (1).
3. Remove the last observation from your data and construct the analysis of variance table for type I and type III sum of squares from the two-way main-effects model, perform the hypothesis testing for H_0, H_0^A, and H_0^B, and draw your conclusions.

R Program for Lesson 4.5

Please refer to the R program and videos for more details.

MA4720: Design and Analysis of Experiments

Week 5 At-a-Glance

Title: Experiments with Three or More Crossed Treatment Factors

Overview

This week, we will discuss the analysis of completely randomized experiments with three or more crossed treatment factors. We will present different statistical models. We will describe the methods that detect the effects of each treatment factor and the interactions among the treatment factors. We will cover the statistical inference including point estimation, confidence interval, and hypothesis testing of a single contrast without or with methods of multiple comparisons under different models. We will discuss how to use Type I and Type III sum of squares to appropriately perform hypothesis testing of main effects of each factor and interaction effects among factors. We will describe how to perform the sample sizes and power calculation for experiments with two or more crossed treatment factors.

Objectives

When you complete this module, you should be able to do the following:

1. Construct and evaluate different analysis of variance models for experiments with three or more crossed treatment factors.
2. Calculate point estimation, confidence interval, and perform hypothesis testing of a single contrast without or with methods of multiple comparisons under different models for experiments with three or more crossed treatment factors.
3. Use appropriate plots and tests to check if the assumptions of models for experiments with three or more crossed treatment factors.
4. Obtain type I or Type III sum of squares and use them to test main effect of each factor and interaction effects between them.

Reading (Optional)

1. Sections 6.6 and 6.7 from Chapter 6.
2. Sections 7.1, 7.2, 7.3, and 7.4 in Chapter 7.

Instruction Content

See other files for details.

Quiz

Two quizzes. See other files for details.

Homework

One homework. See other files for details.

Final Project

One component for the final project. See other files for details.

MA4720: Design and Analysis of Experiments

Week 5 Instruction Contents

Lesson 5.1 Experiments with Three or More Crossed Treatment Factors: Introduction

Required Reading: Sections 7.1 and 7.2 in Chapter 7.

Notations

In this week, we will consider experiments with three or more crossed treatment factors. The manual calculation will be more complicated so we will rely more on the use of computer and R software packages for the calculation. Such calculation will be similar with those presented in Lessons 4.3 and 4.4 in Week 4. We continue to label the factors with uppercase Roman letters and the number of levels with lowercase Roman letters. For example, if an experiment involves four factors, then the four factors can be labeled as A, B, C, and D and have a, b, c, and d levels, respectively. Such experiment is generally denoted by an "$a \times b \times c \times d$ factorial experiment" (read "a by b by c by d") and has a total of $v = abcd$ treatment combinations.

Here, we will use the analysis of experiments with three crossed treatment factors as the illustration. The analysis of experiments with four or more crossed treatment factors are similar. For an experiment with three crossed treatment factors, we use the following notations:

First treatment factor A with a levels and are coded as: $1,2,\cdots,a$

Second treatment factor B with b levels and are coded as: $1,2,\cdots,b$

Third treatment factor C with c levels and are coded as: $1,2,\cdots,c$

For three crossed treatment factors, the total number of combinations is $v = a * b * c$ and they are coded as $111, 112, \cdots, 11c, 121, 122, \cdots, 12b, \cdots, ab1, ab2, \cdots, abc$. Same as in Lesson 4.2, we can recode the treatment combinations as $1,2,\cdots,v$ with the **lexical order** and use the one-way analysis of variance model. In this lesson, we need to investigate the contributions of that each of the factors make individually to the response, so it is more convenient to retain the 3-digit code ijk for a treatment combination in which factor A is at level i, factor B is at level j, and factor C at level k, respectively.

Several related notations are listed below:

The tth observation for the treatment combination ijk: y_{ijkt} (or Y_{ijkt} as the random variable).

The number of samples for the treatment combination ijk: r_{ijk}.

The total number of observations: $n = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} r_{ijk}$

Example 5.1: 3-digit Codes and 1-digit Codes

For an experiment with 3 crossed treatment factors:

Factor A: c= 3 levels; Factor B $b = 2$ level; Factor C $c = 3$ level

There are $v = 3 * 2 * 3 = 18$ treatment combinations, and these 18 treatment combinations are coded as:

1-digit code	1	2	3	4	5	6
3-digite code	111	112	113	121	122	123
1-digit code	7	8	9	10	11	12
3-digite code	211	212	213	221	222	223
1-digit code	13	14	15	16	17	18
3-digite code	311	312	313	321	322	323

Note that such order is important since it is needed for the specification of contrasts and their corresponding contrasts coefficients later.

Exercise 5.1: 3-digit Codes and 1-digit Codes

For an experiment with 3 crossed treatment factors:

Factor A: $a = 2$ levels; Factor B: $b = 3$ levels; Factor C: $c = 2$ levels

There are $v = 2 * 3 * 2 = 12$ treatment combinations, and these 12 treatment combinations are coded as:

1-digit code	1	2	3	4	5	6	7	8	9	10	11	12
3-digite code												

Solution:

1-digit code	1	2	3	4	5	6	7	8	9	10	11	12
3-digite code	111	112	121	122	131	132	211	212	221	222	231	232

Popcorn Microwave Experiment. We will use Popcorn Microwave Experiment to illustrate the analysis of experiments with three crossed treatment factors. The data can be found in the file **popcorn.microwave.txt**. The more detailed discretion of this experiment can be found from the file **week-5-popcorn-microwave.docx**. Here we give a brief introduction of this experiment.

Experiment – Popcorn Microwave. The experiment was run by Jianjian Gong, Chongqing Yan, and Lihua Yang in 1992 to compare brands of microwave popcorn.

The objectives of the experiment was to find out which brand gives rise to the best popcorn in terms of the proportion of popped kernels. The experiment was restricted to popcorn produced in a microwave oven.

Responsible variable. The percentage, rather than number, of successfully popped kernels for each package was selected as the response variable.

Treatment levels. There are three treatment factors.

Factor A: brand, three levels with two national brands and one local brand

Levels 1 and 2: two national brands

Level 3: local brand

Factor B: power of the microwave oven with two levels (500 and 625W, and coded as 1, and 2 in the data file respectively).

Factor C: popping time with three levels (4, 4.5, and 5 minutes and coded as 1, 2, and 3 in the data file, respectively)

Deign: This is a completely randomized and balanced experiment. The number of replicates for each treatment combination is 2.

The Meaning of Three-Way Interactions

In Lesson 4.2 in Week 4, we discussed the two-way interaction for experiments with two crossed treatment factors. A **two-way interaction effect** is change in the main effect of one factor over levels of the other factor and the interaction plots can be used to evaluate such interaction. For experiments with three crossed treatment factors, in addition to the **main effects** and the **two-way interaction effects,** the **three-way interactions** need to be considered too. A three-way **interaction effect** can be considered as change in the two-way interaction effects of two factors over levels of the third factor and the interaction plots can still be used to evaluate such interactions.

In interaction plots for the three-way interaction effects, the y-axis represents $\bar{y}_{ijk\cdot}$, the mean response from the level i of factor A, the level j of factor B, and the level k of factor C. The x-axis represents the levels of one factor while the different symbols represent the levels of another factor, and different plots for each level of the third factor. Here we use the following hypothetical example to illustrate the use of interaction plot to understand the interaction effects.

Example 5.2: Interaction Plot for Three Factors

Suppose the sample mean $\bar{y}_{ijk\cdot}$ for a hypothetical $3 \times 2 \times 2$ experiment involving factors A, B, and C at 3, 2, and 2 levels, respectively:

ijk	111	112	121	122	211	212	221	222	311	312	321	322
$\bar{y}_{ijk\cdot}$	3.0	4.0	1.5	2.5	2.5	3.5	3.0	4.0	3.0	4.0	1.5	2.5

Use the interaction plots to determine if there are three-way interactions.

Solution: The corresponding R coded for this example cab be found in file **week-5-prog-example.R**.

Plot for the two-way interaction: the following plot is for the two-way interactions for factors A and B. Similar as before, the x-axis represents the levels of factor A while the different symbols represent the levels of factor B. The y-axis represents $\bar{y}_{ij\cdot\cdot}$, the mean response from the level i of factor A and the level j of factor B over the levels of factor C. The interaction plots for factors A and C and for factors B and C can be drawn similarly. From the plot, we can there the lines are not parallel, indicating that the factors A and B possibly interact. The perceived AB interaction is averaged over the levels of C and may have no practical meaning if there is an ABC interaction. Consequently, the three-way interaction should be investigated first.

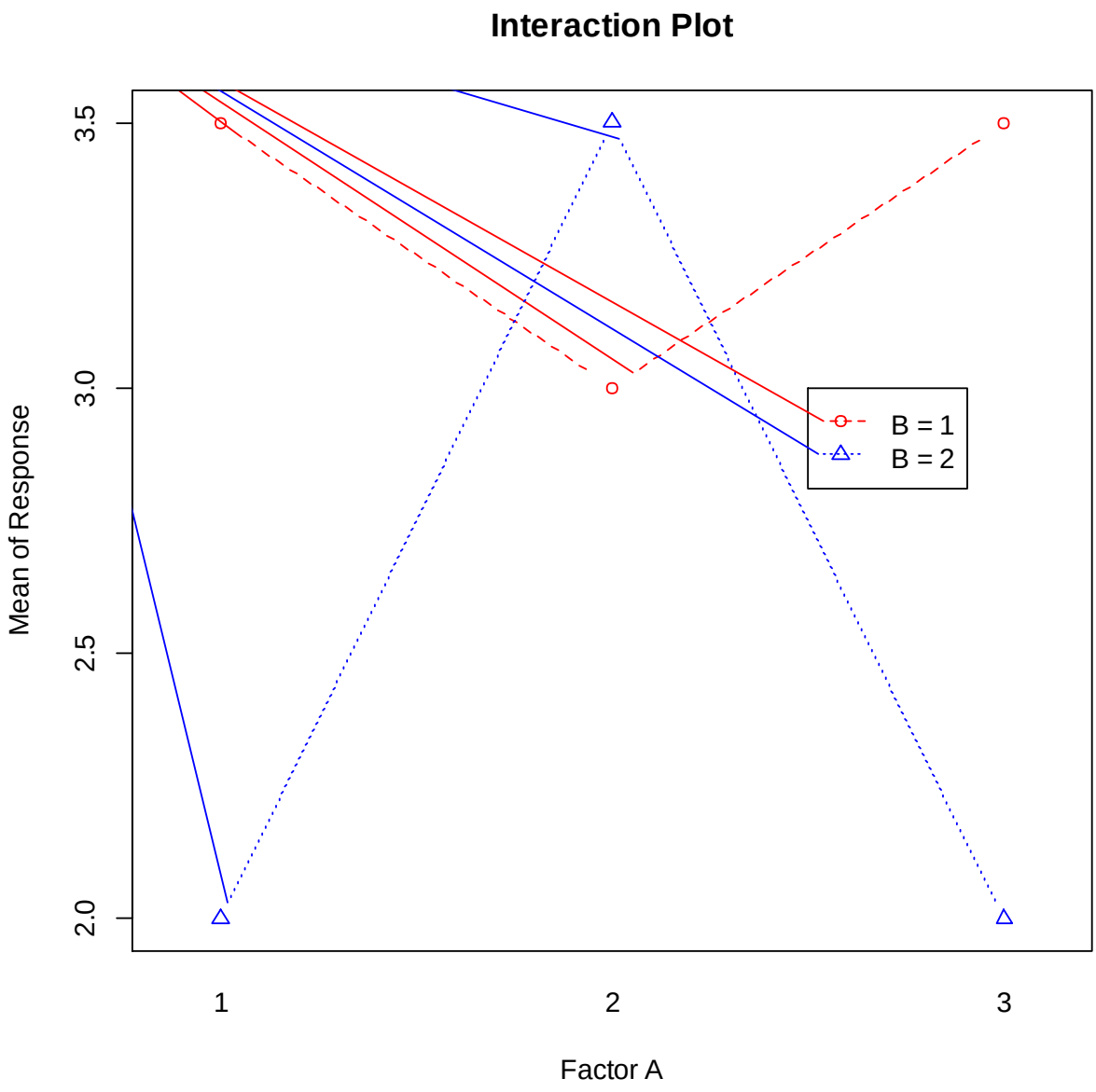


Figure 5.1. AB-interaction plot (averaged over levels of factor C)

Plot for the three-way interaction: the plot contains the two-interactions of two factors at each level of the third factors. Factors $C, B,$ and A are used as the third factor in the following plots, respectively.

In Figure 5.2, the factor C is used as the third factor. Each of the two plots suggests the presence of an AB-interaction effect, but the patterns in the two plots are the same. In other words, the factors A and B apparently interact in the same way at each level of factor C. This indicates a

negligible ABC-interaction effect. The shift in the interaction plot as the level of C changes from one plot to the other indicates a possible main effect of factor C.

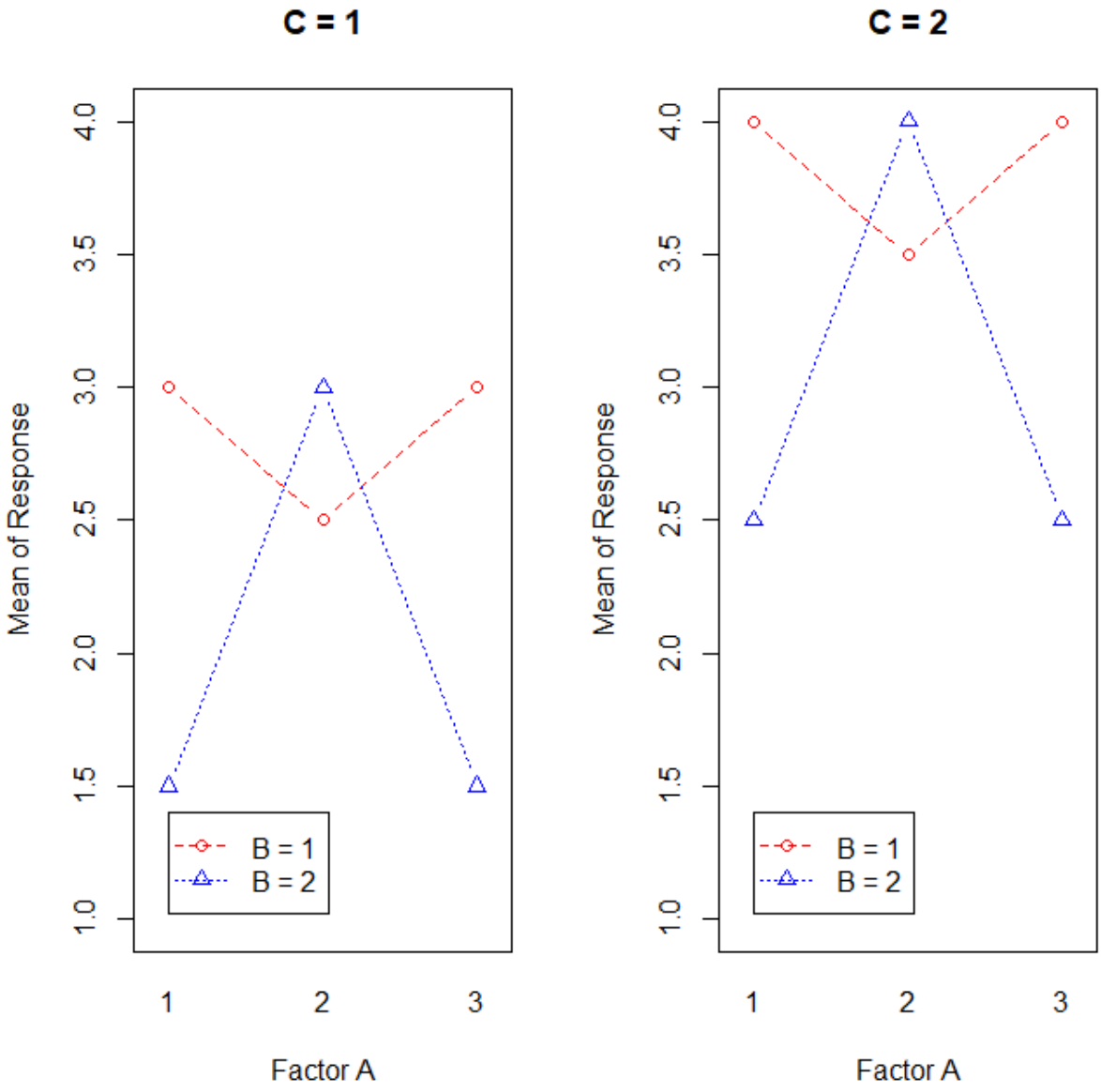


Figure 5.2. ABC-interaction plots with C as the third factor.

In Figure 5.3, the factor B is used as the third factor. Each of the two plots suggests the absence of an AC-interaction effect at each level of B, although the patterns differ from plot to plot. Ff

there is no AC-interaction at either of the levels of B, there is no change in the AC-interaction from one level of B to the other. So, again the ABC-interaction effect appears to be negligible.

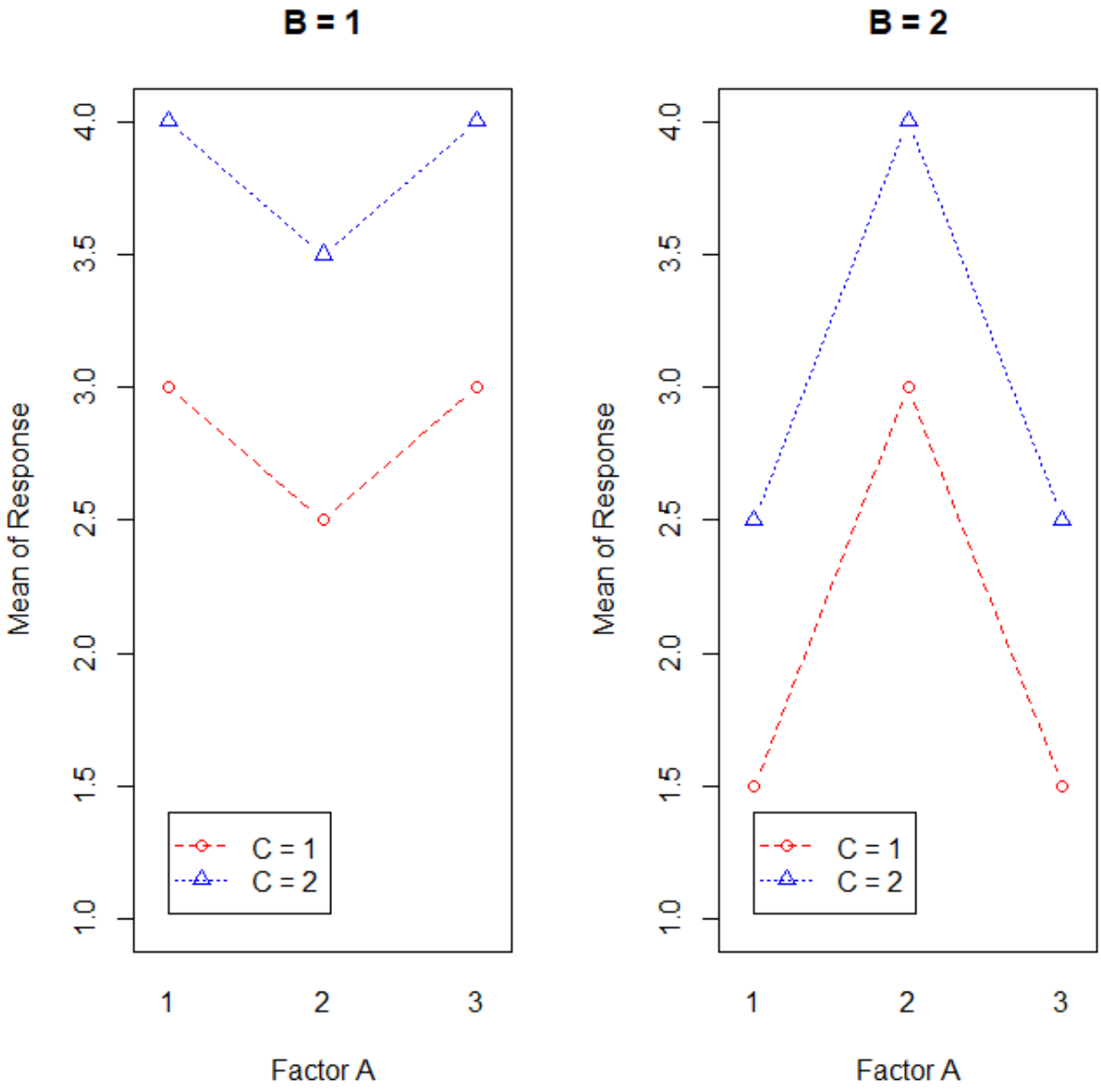


Figure 5.3. ABC-interaction plots with B as the third factor.

In Figure 5.4, the factor A is used as the third factor. The plots are similar with those in Figure 5.3. Each of the three plots suggests the absence of an BC-interaction effect at each level of A, although the patterns differ from plot to plot. Ff there is no BC-interaction at either of the levels

of A, there is no change in the BC-interaction from one level of A to the other. So, again the ABC-interaction effect appears to be negligible.

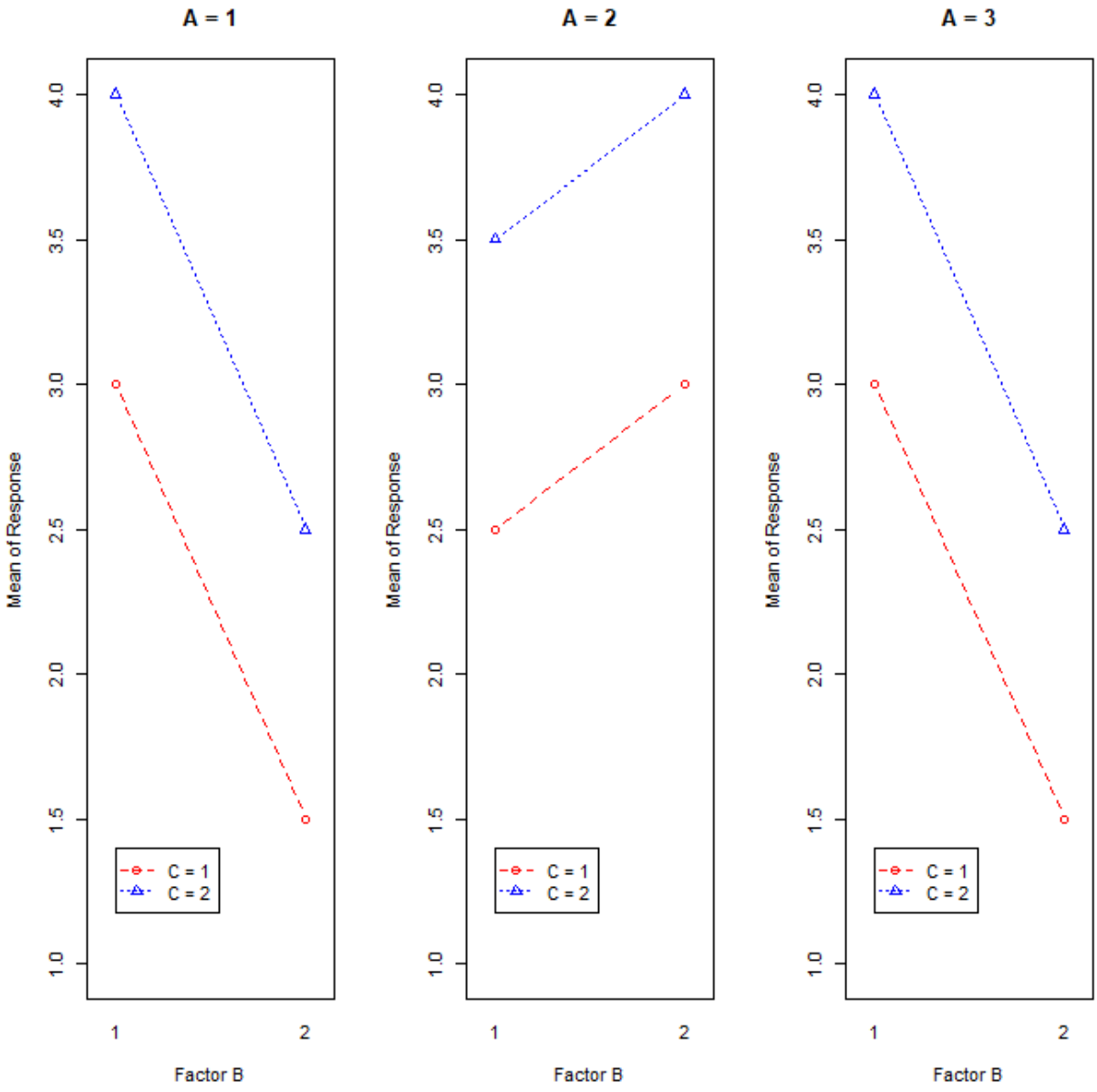


Figure 5.4. ABC-interaction plots with A as the third factor.

Example 5.3: Plots for Three-Way Interactions

Suppose the sample mean $\bar{y}_{ijk\cdot}$ for a hypothetical $3 \times 2 \times 2$ experiment involving factors A, B, and C at 3, 2, and 2 levels, respectively:

ijk	111	112	121	122	211	212	221	222	311	312	321	322
$\bar{y}_{ijk\cdot}$	3.0	2.0	1.5	4.0	2.5	3.5	3.0	4.0	3.0	5.0	3.5	6.0

Use the interaction plots to determine if there are three-way interactions.

Solution: The corresponding R coded for this example cab be found in file **week-5-prog-example.R**.

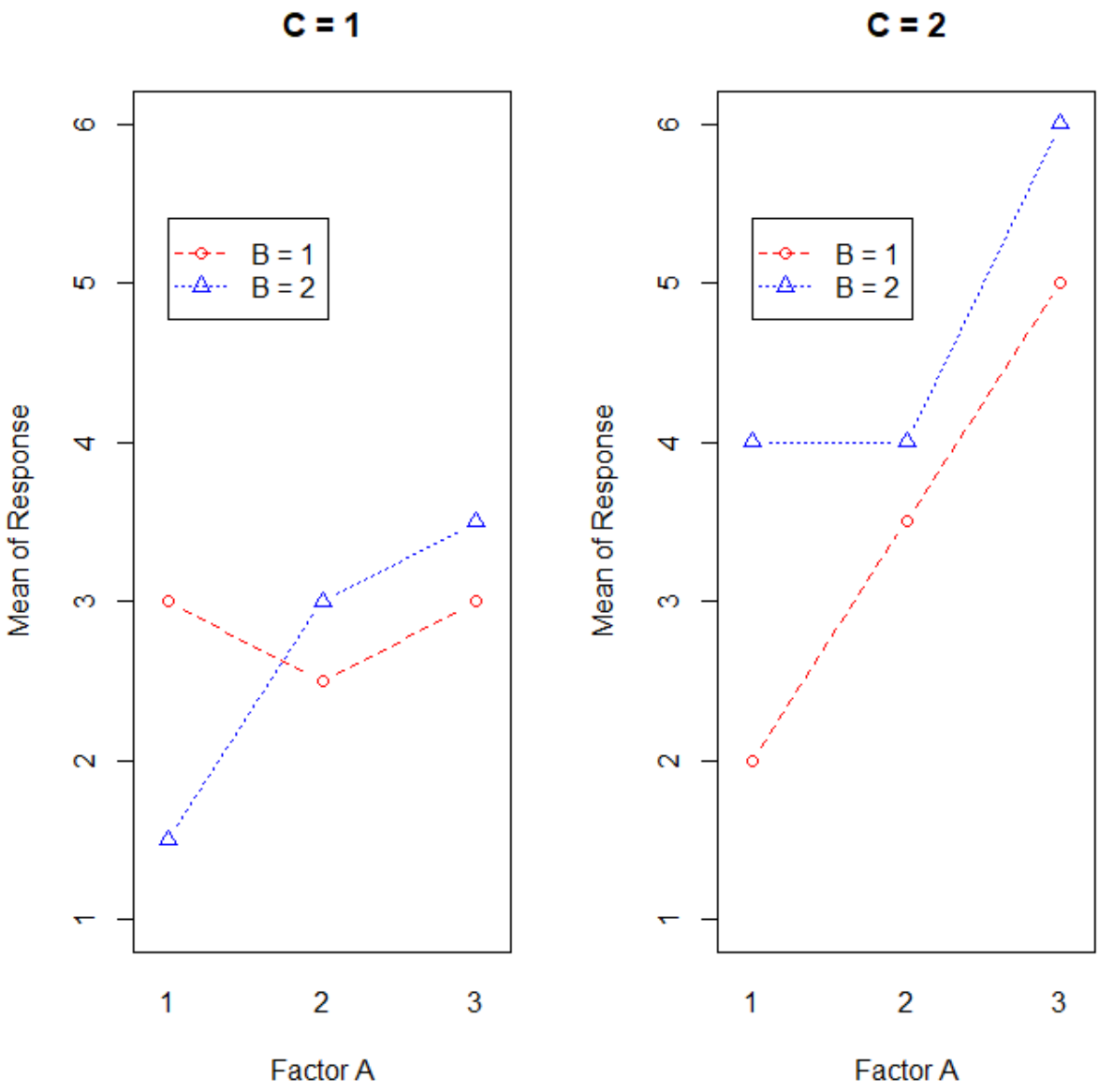


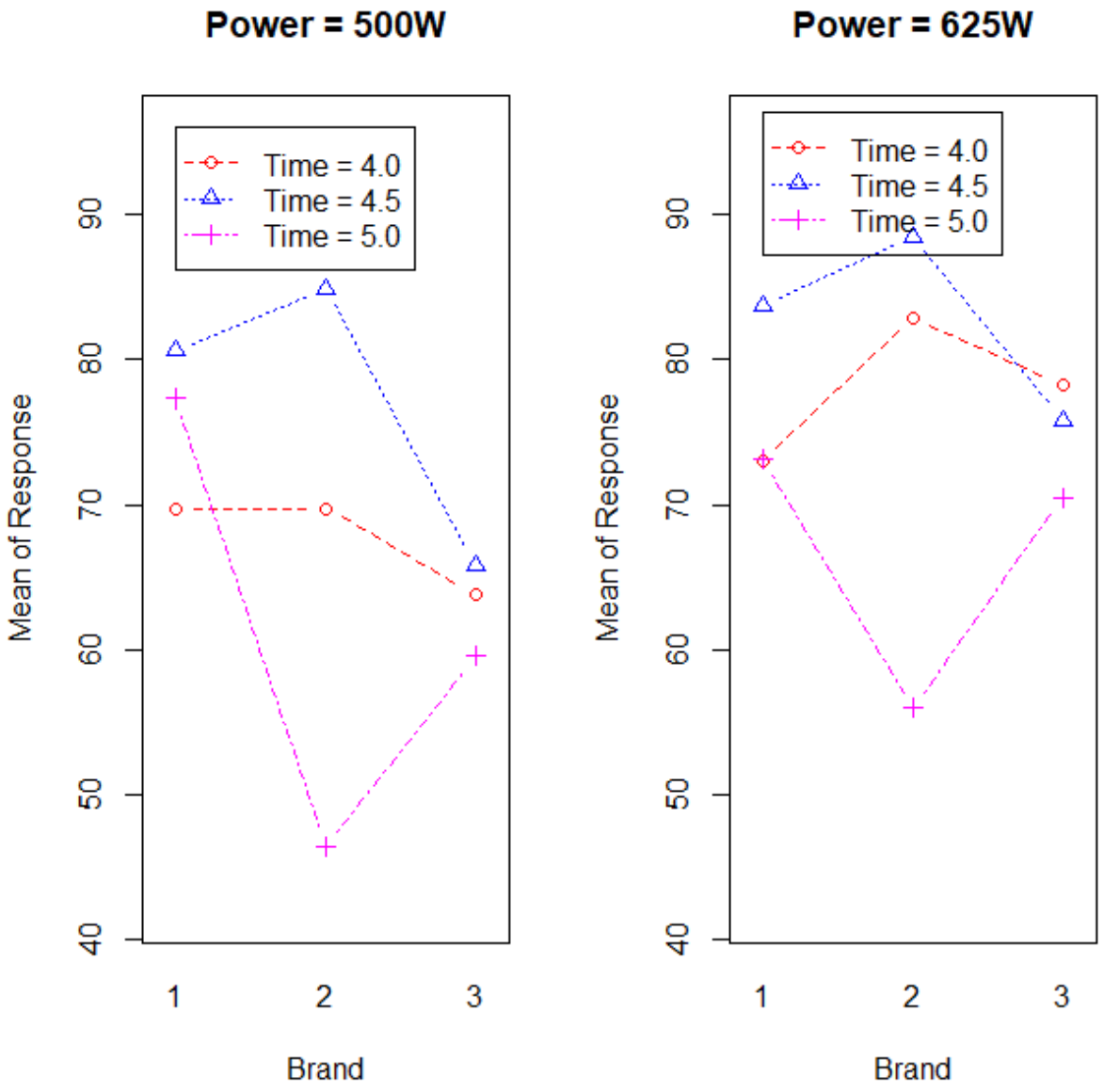
Figure 5.5. ABC-interaction plots with C as the third factor.

In Figure 5.5, the factor C is used as the third factor. Each of the two plots suggests the presence of an AB-interaction effect, and **the patterns are different** in the two plots. In other words, the factors A and B apparently interact in the different way at each level of factor C. This indicates an ABC-interaction effect.

Exercise 5.2: Interaction Plot – Popcorn Microwave Experiment

Draw the three-way interaction plot using the data from Popcorn-Microwave Experiment and factor *B* (power of the microwave oven with two levels) as the third factor in your plot, then interpret the plot.

Solution: Patterns are similar so the three-way interaction is negligible.



Models for Three Treatment Factors

Several models will be introduced here, including the **cell-means model**, **three-way complete model**, and other models such **three-way main-effects model**, etc.

Cell-Means Model

From Lessons 2.1 and 2.2 in Week 2, we have learned how to use the one-way analysis of variance model to analyze experiments with several crossed treatment factors. In that model, the combinations of levels from several treatment factors are considered as the levels of a new treatment factor. The cell-means model is same as the one-way analysis model. However, for convenience and for experiments with three factors, the treatment combinations are coded with three-digit codes and the **cell-means model** is:

$$Y_{ijkt} = \mu + \tau_{ijk} + \epsilon_{ijkt}, t = 1, \cdots, r_{ijk}; i = 1, \cdots, a; j = 1, \cdots, b; k = 1, \cdots, c$$

with the following assumptions:

- μ and τ_{ijk} $(i = 1, \cdots, a; j = 1, \cdots, b; k = 1, \cdots, c)$ are unknown population parameters
- $\epsilon_{ijkt} (t = 1, \cdots, r_i; i = 1, \cdots, a; j = 1, \cdots, b; k = 1, \cdots, c)$ are independent
- $\epsilon_{ijkt} \sim N(0, \sigma^2)$: normal distribution with mean 0 and variance σ^2

Here, i, j, and k are the levels of factors A, B, and C, respectively.

Three-way Complete Model

It is often useful to model the effect on the response of treatment combinations ijk to be the sum of the individual effect of the three factors, together with their interactions; that is:

$$\tau_{ijk} = \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$$

Here, $\alpha_i, \beta_j, \gamma_k$ are the effect (positive on negative) on the response of factors A, B, C at levels i, j, k, respectively, $(\alpha\beta)_{ij}, (\alpha\gamma)_{ik}, (\beta\gamma)_{jk}$ are the additional effects of the pairs of factors together at the specified levels, and $(\alpha\beta\gamma)_{ijk}$ is the additional effect of all three factors together at levels i, j, k. The three sets of factorial effects are called the main-effect parameters, the two-factor interaction parameters, and the three-factor interaction parameter, respectively. The interpretation of these parameters has been discussed. Thus, the corresponding model, which is called **three-way complete model**, is as follows:

$$Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkt},$$

$$t = 1, \cdots, r_{ij}; i = 1, \cdots, a; j = 1, \cdots, b; k = 1, \cdots, c$$

with the following assumptions:

- $\mu, \alpha_i, \beta_j, \gamma_k, (\alpha\beta)_{ij}\ (\alpha\gamma)_{ik}, (\beta\gamma)_{jk}$ and $(\alpha\beta\gamma)_{ijk} (i = 1, \cdots, a; j = 1, \cdots, b;\ k = 1, \cdots, c)$ are unknown population parameters
- $\epsilon_{ijkt} (t = 1, \cdots, r_{ijk}; i = 1, \cdots, a; j = 1, \cdots, b; k = 1, \cdots, c)$ are independent
- $\epsilon_{ijkt} \sim N(0, \sigma^2)$: normal distribution with mean 0 and variance σ^2

Here, the phrase ”three-way” refers to the fact that there are three primary sources of variance, namely, the three treatment factors. We can see that the cell-means model and the three-way complete model just use different parameters but are actually equivalent. These two models are often called the **full model** since every possible effects are included in these models.

Other Reduced Models

If some interaction effects are not observed from the interaction plot, then they can be removed from the three-way complete model. For example, we can have a model with only main effects:

$$Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijkt}, t = 1, \cdots, r_{ijk}; i = 1, \cdots, a; j = 1, \cdots, b; k = 1, \cdots c$$

We can have a model with all two-way interactions but not three-way interactions:

$$Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \epsilon_{ijkt},$$

$$t = 1, \cdots, r_{ijk}; i = 1, \cdots, a; j = 1, \cdots, b; k = 1, \cdots, c$$

We can also have a model with some two-way interactions:

$$Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + \epsilon_{ijkt}, t = 1, \cdots, r_{ijk}; i = 1, \cdots, a; j = 1, \cdots, b; k = 1, \cdots, c$$

These models are "submodels" or "reduced models" of the three-way complete model since only a subset of possible effects are included in the models.

Checking Assumptions on the Model

The assumptions implicit in all models are that the error terms have equal variance, are mutually independent, and are normally distributed. The strategy and the methods for checking the model assumptions are the same as those in Lesson 3.3 in Week 3. The standardized residuals are calculated as

$$z_{ijkt} = \frac{y_{ijkt} - \hat{y}_{ijkt}}{\sqrt{ssE/(n-1)}}$$

where $\hat{y}_{ijkt}$ is the estimated response and is model-dependent. The standardized residuals are plotted against:

(1) the order of observation to check the independence
(2) the levels of each factors or treatment combination to check the model fit
(3) the levels of each factors or treatment combination to check for the outlies
(4) $\hat{y}_{ijkt}$ to check for equality of variances
(5) the normal scores to check the normality assumption

Statistical Inferences

As before, for a specific model (e.g, the cell-means model, the three-way complete model, or a reduced model), we can perform the statistical inference in the following order:

(1) Find the least square estimates (or estimators) of the model parameters (more precisely, estimable functions of parameters of $\tau_{ijk}, \alpha_i, \beta_j, \gamma_k, (\alpha\beta)_{ij}, (\alpha\gamma)_{ik}, (\beta\gamma)_{jk}, (\alpha\beta\gamma)_{ijk}$) and σ^2.
(2) Perform the overall analysis to test if there are any effects on the response including the interaction and main effects from three treatment factors.
(3) Perform the analysis to test if there are any interaction effects on the response from three treatment factors.
(4) Perform the analysis to see if there is any effect on the response from any one of three treatment factors.
(5) Perform the statistical inference for individual contrasts including the point estimation, the confidence interval and the hypothesis testing with appropriate methods for multiple comparisons.

Model Selection/Building

This number of possible models for experiments with several crossed treatment factors can be large. In many situations, we need to find a reduced model instead of the full model for our analysis. Use of a reduced, when appropriate, is advantageous, because simpler models generally yield tighter confidence intervals and more powerful tests of hypotheses. However, if interaction terms are removed from the model when the factors do, in fact, interact, then the resulting analysis and conclusions may be totally incorrect. Here we introduce on principle and two different ways for model building.

The principle here is that when a model includes an interaction between a specific set of m factors, then all interaction terms involving subsets of those m factors should be included in the model. For example, a model that includes the effect of the three-factor interaction ABC would

also include the effects of the AB, AC, and BC interactions as well as the main effects A, B, and C.

One way of building an appropriate model is based on the prior knowledge of effects. If prior to the experiment certain interaction effects are known to be negligible, the corresponding parameters can be removed from the complete model to give a reduced. For example, if the factors A and B are known not to interact in a three-factor experiment, then the AB and ABC interaction effects are negligible, so the terms $(\alpha\beta)_{ij}$ and $(\alpha\beta\gamma)_{ijk}$ are excluded from model. In the extreme case, if no factors are expected to interact, then a main-effects model (which includes no interaction terms) can be used.

Another way of building an appropriate model is based on experiments. In some experiments, the primary objective is to find a model that gives an adequate representation of the experimental data. Such experiments are called experiments for model building. If there are three crossed treatment factors, it is legitimate to use the three-way complete model as a preliminary model. Then, if we fail to reject to the null hypothesis of that AB-interaction and ABC −interaction are negligible based on the experimental data, then a model without $(\alpha\beta)_{ij}$ and $(\alpha\beta\gamma)_{ijk}$ terms can be accepted as a reasonable model to represent the same type of experimental data in future experiments. **Note** that it is not legitimate to use such reduced model as the final model to test further hypotheses or calculate confidence intervals using the same set of data. If this is done, the model is changed based on the data, and the quoted significance levels and confidence levels associated with further inferences will not be correct. Model building should be regarded as a completely different exercise from confidence interval calculation. They should be done using different experimental data.

Example 5.4: Model Specification

For an $a \times b \times c$ factorial experiment, find the appropriate model if

(1) There is no three-way interaction.
(2) There is no interaction at all.

(3) There is no interaction between factors B and C.

Solution:

(1) The model includes three main effects and three two-way interactions. It is:

$$Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \epsilon_{ijkt}$$

(2) The model only includes three main effects. It is:

$$Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijkt}$$

(3) The model includes three main effects and two two-way interactions. It is:

$$Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + \epsilon_{ijkt}$$

Lesson 5.2 Experiments with Three or More Crossed Treatment Factors: Contrasts

In this lesson, we will learn how to perform the statistical inference of contrasts, including point estimation, confidence intervals, and hypothesis testing with appropriate methods of multiple comparisons. For the cell-means model (or equivalent three-way complete model), the least squares estimator of $\mu + \tau_{ijk}$ is:

$$\hat{\mu} + \hat{\tau}_{ijk} = \bar{Y}_{ijk\cdot}$$

where $\bar{Y}_{ijk\cdot}$ is the sample of the response for the treatment combination ijk. For reduced models, the simple algebra expression of least squares estimator of an estimable function of parameters may be obtained for balanced experiments but generally does not exist for unbalanced experiments. Therefore, we will rely on R software to perform the calculation here. However, you still need to able to write out the contrast coefficients first so you can then use R for such calculations.

Estimable Functions and Contrasts from Cell-Means Model

In this section, we discuss the estimable functions and contrasts with the cell-means model (or equivalently, the three-way complete model). This can be easily extended to other reduced models and examples with the cell-means model and reduced models will be given.

For the cell-means model:

$$Y_{ijkt} = \mu + \tau_{ijk} + \epsilon_{ijkt}, t = 1, \cdots, r_{ijk}; i = 1, \cdots, a; j = 1, \cdots, b; k = 1, \cdots, c$$

Or the equivalent three-way complete model:

$$Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkt},$$

A contrast of the parameters must have the following form:

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} h_{ijk}\tau_{ijk}, and \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} h_{ijk} = 0$$

here $h_{ijk}(i = 1, \cdots, a; j = 1, \cdots, b; k = 1, \cdots, c)$ are constants and the sum of them is 0 (in other words, $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c} h_{ijk} = 0$).

Among all possible contrasts, we are particularly interested in the following four types of contrasts:

- **Contrasts for the pairwise difference in response between two treatment combinations ijk and shl:**

$$\tau_{ijk} - \tau_{shl}$$

- **Contrasts for the main effects of a factor.** For example, a contrast for the main effect of A can be written as:

$$\sum_{i=1}^{a} p_i \bar{\tau}_{i\cdot\cdot} = \sum_{i=1}^{a} p_i \left[\alpha_i + \left(\overline{\alpha\beta}\right)_{i\cdot} + \left(\overline{\alpha\gamma}\right)_{i\cdot} + \left(\overline{\alpha\beta\gamma}\right)_{i\cdot\cdot}\right] = \sum_{i=1}^{a} p_i \alpha_i^*,$$

$$and \sum_{i=1}^{a} p_i = 0$$

As another example, a contrast for the main effect of C can be written as:

$$\sum_{k=1}^{c} p_k \bar{\tau}_{\cdot\cdot k} = \sum_{k=1}^{c} p_k \left[\gamma_k + \left(\overline{\alpha\gamma}\right)_{\cdot k} + \left(\overline{\beta\gamma}\right)_{\cdot k\left(\overline{\alpha\beta\gamma}\right)_{\cdot\cdot k}}\right] = \sum_{k=1}^{c} p_k \gamma_k^*,$$

$$and \sum_{k=1}^{c} p_k = 0$$

Note that any contrast for the main effect of a factor can be written as a function of contrasts between two levels of that factor. For example, any contrast of factor B can be written as the function of the following $b - 1$ contrasts:

$$\bar{\tau}_{\cdot 1\cdot} - \bar{\tau}_{\cdot 2\cdot} = \beta_1^* - \beta_2^*, \bar{\tau}_{\cdot 2\cdot} - \bar{\tau}_{\cdot 3\cdot} = \beta_2^* - \beta_3^*, \cdots, \bar{\tau}_{\cdot b-1\cdot} - \bar{\tau}_{\cdot b\cdot} = \beta_{b-1}^* - \beta_b^*$$

- **Contrasts for the two-interaction between two levels of a factor and two levels of another Factor.** For example, the contrast for the two-way interaction between two level s and u of factor B and two levels h and q of factor C can be written as:

$$(\tau_{\cdot sh} - \tau_{\cdot uh}) - \left(\tau_{\cdot sq} - \tau_{\cdot uq}\right) = ((\beta\gamma)_{sh}^* - (\beta\gamma)_{uh}^*) - ((\beta\gamma)_{sq}^* - (\beta\gamma)_{uq}^*)$$

Where $(\beta\gamma)^*_{sh} = (\beta\gamma)_{sh} + (\overline{\alpha\beta\gamma})_{\cdot sh}$. As we have mentioned in Lesson 4.3 in Week 4, the above contrast is the difference of two pairwise differences between two treatment combinations. Note the subscript for the level of factor C (here h) is fixed in the first pairwise difference, the subscript for the level of factor C (here q) is fixed in the second pairwise difference, and the order of subscripts of factor B (here s and u) is same for two pairwise differences. Based on these rules, the following contrasts can also be used as ta two-way interaction between two Levels s and u of factor B and two Levels h and q of factor C:

$$(\tau_{\cdot sh} - \tau_{\cdot sq}) - (\tau_{\cdot uh} - \tau_{\cdot uq}) = ((\beta\gamma)^*_{sh} - (\beta\gamma)^*_{sq}) - ((\beta\gamma)^*_{uh} - (\beta\gamma)^*_{uq})$$

$$(\tau_{\cdot uh} - \tau_{\cdot sh}) - (\tau_{\cdot uq} - \tau_{\cdot sq}) = ((\beta\gamma)^*_{uh} - (\beta\gamma)^*_{sh}) - ((\beta\gamma)^*_{uq} - (\beta\gamma)^*_{sq})$$

While the following contrast is not an interaction contrast:

$$(\tau_{\cdot sh} - \tau_{\cdot uh}) - (\tau_{\cdot uq} - \tau_{\cdot sq}) = ((\beta\gamma)^*_{sh} - (\beta\gamma)^*_{uh}) - ((\beta\gamma)^*_{uq} - (\beta\gamma)^*_{sq})$$

Again, this is a simplest two-way interaction and any contrast of a two-way interaction between two factors can be written as a function of such contrasts. For example, any contrast of two-way interaction between factors A and B can be written as a function of $(a-1)(b-1)$ such contrasts.

- **Contrasts for three-interaction between two levels of each factor.** For example, a three-way interaction between two level s and u of factor A, two levels h and q of factor B, and two levels x and y of factor C can be written as:

$$\begin{aligned}&[(\tau_{shx} - \tau_{sqx}) - (\tau_{uhx} - \tau_{uqx})] - [(\tau_{shy} - \tau_{sqy}) - (\tau_{uhy} - \tau_{uqy})] \\ &= [((\alpha\beta\gamma)_{shx} - (\alpha\beta\gamma)_{sqx}) - ((\alpha\beta\gamma)_{uhx} - (\alpha\beta\gamma)_{uqx})] \\ &\qquad - [((\alpha\beta\gamma)_{shy} - (\alpha\beta\gamma)_{sqy}) - ((\alpha\beta\gamma)_{uhy} - (\alpha\beta\gamma)_{uqy})]\end{aligned}$$

Note that the above contrast is just the difference of two-way interactions between two level s and u of factor A and two levels h and q of factor B at levels x and y of factor C.

Again, this is a simplest three-way interaction and any contrast of a three-way interaction can be written as a function of such contrasts. For example, any contrast of three-way interaction can be written as a function of $(a-1)(b-1)(c-1)$ such contrasts.

Writing Contrasts as Coefficient Lists

Instead of writing out a contrast explicitly, it is sometimes sufficient, and more convenient, to list the contrast coefficients only. For the cell-mean or three-way complete model, we can refer to contrasts as either a list of coefficients of the parameters $\alpha_i^*, \beta_j^*, \gamma_k^*, (\alpha\beta)_{ij}^*, (\alpha\gamma)_{ik}^*, (\beta\gamma)_{jk}^*, (\alpha\beta\gamma)_{ijk}$ or as a list of coefficients of the τ_{ijk}. For any parameters, they are arranged by the lexical order. For example, $(\beta\gamma)_{jk}^*$ are arranged by the following order:

$$(\beta\gamma)_{11}^*, (\beta\gamma)_{12}^*, \cdots, (\beta\gamma)_{1c}^*, (\beta\gamma)_{21}^*, (\beta\gamma)_{22}^*, \cdots, (\beta\gamma)_{2c}^*, \cdots, (\beta\gamma)_{b1}^*, (\beta\gamma)_{b2}^*, \cdots, (\beta\gamma)_{bc}^*$$

As another example, τ_{ijk} are arranged by the following order:

$$\tau_{111}, \tau_{112}, \cdots, \tau_{11c}, \tau_{121}, \tau_{122}, \cdots, \tau_{12c}, \cdots, \tau_{ab1}, \tau_{ab2}, \cdots, \tau_{abc}$$

This is illustrated in the following example.

Example 5.5: Writing Contrasts as Coefficient Lists for Popcorn Microwave Experiment

Note that the experiment has three factors, brand, power, and popping time with 3, 2, and 3 levels, respectively. For the cell-means or the three-way complete model, find the contrast coefficients for the following contrasts:

(1) Pairwise difference between treatment combination 121 and 211 in terms of τ_{ijk}.
(2) Interaction between levels of 1 and 2 from factor A and levels 1 and 3 from factor C in terms of τ_{ijk} or $(\alpha\gamma)_{ik}^*$.
(3) The difference of response of level 3 and the average response of levels 1 and 2 from factor A in terms of τ_{ijk} or $(\alpha)_i^*$.
(4) The linear trend of factor C in terms of τ_{ijk} or γ_k^*.

Solution:

(1) The contrast for the difference between treatment combination 121 and 211 is:

$$\tau_{121} - \tau_{211}$$

To write out the contrast coefficients, you can write out τ_{ijk} with the lexical order first:

$$\tau_{111}, \tau_{112}, \tau_{113}, \tau_{121}, \tau_{122}, \tau_{123}, \tau_{211}, \tau_{212}, \tau_{213}, \tau_{221}, \tau_{222}, \tau_{223}, \tau_{311}, \tau_{312}, \tau_{313}, \tau_{321}, \tau_{322}, \tau_{323}$$

So the corresponding contrast coefficients in terms of τ_{ijk} are (only the 4th and 7th numbers are not equal to 0):

$$[0, 0, 0, 1, 0, 0, -1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0]$$

(2) One possible contrast for the interaction between levels of 1 and 2 from factor A and levels 1 and 3 from factor C is:

$$(\bar{\tau}_{1\cdot 1} - \bar{\tau}_{2\cdot 1}) - (\bar{\tau}_{1\cdot 3} - \bar{\tau}_{2\cdot 3}) = [(\alpha\gamma)^*_{11} - (\alpha\gamma)^*_{21}] - [(\alpha\gamma)^*_{13} - (\alpha\gamma)^*_{23}]$$

Note that $\bar{\tau}_{i\cdot k} = \frac{\tau_{111}+\tau_{121}}{2}$, we have

$$(\bar{\tau}_{1\cdot 1} - \bar{\tau}_{2\cdot 1}) - (\bar{\tau}_{1\cdot 3} - \bar{\tau}_{2\cdot 3}) = \left(\frac{\tau_{111} + \tau_{121}}{2} - \frac{\tau_{211} + \tau_{221}}{2}\right) - \left(\frac{\tau_{113} + \tau_{123}}{2} - \frac{\tau_{213} + \tau_{223}}{2}\right)$$

so the corresponding contrast coefficients in terms of τ_{ijk} are:

$$[0.5, 0, -0.5, 0.5, 0, -0.5, -0.5, 0, 0.5, -0.5, 0, 0.5, 0, 0, 0, 0, 0, 0]$$

To find the corresponding contrast coefficients in terms of $(\alpha\gamma)^*_{ik}$, you can write out all $(\alpha\gamma)^*_{ik}$ in the lexical order first:

$$(\alpha\gamma)^*_{11}, (\alpha\gamma)^*_{12}, (\alpha\gamma)^*_{13}, (\alpha\gamma)^*_{21}, (\alpha\gamma)^*_{22}, (\alpha\gamma)^*_{23}, (\alpha\gamma)^*_{31}, (\alpha\gamma)^*_{32}, (\alpha\gamma)^*_{33}$$

Thus the corresponding contrast coefficients in terms of $(\alpha\gamma)^*_{ik}$ are:

$$[1, 0, -1, -1, 0, 1, 0, 0, 0]$$

(3) The contrast for the difference of response of level 3 and the average response of levels 1 and 2 from factor A is:

$$\alpha^*_3 - \frac{1}{2}(\alpha^*_1 + \alpha^*_2) = \bar{\tau}_{3\cdot\cdot} - \frac{1}{2}(\bar{\tau}_{1\cdot\cdot} + \bar{\tau}_{2\cdot\cdot})$$

$$= \frac{1}{6}(\tau_{311} + \tau_{312} + \tau_{313} + \tau_{321} + \tau_{322} + \tau_{323})$$

$$-\frac{1}{2}\left(\frac{1}{6}(\tau_{111} + \tau_{112} + \tau_{113} + \tau_{121} + \tau_{122} + \tau_{123})\right.$$

$$\left. + \frac{1}{6}(\tau_{211} + \tau_{212} + \tau_{213} + \tau_{221} + \tau_{222} + \tau_{223})\right)$$

So the corresponding contrast coefficients in terms of α^*_i are:

$$[-0.5, -0.5, 1]$$

So the corresponding contrast coefficients in terms of τ_{ij} are (first 12 numbers are $-\frac{1}{12}$ and the last six numbers are $\frac{1}{6}$):

$$[-\frac{1}{12},-\frac{1}{12},-\frac{1}{12},-\frac{1}{12},-\frac{1}{12},-\frac{1}{12},-\frac{1}{12},-\frac{1}{12},-\frac{1}{12},-\frac{1}{12},-\frac{1}{12},-\frac{1}{12},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6}]$$

(4) The factor C has three equally spaced levels, the contrast coefficients for the linear trend in terms of γ_k^* for the balanced experiments, can be obtained from Table A.2:

$$[-1,0,1].$$

Therefore, the contrast is:

$$-\gamma_1^* + \gamma_3^* = -\bar{\tau}_{..1} + \bar{\tau}_{..3}$$

$$= -\frac{\tau_{111}+\tau_{121}+\tau_{211}+\tau_{221}+\tau_{311}+\tau_{321}}{6} + \frac{\tau_{113}+\tau_{123}+\tau_{213}+\tau_{223}+\tau_{313}+\tau_{323}}{6}$$

So the corresponding contrast coefficients in terms of τ_{ijk} are (six of $[-\frac{1}{6},0,\frac{1}{6}]$):

$$[-\frac{1}{6},0,\frac{1}{6},-\frac{1}{6},0,\frac{1}{6},-\frac{1}{6},0,\frac{1}{6},-\frac{1}{6},0,\frac{1}{6},-\frac{1}{6},0,\frac{1}{6}]$$

If the experiments are unbalanced, we need to calculate the contrast coefficients for the linear trend in terms of γ_k^* first then re-write them in in terms of τ_{ijk}

Exercise 5.3: Writing Contrasts as Coefficient Lists - Popcorn Microwave Experiment

Note that the experiment has three factors, brand, power, and popping time with 3, 2, and 3 levels, respectively. For the cell-mean or the three-way complete model, find the contrast coefficients for the following contrasts:

(1) Pairwise difference between treatment combination 213 and 123 in terms of τ_{ijk}.

(2) Interaction between levels of 1 and 2 from factor B and levels 2 and 3 from factor C in terms of τ_{ijk} or $(\beta\gamma)_{jk}^*$.

(3) The difference of response of level 3 and the average response of levels 1 and 2 from factor C in terms of τ_{ijk} or $(\gamma)_k^*$.

(4) The quadratic trend of factor C in terms of τ_{ijk} or γ_k^*.

Solution:

(1) The contrast for the difference between treatment combination 213 and 123 is:

$$\tau_{213} - \tau_{123}$$

So the corresponding contrast coefficients in terms of τ_{ijk} are:

$$[0,0,0,0,0,-1,0,0,1,0,0,0,0,0,0,0,0,0]$$

(2) One contrast for the interaction between levels of 1 and 2 from factor B and levels 2 and 3 from factor C is:

$$(\bar{\tau}_{\cdot 12} - \bar{\tau}_{\cdot 22}) - (\bar{\tau}_{\cdot 13} - \bar{\tau}_{\cdot 23}) = ((\beta\gamma)^*_{12} - (\beta\gamma)^*_{22}) - ((\beta\gamma)^*_{13} - (\beta\gamma)^*_{23})$$

Note that $\tau_{\cdot jk} = \frac{\tau_{1jk}+\tau_{2jk}+\tau_{3jk}}{3}$, so the corresponding contrast coefficients in terms of τ_{ijk} are:

$$[0,\frac{1}{3},-\frac{1}{3},0,-\frac{1}{3},\frac{1}{3},0,\frac{1}{3},-\frac{1}{3},0,-\frac{1}{3},\frac{1}{3},0,\frac{1}{3},-\frac{1}{3},0,-\frac{1}{3},\frac{1}{3}]$$

So the corresponding contrast coefficients in terms of $(\beta\gamma)^*_{jk}$ are also:

$$[0,1,-1,0,-1,1]$$

(3) The difference of response of level 3 and the average response of levels 1 and 2 from factor C in terms of τ_{ijk} or $(\gamma)^*_k$.

$$\gamma^*_3 - \frac{1}{2}(\gamma^*_1 + \gamma^*_2) = \bar{\tau}_{\cdot\cdot 3} - \frac{1}{2}(\bar{\tau}_{\cdot\cdot 1} + \bar{\tau}_{\cdot\cdot 2})$$

So the corresponding contrast coefficients in terms of $(\gamma)^*_k$ are:

$$[-0.5,-0.5,1]$$

Note that

$$\bar{\tau}_{\cdot\cdot k} = \frac{\tau_{11k} + \tau_{12k} + \tau_{21k} + \tau_{22k} + \tau_{31k} + \tau_{31k}}{6}$$

So the corresponding contrast coefficients in terms of τ_{ijk} are (six of $[-\frac{1}{12},-\frac{1}{12},\frac{1}{6}]$):

$$[-\frac{1}{12},-\frac{1}{12},\frac{1}{6},-\frac{1}{12},-\frac{1}{12},\frac{1}{6},-\frac{1}{12},-\frac{1}{12},\frac{1}{6},-\frac{1}{12},-\frac{1}{12},\frac{1}{6},-\frac{1}{12},-\frac{1}{12},\frac{1}{6},-\frac{1}{12},-\frac{1}{12},\frac{1}{6}]$$

(4) The factor C has three equally spaced levels, so the contrast coefficients for the quadratic trend in terms of γ^*_k for the balance experiments, can be obtained from Table A.2:

$$[1,-2,1].$$

Therefore, the contrast is:

$\gamma^*_1 - 2 * \gamma^*_2 + \gamma^*_3 = \bar{\tau}_{\cdot\cdot 1} - 2\bar{\tau}_{\cdot\cdot 2} + \bar{\tau}_{\cdot\cdot 3}$So the corresponding contrast coefficients in terms of τ_{ij} are (six of $[\frac{1}{6},-\frac{1}{3},\frac{1}{6}]$):

$$[\frac{1}{6},-\frac{1}{3},\frac{1}{6},\frac{1}{6},-\frac{1}{3},\frac{1}{6},\frac{1}{6},-\frac{1}{3},\frac{1}{6},\frac{1}{6},-\frac{1}{3},\frac{1}{6},\frac{1}{6},-\frac{1}{3},\frac{1}{6},\frac{1}{6},-\frac{1}{3},\frac{1}{6},\frac{1}{6},-\frac{1}{3},\frac{1}{6}]$$

Least Squares Estimators and Estimates

Since the three-complete model is equivalent to the cell-means model and the cell-means model is actually a one-analysis of variance model, we can easily use the results from the one-way analysis of variance model to obtain the least squares estimators of $\mu + \tau_{ijk}$ from the cell-means model and corresponding parameters from the three-way complete model and σ^2.

The least squares estimator of $\mu + \tau_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$ is just the sample mean of response for treatment combination ijk, which is:

$$\hat{\mu} + \hat{\tau}_{ijk} = \bar{Y}_{ijk\cdot}$$

The unbiased estimate of σ^2 is:

$$\hat{\sigma}^2 = msE = \frac{ssE}{n - a * b * c}$$

The sum of squares for error (ssE) is:

$$ssE = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{t=1}^{r_{ijk}}(y_{ijkt} - \bar{y}_{ijk\cdot})^2$$

Since $\bar{Y}_{ijk\cdot}$ has a normal distribution with mean of $\mu + \tau_{ijk}$ and variance of $\frac{\sigma^2}{r_{ijk}}$, we can easily derive the point estimate, the confidence interval, and the hypothesis testing of a contrast

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}h_{ijk}\tau_{ijk}\,, and\ \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}h_{ijk} = 0$$

The formulas are similar with those in Lesson 4.2 in Week 4. Since we will rely on R for such statistical inference, we skip the detailed formulas here but focus on the estimation of σ^2. Note that Simple algebra formulas may not exist for reduced models and/or unbalanced experiments.

For any model, the unbiased estimate of σ^2 can written as:

$$\hat{\sigma}^2 = msE = \frac{ssE}{degrees\ of\ freedom\ for\ error}$$

Here, we outline the calculation of the degrees of freedom of error in a given model for experiments with three crossed treatment factors:

(1) degrees of freedom for error = total degrees of freedom – degrees of freedom for model
(2) total degrees of freedom = $n - 1$, where n is the total number of samples.
(3) degrees of freedom for model is the sum of degrees freedom of each factorial effect presented in the model:

degrees of freedom for ABC-interaction: $(a-1)(b-1)(c-1)$

degrees of freedom for AB-interaction: $(a-1)(b-1)$

degrees of freedom for AB-interaction: $(a-1)(c-1)$

degrees of freedom for BC-interaction: $(b-1)(c-1)$

degrees of freedom for main effects of A: $a-1$

degrees of freedom for main effects of B: $b-1$

degrees of freedom for main effects of C: $c-1$

Example 5.6: Degrees of Freedom – Popcorn Microwave Experiment

Find the degrees of freedom for error for a $a \times b \times c$ factorial experiment and Popcorn Microwave Experiment with the following models:

(1) The cell-mean or three-way complete model.
(2) $Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + \epsilon_{ijkt}$

Solution:

(1) Based on the one-way analysis of variance model, the degrees of freedom for error is:

$$n - abc$$

We can also calculate it with the aforementioned method too:

total degrees of freedom = $n - 1$

degrees of freedom for model

$$= (a-1)(b-1)(c-1) + (a-1)(b-1) + (a-1)(c-1)$$

$$+(b-1)(c-1) + a - 1 + b - 1 + c - 1$$

$$= (a-1)(b-1)c + (a-1)c + (b-1)c + c - 1$$

$$= c\big((a-1)(b-1) + a - 1 + b - 1 + 1\big) - 1 = abc - 1$$

degrees of freedom for error $= (n-1) - (abc-1) = n - abc$

For Popcorn Microwave Experiment, $n = 36, a = 3, b = 2, c = 3$, so

total degrees of freedom = $n - 1 = 35$

degrees of freedom for model $= abc - 1 = 3 * 2 * 3 - 1 = 17$

degrees of freedom for error $= 35 - 17 = 18$

(2) Similar with (1), we have:

total degrees of freedom = $n - 1$

degrees of freedom for model

$$= a - 1 + b - 1 + c - 1 + (a-1)(b-1) + (a-1)(c-1) = ab + ac - a - 1$$

degrees of freedom for error $= (n-1) - (ab - ac - a - 1) = n - ab - ac + a$

For Popcorn Microwave Experiment, $n = 36, a = 3, b = 2, c = 3$, so

total degrees of freedom = $n - 1 = 35$

degrees of freedom for model

$$= 3 - 1 + 2 - 1 + 3 - 1 + (3-1)(2-1) + (3-1)(3-1) = 11$$

degrees of freedom for error $= 35 - 11 = 24$

Exercise (Optional): Degrees of Freedom – Popcorn Microwave Experiment

Find the degrees of freedom for error for a $a \times b \times c$ factorial experiment and Popcorn Microwave Experiment with the following models:

(1) $Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijkt}$

(2) $Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik} + \epsilon_{ijkt}$

Solution:

(1) We have:

total degrees of freedom = $n - 1$

degrees of freedom for model

$$= a - 1 + b - 1 + c - 1 = a + b + c - 3$$

degrees of freedom for error $= (n - 1) - (a + b + c - 3) = n - a - b - c + 2$

For Popcorn Microwave Experiment, $n = 36, a = 3, b = 2, c = 3$, so

total degrees of freedom = $n - 1 = 35$

degrees of freedom for model $= a + b + c - 3 = 3 + 2 + 3 - 3 = 5$

degrees of freedom for error $= 35 - 5 = 30$

(2) Similar with (1), we have:

total degrees of freedom = $n - 1$

degrees of freedom for model

$$= a - 1 + b - 1 + c - 1 + (a - 1)(c - 1) = ac + b - 2$$

degrees of freedom for error $= (n - 1) - (ac + b - 2)$

For Popcorn Microwave Experiment, $n = 36, a = 3, b = 2, c = 3$, so

total degrees of freedom = $n - 1 = 35$

degrees of freedom for model

$$= 3 - 1 + 2 - 1 + 3 - 1 + (3 - 1)(3 - 1) = 9$$

degrees of freedom for error $= 35 - 9 = 26$

Methods for Multiple Comparisons

When planning the experiment, the experimenter should specify which treatment contrasts are of interest, together with overall error rates for hypothesis tests and overall confidence levels for confidence intervals. Depending on the model used, the following comparisons may be of interest: (1) treatment combinations (only applicable when the cell-means model is used); (2) main effects; (3) two-way interactions; and (4) three-way interactions (only applicable when the cell-means model or the three-way complete model is used). If α is the overall significance, then the experimenter needs to set the significance level of ith part, $\alpha_i (i = 1,2,3,4)$ and $\alpha = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$. According to the Bonferroni method, the experimentwise error rate will be not greater than $\alpha = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$. For the ith part, the experimenter can select the appropriate methods for multiple comparisons to control the overall error rate is not greater than α_i. Note the selection of $\alpha_i$should depend on experiments and there is no optimal way.

For this sub-section, we outline how to perform the methods for multiple comparisons for a given part and a given model (in other words, we are not limited to the cell-mean or the three-way complete model).

For any contrast, its contrast coefficients, which depend on parameters used, can be obtained accordingly. In addition, which methods of multiple comparisons can be used also depends on parameters used to determine the contrast coefficients. The simultaneous $(1-\alpha)100\%$ confidence interval of a contrast is:

$$point\ estimate \pm w * standard\ error$$

where the critical coefficient w for each of the four methods is, respectively,

Bonferroni Method: $w_B = t_{df,\alpha/(2*m)}$, m is the number pre-specified contrasts

Scheffe Method: $w_s = \sqrt{(x-1)F_{x-1,df,\alpha}}$

Tukey Method: $w_T = \frac{q_{x,df,\alpha}}{\sqrt{2}}$ (best and only for pair-wise comparison)

Dunnett Method: $w_{D2} = |t|^{(0.5)}_{x-1,df,\alpha}$ (best and only for treatment-versus-control)

where x is the length of contrast coefficients and df is the degrees of freedom for error. Let us use some examples to show how these methods can be applied.

Example 5.7: Multiple Comparison – Cell-Means or Three-Way Complete Model

Suppose that we use the cell-means model (or equivalently, the three-way complete model) an $a \times b \times c$ experiment and interested in calculating the simultaneous $(1-\alpha)100\%$ confidence interval of following groups of comparisons (hypothetically and pre-planned):

(1) All pairwise comparisons of main effects of factor A.
(2) All pairwise comparisons of main effects of factor B.
(3) Three two-way interactions between factors A and C:
    a. Two-way interaction between levels 1 and 2 from factor A and levels 1 and 2 from factor C
    b. Two-way interaction between levels 1 and 3 from factor A and levels 1 and 2 from factor C
    c. Two-way interaction between levels 2 and 3 from factor A and levels 1 and 2 from factor C

Solution: First, for the cell-mean or the three-way complete model, we have

$$degrees\ of\ freedom\ for\ error\ =\ df\ =\ n-abc$$

There are different ways to solve this problem.

First, we can use the cell-means model and the contrast coefficients are based on τ_{ijk}. In this situation, the length of contrast coefficients x, is, abc. We have:

Bonferroni Method: $w_B = t_{n-abc,\alpha/(2*m)}$, $m = \frac{a*(a-1)}{2} + \frac{b*(b-1)}{2} + 3$

Since we have $\frac{a*(a-1)}{2}$ pairwise comparisons of main effects of factor A, $\frac{b*(b-1)}{2}$ pairwise comparisons of main effects of factor B, and 3 contrasts for two-way interactions.

Scheffe Method: $w_s = \sqrt{(abc-1)F_{abc-1,n-abc,\alpha}}$

Tukey Method: not applicable since these comparisons are not pair-wise comparisons in terms of τ_{ijk}

Dunnett Method: not applicable

The second strategy is to set three significance levels α_1, α_2, and α_3 for three groups of comparisons, respectively. Then the simultaneous $(1-\alpha_i)100\%$ confidence interval of a contrast can be constructed within each group accordingly. Note to have the overall simultaneous $(1-\alpha)100\%$ confidence intervals, we have $\alpha = \alpha_1, +\alpha_2 + \alpha_3$. For example, we can set (again, hypothetically) $\alpha_1 = \frac{\alpha}{2}$, $\alpha_2 = \frac{\alpha}{4}$, and $\alpha_3 = \frac{\alpha}{4}$. The methods of multiple comparisons that can be used for the pairwise comparisons of main effects of factor A are:

Bonferroni Method: $w_B = t_{n-abc,\alpha_1/(2*m)}$, $m = \frac{b*(b-1)}{2}$

Scheffe Method: $w_s = \sqrt{(a-1)F_{a-1,n-abc,\alpha_1}}$ in terms of $(\alpha)_i^*$ (preferred)

$w_s = \sqrt{(abc-1)F_{abc-1,n-abc,\alpha_1}}$ in terms of τ_{ijk}

Tukey Method: $w_T = \frac{q_{a,n-abc,\alpha_1}}{\sqrt{2}}$ in terms of $(\alpha)_i^*$

Not applicable in terms of τ_{ijk}

Dunnett Method: not applicable.

In general, unless it involves the comparison of treatment combinations, the second strategy should be used.

Exercise 5.4: Multiple Comparison – Cell-Means or Three-Way Complete Model

Using the second strategy outlined in Example 5.7, determine which methods of multiple comparisons that can be used for three two-way interactions between factors A and C.

Solution:

For a given α_3, we have

Bonferroni Method: $w_B = t_{n-abc,\alpha_3/(2*m)}$, $m = 3$ since there are 3 contrasts

Scheffe Method: $w_s = \sqrt{(abc-1)F_{abc-1,n-abc,\alpha_3}}$ in terms of τ_{ijk}

$w_s = \sqrt{(ac-1)F_{ac-1,n-abc,\alpha_3}}$ in terms of $(\alpha\gamma)^*_{ik}$

Tukey Method: not applicable in terms of $(\alpha\gamma)^*_{ik}$, also not applicable in terms of τ_{ijk}

Dunnett Method: not applicable.

Example 5.8: Multiple Comparison – Reduced Model

Suppose there is an $a \times b \times c$ experiment and we are interested in calculating the simultaneous $(1-\alpha)100\%$ confidence interval of following groups of comparisons (hypothetically and pre-planned):

(1) All pairwise comparisons of main effects of factor A
(2) All pairwise comparisons of main effects of factor B
(3) Three two-way interactions between factors A and C:
    a. Two-way interaction between levels 1 and 2 from factor A and levels 1 and 2 from factor C
    b. Two-way interaction between levels 1 and 3 from factor A and levels 1 and 2 from factor C
    c. Two-way interaction between levels 2 and 3 from factor A and levels 1 and 2 from factor C

However, we know the three-way interaction negligible and use the following model:

$$Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{ik} + \epsilon_{ijkt}$$

Solution:

First, for the above model, we have

$$degrees\ of\ freedom\ for\ error\ =\ df\ =$$

$$(n-1) - (a-1+b-1+c-1+(a-1)(b-1)+(a-1)(c-1)+(b-1)(c-1))$$

For such reduced model, we can set $\alpha = \alpha_1, +\alpha_2 + \alpha_3$. For a given α_2, the methods of multiple comparisons that can be used for the pairwise comparisons of main effects of factor B are:

Bonferroni Method: $w_B = t_{df,\alpha_2/(2*m)}$, $m = \frac{b*(b-1)}{2}$

Scheffe Method: $w_s = \sqrt{(b-1)F_{b-1,df,\alpha_2}}$ in terms of $(\beta)_j^*$

Tukey Method: $w_T = \frac{q_{b,df,\alpha_2}}{\sqrt{2}}$ in terms of $(\beta)_j^*$

Dunnett Method: not applicable.

Exercise 5.5: Multiple Comparison – Reduced Model

Using the second strategy outlined in Example 5.7, determine which methods of multiple comparisons that can be used for three two-way interactions between factors A and C from Example 5.8.

Solution: For a given α_3, we have

$$degrees\ of\ freedom\ for\ error\ =\ df\ =$$

$$(n-1)-(a-1+b-1+c-1+(a-1)(b-1)+(a-1)(c-1)+(b-1)(c-1))$$

Bonferroni Method: $w_B = t_{df,\alpha_3/(2*m)}$, $m = 3$ since there are 3 contrasts

Scheffe Method: $w_s = \sqrt{(ac-1)F_{ac-1,df,\alpha_3}}$ in terms of $(\alpha\gamma)_{ik}^*$

Tukey Method: not applicable in terms of $(\alpha\gamma)_{ik}^*$

Dunnett Method: not applicable.

Example 5.9: Simultaneous Confidence Intervals - Popcorn Microwave Experiment

Since the three-way interaction seems to be negligible, we will use the following model in the analysis:

$$Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{ik} + \epsilon_{ijkt}$$

For Popcorn Microwave Experiment, we have:

Total sample size: $n = 36$

Factor A (Brand): $a = 3$

Factor B (Power): $b = 2$

Factor C (Popping time): $c = 3$

Suppose that we are interested in the simultaneous $(1-\alpha)100\%$ confidence intervals of the following groups of contrasts:

(1) All pairwise comparisons of main effects of factor A and the difference in response between two national brands and the local brand
(2) All pairwise comparisons of main effects of factor C
(3) Three two-way interactions between factors A and B:
    a. Two-way interaction between levels 1 and 2 from factor A and levels 1 and 2 from factor B
    b. Two-way interaction between levels 1 and 3 from factor A and levels 1 and 2 from factor B
    c. Two-way interaction between levels 2 and 3 from factor A and levels 1 and 2 from factor B
(4) The linear trend and quadratic trend for factor C.

Solution:

We will rely on R program for such calculations. The R codes for this example can be found in the file **week-5-prog-example.R**. Here, we illustrate how the simultaneous $(1-\alpha_1)100\%$ confidence interval of one contrast from group (1) is obtained. Since $\alpha = 0.05$, we set $\alpha_1 = 0.02$ here although other values of α_1 can be used too.

To use R to find the simultaneous $(1-0.02)100\% = 98\%$ confidence interval of contrasts from group (1), we need to specify the contrasts and their corresponding contrast coefficients in terms of $(\alpha)_i^*$:

Difference in Response	Contrasts	Contrast Coefficients
Levels 1 and 2	$(\alpha)_1^* - (\alpha)_2^*$	[1,-1,0]
Levels 1 and 3	$(\alpha)_1^* - (\alpha)_3^*$	[1,0,-1]
Levels 2 and 3	$(\alpha)_2^* - (\alpha)_3^*$	[0,1,-1]

Average of Levels 1 and 2 and Level 3	$\frac{(\alpha)_1^*+(\alpha)_2^*}{2}-(\alpha)_3^*$	[0.5,0.5,-1]

Note that τ_{ijk} can not be used since the cell-means model is not used here. From R program, we can get the following point estimates and confidence intervals for contrast $\frac{(\alpha)_1^*+(\alpha)_2^*}{2}-(\alpha)_3^*$:

Multiple Comparison	**Point Estimate**	**Standard Error**	**98% Confidence Interval**
None	4.88	3.04	(-2.74, 12.5)
Bonferroni	4.88	3.04	(-4.60, 14.4)
Scheffe	4.88	3.04	(-4.44, 14.2)

Note that Tukey and Dunnett methods are not applicable and the point estimate and standard error are same for all methods. For the model,

$$Y_{ijkt}=\mu+\alpha_i+\beta_j+\gamma_k+(\alpha\beta)_{ij}+(\alpha\gamma)_{ik}+(\beta\gamma)_{ik}+\epsilon_{ijkt}$$

The degrees of freedom of error (which can be obtained from R too) is:

$$(n-1)-\big(a-1+b-1+c-1+(a-1)(b-1)+(a-1)(c-1)+(b-1)(c-1)\big)$$

$$=(36-1)-\big(3-1+2-1+3-1+(3-1)(2-1)+(3-1)(3-1)+(2-1)(3-1)\big)$$

$$=35-(2+1+2+2+4+2)=22$$

The 98% confidence intervals of $\frac{(\alpha)_1^*+(\alpha)_2^*}{2}-(\alpha)_3^*$ is:

$$point\ estimate\ \pm w*standard\ error=4.88\pm w*3.04$$

Without multiple correction:

$$w=t_{22,\frac{0.02}{2}}=2.508$$

$$4.88\pm 2.508*3.04=(-2.74,12.50)$$

Bonferroni Method:

$$w_B=t_{22,\frac{0.02}{2*4}}=3.119$$

$$4.88 \pm 3.119 * 3.04 = (-4.60, 14.36)$$

Scheffe Method:

$$w_s = \sqrt{(3-1)F_{3-1,22,0.02}} = 3.065$$

$$4.88 \pm 3.065 * 3.04 = (-4.44, 14.20)$$

Since the confidence interval contains 0, therefore, we do not have sufficient evidence to show that there is a difference in response between two national brands and the local brand at the significance level of 0.05. Note we use α =0.05 instead of $\alpha_1 = 0.02$ here since we need to report the results at the overall significance level.

Week 5 Practice Problem Discussion

Consider the following practice problem and participate in the Week 5 Practice Problem Discussion for assistance, or to provide your own suggestions.

Set $\alpha_3 = 0.01$ and use R program to find simultaneous $(1 - 0.01) * 100\% = 99\%$ confidence intervals of three contrasts from (3) in **Example 5.9**. Fill the following tables and manually verify the second table. **Note that although $\alpha_3 = 0.01$ is used, the overall significance is still 0.05 since this group of contrasts is only one of four groups of contrasts considered.**

Interaction	**Contrast**	**Contrast Coefficients**
Levels 1 and 2 of *A* and Levels 1 and 2 of *B*		
Levels 1 and 3 of *A* and Levels 1 and 2 of *B*		
Levels 2 and 3 of *A* and Levels 1 and 2 of *B*		

Multiple Comparison	**Point Estimate**	**Standard Error**	**95% Confidence Interval**
None			
Bonferroni			
Scheffe			

Solution:

Interaction	Contrast	Contrast Coefficients
Levels 1 and 2 of A and Levels 1 and 2 of B	$[(\alpha\beta)_{11} - (\alpha\beta)_{12}] - [(\alpha\beta)_{21} - (\alpha\beta)_{22}]$	[1, -1, -1, 1, 0, 0]
Levels 1 and 3 of A and Levels 1 and 2 of B	$[(\alpha\beta)_{11} - (\alpha\beta)_{12}] - [(\alpha\beta)_{31} - (\alpha\beta)_{32}]$	[1, -1, 0, 0, -1, 1]
Levels 2 and 3 of A and Levels 1 and 2 of B	$[(\alpha\beta)_{21} - (\alpha\beta)_{22}] - [(\alpha\beta)_{31} - (\alpha\beta)_{32}]$	[0, 0, 1, -1, -1, 1]

From R program, we can get the following point estimates and confidence intervals for contrast for $[(\alpha\beta)_{21} - (\alpha\beta)_{22}] - [(\alpha\beta)_{31} - (\alpha\beta)_{32}]$:

Multiple Comparison	Point Estimate	Standard Error	95% Confidence Interval
None	3.05	7.02	(-16.74, 22.84)
Bonferroni	3.05	7.02	(-20.05, 26.15)
Scheffe	3.05	7.02	(-28.29, 34.39)

Without multiple correction:

$$w = t_{22,\frac{0.01}{2}} = 2.819$$

$$3.05 \pm 2.508 * 7.02 = (-16.74, 22.84)$$

Bonferroni Method:

$$w_B = t_{22,\frac{0.01}{2*3}} = 3.291$$

$$3.05 \pm 3.291 * 7.02 = (-20.05, 26.15)$$

Scheffe Method:

$$w_s = \sqrt{(6-1)F_{6-1,22,0.01}} = 4.465$$

$$3.05 \pm 4.465 * 7.02 = (-28.29, 34.39)$$

Lesson 5.3 Experiments with Three or More Crossed Treatment Factors: Analysis of Variance

Required Reading: Sections 7.3 and 7.4 from Chapter 7

Analysis of Variance Model

Similar with the analysis of experiments with two crossed treatment factors, the hypothesis testing of H_0 for a model of experiment with three or more crossed treatment factors can be done by the following F-test statistic:

$$F = \frac{(SSE_{Reduced} - SSE_{Full})/(df\ of\ SSE\ in\ Reduced\ Model - df\ of\ SSE\ in\ Full\ Model)}{SSE_{Full}/(df\ of\ SSE\ in\ Full\ Model)}$$

In the formula. df refers to the degrees of freedom, and the reduced model refers to the model when the null hypothesis is true. In some situations, the reduced model can be written out explicitly. In other situations, the reduced model cannot be described by simple algebra formulas. Under the null hypothesis (H_0 is true), the above F-test statistic has F-distribution with $df_1 = df\ of\ SSE\ in\ Reduced\ Model - df\ of\ SSE\ in\ Full\ Model$ and $df_2 = df\ of\ SSE\ in\ Full\ Model$ degrees of freedom, and H_0 is rejected at the significance level of α if

$$F = \frac{(ssE_{Reduced} - ssE_{Full})/df_1}{ssE_{Full}/df_2} > F_{df_1, df_2, \alpha}$$

and the corresponding $p - value$ is:

$$p - value = P(F_{df_1, df_2} \geq F = \frac{(ssE_{Reduced} - ssE_{Full})/df_1}{ssE_{Full}/df_2})$$

where F_{df_1, df_2} is a random variable with F-distribution with df_1 and df_2 degrees of freedom.

Similarly, the following analysis of variance table can be obtained for the three-way complete model:

Source of Variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**F-test Statistic**
Factor A	$a-1$	ssA	$msA = \frac{ssA}{a-1}$	$\frac{msA}{msE}$
Factor B	$b-1$	ssB	$msB = \frac{ssB}{b-1}$	$\frac{msB}{msE}$
Factor C	$c-1$	ssC	$msC = \frac{ssC}{c-1}$	$\frac{msC}{msE}$
AB	$(a-1)(b-1)$	$ssAB$	$msAB = \frac{ssAB}{(a-1)(b-1)}$	$\frac{msAB}{msE}$
AC	$(a-1)(c-1)$	$ssAC$	$msAC = \frac{ssAC}{(a-1)(c-1)}$	$\frac{msAC}{msE}$
BC	$(b-1)(c-1)$	$ssBC$	$msBC = \frac{ssBC}{(b-1)(c-1)}$	$\frac{msBC}{msE}$
ABC	$(a-1)(b-1)(c-1)$	$ssABC$	$msABC = \frac{ssABC}{(a-1)(b-1)(c-1)}$	$\frac{msABC}{msE}$
Error	$n-abc$	ssE	$msE = \frac{ssE}{n-abc}$	
Total	$n-1$	$sstot$		

Type I and Type III Sum of Squares

Again there are different ways to calculate ssA, ssB, etc. in the model. As we have discussed in Lesson 4.4 in Week 4, we focus on two type of sum of squares.

Type I sum of squares is also called sequential sum of squares. Type I sum of squares of each term is adjusted for only term that precedes it in the model. For a three-way complete model,

$$Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkt}$$

We assume that the factor A, factor B, factor C, AB interaction, AC interaction, and ABC interaction are added to the model one by one, then we can have the following seven models:

$$M0: Y_{ijkt} = \mu + \epsilon_{ijkt}$$

$$M1: Y_{ijkt} = \mu + \alpha_i + \epsilon_{ijkt}$$

$$M2: Y_{ijkt} = \mu + \alpha_i + \beta_j + \epsilon_{ijkt}$$

$$M3: Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijkt},$$

$$M4: Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + \epsilon_{ijkt},$$

$$M5: Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + \epsilon_{ijkt},$$

$$M6: Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \epsilon_{ijkt},$$

$$M7: Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkt},$$

Then the type I sum of squares for the factor A, factor B, factor C, AB interaction, AC interaction, and ABC interaction can be calculated by the following way:

$$ssA = ssE\ of\ model\ M0 - \ ssE\ of\ model\ M1$$

$$ssB = ssE\ of\ model\ M1 - \ ssE\ of\ model\ M2$$

……

$$ssAB = ssE\ of\ model\ M3 - \ ssE\ of\ model\ M4$$

……

$$ssBC = ssE\ of\ model\ M5 - \ ssE\ of\ model\ M6$$

$$ssABC = ssE\ of\ model\ M6 - \ ssE\ of\ model\ M7$$

For example, ssA is the difference of the sum of squares for error between $M0$ and $M1$ since the factor A is first term in the model, while ssB is the difference of the sum of squares for error between $M1$ and $M2$ since the factor B is second term after the factor A in the model. Therefore,

the order of terms in the model affect the type I sum of squares. In addition, we have the following equation for the type I sum of squares:

$$sstot = ssA + ssB + ssC + ssAB + ssAC + ssBC + ssABC + ssE$$

And

$$ssT = ssA + ssB + ssC + ssAB + ssAC + ssBC + ssABC$$

Type III sum of squares for each term is adjusted for any other terms that don't contain the effect and orthogonal to any effects that contain the effect. For example, for a three-way complete model, ssA is the difference of the sum of squares for error between the three-way complete model and a reduced model. Since the factor B and BC interaction do not contain the effects of factor A and AB interaction, AC interaction, and ABC interaction contain the effects of factor A, here the reduce model contain the effects of factor B, BC interactio and the effects from the interactions that are orthogonal to the interaction effects that contain the effects of factor A. Therefore, the order of terms in the model does not affect the type III sum of squares. To test the an individual hypotheses (e.g., H_0^{BC}, H_0^A, or H_0^C), the type III sum of squares should be used.

For balanced experiments, the type I sum of squares of a term is same its type III sum of squares; while for unbalance experiments, the type I sum of squares of a term is generally different from its type III sum of squares.

Example 5.10: Type I and Type III Sum of Squares - Popcorn Microwave Experiment, Balanced

The following three-way complete model is used in the example:

$$Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \epsilon_{ijkt}$$

The following analysis of variance table for the type I sum of squares can be obtained:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F-test Statistic	*p*-value

A (Brand)	2	333.10	165.55	1.889	0.180
B (Power)	1	455.11	455.11	5.192	0.035
C (Time)	2	1554.58	777.29	8.867	0.002
AB	2	196.04	98.02	1.118	0.349
AC	4	1433.86	358.46	4.089	0.016
BC	2	47.71	23.85	0.272	0.765
ABC	4	47.33	11.83	0.135	0.967
$Error$	18	1577.87	87.66		
$Total$	35	5643.60			

For the null hypothesis of

$$H_0: \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} \text{ are equal}$$

We can calculate F-test statistic from the type I sum of squares:

$$F = \frac{msT}{msE} = \frac{(333.10 + 455.11 + \cdots + 47.33)/17}{87.66} = 2.73$$

The corresponding p-value is:

$$p - value = \Pr(F_{17,18} \geq 2.73) = 0.021$$

The F-test statistic and p-value can be directly obtained by either using a cell-means model or the comparison of the three-way complete model and the reduced model under H_0. Our conclusion here is that the different treatment combinations have different effects on the percentage of successfully popped kernels.

For this balanced experiment, the type III sum of squares is same as the type I sum of squares. Based on p-values in the table, we can make conclusions about those effects. For example,

- At the significance level of 0.05, we fail to reject H_0^{ABC} so we do not have sufficient evidence to reject that the three-way interaction is negligible.
- We reject H_0^{AC} at the significance level of 0.05 and conclude that there is a significant difference on the percentage of successfully popped kernels of combinations of brand and popping time (interaction on the response between brand and popping time) averaged over two power setting.

Example 5.11: Type I and Type III Sum of Squares - Popcorn Microwave Experiment, Unbalanced

We use the data from the file: **popcorn.microwave.drop.one.txt**. This data is obtained by removing the last sample from the original data.

Instead of three-way complete model, we remove ABC interaction and BC interaction from the model and use the following model:

$$Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + \epsilon_{ijkt}$$

In R program, the order of the terms in the model is: A (brand), B (power), C (popping time, AB interaction, and AC interaction.

The analysis of variance table for the type I sum of squares is:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F-test Statistic	*p*-value
A	2	378.19	189.09	2.689	0.0893
B	1	407.08	407.08	5.788	0.0245
C	2	1730.86	865.43	12.304	0.00023
AB	2	139.45	69.73	0.991	0.386
AC	4	1346.56	336.64	4.786	0.0059
Error	23	1617.71	70.34		
Total	35	5619.85			

The analysis of variance table for the type III sum of squares is:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F-test Statistic	*p*-value
A	2	375.34	174.67	2.668	0.091
B	1	376.99	376.99	5.360	0.0299
C	2	1602.39	1801.20	11.391	0.00036
AB	2	153.10	76.65	1.088	0.354
AC	4	1346.56	336.64	4.786	0.0059
Error	23	1617.71	70.34		
Total	35	5619.85			

Note that Type I and III sum of squares for A, B, C, and AB interaction are different but Type I and III sum of squares for AC interaction are same since AC interaction is the last term in both models. However, all conclusions based on the type I sum of squares and Type III sum of squares are similar for this data since although the data is unbalanced but it is very close to a balanced data.

Exercise 5.7 - Type I and Type III Sum of Squares for Unbalance Experiments - Popcorn Microwave Experiment

Use the data from **popcorn.microwave.drop.one.txt** and obtain the tables for Type I and III sum of squares with the following model:

$$Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\gamma)_{ik} + (\alpha\beta)_{ij} + \epsilon_{ijkt}$$

Note that the order of the terms in the model is: A (brand), B (power), C (popping time, AC interaction, and AB interaction. Are your tables same as the tables from Example? Can you explain where some of sum of squares are same while the others are different?

Solution:

The analysis of variance table for the type I sum of squares is:

Source of Variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**F-test Statistic**	***p*-value**
A	2	378.19	189.09	2.689	0.0893
B	1	407.08	407.08	5.788	0.0245
C	2	1730.86	865.43	12.304	0.00023
AC	4	1332.92	333.23	4.738	0.0617
AB	2	153.10	76.55	1.088	0.354
Error	23	1617.71	70.34		
Total	35	5619.85			

The analysis of variance table for the type III sum of squares is:

Source of Variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**F-test Statistic**	***p*-value**
A	2	375.34	174.67	2.668	0.091
B	1	376.99	376.99	5.360	0.0299
C	2	1602.39	1801.20	11.391	0.00036
AC	4	1346.56	336.64	4.786	0.0059
AB	2	153.10	76.65	1.088	0.354
Error	23	1617.71	70.34		
Total	35	5643.60			

Comparison of Type I sum of Squares between the model in Example 5.10 and model in this exercise:

$ssA, ssB, ssC, ssE, sstot$: same

$ssAB, ssAC$: different

Comparison of Type III sum of Squares between the model in Example 5.10 and model in this exercise: all sums of squares are same.

Lesson 5.4 Sample Sizes and Power

Required Reading: Sections 6.6 in Chapter 6.

In Lesson 2.2 in Week 2, we discussed how to perform the power and sample size calculation for experiments with one treatments factor. The following notations are used:

- α: significance level
- v: number of treatment levels
- $r = r_1 = \cdots = r_v$: number of replicates for each level
- π: power
- σ^2: variance in the model
- Δ: a threshold for the differences in the effects of the treatments

To use Table A.7 to calculate the sample sizes or power, we have:

- $v_1 = v - 1$
- $v_2 = degrees\ of\ freedom\ for\ error$. For one-way analysis of variance model, we have $v_2 = n - v = v * r - v$
- The formula for the sample size (r) and ϕ that is used in Table A.7 for power calculation:

$$\phi = \sqrt{\frac{r * \Delta^2}{2 * v * \sigma^2}}$$

For experiments with two or more crossed treatment factors, such formula can be modified and used to calculate the power/sample sizes of each factor separately. To do so, we need to the following notations and formulas:

- v: number of levels of corresponding treatment factor. For example,
  - $v = a$ if we are interested in the power/sample size of factor A
  - $v = c$ if we are interested in the power/sample size of factor C
- $v_1 = v - 1$
- r: number of replicates for each treatment combination
- π: power
- σ^2: variance in the model
- Δ: a threshold for the differences in the effects of the treatments

- $v_2 = degrees\ of\ freedom\ for\ error$. This depends on the models since there are more than one model can be used to experiments with two or more factors. For example, for an experiment with two treatment factors,

  $v_2 = n - a * b = a * b * r - a * b$ if two-way complete model is used.

  $v_2 = n - a - b + 1 = a * b * r - a - b + 1$ if main-effects model is used.

- The formula for the sample size (r) and ϕ that is used in Table A.7 for power calculation:

$$\phi = \sqrt{\frac{r^* * \Delta^2}{2 * v * \sigma^2}}$$

  where r^* is number of samples for each level of corresponding treatment factor. For example, suppose we are interested in the power/sample size of factor B, then

  $r^* = a * r, v = b$ for an $a \times b$ factorial experiment,

  $r^* = a * c * r, v = b$ for an $a \times b \times c$ factorial experiment,

Example 5.12: Power and Sample Size - Popcorn Microwave Experiment

Suppose that we are interested in if the current study has the sufficient power to detect the pairwise difference for factor A (brand). Since the three-way interaction is negligible, we use the following model:

$$Y_{ijkt} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{ik} + \epsilon_{ijkt}$$

Solution:

From the model, the least squares estimates are:

$$\hat{\alpha}_1^* - \hat{\alpha}_2^* = 4.84, \hat{\alpha}_1^* - \hat{\alpha}_3^* = 7.30, \hat{\alpha}_2^* - \hat{\alpha}_3^* = 2.46, \hat{\sigma}^2 = 73.87$$

We can take:

$$\Delta = 2.46, \sigma^2 = 73.87, \alpha = 0.05, a = 3, b = 2, c = 3, r = 2$$

And we have:

$$v_1 = a - 1 = 3 - 1 = 2$$

$$v_2 = 22$$

$$r^* = b * c * r = 2 * 3 * 2 = 12$$

$$\phi = \sqrt{\frac{r^* * \Delta^2}{2 * a * \sigma^2}} = \sqrt{\frac{12 * 2.46^2}{2 * 3 * 73.87}} = 0.40$$

From Table A.7, the power is 0.282 for $\alpha = 0.05$, $v_1 = 2$, $v_2 = 20$ (since $v_2 = 18$ are not in the table), and $\phi = 1.00$. So we would see the power of current experiment is quite low.

Exercise Problem 5.8 (Popcorn Microwave Experiment). Suppose that we are interested in if the current study has the sufficient power to detect the pairwise difference for factor C (popping time) and the three-way complete model is used, perform the power analysis for the current experiment.

Solution:

From the model, the least squares estimates are:

$$\gamma_1^* - \hat{\gamma}_2^* = -6.97, \hat{\gamma}_1^* - \hat{\gamma}_3^* = 9.08, \hat{\gamma}_2^* - \hat{\gamma}_3^* = 16.05, \sigma^2 = 87.66$$

We can take:

$$\Delta = -6.97, \sigma^2 = 87.66, \alpha = 0.05, a = 3, b = 2, c = 3, r = 2$$

And we have:

$$v_1 = c - 1 = 3 - 1 = 2$$

$$v_2 = 18$$

$$r^* = a * b * r = 3 * 2 * 2 = 12$$

$$\phi = \sqrt{\frac{r^* * \Delta^2}{2 * c * \sigma^2}} = \sqrt{\frac{12 * (-6.97)^2}{2 * 3 * 87.66}} = 1.05$$

From Table A.7, the power is 0.282 for $\alpha = 0.05$, $v_1 = 2$, $v_2 = 20$ (since $v_2 = 18$ are not in the table), and $\phi = 1.00$. So we would see the power of current experiment is quite low.

MA4720: Design and Analysis of Experiments

Week 6 At-a-Glance

Title: Small Experiment and Complete Block Designs

Overview

This, we will discuss the analysis of single replicate experiments and design and analysis of experiments with blocking factors. We will present several analytical methods for single replicate experiments, including the analysis based on orthogonal contrasts, the analysis assuming that certain interaction effects are negligible, and the analysis using half-normal probability plot of effect estimates. We will discuss design issues of experiments with blocking factors. We will discuss the statistical models for the randomized complete block designs and general complete block designs.

Objectives

When you complete this module, you should be able to do the following:

1. Construct orthogonal contrasts and use them to analyze single replicate experiments.
2. Use half-normal probability plot of effect estimates to determine the appropriate model for the analysis of single replicate experiments.
3. Design experiments with randomized complete block deigns and general complete block designs.
4. Use appropriate models for the analysis of randomized complete block deigns and general complete block designs.
5. Calculate sample size and power and check model assumptions for randomized complete block deigns and general complete block designs

Reading (Suggested)

1. Section 6.7 in Chapter 6.
2. Section 7.5 in Chapter 7.
3. Sections 10.1, 10.3, 10.4, 10.5, 10.6, 10.7, and 10.8 in Chapter 10.

Instruction Content

See other files for details.

Quiz

Two quizzes. See other files for details.

Homework

One homework. See other files for details.

Final Project

One component for the final project. See other files for details.

MA4720: Design and Analysis of Experiments

Week 6 Instruction Contents

Lesson 6.1 Analysis of Single Replicate Experiments

Required Reading: Section 6.7 from Chapter 6 and Section 7.5 from Chapter 7

When observations are extremely time-consuming or expensive to collect, an experiment may be designed to have $r = 1$ observation on each treatment combination. Such experiments are called ***experiments with one observation per cell*** or ***single replicate experiments***. Since the ability to choose the sample sizes is lost, it should be recognized that confidence intervals may be wide and hypothesis tests not very powerful.

For such experiments, the cell-means model or the equivalent model with all possible interaction terms cannot be used since the degrees of freedom for error will be 0 for such models. Thus, a single replicate experiment with the cell-means model (or the equivalent model with all possible interaction terms) can be analyzed only if one of the following is true:

(1) σ^2 is known in advance.
(2) The interaction is expected to be of a certain form that can be modeled with fewer than degrees of freedom.
(3) The number of treatment combinations is large, and only a few contrasts are likely to be nonnegligible (***effect sparsity***).

However, it is unlikely to know σ^2 in advance, so we will not pursue it in this course. In this lesson, we will introduce several methods for the second and third situations.

Analysis Based on Orthogonal Contrasts

Two estimable contrasts are called orthogonal contrasts if and only if their least squares estimators are uncorrelated or, equivalently, have zero covariance. For the moment, we recode the treatment combinations with the lexical order to obtain a single-digit code. For example, for $a \times b$ factorial experiments, we code τ_{ij}'s as $\tau_1, \tau_2, \cdots, \tau_v$ where $v = ab$. While for $a \times b \times c$ factorial

experiments, we code τ_{ijk}'s as $\tau_1, \tau_2, \cdots, \tau_v$ where $v = abc$. It can be verified that two contrasts, $\sum_{i=1}^{v} c_i\tau_i$ and $\sum_{i=1}^{v} k_i\tau_i$ are orthogonal if and only if

$$\sum_{i=1}^{v} \frac{c_i k_i}{r_i} = 0$$

If the sample sizes are equal, then this reduces to

$$\sum_{i=1}^{v} c_i k_i = 0$$

In other words, if the sample sizes are equal, then two contrasts are orthogonal is and only if the dot product of their contrast coefficients is 0.

Note that experiments with one observation per cell, the sample sizes are equal since there is only one observation per cell. In addition, for v treatments, or treatment combinations, a set of $v - 1$ orthogonal contrasts is called a complete set of orthogonal contrasts. It is not possible to find more than $v - 1$ contrasts that are mutually orthogonal. We write the sum of squares for the qth orthogonal contrast in a complete set as ssc_q, then the sum of squares for treatments, ssT, can be partitioned into the sums of squares for the $v - 1$ orthogonal contrasts in a complete set; that is,

$$ssT = ssc_1 + ssc_2 + \cdots + ssc_{v-1}$$

In addition, if all the $v - 1$ orthogonal contrasts are in the model, then the sum of squares for treatments, ssT, is the total sum of squares, in other words,

$$ssT = ssc_1 + ssc_2 + \cdots + ssc_{v-1} = sstot$$

Therefore, if we know *in advance* that *e* orthogonal contrasts are likely to be negligible, then the sums of squares for these *e* negligible contrasts can be pooled together to obtain an estimate of error variance, based on *e* degrees of freedom,

$$ssE = \sum_{h=1}^{e} ssc_h \, , msE = \frac{ssE}{e}$$

Then the sums for the remaining interaction contrast can be used to test any individual contrasts and combinations of contrasts.

Example 6.1: Example of Orthogonal Contrasts – Trend Contrasts

For experiments with v treatments, or treatment combinations, trend contrasts provide a set of $v - 1$ contrasts that are mutually orthogonal.

For example, for $v = 4$, the trend contrasts are (from Table A.2):

Trend	Contrast Coefficients in Terms of $\tau_1, \tau_2, \cdots, \tau_v$
Linear	[-3,-1,1,3]
Quadratic	[1, -1, -1, 1]
Cubic	[-1, 3, – 3, 1]

We can verify:

Linear and Quadratic: $(-3) * 1 + (-1) * (-1) + 1 * (-1) + 3 * 1 = 0$

Linear and Cubic: $(-3) * (-1) + (-1) * 3 + 1 * (-3) + 3 * 1 = 0$

Quadratic Cubic: $1 * (-1) + (-1) * 3 + (-1) * (-3) + 1 * 1 = 0$

Exercise 6.1: An Example of Orthogonal Contrasts

For experiments with $v = 4$ treatments, or treatment combinations, verify the following set of 3 contrasts in terms of $\tau_1, \tau_2, \cdots, \tau_v$ are mutually orthogonal:

$$c_1 = [1, -1, 0, 0]$$
$$c_2 = [1, 1, -2, 0]$$
$$c_3 = [1, 1, 1, -3]$$

Solution: The dot products of these contrast coefficients:

c_1 and c_2: $1 * 1 + (-1) * 1 + 0 * (-2) + 0 * 0 = 0$

c_1 and c_3: $1 * 1 + (-1) * 1 + 0 * 1 + 0 * (-3) = 0$

c_2 and c_3: $1 * 1 + 1 * 1 + (-2) * 1 + 0 * (-3) = 0$

It is also worth mentioning the meaning of these three contrasts:

- c_1 is the difference between the first and second level.
- c_2 is the difference between the average of first and second level and the third level (up to a constant) since $c_2 = [1,1,-2,0] = \frac{[0.5,0.5,-1,0]}{2}$.
- Similarly, c_3 is the difference between the average of first to third level and the fourth level (up to a constant).

Exercise 6.2: An Example of Non-Orthogonal Contrasts

For experiments with $v = 4$ treatments, or treatment combinations, verify the following set of 3 pairwise contrasts in terms of $\tau_1, \tau_2, \cdots, \tau_v$ are not orthogonal:

$$c_1 = [1,-1,0,0]$$
$$c_2 = [1,0,-1,0]$$
$$c_3 = [1,0,0,-1]$$

Solution: Skipped.

Example 6.2: A Real Experiment – Air Velocity Experiment

The data can be found in file **air.velocity.txt** and its description can be found in file **Data-Sets-Summary.docx**. The data is part of an experiment described by D.Wilkie in the 1962 issue of Applied Statistics (volume 11, pages 184–195). The experiment was designed to examine the position of maximum velocity of air blown down the space between a roughened rod and a smooth pipe surrounding it. The treatment factors were the height of ribs on the roughened rod (factor A) at equally spaced heights 0.010, 0.015, and 0.020 inches (coded 1, 2, 3) and Reynolds number (factor B) at six levels (coded 1–6) equally spaced logarithmically over the range 4.8–5.3. The responses were measured as $y = (d - 1.4) \times 103$, where d is the distance in inches from the center of the rod. **Figure 6.1** shows very little interaction between the factors. However, prior to the experiment, the investigators had thought that the factors would interact to some extent. They wanted to use the set of orthogonal polynomial trend contrasts for the AB interaction and were reasonably sure that the contrasts $A_Q B_{qr}$, $A_L B_{qn}$, $A_Q B_{qn}$ would be negligible.

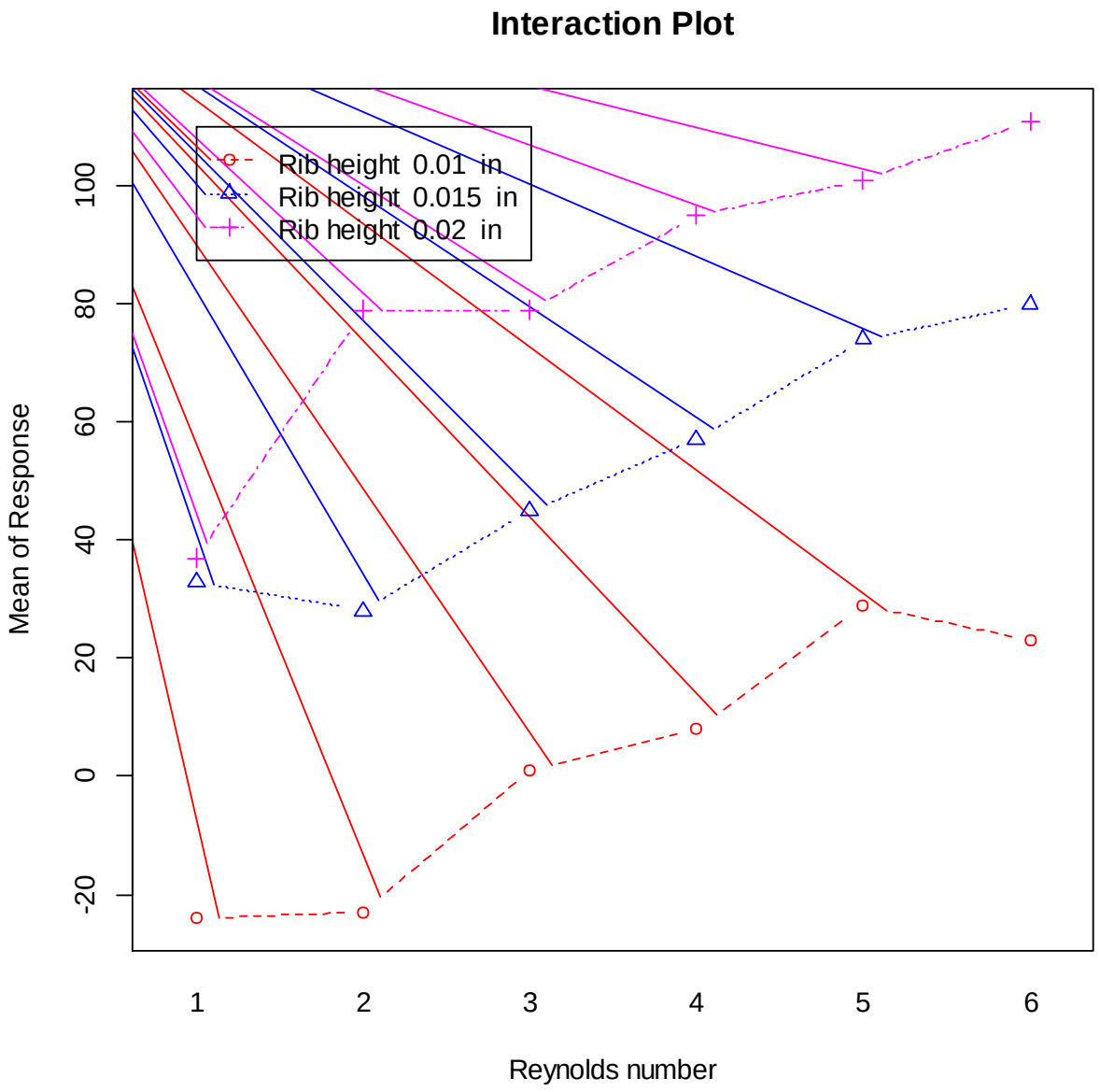


Figure 6.1. Interaction plot for the air velocity experiment.

Solution: We have $v = 18$ so we can construct a set of 17 orthogonal contrasts. For experiments with two or more crossed experiments, there are multiple ways to construct a set of orthogonal contrasts. Here we introduce the method used in this example. For this 3×6 experiment, we can have:

$2 = 3 - 1$ orthogonal contrasts for factor A (the height of ribs on the roughened rod)

$5 = 6 - 1$ orthogonal contrasts for factor B (Reynolds number)

$10 = 2 * 5$ orthogonal contrasts for AB-intreaction

In this example, since both factors A and B are equally spaced, the polynomial trend contrasts are used. For factor A, the linear trend (A_L) and the quadratic trend (A_Q) are used. For factor B, the linear trend (B_L), the quadratic trend (B_Q), the cubic trend (B_C), the quadratic trend (B_{qr}), and the quantic trend (B_{qn}) are used. We are using "L, Q, C, qr, qn" as shorthand notation for linear, quadratic, cubic, quartic, and quintic contrasts, respectively. The coefficients for these contrasts can be obtained from Table A.2. The contrasts for interaction terms are obtained by multiplying the corresponding main-effect coefficients.

These coefficients can be found in file **air.velocity.contrasts.txt**. Here we use $A_Q B_L$ as an example to illustrate how these coefficients are obtained:

The coefficients for the quadratic trend contrast of factor A in terms of α_i^* are

$$A_Q: [1, -2, 1]$$

These coefficients in terms of $\tau_1, \tau_2, \cdots, \tau_{18}$ are:

$$A_Q: \frac{[1,1,1,1,1,1, -2, -2, -2, -2, -2, -2, 1,1,1,1,1,1]}{6}$$

The divisor 6 can be omitted if we are only interested in hypothesis testing, so:

$$A_Q: [1,1,1,1,1,1, -2, -2, -2, -2, -2, -2, 1,1,1,1,1,1]$$

The coefficients for the linear trend contrast of factor B in terms of β_j^* are

$$B_L: [-5, -3, -1, 1, 3, 5]$$

These coefficients in terms of $\tau_1, \tau_2, \cdots, \tau_{18}$ are (up to a constant):

$$B_L: [-5, -3, -1, 1, 3, 5, -5, -3, -1, 1, 3, 5, -5, -3, -1, 1, 3, 5]$$

Therefore, the contrast coefficients for $A_Q B_L$ interaction are (in terms of $\tau_1, \tau_2, \cdots, \tau_{18}$):

$$A_Q B_L: [-5, -3, -1, 1, 3, 5, 10, 6, 2, -2, -6, -10, -5, -3, -1, 1, 3, 5]$$

Once these contrast coefficients are obtained, they can be used in the linear model. Since the contrasts $A_Q B_{qr}$, $A_L B_{qn}$, $A_Q B_{qn}$ were considered as negligible in advance, the sum of squares for error is 175.74 with 3 degrees of freedom. This can be used to test each contrast is zero. For example, if we would like to test if the linear contrast of factor A is zero at the significance level of 0.05, we have the F-test statistic:

$$F = \frac{ss(A_L)/1}{ssE/3} = \frac{19845.33/1}{175.74/3} = 338.77$$

The p-value is:

$$p - value = P\left(F_{1,3} > 338{,}77\right) = 0.0003 < \frac{0.05}{14}$$

We reject the null hypothesis of the linear contrast of factor A with the Bonferroni method, This indicates there is a significant linear trend of factor A and validates our observation from Figure 6.1. Also as expected from Figure 6.1, none of interaction effects is significant.

Exercise 6.3: Another Example of Orthogonal Contrasts

Verify the three contrast, A_Q, B_L, and $A_Q B_L$ are mutually orthogonal.

Solution: Skipped.

Exercise 6.4: Hypothesis Testing and Checking Assumptions - Air Velocity Experiment

Use the model in Example 6.2:

(1) Perform the hypothesis test H_0: Both A_L and A_Q are 0 at the significance level of 0.05. This is corresponding to test if there is a main effect of factor A.

(2) Check the assumptions on the model by plotting the standardized residuals against the predicted responses, the treatment factor levels, and the normal scores. State your conclusions.

Solution:

(1) From the linear model, we have:

$$ss(A_L) = 19845.33, ss(A_Q) = 386.78, ssE = 175.74$$

So the test statistic is:

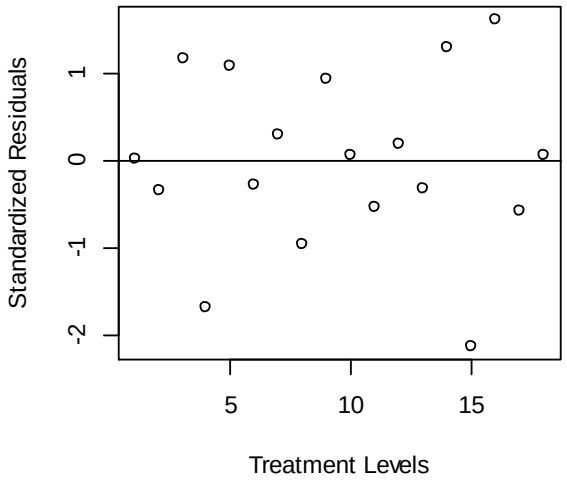
$$F = \frac{(19845.33 + 386.78)/2}{175.74/3} = 172.69$$

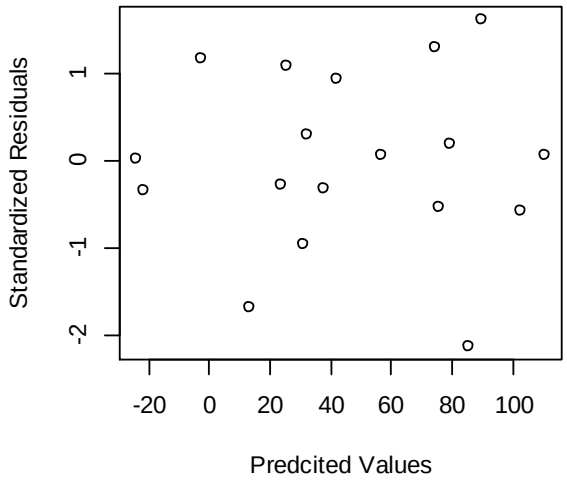
The p-value is:

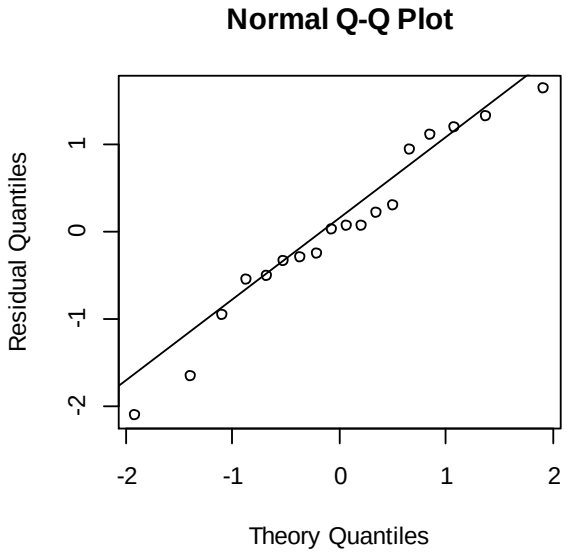
$$p - value = P(F_{2,3} > 172.69) = 0.000799$$

So we reject the null hypothesis, indicating there is a main effect of A.

(2) The following plots are obtained: all assumptions are satisfied.







Analysis Assuming that Certain Interaction Effects are Negligible

For a single replicate factorial experiment, if the experimenter knows ahead of time that certain interactions are negligible, then by excluding those interactions from the model, the corresponding degrees of freedom can be used to estimate the error variance. It must be recognized, however, that if interactions are incorrectly assumed to be negligible, then msE will be inflated, in which case the results of the experiment may be misleading. Actually, Example 6.2 can be considered as

an example of this analytical method: the interaction contrasts $A_Q B_{qr}$, $A_L B_{qn}$, $A_Q B_{qn}$ were considered as negligible thus excluded from the analysis.

Example 6.3: Models without Negligible Interactions - Drill Advance Experiment

The data can be found in file **drill.advance.txt** and the description can be found in file **Data-Sets-Summary.docx**. Daniel (1976) described a single replicate $2 \times 2 \times 2 \times 2$ experiment to study the effects of four treatment factors on the rate of advance of a small stone drill. The treatment factors were "load on the drill" (A), "flow rate through the drill" (B), "speed of rotation" (C), and "type of mud used in drilling" (D). Each factor was observed at two levels, coded 1 and 2. The author examined several different transformations of the response and concluded that the **log transform** (to the base 10) was one of the more satisfactory ones. **In many experiments with a number of treatment factors, experimenters are willing to believe that some or all of the interactions are very small.** For model with only main-effects, the following analysis of variance table are obtained:

Source of variation	Degrees of freedom	Sum of Squares	Mean Square	Ratio	p-value
A	1	0.01275	0.01275	7.02	0.0226
B	1	0.25837	0.25837	139.74	<0.0001
C	1	1.00550	1.00550	553.46	0.0001
D	1	0.08045	0.08045	44.28	<0.0001
Error	11	0.01998	0.00182		
Total	15	1.37254			

We can see that the degrees of freedom are $11 = 15 - 4$.

Simultaneous confidence intervals for the $m = 4$ main-effect contrasts using Bonferroni's method at an overall level of at least 95% can be calculated from R program with $\alpha_A = \alpha_B = \alpha_C = \alpha_D = \frac{0.05}{4} = 0.0125$:

$$A\text{: } (-0.0071, 0.120)$$

$$B: (0.188, 0.315)$$

$$C: (0.438, 0.565)$$

$$D: (0.078, 0.205)$$

We see that the high levels of B, C, D give a somewhat higher response in terms of log drill advance (with overall confidence level at least 95%), whereas the interval for A includes zero, indicating that we do not have sufficient evidence to reject the two levels of A have same effect on the response.

Analysis Using Half-Normal Probability Plot of Effect Estimates

The aforementioned methods for the analysis of single replicate experiment require to know the contrasts that are negligible in advance. However, in many situations, we may have no information about which contrasts are negligible. In this section, we discuss a method to identify of nonnegligible contrasts. This method works well under effect sparsity, that is, when most of the treatment contrasts under examination are negligible.

For a single replicate factorial experiment, with v treatment combinations, one can find a set of $v - 1$ orthogonal contrasts. When these are normalized, the contrast estimators all have variance σ^2. If the assumptions of normality, equal variances, and independence of the response variables are approximately satisfied, the estimates of negligible contrasts are like independent observations from a normal distribution with mean zero and variance σ^2. If we were to plot the normalized contrast estimates against their normal scores (in the same way that we checked for normality of the error variables in Lesson 3.3), the estimates of negligible effects would tend to fall nearly on a straight line. Any contrast for which the corresponding estimate would appear to be far from the line would be considered to be nonnegligible. The sign of such a nonnegligible contrast could be positive or negative and this depends upon which level of the factor is labeled as the high level and which is labeled as the low level. For many factors (especially qualitative factors), these designations are arbitrary, and so it is common to use a **half-normal probability**

plot for detecting nonnegligible contrasts. This is obtained by plotting the absolute values of the normalized contrast estimates against their half-normal scores.

Specifically, the steps to use the half-normal plot to a single replicate factorial experiment with v treatment combinations, are:

Step 1: Find $v - 1$ orthogonal contrasts and their contrast coefficients, denote them as:

$$C_1: [c_{11}, c_{12}, \cdots, c_{1v}]$$

$$C_2: [c_{21}, c_{22}, \cdots, c_{2v}]$$

$$\cdots$$

$$C_{v-1}: [c_{v-1,1}, c_{v-1,2}, \cdots, c_{v-1,v}]$$

Step 2: Normalize the contrasts according to $\sqrt{\sum_{k=1}^{v} c_{ik}^2}$ since the variance of the least square estimator of a contrast is proportional to $\sum_{k=1}^{v} c_{ik}^2$. The purpose of this step is to make sure that the least square estimators of these contrasts have the same variance. We denote the normalized coefficients of ith contrast as:

$$C_i: [h_{i1}, h_{i2}, \cdots, h_{iv}]$$

Note that $h_{ij} = \frac{c_{ij}}{\sqrt{\sum_{k=1}^{v} c_{ik}^2}} (i = 1, \cdots, v-1; j = 1, \cdots v)$ and for this new set of coefficients, all $\sum_{k=1}^{v} h_{ik}^2 = 1$.

Step 3: Find the least square estimate of each contrast. For a model with $v - 1$ orthogonal contrasts (or equivalently the cell-means model) for a single replicate experiment, the least squares estimate of the ith contrast is just the dot product of the contrast coefficients and the corresponding values of response.

Step 4. Calculate the Half-normal scores. The half normal scores are percentiles of the half-normal distribution with $\sigma^2 = 1$, corresponding to the distribution of the absolute value of a standard normal random variable. In particular, the qth half-normal score for $m = v - 1$ contrast absolute estimates is the value ξ_q for which

$$P(Z \leq \xi_q) = 0.5 * [1 + q/(m + 1)]$$

where Z is a standard normal random variable. Hence, the qth half-normal score is

$$\xi_q = \Phi^{-1}\left[0.5 * \left(1 + \frac{q}{m+1}\right)\right] = z_{1-0.5*(1+\frac{1}{m+1})}$$

where Φ is the cumulative distribution function (cdf) of the standard normal distribution.

Step 5. Draw the half-normal plot and find the estimates that stand out as relatively large, creating a nonlinear pattern. The corresponding contrasts are not eligible. The y-axis will be the ordered absolute values of the least square estimates from Step 3. The x-axis will be the half-normal scores from step 4.

Example 6.4: Half-Normal Plot - Drill Advance Experiment

The experiment involved treatment factors "load" (A), "flow" (B), "speed" (C), and "mud" (D) and response "log(drill advance)" (Y). If we have no information about which factors are likely to be negligible, we would use the four-way complete model or the equivalent cell-means model for the analysis. To use the half-normal plot to find which contrasts are not negligible, we use the following steps:

Step 1: Find 15 orthogonal contrasts. Since each factor have two levels, so we can use the pairwise contrasts for the four main effects and then obtain the coefficients for interactions by multiplying together the corresponding main-effect coefficients. The coefficients of these contrasts can be found in Table 7.1 on Page 208.

Step 2. Normalize the coefficients. Since each coefficient is either 1 or -1, we use $\frac{1}{\sqrt{16}} = \frac{1}{4}$ as the divisor for all contrasts. For example, we can use 1 as the divisor, then the contrast coefficients for the main effect of B is:

$$\frac{1}{4}[-1,-1,-1,-1,1,1,1,1,-1,-1,-1,-1,1,1,1,1]$$

Step 3. Obtain the least squares estimates. For example, the least squares estimate of AB-interaction is:

$$\frac{1}{4}(1 * 0.2253 + 1 * 0.3160 + \cdots + 1 * 1.2122) = -0.0298$$

Step 4. Calculate the half-normal scores. Note in this table, the least squares estimates are shown but are ranked in an ascending order according to their absolute values. Here $m = 15$, so we can calculate:

$$\xi_1 = \Phi^{-1}\left[0.5 * \left(1 + \frac{1}{15+1}\right)\right] = \Phi^{-1}(0.53125) = 0.0784$$

Rank	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Effect	AC	ABC	BD	AB	BCD	$ABCD$	BC	ABD
Estimate	0.0090	0.0090	-0.0130	-0.0298	-0.0300	0.0335	-0.0436	0.0454
Half-Normal Score (ξ_q)	0.0784	0.1573	0.2372	0.3186	0.4023	0.4888	0.5791	0.6745
Rank	**9**	**10**	**11**	**12**	**13**	**14**	**15**	
Effect	ACD	AD	CD	A	D	B	C	
Estimate	0.0462	0.0581	0.0852	0.1129	0.2836	0.5039	1.0027	
Half-Normal Score (ξ_q)	0.7764	0.8871	1.0010	1.1503	1.3180	1.5341	1.8627	

Step 5. Draw the half-normal plot according to the aforementioned table in Step 4. From Figure 6.2, we can observe that all the estimates fall roughly on a straight line, except for the estimates for the main-effects of factors D, B, and C. Hence, these three main effects appear to be nonnegligible, and the final model should include these three main effects.

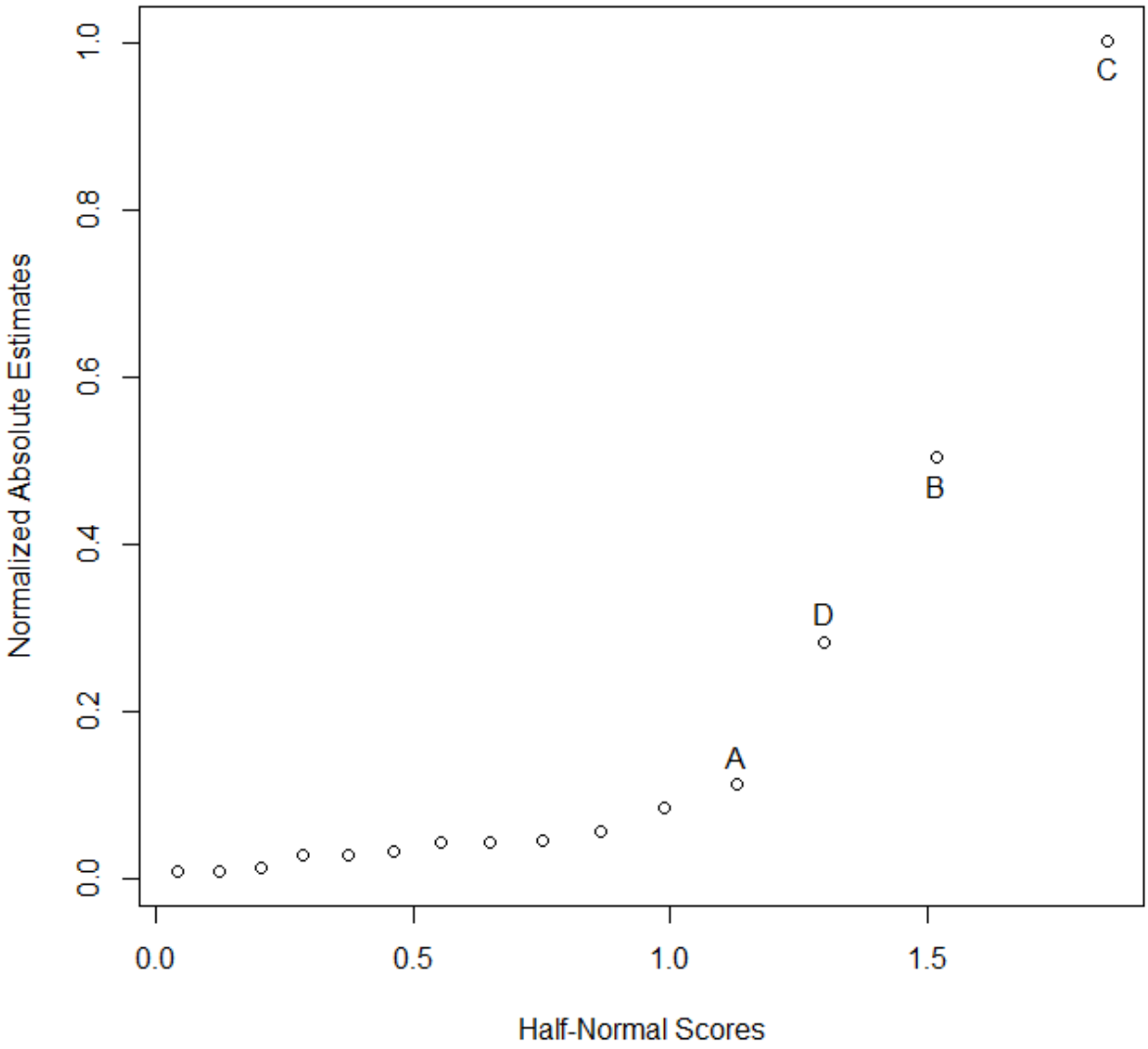
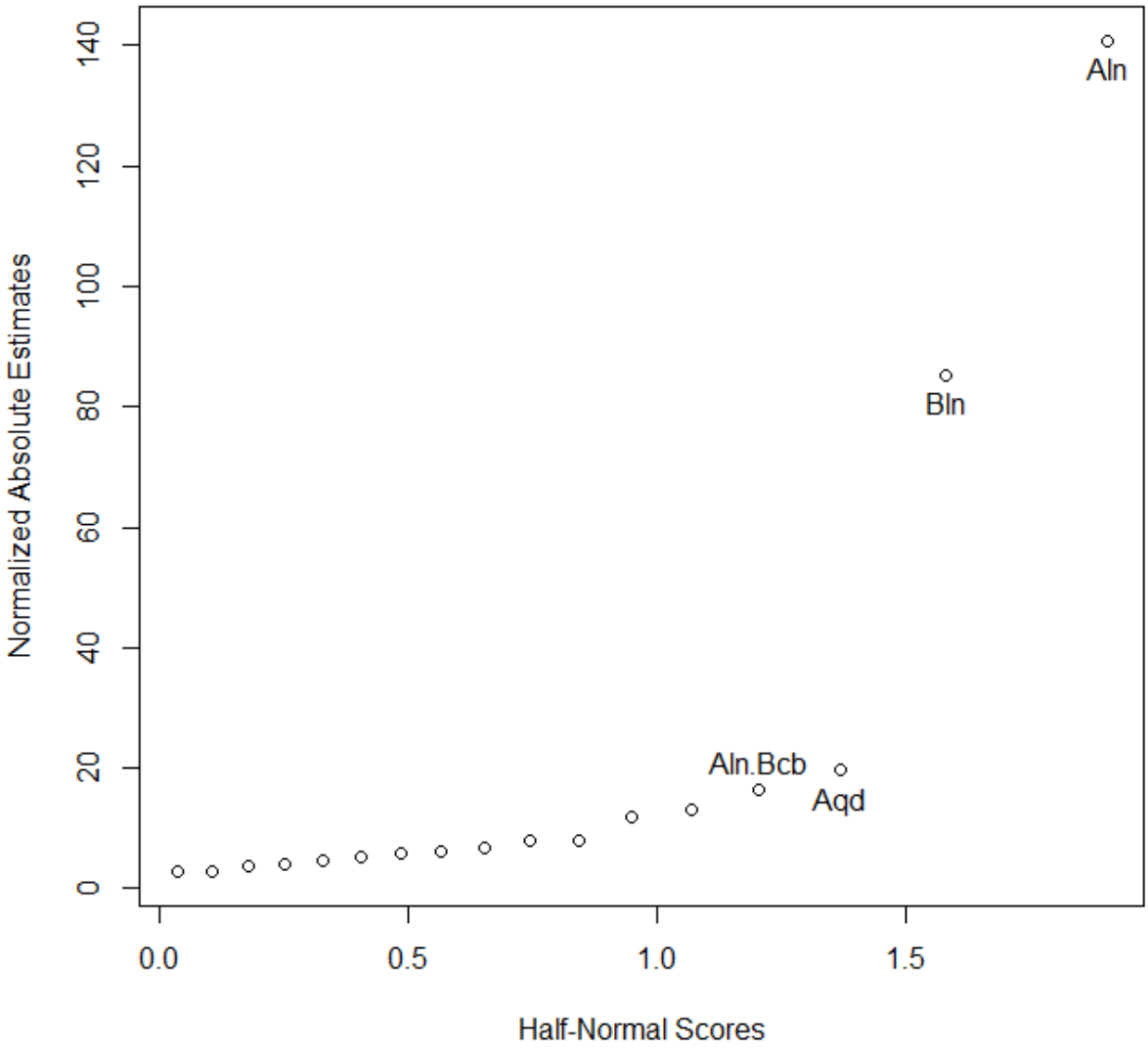


Figure 6.2. Half-normal probability plot of normalized contrast absolute estimates for the drill advance experiment.

Week 6 Practice Problem Discussion: Half-Normal Plot - Air Velocity Experiment

Use the half-normal plot to find which contrast are nonnegligible. Use the set of complete 17 orthogonal contrasts based on trends (the contrasts for factors *A* and *B* are already in file **air.velocity.contrasts.txt**) and R function qqnorm.aov.

Solution: We can observe that all the estimates fall roughly on a straight line, except for the estimates for the linear trend of factors A, and B. Hence, these two effects appear to be nonnegligible, and the final model should include these two effects.



Lesson 6.2 Introduction of Complete Block Designs

Required Reading: Sections 10.1, 10.2, and 10.3 in Chapter 10.

In Lesson 1.1, we discussed the nuisance factors and classified these nuisance factors into three types: blocking factors, noise factors, and covariates. Different types of nuisance factors lead to different types of designs and analyses. In this lesson, we focus on the blocking factors. Specifically, we focus on the experiments with one blocking factor and the corresponding design issues. The most commonly used block designs are the complete block designs.

A block is appropriate when the goal of the experiment is to compare the effects of different treatments averaged over a range of different conditions. The experimental units are grouped into sets in such a way that two experimental units in the same set are similar and can be measured under similar experimental conditions, but two experimental units in different sets are likely to give rise to quite different measurements even when assigned to the same treatment. The sets of similar experimental units are called blocks, and the conditions that vary from block to block form the levels of the blocking factor.

The intent of blocking is to prevent large differences in the experimental units from masking differences between treatment effects, while at the same time allowing the treatments to be examined under different experimental conditions. For example, suppose that a researcher would like to know which one of two types of berry trees can give more yield and two fields are available to plant those trees, then the filed should be considered as a blocking factor.

The levels of a blocking factor may be the values of a covariate that has been measured prior to the experiment and whose values are used to group the experimental units. More often, however, the levels of a blocking factor are groupings of characteristics that cannot be conveniently measured. For example, grouping the time slots in the same day into the same block can insure that environmental conditions within a block are fairly similar without the necessity of measuring them. For this reason, the block design is very popular.

Notations and Block Sizes

In this lesson, we consider experiment with one treatment factor and one blocking factor. Although it is perfectly possible for the numbers of experimental units in each block to be unequal, the most common setting, and the only one that we will examine here, is when the blocks are of the same size.

We use the following notations:

- v: the number of levels of treatment (or the number of treatment combinations from two or more treatments)
- b: the number of blocks, which is the number of levels of blocking factor (or the number of combined levels of two or more blocking factors)
- k: the block size, which is the number of experimental units in each block
- $n = b * k$: the total number of samples

Sometimes the block sizes are naturally defined, and sometimes they need to be specifically selected by the experimenter.

Example 6.5: Block Sizes - Bread-baking Experiment

An experimenter wishes to compare the shelf life of loaves made from $v = 4$ different bread doughs, coded 1, 2, 3, 4. An oven with three shelves will be used, and each shelf is large enough to take four baking tins. A temperature difference is anticipated between the different shelves but not in different positions within a shelf. The oven will be used twice. Then what block size should be sued here?

Solution: The experimental units are the baking tins in different positions in the oven. Since the temperature difference is anticipated between the different shelves but not in different positions within a shelf, if the measured response is affected by temperature, then experimental units on the same shelf are alike, but those on different shelves are different. Therefore, a natural choice for the number of blocks and the block size would be:

- The number of blocks $b = 6$ which is the number of shelves used (3 shelfs used twice)
- The block size $k = 4$ which is the number of baking tins that can be accommodated on each shelf

Block size is not always dictated by the experimental equipment. The size often needs to be determined by the judgment of the experimenter. Let us look at the example from Breathalyzer Experiment.

Example 6.6: Block Sizes - Breathalyzer Experiment

Bob Belloto in the Department of Pharmacy at The Ohio State University conducted a pilot experiment and obtained 20 readings by a breathalyzer for a given concentration of alcohol. The plot shows the readings decrease over time. Likely causes for this decrease include changes in atmospheric conditions, evaporation of alcohol, and deterioration of the breathalyzer filters. Based on this plot, can you provide some suggestion about the block size?

Solution: Based on the plot alone, the block size can be chosen to five, that is, the first five observations would be in one block, the next five in the next block, and so on. The reason for the choice is that the observations in the pilot experiment seem to be fairly stable in groups of five. Another reason that the research chose 5 as the block size was that the experiment was to be run by two different technicians, who alternated shifts and took five observations per shift. Thus the levels of blocking factors represented combinations of time and technicians.

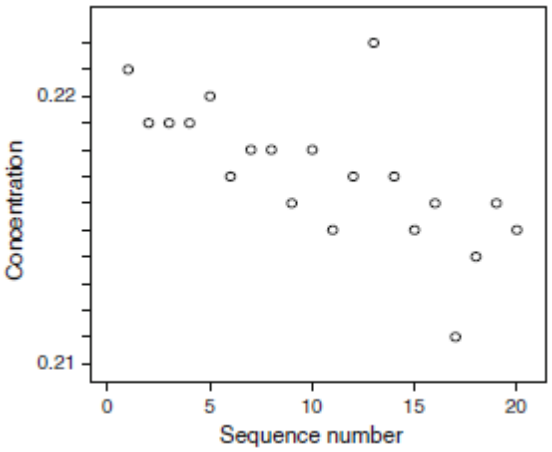


Figure 6.3 Pilot data for the breathalyzer experiment.

Complete Block Design Definitions

Once the block size are determined and the experimental units are grouped into blocks, the next step is to assign the units to the levels of the treatment factors. The best possible assignment is one that allocates to every treatment the same number of experimental units per block. This can be achieved only when the block size k is a multiple of v, the number of treatments. Therefore, we can define the following different types of block designs based on the relationship between k and v:

- **Complete block designs:** k is a multiple of v.
- **Randomized complete block designs:** $k = v$. These designs have historically been called randomized complete block designs or, simply, randomized block designs. The historical name is unfortunate, since all block designs need to be randomized. Nevertheless, we will retain the name randomized complete block design for block size $k = v$.
- **General complete block designs**: $k = v * s$ and $s \geq 2$.
- **Incomplete block designs**: the block size k is not a multiple of v. This term is sometimes reserved for the smaller designs where $k < v$, but we will find it convenient to classify all designs as either complete or incomplete. Incomplete block designs are more complicated to design and analyze than complete block designs, and we will not discuss them in our course.

Randomized Complete Block Design

A **randomized complete block design** is a design with v treatments (which may be factorial treatment combinations) and with $n = bv$ experimental units grouped into b blocks of $k = v$ units in such a way that units within a block are alike and units in different blocks are substantially different. The $k = v$ experimental units within each block are randomly assigned to the v treatments so that each treatment is assigned one unit per block. Thus, each treatment appears once in every block ($s = 1$) and $r = b$ times in the design.

Example 6.7: Randomization for Randomized Complete Block Designs - Bread-baking Experiment

An experimenter wishes to compare the shelf life of loaves made from $v = 4$ different bread doughs, coded 1, 2, 3, 4. An oven with three shelves will be used, and each shelf is large enough to take four baking tins. A temperature difference is anticipated between the different shelves but not in different positions within a shelf. The oven will be used twice. How do you perform the randomization here?

Solution: We have $b = 6, k = v = 4$. Within each block, the numbers 1, 2, 3, and 4 are randomly rearranged using R function "sample". We can then allocate the experimental units in the order FL, FR, BL, BR to the randomly sorted treatments. A randomized complete design is obtained and presented in the following table:

Block	Run	Shelf	FL	FR	BL	BR
1	**1**	**1**	2	1	4	3
2		**2**	2	4	3	1
3		**3**	3	1	2	4
4	**2**	**1**	2	4	3	1
5		**2**	1	2	4	3
6		**3**	3	4	2	1

Notice that in the randomization that we have obtained, the dough 4 is never observed FL position while the doughs 1 and 3 are never observed FR position. If a temperature difference in positions is present, then this could cause problems in estimating treatment differences, and the randomized complete block design is not the correct design to use. Instead, an incomplete block design should be used.

Exercise 6.5: Randomization for Randomized Complete Block Designs

Conduct a randomization for a randomized complete block design with $v = 6$ treatments observed once ($s = 1$) in each of $b = 5$ blocks.

Solution: Please refer to the R program.

General Complete Block Designs

The **general complete block design** is a design with v treatments (which may be factorial treatment combinations) and with $n = bvs$ experimental units grouped into b blocks of $k = vs$ units in such a way that units within a block are alike and units in different blocks are substantially different. The $k = vs$ experimental units within each block are randomly assigned to the v treatments so that each treatment is assigned s units per block. Thus, each treatment appears s times in every block and $r = bs$ times in the design.

Example 6.8: Randomization for General Complete Block Designs - DCIS Experiment

An experiment was run by Matthew Darr, David Holman, Nasser Kashou, and AngelaWendel in 2006 to examine the variability of a Dynamic Inline Conveyor Scale (DCIS). The DCIS is an automated system for weighing individual pieces of large fruit (for example, watermelons) while they are conveyed from one location to another. The device uses an optical switch to trigger the weighing system, a load cell to perform the weighing operation and a computer-based data recording system. The objective of the experiment was to reduce the variability associated with the recorded weight of each piece of fruit.

The researchers decided to examine the effects of two treatment factors. Treatment factor A was

the length of time during which the weight was recorded (with three levels: 50, 75, 100, milliseconds; coded 1, 2, 3). Treatment factor B was the position of the optical switch (with two levels; 1 inch and 2 inches from the end of the scale plate; coded 1, 2). Thus there were six treatments (treatment combinations), coded as follows:

(50 millisec, 1 inch) = 1, (50 millisec, 2 inch) = 2,

(75 millisec, 1 inch) = 3, (75 millisec, 2 inch) = 4,

(100 millisec, 1 inch) = 5, (100 millisec, 2 inch) = 6,

The currently employed setting was treatment 4 (time 75milliseconds, switch at 2 inches from the scale plate). The experimenters wished to see if there was a better setting, while taking into account a range of possible lubrication levels of the conveyor system. Changing the lubrication levels was difficult and time consuming, so the experiment was run in two blocks at the extreme levels of lubrication. In block 1, the conveyor pan was completely saturated with oil lubricant and, in block 2, all lubricant was removed from the conveyor pan. In each block, there were s = 2 observations on each treatment. A random assignment of the $k = sv = 12$ experimental units (time slots) to the treatments within each block gave the following observation order from the R program:

Block 1: 1 6 6 5 4 2 5 1 3 3 4 2

Block 2: 2 2 4 3 5 1 6 4 6 1 3 5

For each observation on a specified treatment in a block, the response was a function, called "uncertainty", of the standard deviation of 30 weighings of a watermelon. The data are can find shown in file **DCIS.txt**.

Exercise 6.6: Randomization for General Complete Block Designs

Conduct a randomization for a general complete block design with $v = 6$ treatments observed once ($s = 2$) in each of $b = 5$ blocks.

Solution: Please refer to the R program.

Lesson 6.3 Analysis of Randomized Complete Block Designs

Required Readings: Section 10.4 and 10.5 from Chapter 10.

The standard model for a randomized complete block design (with s = 1 observation on each treatment in each block) is:

$$Y_{hi} = \mu + \theta_h + \tau_i + \epsilon_{hi}, h = 1, \cdots, b; i = 1, \cdots, v$$

with the following assumptions:

- μ, θ_h, τ_i $(h = 1, \cdots, b; i = 1, \cdots, v)$ are unknown population parameters
- θ_h is the effect of the hth block
- τ_i is the effect of the ith treatment
- $\epsilon_{hi} (h = 1, \cdots, b; i = 1, \cdots, v)$ are independent and $\epsilon_{hi} \sim N(0, \sigma^2)$: normal distribution with mean 0 and variance σ^2

We will call this standard model the *block–treatment model*. Notice that the randomized complete block designs are actually single replicate experiments, therefore such block–treatment model does not include a term for the interaction between blocks and treatments. If interaction effects were to be included in the model, there would be no degrees of freedom for error with which to estimate the error variance. In many blocked experiments, absence of $block \times treatment$ interaction is a reasonable assumption. However, if interaction is suspected in a given experiment, then the block size must be increased to allow its estimation. The block–treatment model looks similar to the two-way main-effects model for single replicate experiments with two treatment factors. Not surprisingly, then, the analysis of variance table or the randomized complete block design looks similar to the two-way analysis of variance table for single replicate experiments with two treatment factors:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F-test Statistic

Block	$b-1$	$ss\theta$	$ms\theta = \frac{ss\theta}{h-1}$	-
Treatment	$v-1$	ssT	$msT = \frac{ssT}{v-1}$	$\frac{msT}{msE}$
Error	$bv-b-v+1$	ssE	$msE = \frac{ssE}{bv-b-v+1}$	
Total	$bv-1$	$sstot$		

However, there are several differences between the completely randomized design and the complete block designs.

First, in the completely randomized designs, the experimental units are randomly assigned to the treatment combinations, and so to the levels of both factors. On the other hand, in a block design, although observations are taken on all combinations of treatments and blocks, the experimental units are randomly assigned to the levels of the treatment factor only. The levels of the blocking factor represent intentional groupings of the experimental units.

Second, we are generally more interested in of the use of blocking factors can reduce the mean square for error than in testing for equality of block effects. if $ss\theta$ is much less than ssE, then the creation of blocks was not helpful. in fact, has lowered the power of hypothesis tests and increased the lengths of confidence intervals for treatment contrasts. This ca be explained by comparing the F-test statistic for the treatment in the block-treatment model and one-way analysis of variance model:

Block-treatment model: $F = \frac{ssT/(v-1)}{ssE/(bv-b-v+1)}$

One-way analysis of variance model: $F = \frac{ssT/(v-1)}{ssE/(bv-v)}$

If $ss\theta$ is much less than ssE, then ssE from two model are similar while the degrees of freedom for error in the one-way analysis of variance model, results in the smaller msE in the one-way ANOVA model. However, for other reasons, we more often compare $ms\theta$ and msE: the

blocking factor should be included as long as $ms\theta$ is much larger msE since this indicates the effects of the blocking factor are statistically significant thus should be included in the model.

Example 6.9: Block-Treatment Model - Resting Metabolic Rate Experiment

The data can be found in file **resting.metabolic.rate.txt** and its description can be found in **Data-Sets-Summary.docx**. In the 1993 issue of Annals of Nutrition and Metabolism, R. C. Bullough and C. L. Melby describe an experiment that was run to compare the effects of inpatient and outpatient protocols on the in laboratory measurement of resting metabolic rate (RMR) in humans. A previous study had indicated measurements of RMR on elderly individuals to be 8% higher using an outpatient protocol than with an inpatient protocol. If the measurements depend on the protocol, then comparison of the results of studies conducted by different laboratories using different protocols would be difficult. The experimenters hoped to conclude that the effect on RMR of different protocols was negligible.

The experimental treatments consisted of three protocols: (1) an inpatient protocol in which meals were controlled—the patient was fed the evening meal and spent the night in the laboratory, then RMR was measured in the morning; (2) an outpatient protocol in which meals were controlled—the patient was fed the same evening meal at the laboratory but spent the night at home, then RMR was measured in the morning; and (3) an outpatient protocol in which meals were not strictly controlled—the patient was instructed to fast for 12 hours prior to measurement of RMR in the morning. The three protocols formed the $v = 3$ treatments in the experiment.

Since subjects tend to differ substantially from each other, error variability can be reduced by using the subjects as blocks and measuring the effects of all treatments for each subject. In this experiment, there were nine subjects (healthy, adult males of similar age) and they formed the $b = 9$ levels of a blocking factor "subject." Every subject was measured under all three treatments (in a random order), so the blocks were of size $k = 3 = v$. RMR readings were taken over a one-hour period shortly after the subject arrived in the laboratory. For this experiment, we are interested in the following questions:

(1) If the interaction effects between the treatment factor and the blocking factor are negligible so the standard model for the complete block designs can be used.

(2) If the use of blocks reduces ssE substantially thus increase the statistical power.

(3) If the difference in response for three protocols is negligible.

(4) The pairwise comparison on the response for the treatment factor.

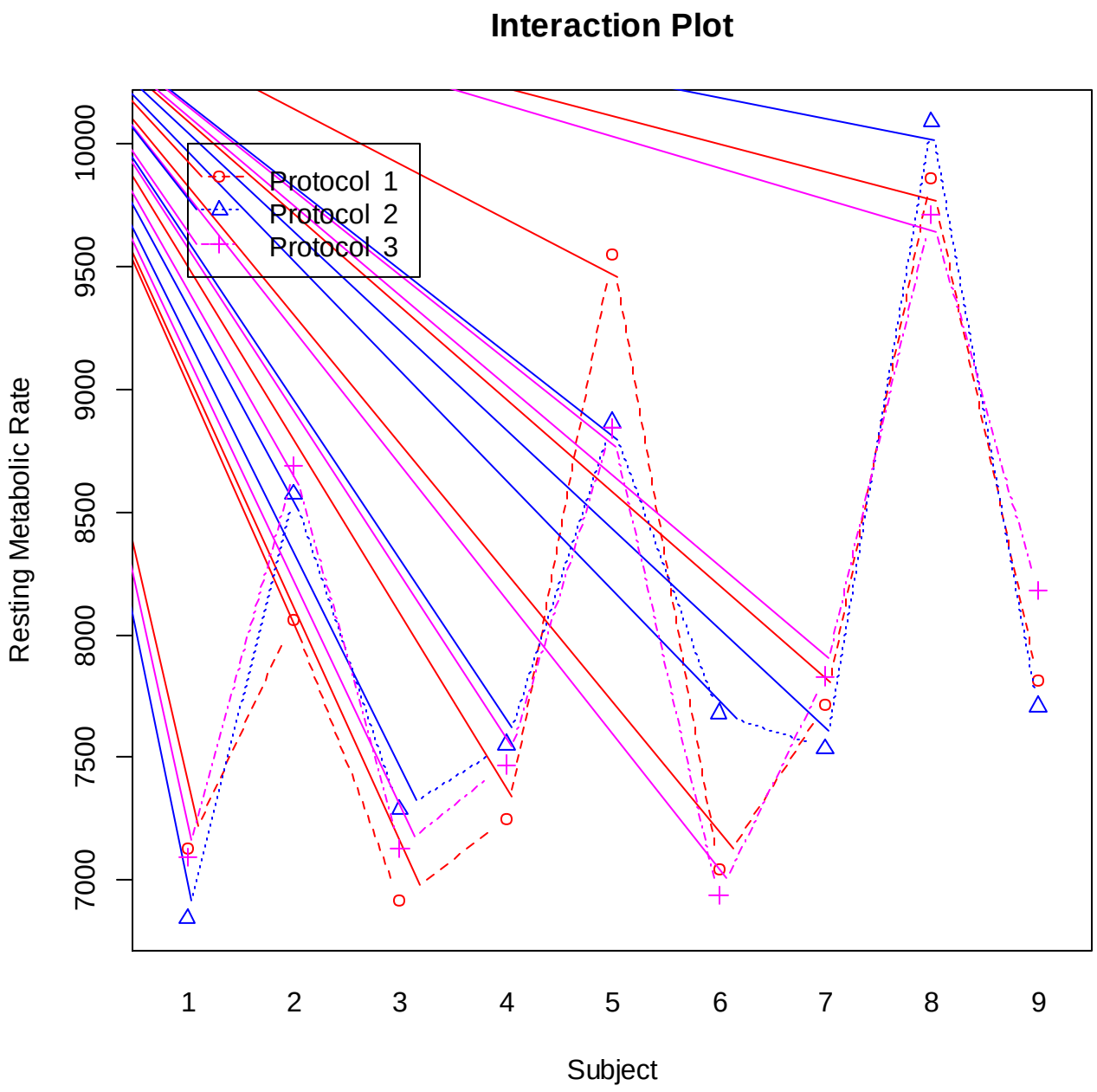


Figure 6.4 Interaction plot for Resting Metabolic Rate Experiment.

Solution: The interaction effects between the treatment factor and the blocking factor are assessed using the interaction plot. It seems that the interaction effects are negligible so the block-treatment model are used and the following ANOVA table are obtained:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F-test Statistic
Block	8	23,117,462.30	2,889,682.79	-
Treatment	2	35948.74	17,974.37	0.23
Error	16	1,235,483.26	77217.70	
Total	27	24,388,894.30		

From this table, we can see that:

(1) $ss\theta$ is much larger than ssE, and $ms\theta$ is much larger than msE indicating that the use of blocking factor indeed reduces msE. So a block design is a preferable choice for this experiment.

(2) The null hypothesis of no difference in the protocols cannot be rejected at any reasonable selection of α, since $msT/msE = 0.23$ and the p-value$= 0.7950$. The ratio tells us that the average variability of the measurements from one protocol to another was four times smaller than the measurement error variability. This is unusual, since measurements from one protocol to another must include measurement error. The p-value is 0.7950, indicating that there is only a 20% chance that we would see a value this small or smaller when there is no difference whatsoever in the effects of the protocols.

To perform the pairwise comparison on the response for the treatment factors, methods of multiple comparisons should be used. The critical coefficients for the Bonferroni, Tukey, and Dunnett methods are, respectively:

Bonferroni Method: $w_B = t_{bv-b-v+1,\alpha/(2*m)}$, m is the number pre-specified contrasts

Scheffe Method: $w_s = \sqrt{(v-1)F_{v-1,bv-b-v+1,\alpha}}$

Tukey Method: $w_T = \frac{q_{v,bv-b-v+1,\alpha}}{\sqrt{2}}$ (best and only for pair-wise comparison)

Dunnett Method: $w_{D2} = |t|^{(0.5)}_{v-1,bv-b-v+1,\alpha}$ (best and only for treatment-versus-control)

These coefficients are same as those for the main effect from a two-way main-effects model. For this experiment, the pairwise comparison of the treatment, and the overall significance of 0.05, we have

Bonferroni Method: $w_B = t_{16,0.05/(2*3)} = 2.673$

Scheffe Method: $w_s = \sqrt{(3-1)F_{3-1,16,0.05}} = 2.696$

Tukey Method: $w_T = \frac{q_{3,16,0.05}}{\sqrt{2}} = 2.580$

The simultaneous 95% confidence intervals of pairwise comparison for protocols are:

Contrast	Point Estimate	Standard Error	Simultaneous 95% confidence intervals		
			Bonferroni	Scheffe	Tukey
$1-2$	-87.6	131	(-438, 263)	(-441, 266)	(-426, 250)
$1-3$	-59.3	131	(-409, 291)	(-412, 294)	(-397, 279)
$2-3$	28.2	131	(-322, 378)	(-325, 381)	(-310, 366)

All intervals include 0, indicating that we do not have sufficient evidence to reject there is no difference on the response for three protocols. Here, Tukey method is preferred since it gives the shortest confidence intervals. In addition to the pairwise comparison of protocols, another interest comparison is the comparison between the inpatients protocols with two outpatients protocols. The corresponding contrast coefficients are $[1, -0.5, -0.5]$ in terms of τ_i. Its 99% confidence interval is $(-405, 258)$, which contains 0 too.

In the paper from the experimenters, the experimenters discussed possible reasons for the fact that their conclusions differed from those of previous studies. Reasons included the different age of the subjects (27–29 years rather than 64–67 years) and the fact that they provided transport to the laboratory for the outpatients, whereas previous studies had not.

Exercise 6.7: Checking Assumptions for Block-Treatment Model - Resting Metabolic rate Experiment

(1) Check outliers, equal-variance assumption, and normality assumption.
(2) Remove outliers and perform the same analysis as it is in Example 6.9 and draw conclusions.

Solution: Please refer to the R program.

Exercise 6.8: Block-Treatment Model - Cotton-Spinning Experiment

The experiment was reported in the November 1953 issue of the journal *Applied Statistics* by Robert Peake, of the British Cotton Industry Research Association. The purpose of the experiment was twofold; first, to investigate the way in which different degrees of twist (measured in turns per inch) affected the breakage rate of the roving, and secondly, to compare the ordinary flyer with the newly devised special flyer.

A randomized complete block designs was used. Each experimental unit was the production of one full set of bobbins on a single machine with a single operator. A block consisted of a group of experimental units with the same machine, the same operator, and observed in the same week. Thus, the different levels of the blocking factor represented differences due to combinations of machines, operators, environmental conditions, and raw material. The block size was chosen to be six, as this was equal to the number of treatment combinations and also to the number of observations that could be taken on one machine in one week.

There were tow treatment factors: “flyer” and “degree of twist.” Flyer had two levels, “ordinary” and “special.” Twist had four levels, 1.63, 1.69, 1.78, and 1.90. For practical reasons, the combinations of flyer and twist equal to (ordinary, 1.63) and (special, 1.90) were not observed. We will recode the six treatment combinations that were observed as follows:

(ordinary, 1.69) = 1, (ordinary, 1.78) = 2, (ordinary, 1.90) = 3,

(special, 1.63) = 4, (special, 1.69) = 5, (special, 1.78) = 6.

The data and the description of this experiment can be found from files and, respectively. Perform the following analysis to confirm the results in Section 10.5 of Chapter 10:.

(1) Draw an interaction plot to verify if the block-treatment model is appropriate.
(2) Construct the analysis of variance table from the block-treatment model and comment on $ss\theta$ and ssE.
(3) Test the main effects of factors “flyer” and “degree of twist”.
(4) Check model assumptions.

Lesson 6.4 Analysis of General Complete Block Designs

Required Reading: Sections 10.6, 10.7, and 10.8 from Chapter 10.

Model and Analysis of Variance

For general complete block designs, we have $n = bk = bvs, k = vs$ and $s > 1$. Having every level of the treatment factor observed more than once per block gives sufficient degrees of freedom to be able to measure a $block \times treatment$ interaction if one is anticipated.

Therefore, there are two standard models for the general complete block design, the block–treatment model (without interaction):

$$Y_{hit} = \mu + \theta_h + \tau_i + \epsilon_{hit}, t = 1, \cdots, s; h = 1, \cdots, b; i = 1, \cdots, v$$

and the *block–treatment interaction* model:

$$Y_{hit} = \mu + \theta_h + \tau_i + (\theta\tau)_{hi} + \epsilon_{hit}, t = 1, \cdots, s; h = 1, \cdots, b; i = 1, \cdots, v$$

In each case, the model includes the error assumptions:

- $\epsilon_{hit} (t = 1, \cdots, s; h = 1, \cdots, b; i = 1, \cdots, v)$ are independent and $\epsilon_{hit} \sim N(0, \sigma^2)$: normal distribution with mean 0 and variance σ^2

The block–treatment model for a general complete block design is similar to the two-way main-effects model, and the block–treatment interaction model is like the two-way complete model for two treatment factors in a completely randomized design, each with s observations per cell. Analogously, the analysis of variance tables for the block-treatment models, with and without interaction, look similar to those for the two-way main-effects and two-way complete models.

Here we list topics to use the block-treatment model or the block-treatment interaction model for the analysis of general complete block designs.

Analysis of Variance Table

Once the analysis of variance table is obtained, we can look at and perform the hypothesis of:

$$H_0: \tau_1 = \cdots = \tau_v$$

with the following:

$$Reject\ H_0\ if\ F = \frac{ssT/(v-1)}{ssE/df} > F_{v-1,df,\alpha}$$

Where df is the degrees of freedom for error and $df = bvs - b - v + 1$ for the block-treatment model and $df = bvs - bv$ for the block-treatment interaction model.

If the $block \times treatment$ interaction term is included in the model, a test of the hypothesis of there is no interaction between the blocking factor and the treatment factor can be achieved by the use of the following formula:

Reject the null hypotheses $if\ F = \frac{ss\theta T/[(b-1)*(v-1)]}{ssE/(bvs-bv)} > F_{(b-1)*(v-1),bvs-bv,\alpha}$

Methods of Multiple Comparisons

For any contrast, its contrast coefficients, which depend on parameters used, can be obtained accordingly. In addition, which of multiple comparison methods can be used also depends on parameters used to determine the contrast coefficients. The simultaneous $(1-\alpha)100\%$ confidence interval of a contrast is:

$$point\ estimate \pm w * standard\ error$$

where the critical coefficient w for each of the four methods is, respectively,

Bonferroni Method: $w_B = t_{df,\alpha/(2*m)}$, m is the number pre-specified contrasts

Scheffe Method: $w_s = \sqrt{(x-1)F_{x-1,df,\alpha}}$

Tukey Method: $w_T = \frac{q_{x,df,\alpha}}{\sqrt{2}}$ (best and only for pair-wise comparison)

Dunnett Method: $w_{D2} = |t|^{(0.5)}_{x-1,df,\alpha}$ (best and only for treatment-versus-control)

where x is the length of contrast coefficients and df is the degrees of freedom for error. For the contrasts for the treatment factor, $x = v$ is the number of levels of the treatment factor. Again, $df = bvs - b - v + 1$ for the block-treatment model and $df = bvs - bv$ for the block-treatment interaction model. We rely on R program to obtain the point estimate and the standard error of a contrast.

Sample Sizes and Power

A complete block design has $n = bvs$ experimental units divided into b blocks of size $k = vs$.The block size k and the number of blocks b must be chosen to accommodate the experimental conditions, the budget constraints, and the requirements on the lengths of confidence intervals or powers of hypothesis tests in the usual way. If the number of blocks is fixed, one can calculate the number of observations s per treatment per block needed to achieve a prescribed power of a test of no treatment differences.

The sample size and power calculation would be similar with those described in Lesson 5.4. Here, we are interest in the sample size and/or power to test the effect of treatment factor. Therefore, the following notations are used:

- α: significance level
- v: number of treatment levels
- b: number of levels from the blocking factor
- s: sample size, number of replicates for each block and treatment combination
- π: power
- σ^2: variance in the model
- Δ: a threshold for the differences in the effects of the treatment factor

To use Table A.7 to calculate the sample sizes or power, we have:

- $v_1 = v - 1$
- $v_2 = degrees\ of\ freedom\ for\ error$. Again, $df = bvs - b - v + 1$ for the block-treatment model and $df = bvs - bv$ for the block-treatment interaction model.
- The formula for the sample size (s) and ϕ that is used in Table A.7 for power calculation:

$$\phi = \sqrt{\frac{b * s * \Delta^2}{2 * v * \sigma^2}}$$

Checking Model Assumptions

The assumptions on the block–treatment models and on the block–treatment interaction model for complete block designs need to be checked as usual. The assumptions on the error variables are that they have equal variances, are independent, and have a normal distribution. The form of the model must also be checked.

A visual check of an assumption of no $block \times treatment$ interaction can be made by plotting the average of response of each combination of the blocking factor and the treatment factors against the treatment factor levels i for each block h in turn. If the lines plotted for each block are parallel, then $block \times treatment$ interaction is likely to be negligible and error variability is small. If the lines are not parallel, then either $block \times treatment$ interaction is present or error variability is large.

Checking the independence can be done by plotting the standardized residuals against the order of observations (in space or tome). Checking the equal variance can be done by plotting the standardized residuals against the predicted values of the response. This is equivalent to plot the standardized residuals against the levels of treatment factor, or the levels of blocking factor, or their combinations. Checking the outliers can be done by investigating the absolute value of the standardized residuals and by plotting the standardized residuals against the levels of treatment factor, or the levels of blocking factor, or their combinations. Checking the normality can be done by plotting the standardized residuals against normal scores.

Factorial Experiments

When the treatments are factorial in nature, the treatment parameter t_i in the complete block design models can be replaced by main-effect and interaction parameters. Suppose, for example, we have an experiment with two treatment factors that is designed as a randomized

complete block design. In order not to confuse the number b of blocks with the number of levels of a treatment factor, we will label the two treatment factors as C and D with c and d levels respectively. If we retain the two digit codes for the treatment combinations, then the block–treatment model is

$$Y_{hijt} = \mu + \theta_h + \tau_{ij} + \epsilon_{hijt}, t = 1, \cdots, s; h = 1, \cdots, b; i = 1, \cdots, c; j = 1, \cdots, d$$

with the usual assumptions on the error variables. We can then express τ_{ij}, the effect of treatment combination $i\,j$, in terms of γ_i (the effect of C at level i), δ_j (the effect of D at level j), and $(\gamma\delta)_{ij}$(the effect of their interaction when C is at level i and D at level j); that is,

$$Y_{hijt} = \mu + \theta_h + \gamma_i + \delta_j + (\gamma\delta)_{ij} + \epsilon_{hijt}$$

In a general complete block design with $s > 1$ observations per treatment combination per block, we may include in the model some or all of the $block \times treatment$ interactions. For example, with two treatment factors, the block–treatment interaction model can be expressed as

$$Y_{hijt} = \mu + \theta_h + \gamma_i + \delta_j + (\gamma\delta)_{ij} + (\theta\gamma)_{hi} + (\theta\delta)_{hj} + (\theta\gamma\delta)_{hij} + \epsilon_{hijt}$$

If there are more than two factors, the additional main effects and interactions can be added to the model in a similar way.

Example 6.10: Block-Treatment Interaction Model - DCIS Experiment

For this experiment, the response was a function, called "uncertainty", of the standard deviation of 30 weighings of a watermelon. There was a blocking factor with two levels. There were two treatment factors, the length of time (Factor A) and the position of the optical switch (Factor B). Six treatments (treatment combinations) were coded as follows:

(50 millisec, 1 inch) = 1, (50 millisec, 2 inch) = 2,

(75 millisec, 1 inch) = 3, (75 millisec, 2 inch) = 4,

(100 millisec, 1 inch) = 5, (100 millisec, 2 inch) = 6

For this experiment, 24 observations were collected so $s = 2$. The data and its description can be found in files **DCIS.txt** and **Data-Sets-Summary.docx**, respectively. Here we outline the analysis we will perform.

First, we can draw an interaction plot to see if the interaction terms between the blocking factors and the treatment factors. Figure 6.4 suggests that there might be a small interaction between block and treatment combination. Therefore, a block-treatment interaction model is fitted. combination.

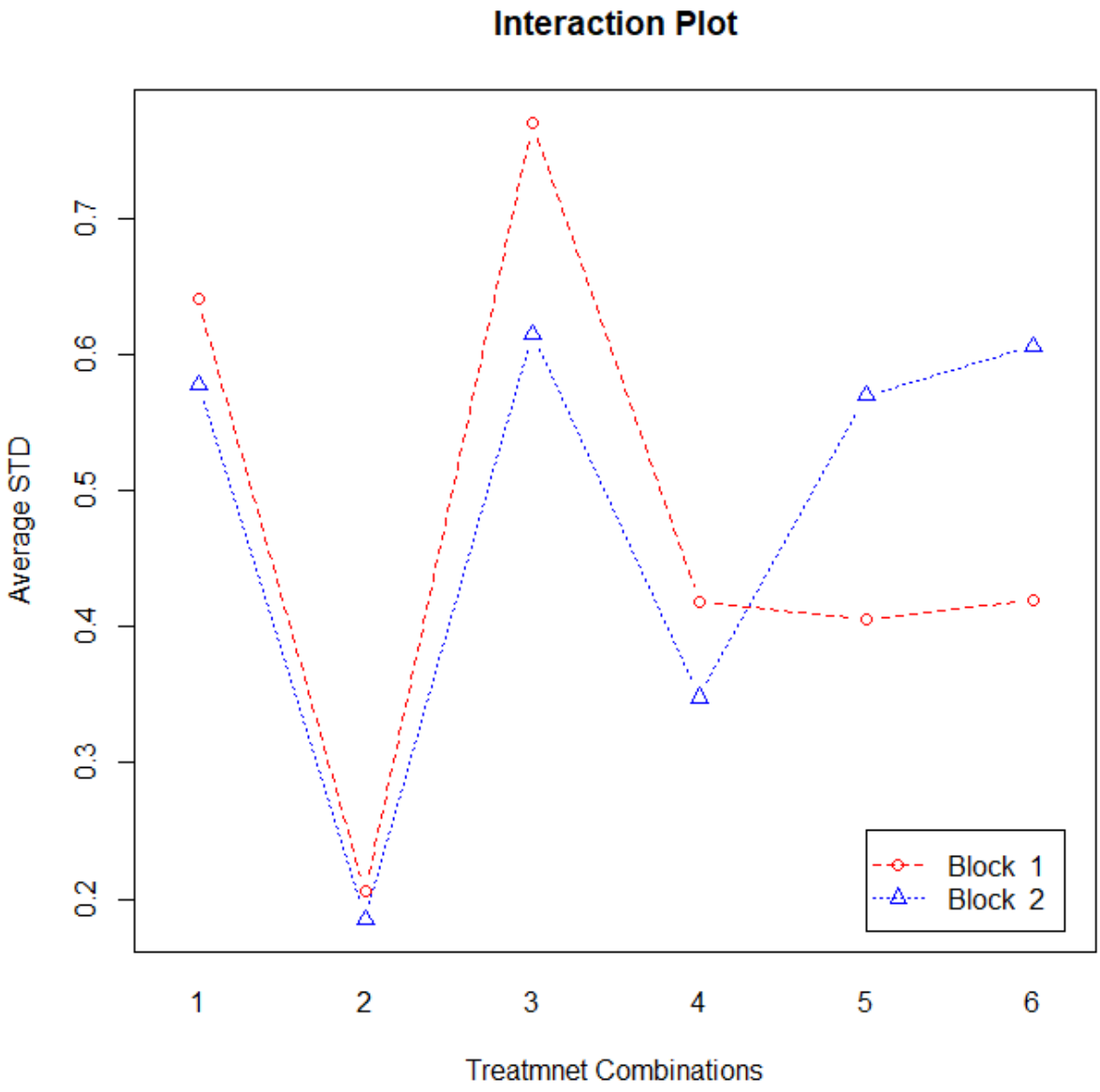


Figure 6.5 Interaction plot for DCIS Experiment.

Second, the following block-treatment interaction model is used:

$$Y_{hit} = \mu + \theta_h + \tau_i + (\theta\tau)_{hi} + \epsilon_{hit}$$

Then the corresponding analysis of variance table is obtained:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F-test Statistic	p-value
Block	1	0.003		-	
Treatment	5	0.6132	0.1226	9.4	0.0008
Interaction	5	0.0952	0.0190	1.46	0.2726
Error	12	0.1563	0.0130		
Total	23	0.8649			

The p-value for the interaction between the blocking factor and the treatment combination is 0,2726, so we fail to reject the null hypothesis of the interaction terms are negligible. The p-value for the treatment combination is 0.0008, so we conclude over the two blocks (levels of pan lubrication), there is a significant difference between the treatment combinations at level $\alpha = 0.01$. We also find the $ss\theta + ss\theta T = 0.1252$ is comparable to $ssE = 0.1563$, so it is not clear if it is beneficial to include the blocking factor here.

Third, we can examine the treatment effects on the weight "uncertainty" averaged over blocks or examining the difference in the treatment effects for each block separately. The former would be of most interest if it is not possible to control the level of conveyor pan lubrication in an industrial setting, while the latter would be of interest if the amount of lubrication could be fixed. Here, as an example, we investigate the latter and are interested in the following comparisons:

block 1, treatment 4 – 2

block 2, treatment 4 – 2

block 1, switch 1in vs 2in

block 2, switch 1in vs 2in

To use R to obtain, for example, the simultaneous 96% confidence intervals for these contrasts. We need to write their contrasts coefficients in terms of $(\theta\tau)_{hi}(h = 1,2; i = 1, \cdots, 6)$. As an example, the contrast for block 2, switch 1in vs 2in is:

$$\frac{1}{3}((\theta\tau)_{21} + (\theta\tau)_{23} + (\theta\tau)_{25}) - \frac{1}{3}((\theta\tau)_{22} + (\theta\tau)_{24} + (\theta\tau)_{26})$$

The contrast coefficients are:

$$[0,0,0,0,0,0,\frac{1}{3}, -\frac{1}{3}, \frac{1}{3}, -\frac{1}{3}, \frac{1}{3}, -\frac{1}{3}]$$

The following simultaneous 96% confidence intervals can be obtained:

Contrast	Point Estimate	Standard Error	Simultaneous 96% confidence intervals	
			Bonferroni	Scheffe
block 1, treatment 4 – 2	0.212	0.1141	(-0.136, 0.561)	(-0.433, 0.858)
block 2, treatment 4 – 2	0.163	0.1141	(-0.186, 0.512)	(-0.482, 0.808)
block 1, switch 1in vs 2in	0.258	0.0659	(0.056, 0.459)	(-0.115, 0.630)
block 2, switch 1in vs 2in	0.207	0.0659	(0.006, 0.409)	(-0.165, 0.580)

The critical coefficient w are,

Bonferroni Method: $w_B = t_{12,0.04/(2*4)} = 3.0545$

Scheffe Method: $w_s = \sqrt{(12-1)F_{12-1,12,0.04}} = 5.6534$

We can see that Bonferroni method is better than Scheffe'e method here. From the results of Bonferroni method, at overall confidence level of at least 96%, we cannot detect a difference between the effects of treatment 2 and the current treatment 4 in either block. However, switch position 2 inches from the end of the scale plate (averaged over weighing times) reduces "uncertainty" as compared with switch position 1 inch by up to .4 in both blocks, and hence seems to be the better position.

Fourth, if this is a pilot study and we would like to perform a power calculation for this experiment. From this data with the block-treatment interaction model, we can obtain $\hat{\sigma}^2 = 0.156$ and use $\Delta = 0.198$, which the median of difference on the response between treatments. We would like to test the null hypothesis of no treatment differences whether or not there is $block \times treatment$ interaction and use a significance level of 0.05. We can use the following formulas:

$$b = 2, v = 6, s = 2,$$

$$v_1 = v - 1 = 5, v_2 = bvs - bv = 12$$

$$\phi = \sqrt{\frac{b * s * \Delta^2}{2 * v * \sigma^2}} = \sqrt{\frac{2 * 2 * 0.188^2}{2 * 6 * 0.013}} = 0.952$$

From Table A.7, we can see the power to detect the treatment differences of $\Delta = 0.188$ is less than 0.289.

Fifth, we plot the following residual plots to check the model assumptions.

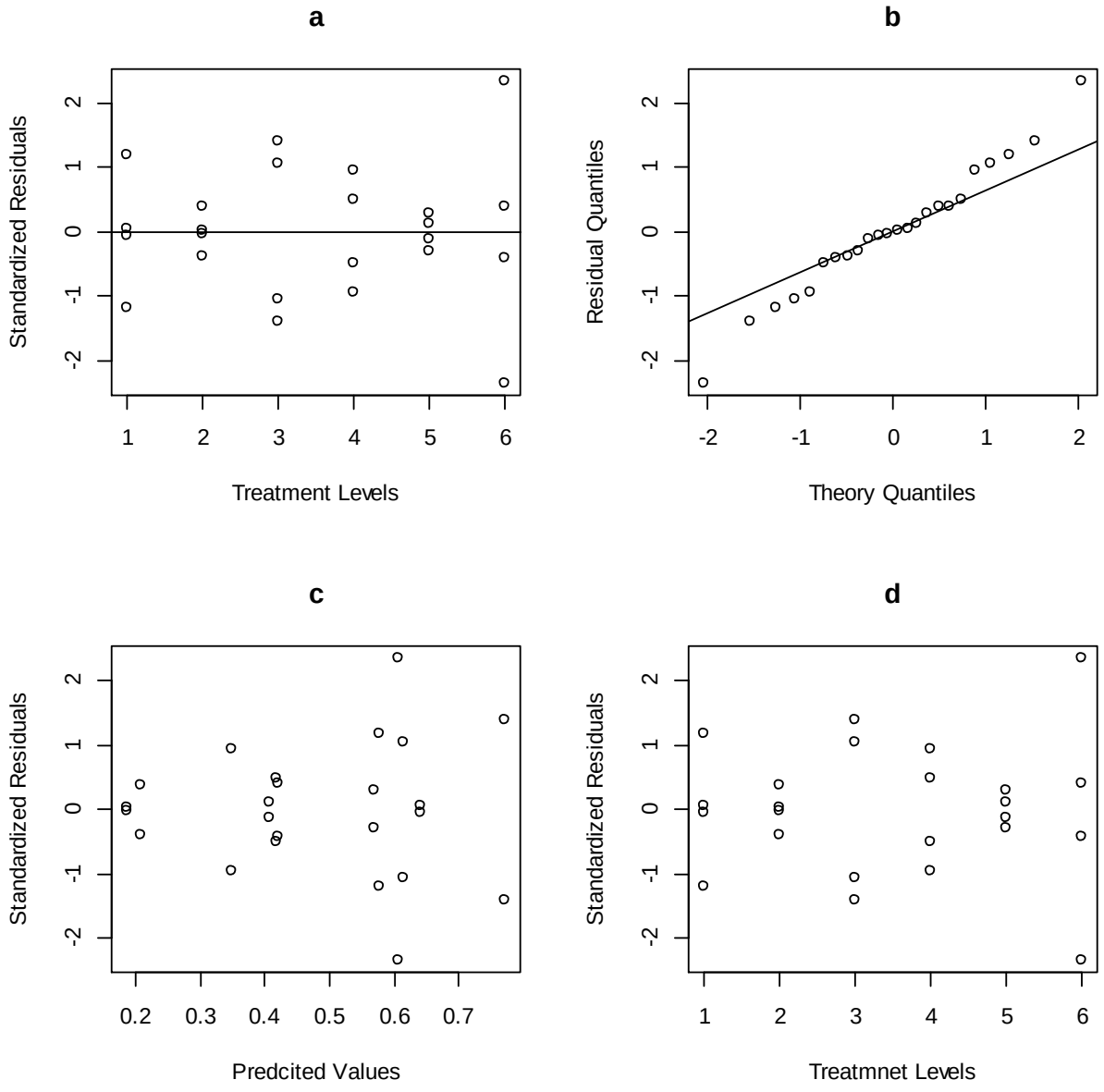


Figure 6.6 Residual plots for DCIS Experiment.

From the plot, is seems that there are two potential outliers with the standardized residual that is either greater than 2 or less than -2. The normality and equal variance assumptions seem to be violated. From plots and d of Figure 6.6, we can see two outliers from treatment combination 6 and the variance for this treatment appears to be larger. We can check if these outliers are incorrect records and may perform a transformation.

Sixth, since there are actually two treatment factors, we can try to fit the following model:

$$Y_{hijt} = \mu + \theta_h + \gamma_i + \delta_j + (\gamma\delta)_{ij} + (\theta\gamma)_{hi} + (\theta\delta)_{hj} + (\theta\gamma\delta)_{hij} + \epsilon_{hijt}$$

Here we use γ_i and $\delta_j$to represent the effects of the length of time and the position of the optical switch. The advantage of this model is that we can easily assess the effects of each factor and their interactions. Although this can be done by the block-treatment interaction model, some complicated programming is needed. From the following analysis of variance table, we can conclude that only the effects of the position of the optical switch and block×position are statistically significant.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F-test Statistic	p-value
Block	1	0.003		-	
Time	2	0.078	0.039	3.003	0.088
Position	1	0.325	0.325	24.92	0.0003
Block×Time	2	0.091	0.045	3.469	0.065
Block×Position	1	0.0038	0.0038	0.290	0.600
Time×Position	2	0.210	0.105	8.074	0.006
Three-way	2	0.001	0.005	0.040	0.961
Error	12	0.1563	0.0130		
Total	23	0.8649			

Exercise 6.9: Block-Treatment Interaction Model - Banana Experiment

The experiment was conducted by by K. Collins, D. Marriott, P. Kobrin, G. Kennedy, and S. Kini in 1995. Its major objective is to determine whether or not any differences in the percentage of black skin exist between bananas that are treated conventionally, i.e., placed on a counter, and bananas that are hung up. A minor objective is to determine whether or not any difference exists

in the percentage of black skin between bananas allowed to ripen in a normal day/night cycle versus those ripening in the dark such
as might occur if placed in a pantry. The blocking factor has three level. There are two treatment factors and each of them has two levels:

Factor C – Light conditions, (1 = day/night cycle, 2 = dark closet).

Factor D - Storage method (1 = hanging, 2 = counter-top).

A general complete block design was used and there were $s = 4$ observations for each treatment combinations. The data and its description can be found in files **banana.txt** and **Data-Sets-Summary.docx**, respectively. Please perform the following analysis:

(1) The experimenters did not anticipate a $block \times treatment$ interaction, use an interaction plot to confirm the experimenters' perception.

(2) Use the block-treatment model to construct the analysis of variance table and draw the conclusions.

(3) Check the model assumptions.

(4) Construct the simultaneous 95% confidence interval for the pairwise comparisons on the response for the treatment combinations and state your conclusions.

(5) Check the model assumptions.

(6) Use the following model to construct the analysis of variance table and draw your conclusions.

$$Y_{hijt} = \mu + \theta_h + \gamma_i + \delta_j + (\gamma\delta)_{ij} + (\theta\gamma)_{hi} + (\theta\delta)_{hj} + \epsilon_{hijt}$$

Here we use θ_h, γ_i and $\delta_j$to represent the effects of the blocking factor and two treatment factors, respectively.

Solution: Please refer to the R program.

MA4720: Design and Analysis of Experiments

Week 7 At-a-Glance

Title: Analysis of Covariance

Overview

In this week's lessons, we will discuss the models for the analysis of covariance. We will first present the statistical models and the assumptions for the analysis of one treatment factor (or treatment combination) and one quantitative factor. Then we will describe the statistical inference for the contrasts, including the least squares estimates, the confidence intervals with and without methods of multiple comparisons. We will discuss the construction and interpretation of the analysis of variance table.

Objectives

When you complete this module, you should be able to do the following:

1. Apply the analysis of covariance model to analyze experiments with nuisance factors that cannot be measured or can only be measured until the experiment is conducted.
2. Calculate the least squares estimates and the confidence interval with appropriate methods of multiple comparisons.
3. Calculate sample size and power and check model assumptions

Reading (Suggested)

1. Sections 9.1 to 9.5 in Chapter 9.

Instruction Content

See other files for details.

Quiz

No quiz for this final week.

Homework

One optional homework. See other files for details.

Final Exam

One final exam. See other files for details.

Final Project

One final report and one final presentation. See other files for details.

MA4720: Design and Analysis of Experiments

Week 7 Instruction Contents

Lesson 7.1 Analysis of Covariance - Models

Required Reading: Sections 9.1 and 9.2 from Chapter 9

Introduction

In Lesson 1.1, we discussed the nuisance factors and classified these nuisance factors into three types: blocking factors, noise factors, and covariates. If nuisance factors are expected to be a major source of variation, they should be taken into account in the design and analysis of the experiment. In Week 6, we described the randomized complete block designs and the general complete block designs and illustrated how to take into account blocking factors at the design stage.

In this lesson, we discuss the **analysis of covariance**, which is a means of adjusting the analysis for nuisance factors that cannot be controlled and that sometimes cannot be measured until the experiment is conducted. The method is applicable if the nuisance factors are related to the response variable but are themselves unaffected by the treatment factors.

Example 7.1: Adjustment of Covariates

Suppose an investigator wants to compare the effects of several diets on the weights of month-old piglets. The response (weight at the end of the experimental period) is likely to be related to the weight at the beginning of the experimental period, and these weights will typically be somewhat variable. To control or adjust for this prior weight variability, there are two possible solutions:

(1) The block deigns can be used. The piglets can be divided into groups (or blocks) of comparable weight, then the effects of diets are compared within blocks.
(2) A completely randomized design with response being the weight gain over the experimental period is used. This loses information, however, since heavier piglets may experience higher weight gain than lighter piglets, or vice versa. It is preferable to include

the prior weight in the model as a variable, called a **covariate**, that helps to explain the final weight.

Such model includes the effects of the treatment factors of interest, together with the effects of any nuisance factors (covariates). **Analysis of covariance** is the comparison of treatment effects, adjusting for one or more covariates.

Statistical Models

Consider an experiment conducted as a completely randomized design to compare the effects of the levels of v treatments (or treatment combinations) on a response variable Y. Suppose that the response is also affected by a nuisance factor (covariate) whose value x can be measured during or prior to the experiment. Then the analysis of covariance model:

$$Y_{it} = \mu + \tau_i + \beta x_{it} + \epsilon_{it}, t = 1, \cdots, r_i; i = 1, \cdots, v$$

with the following assumptions:

- μ, β, τ_i $(i = 1, \cdots, v)$ are unknown population parameters
- τ_i is the effect of the ith treatment
- β is the effect of covariate, which is a slope.
- $\epsilon_{it} (t = 1, \cdots, r_i; i = 1, \cdots, v)$ are independent and $\epsilon_{it} \sim N(0, \sigma^2)$: normal distribution with mean 0 and variance σ^2

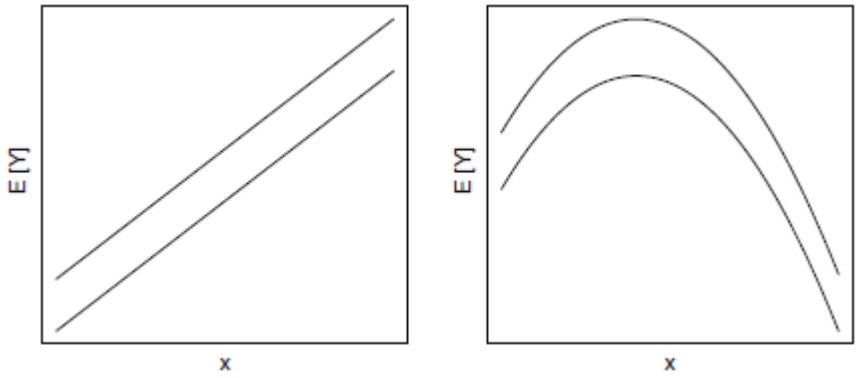
In this model, the effect of the ith treatment is modeled as τ_i, as usual. If there is more than one treatment factor, then τ_i represents the effect of the ith treatment combination and could be replaced by main-effect and interaction parameters. The value of the covariate on the tth time that treatment i is observed is written as x_{it}, and the linear relationship between the response and the covariate is modeled as βx_{it} as in a regression model. It is important for the analysis that

follows that the value x_{it} of the covariate not be affected by the treatment - otherwise, omparison of treatment effects at a common x −value would not be meaningful.

From the above model, we can easily obtain that:

$$E(Y_{it}) = E(\mu + \tau_i + \beta x_{it} + \epsilon_{hi}) = \mu + \tau_i + \beta x_{it}$$

This actually tells us that the model assumes there is a linear relationship between the expected value of the response variable, $E[Y]$ and the covariate, x, with the same slope for each treatment. Then, if we plot $E[Y]$ versus x for each treatment separately, we would see parallel lines, as illustrated for two treatments in Figure 7.1. A comparison of the effects of the two treatments can be done by comparison of mean response at any value of x.



(a) Linear response curves **(b)** Quadratic response curves

Figure 7.1 Linear and quadratic parallel response curves.

Extended and Alternative Models

The analysis of covariance model can be generalized in various ways that we will mention here but not consider further. If the effect of the covariate is not linear, then βx can be replaced with a higher-order polynomial function $\beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p$ to adequately model the common shape of the response curves for each treatment, analogous to the polynomial response curve

models of Chapter 8 in the textbook. For example, parallel quadratic response curves for two treatments are shown in Figure 7.1 (b).

If there is more than one covariate, the single covariate term can be replaced by an appropriate polynomial function of all the covariates. For example, for two covariates x_1 and x_2, the second-order function

$$\beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2$$

might be used, analogous to the polynomial response surface models of Chapter 16 in the textbook.

In this course, we will consider one covariate and the following alternative model will be used:

$$Y_{it} = \mu + \tau_i + \beta(x_{it} - \bar{x}_{..}) + \epsilon_{it}$$

In this model, the covariate values have been "centered" by their overall mean. Although the meaning of $\mu + \tau_i$ in these two models are slightly different, these two models are equivalent for comparison of treatment effects. The slope parameter β has the same interpretation in both models. This model is often used to reduce computational problems and is a little easier to work with in obtaining least squares estimates.

Checking Model Assumptions and Equality of Slopes

In addition to the usual assumptions on the error variables, the analysis of covariance model assumes two additional assumptions: (1) there is a linear relationship between the covariate and the mean response, and (2) the same slope is same for each treatment, as illustrated in Figure 7.1 (a).

To check if the linear relationship is sufficient, the lack of fit can be investigated by plotting the residuals versus the covariate for each treatment on the same scale. If the plot looks nonlinear for

any treatment, then a linear relationship between the response and covariate may not be adequate so the higher orders of the covariate can be included in the model. If each plot does look linear, one can assess whether the slopes are comparable. In next lesson, we will introduce a plot and a formal test and to investigate if the slope is same for each treatment.

If there is no significant lack of fit of the model and the test of equality of slope is not rejected, the model $Y_{it} = \mu^* + \tau_i + \beta(x_{it} - \bar{x}_{..}) + \epsilon_{it}$ can be used, Then plots of the residuals versus run order, predicted values, and normal scores can be used as in lesson 3.3. to assess the assumptions of independence, equal variances, and normality of the random error terms.

Balloon Experiment

This experiment will be used as an example for the analysis of covariance. Prior to 1985, the experimenter (Meily Lin) had observed that some colors of birthday balloons seem to be harder to inflate than others. She ran this experiment to determine whether balloons of different colors are similar in terms of the time taken for inflation to a diameter of 7 inches. Four colors were selected from a single manufacturer. An assistant blew up the balloons and the experimenter recorded the times (to the nearest 1/10 second) with a stop watch. The data, in the order collected, are given in the file **balloon.txt**, where the codes 1, 2, 3, 4 denote the colors pink, yellow, orange, blue, respectively. The description of this experiment can be found in the file **Data-Sets-Summary.docx.**

If we use the one-way analysis of variance model and plot the standardized residuals were plotted against the run order of the observations. The plot, reproduced in Figure 7.2, shows a clear linear decreasing trend in the residuals. This trend indicates a definite lack of independence in the error terms under the one-way analysis of variance model, but the trend can be eliminated by including the run order as a covariate in the model. The analysis of covariance table and the check of the model assumptions will be discussed in next lesson.

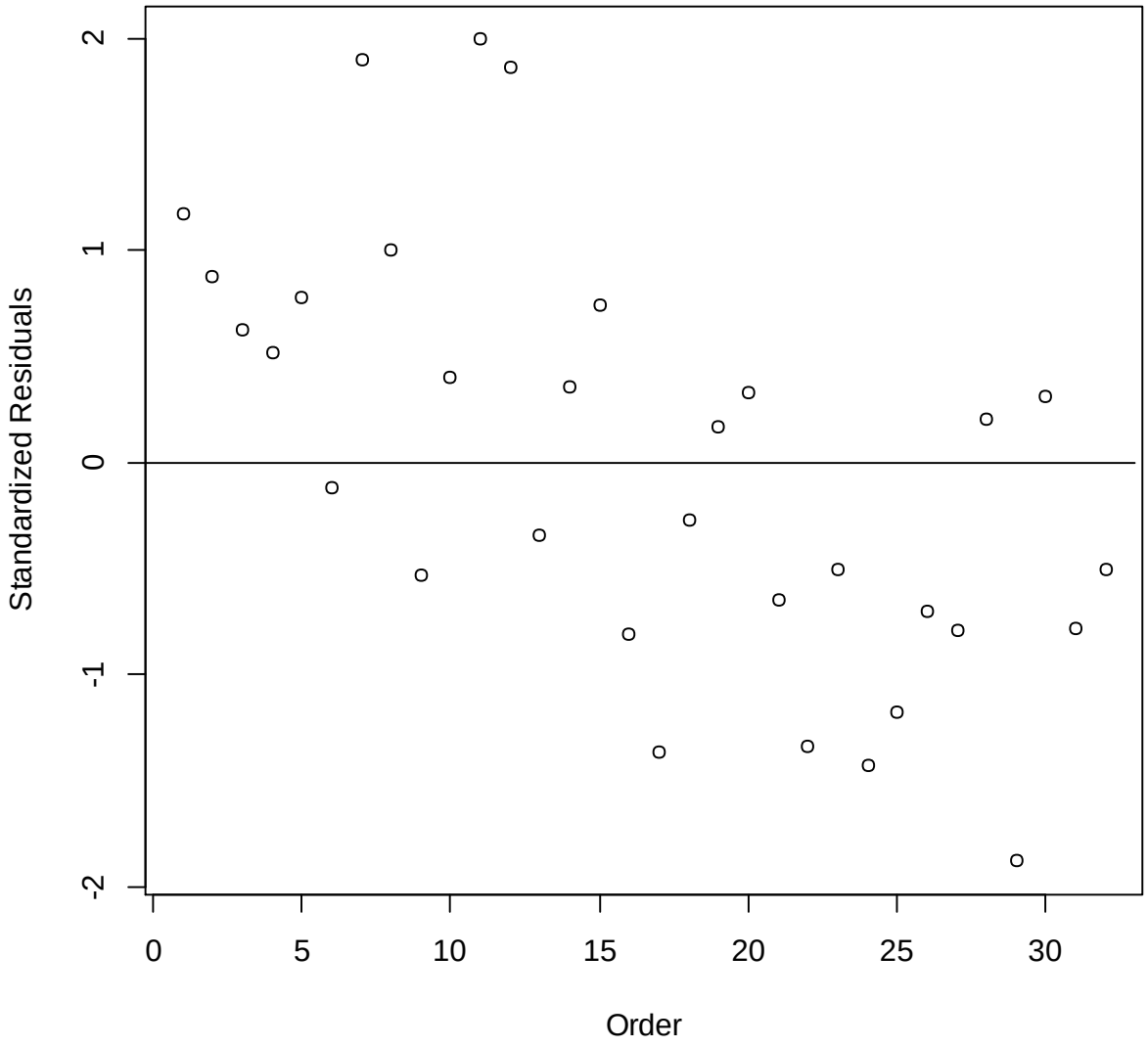


Figure 7.2 Residual plot for Balloon Experiment from the on-way analysis of variance model.

Catalyst Experiment

This experiment will be used as an exercise problem for this week. H. Smith, in the 1969 volume of *Journal of Quality Technology*, described an experiment that investigated the effect of four reagents (factor *A*) and three catalysts (factor *B*) on the production rate in a catalyst plant. He coded the reagents as *A*, *B*, *C*, and *D*, and the catalysts as *X*, *Y* , and *Z*, giving twelve treatment combinations, coded as *AX*, *AY*, . . . , *DZ*. In the data file, the four reagents are recoded as $A = 1$,

$B = 2$, $C = 3$, and $D = 4$. While the three levels pf catalyst are recoded as $X = 1$, $Y = 2$, and $Z = 3$.

Two observations were taken on each treatment combination, and these are shown in the data file, together with the order in which the observations were collected.

The data and the description of this experiment can be found in the files **catalyst.txt** and **Data-Sets-Summary.docx**, respectively.

Lesson 7.2 Least Squares Estimates and Analysis of Covariance

Required Reading: Sections 9.3 and 9.4 from Chapter 9

Least Squares Estimates from Four Models

Before we use the following analysis of covariance model:

$$Y_{it} = \mu + \tau_i + \beta(x_{it} - \bar{x}_{..}) + \epsilon_{it}$$

We would like to know if the slope is same for each treatment. This can be done by the following F-statistic:

$$F = \frac{(SSE_{Reduced} - SSE_{Full})/(df\ of\ SSE\ in\ Reduced\ Model - df\ of\ SSE\ in\ Full\ Model)}{SSE_{Full}/(df\ of\ SSE\ in\ Full\ Model)}$$

Where the full model and reduced model are:

Full Model: $Y_{it} = \mu + \tau_i + \beta_i(x_{it} - \bar{x}_{..}) + \epsilon_{it}$

Reduced Model: $Y_{it} = \mu + \tau_i + \beta(x_{it} - \bar{x}_{..}) + \epsilon_{it}$

The full model assumes that the slopes are not same for all treatments.

If the test of equality of slope is not rejected and the analysis of covariance model is used, we are interested in the following two hypothesis test:

First hypothesis test: $\tau_1 = \tau_2 = \cdots = \tau_v$

Second hypothesis: $\beta = 0$

These can be done by comparing the analysis of covariance model with the following models, respectively:

$$Y_{it} = \mu + \tau + \beta(x_{it} - \bar{x}_{..}) + \epsilon_{it}$$

$$Y_{it} = \mu + \tau_i + \epsilon_{it}$$

In summary, we are interested in the following models and their least squares estimates:

Model	Model	Note
M1	$Y_{it} = \mu + \tau_i + \beta_i(x_{it} - \bar{x}_{..}) + \epsilon_{it}$	Different treatment effects and slopes
M2	$Y_{it} = \mu + \tau_i + \beta(x_{it} - \bar{x}_{..}) + \epsilon_{it}$	Different treatment effects and same slopes
M3	$Y_{it} = \mu + \tau + \beta(x_{it} - \bar{x}_{..}) + \epsilon_{it}$	Same treatment effects and slopes
M4	$Y_{it} = \mu + \tau_i + \epsilon_{it}$	Different treatment effects and no slope

Model M1:

$$Y_{it} = \mu + \tau_i + \beta_i(x_{it} - \bar{x}_{..}) + \epsilon_{it}$$

For this model, it is equivalent to fit a simple linear regression model using data from each treatment separately. Therefore, the least squares estimates of model parameters ($\mu + \tau_i$, β_i, and σ^2) can be obtained based on the linear model. For this course, we will rely on the R function lm() to obtain these estimates. In addition, similar to the analysis of variance table, the following analysis of covariance table can be obtained:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F-test Statistic
$T\|\beta$	$v - 1$	$ss(T\|\beta)$	$ms(T\|\beta) = \frac{ss(T\|\beta)}{v-1}$	$\frac{ms(T\|\beta)}{msE}$
$\beta\|T$	v	$ss(\beta\|T)$	$ms(\beta\|T) = \frac{ss(\beta\|T)}{v}$	$\frac{ms(\beta\|T)}{msE}$
Error	$n - 2v$	ssE	$msE = \frac{ssE}{n-2v}$	
Total	$n - 1$	$sstot$		

Note that we are more interested in the test of equality of slope, and here $ss(\beta|T)$ can be obtained by comparing the ssE from model $M1$ and the ssE from model $M2$:

$$ss(\beta|T) = ssE_{Model\ M2} - ssE_{Model\ M1}$$

Model M2:

$$Y_{it} = \mu + \tau_i + \beta(x_{it} - \bar{x}_{..}) + \epsilon_{it}$$

This is the analysis of covariance model. Again, we will use the R function **lm()** to obtain the least squares estimates of model parameters. Similarly, the following analysis of covariance table can be obtained:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F-test Statistic
$T\|\beta$	$v-1$	$ss(T\|\beta)$	$ms(T\|\beta) = \frac{ss(T\|\beta)}{v01}$	$\frac{ms(T\|\beta)}{msE}$
$\beta\|T$	1	$ss(\beta\|T)$	$ms(\beta\|T) = \frac{ss(\beta\|T)}{1}$	$\frac{ms(\beta\|T)}{msE}$
Error	$n-v-1$	ssE	$msE = \frac{ssE}{n-v-1}$	
Total	$n-1$	$sstot$		

Note $ss(T|\beta)$ and $ss(\beta|T)$ can be obtained by comparing the ssE from model $M2$ with the ssE from model $M3$ and the model $M4$, respectively:

$$ss(T|\beta) = ssE_{Model\ M3} - ssE_{Model\ M2}$$

$$ss(\beta|T) = ssE_{Model\ M4} - ssE_{Model\ M2}$$

Model M3:

$$Y_{it} = \mu + \tau + \beta(x_{it} - \bar{x}_{..}) + \epsilon_{it}$$

For this model, it is equivalent to fit a simple linear regression model using all data. Again, we will use the R function **lm()** to obtain the least squares estimates of model parameters. Similarly, the following analysis of covariance table can be obtained:

Source of Variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**F-test Statistic**
β	1	$ss\beta$	$ms\beta = ss\beta$	$\frac{ms\beta}{msE}$
Error	$n-2$	ssE	$msE = \frac{ssE}{n-2}$	
Total	$n-1$	$sstot$		

Model M4:

$$Y_{it} = \mu + \tau_i + \epsilon_{it}$$

This is just the one-way analysis of variance model. The following one-way analysis of variance table can be obtained:

Source of Variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**F-test Statistic**
Treatment	$v-1$	ssT	$msT = \frac{ssT}{v-1}$	$\frac{msT}{msE}$
Error	$n-v$	ssE	$msE = \frac{ssE}{n-v}$	
Total	$n-1$	$sstot$		

In summary, we can combine the sum of squares of error to the following table:

Model	**Degrees of Freedom**	$\boldsymbol{ssE}$	$\boldsymbol{msE}$

	for Error		
$M1$	$n-2v$	ssE_{M1}	$msE_{M1}=\frac{ssE_{M1}}{n-2v}$
$M2$	$n-v-1$	ssE_{M2}	$msE_{M2}=\frac{ssE_{M2}}{n-v-1}$
$M3$	$n-2$	ssE_{M3}	$msE_{M3}=\frac{ssE_{M3}}{n-2}$
$M4$	$n-v$	ssE_{M4}	$msE_{M4}=\frac{ssE_{M4}}{n-v}$

We reject $H_0: \beta_1=\beta_2=\cdots=\beta_v$ at the significance α if:

$$F=\frac{(ssE_{M2}-ssE_{M1})/(v-1)}{msE_{M1}}>F_{\alpha,v-1,n-2v}$$

We reject $H_0: \tau_1=\tau_2=\cdots=\tau_v$ at the significance α if:

$$F=\frac{(ssE_{M3}-ssE_{M2})/(v-1)}{msE_{M2}}>F_{\alpha,v-1,n-v-1}$$

We reject $H_0: \beta=0$ at the significance α if:

$$F=\frac{(ssE_{M4}-ssE_{M2})/1}{msE_{M2}}>F_{\alpha,1,n-v-1}$$

Example 7.2: Test of Equality of Slopes - Balloon Experiment

Use the order as the covariate and find the least squares estimates of model parameters for the following models:

M1: $Y_{it}=\mu+\tau_i+\beta_i(x_{it}-\bar{x}_{..})+\epsilon_{it}$

M2: $Y_{it}=\mu+\tau_i+\beta(x_{it}-\bar{x}_{..})+\epsilon_{it}$

Draw a plot and perform a formal test of equality of slope using a level of significance $\alpha=0.05$.

Solution: Please refer to R program and the video to find how we can obtain the least squares estimates of model parameters. Here we list the results here:

$$n = 32, v = 4$$

For model M1: $Y_{it} = \mu + \tau_i + \beta_i(x_{it} - \bar{x}_{..}) + \epsilon_{it}$, we have:

$$\hat{\beta}_1 = -0.140, \hat{\beta}_2 = -0.385, \hat{\beta}_3 = -0.156, \hat{\beta}_4 = -0.270$$

$$\hat{\mu} + \hat{\tau}_1 = 18.303, \hat{\mu} + \hat{\tau}_2 = 22.239, \hat{\mu} + \hat{\tau}_3 = 22.031, \hat{\mu} + \hat{\tau}_4 = 18.221$$

$$\hat{\sigma}^2 = 6.543, ssE = \hat{\sigma}^2 * 24 = 157.032$$

$$df\ for\ error: n - 2 * v = 32 - 2 * 4 = 24$$

For model M2: $Y_{it} = \mu + \tau_i + \beta(x_{it} - \bar{x}_{..}) + \epsilon_{it}$, we have

$$\hat{\beta} = -0.211$$

$$\hat{\mu} + \hat{\tau}_1 = 18.285, \hat{\mu} + \hat{\tau}_2 = 22.390, \hat{\mu} + \hat{\tau}_3 = 22.086, \hat{\mu} + \hat{\tau}_4 = 18.214$$

$$\hat{\sigma}^2 = 6.731, ssE = \hat{\sigma}^2 * 27 = 181.742$$

$$df\ for\ error = n - v - 1 = 32 - 4 - 1 = 27$$

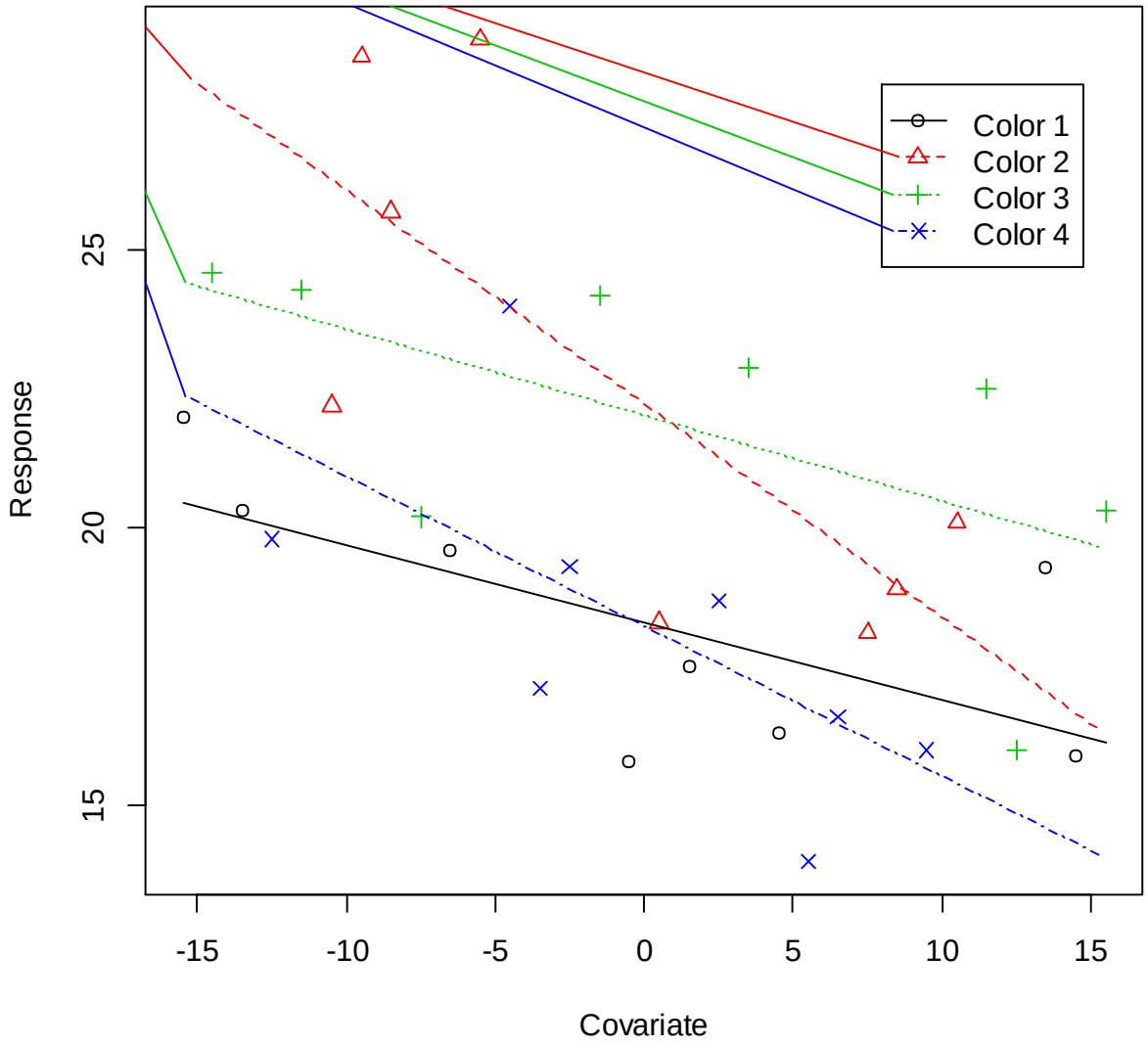


Figure 7.3 Response curves for Balloon Experiment.

From the plot, we can see one some curves are parallel. The F-statistic is:

$$F = \frac{(ssE_{M2} - ssE_{M1})/(v-1)}{msE_{M1}} = F = \frac{(181.742 - 157.032)/(4-1)}{6.653} = 1.2589$$

The p-value is:

$$p - value = P(F_{3,24} > 1.2589) = 0.3108$$

At the significance 0.05, we fail to reject the null hypothesis of equality of slopes, so the analysis of covariance model (model M2) can be used here.

Exercise 7.1: Test of Equality of Slopes - Catalyst Experiment

Use the order as the covariate, use the reagent as the covariate, and find the least squares estimates of model parameters for the following models:

M1: $Y_{it} = \mu + \tau_i + \beta_i(x_{it} - \bar{x}_{..}) + \epsilon_{it}$

M2: $Y_{it} = \mu + \tau_i + \beta(x_{it} - \bar{x}_{..}) + \epsilon_{it}$

Draw a plot and perform a formal test of equality of slope using a level of significance $\alpha = 0.05$.

Solution: R program for this exercise is provided.

Example 7.3: Analysis of Covariance - Balloon Experiment

Use the order as the covariate and find the least squares estimates of model parameters for the following models:

M3: $Y_{it} = \mu + \tau + \beta(x_{it} - \bar{x}_{..}) + \epsilon_{it}$

M4: $Y_{it} = \mu + \tau_i + \epsilon_{it}$

Test $\tau_1 = \tau_2 = \cdots = \tau_v$ and $\beta = 0$ separately, using a level of significance $\alpha = 0.05$.

Solution: Again, please refer to R program for the detailed calculation. Here we list the results here:

$$n = 32, v = 4$$

For model M3: $Y_{it} = \mu + \tau + \beta(x_{it} - \bar{x}_{..}) + \epsilon_{hi}$, we have:

$$\hat{\beta} = -0.210, \hat{\mu} + \hat{\tau} = 20.244$$

$$\hat{\sigma}^2 = 10.314, ssE = \hat{\sigma}^2 * 30 = 309.42$$

$$df\ for\ error: n - 2 = 32 - 2 = 30$$

For model M4: $Y_{it} = \mu + \tau_i + \epsilon_{it}$, we have

$$\hat{\mu} + \hat{\tau}_1 = 18.338, \hat{\mu} + \hat{\tau}_2 = 22.575, \hat{\mu} + \hat{\tau}_3 = 21.875, \hat{\mu} + \hat{\tau}_4 = 18.187$$

$$\hat{\sigma}^2 = 10.806, ssE = \hat{\sigma}^2 * 28 = 302.577$$

$$df\ for\ error = n - v = 32 - 4 = 28$$

The F-statistic for testing H_0: $\tau_1 = \tau_2 = \cdots = \tau_v$:

$$F = \frac{(ssE_{M3} - ssE_{M2})/(v-1)}{msE_{M2}} = \frac{(309.442 - 181.742)/(4-1)}{6.731} = 6.3227$$

The corresponding p-value is:

$$p - value = P(F_{3,27} > 6.3227) = 0.00218$$

At the significance 0.05, we reject the null hypothesis of $\tau_1 = \tau_2 = \cdots = \tau_v$, indicating the treatments have different effects on the response.

The F-statistic for testing H_0: $\beta = 0$ is:

$$F = \frac{(ssE_{M4} - ssE_{M2})/1}{msE_{M2}} = \frac{302.58 - 181.742}{6.731} = 17.9516$$

The corresponding p-value is:

$$p - value = P(F_{1,27} > 17.9516) = 0.0002358$$

At the significance 0.05, we reject the null hypothesis of $\beta = 0$. This conclusion is consistent with Figure 7.2 that shows a clear linear decreasing trend in the residuals.

Exercise 7.2: Catalyst Experiment

Use the order as the covariate, use the reagent as the covariate, and find the least squares estimates of model parameters for the following models:

M3: $Y_{it} = \mu + \tau + \beta(x_{it} - \bar{x}_{..}) + \epsilon_{it}$

M4: $Y_{it} = \mu + \tau_i + \epsilon_{it}$

Test $\tau_1 = \tau_2 = \cdots = \tau_v$ and $\beta = 0$ separately, using a level of significance $\alpha = 0.05$.

Solution: R program for this exercise is provided.

Example 7.4: Checking Model Assumptions - Balloon Experiment

Use the order as the covariate and use the R function lm to fit the model M2

(1) Check the independence and the equal variance assumptions.
(2) Check if the linear relationship is adequate to describe the relationship between the response and the covariate.

Solution: Please refer the R program and video for this example. Here we only present the results of (2). From Figure 7.4, it seems that the linear relationship may not be sufficient for Color 2 (Yellow).

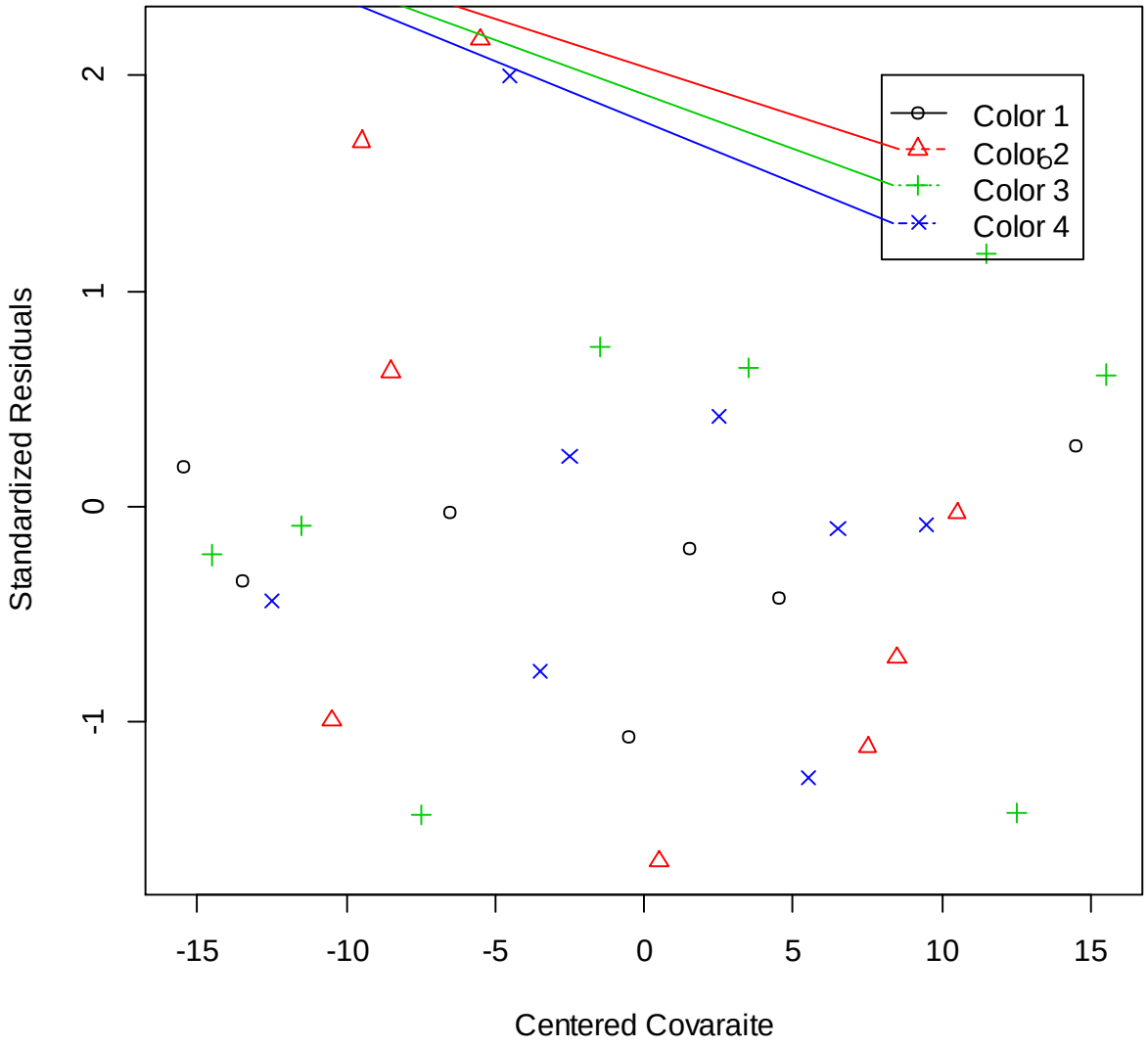


Figure 7.4 Plot of standardized results versus the centered covariate.

Exercise 7.3: Catalyst Experiment

Use the order as the covariate, use the reagent as the covariate, and use the R function lm to fit the model M1, M2, M3, and M4,

(1) Check the independence and the equal variance assumptions.

(2) Check if the linear relationship is adequate to describe the relationship between the response and the covariate.

Solution: R program for this exercise is provided.

Lesson 7.3 Treatment Contrasts and Confidence Intervals

Required Reading: Section 9.5 from Chapter 9

For this lesson, we consider one factor with v treatments (or treatment combinations), one covariate, and the following analysis of covariance model:

$$Y_{it} = \mu + \tau_i + \beta(x_{it} - \bar{x}_{..}) + \epsilon_{it}$$

The conclusions can be easily extended to the designs with more one treatment factor and/or more than one covariate.

Confidence Intervals

From Lesson 7.2, the least squares estimator of the contrast $\sum_{i=1}^{v} c_i\tau_i$, where $\sum_{i=1}^{v} c_i = 0$ is:

$$\sum_{i=1}^{v} c_i\hat{\tau}_i = \sum_{i=1}^{v} c_i(\bar{Y}_{i\cdot} - \hat{\beta}(\bar{x}_{i\cdot} - \bar{x}_{..})) = \sum_{i=1}^{v} c_i(\bar{Y}_{i\cdot} - \hat{\beta}\bar{x}_{i\cdot})$$

Based on this estimator, we can derive the variance of $\sum_{i=1}^{v} c_i\hat{\tau}_i$ and calculate its $(1-\alpha) * 100\%$ confidence interval as:

$$point\ estimate \pm w * standard\ error$$

where the critical coefficient w depends on if and which methods of multiple comparisons that will be used:

No Multiple Comparison: $w = t_{n-v-1,\alpha}$

Bonferroni Method: $w_B = t_{n-v-1,\alpha/(2*m)}$, m is the number pre-specified contrasts

Scheffé Method: $w_S = \sqrt{(v-1)F_{v-1,n-v-1,\alpha}}$

Since $(\bar{Y}_{i\cdot} - \hat{\beta}\bar{x}_{i\cdot})(i = 1, \cdots, v)$ may not be independent, so the methods of Tukey and Dunnett are not know to apply. In this lesson, we will rely on the R program to find the point estimate and the standard error of a contrast.

Example 7.5: Contrasts in Analysis of Covariance Model - Balloon Experiment

Use the order as the covariate and the analysis of covariance model to find the simultaneous 95% confidence intervals for all pairwise treatment comparisons.

Solution: The following table can be obtained:

Contrast	Point Estimate	Standard Error	Simultaneous 95% confidence intervals		
			No Multiple Comparison	Bonferroni	Scheffé
$\tau_1 - \tau_2$	-4.106	1.3	$(-6.77, -1.44)$	$(-7.80, -0.41)$	$(-7.97, -0.238)$
$\tau_1 - \tau_3$	-3.801	1.3	$(-6.47, -1.14)$	$(-7.50, -0.104)$	$(-7.67, 0.069)$
$\tau_1 - \tau_4$	0.071	1.3	$(-2.59, 2.73)$	$(-3.62, 3.76)$	$(-3.80, 3.94)$
$\tau_2 - \tau_3$	0.304	1.3	$(-2.36, 2.97)$	$(-3.40, 4.01)$	$(-3.57, 4.18)$
$\tau_2 - \tau_4$	4.177	1.3	$(1.51, 6.84)$	$(0.48, 7.87)$	$(0.308, 8.05)$
$\tau_3 - \tau_4$	3.872	1.3	$(1.21, 6.54)$	$(0.18, 7.57)$	$(0.0041, 7.74)$

The critical coefficient w is,

No Multiple Comparison: $w = t_{27,0.05/2} = 2.0518$

Bonferroni Method: $w_B = t_{27,0.05/(2*6)} = 2.8469$

Scheffé Method: $w_s = \sqrt{(4-1)F_{4-1,27,0.05}} = 2.9801$

We can verify the results from the above table (up to some round errors). For example, for $\tau_1 - \tau_2$, we have

No Multiple Comparison: $-4.106 \pm 1.3 * 2.0518 = (-6.77, -1.44)$

Bonferroni Method: $-4.106 \pm 1.3 * 2.8469 = (-7.81, -0.41)$

Scheffé Method: $-4.106 \pm 1.3 * 2.9801 = (-7.98, -0.23)$

From the above table, we can see Bonferroni gives the shorter confidence intervals than Shceffe method. Based on the results from Bonferroni method, we can observe that: (1) the mean time to inflate balloons is longer for color 2 (yellow) than for colors 1 and 4 (pink and blue), and the mean inflation time for color 3 (orange) is longer than for colors 1 and 4 (pink and blue).

Exercise 7.4: Contrasts in Analysis of Covariance Model - Catalyst Experiment

Use the order as the covariate and the analysis of covariance model to find the simultaneous 95% confidence intervals for all pairwise treatment comparisons for the reagent. Construct a table that is similar to the table in the solution of Example 7.5 and state your conclusions. Calculate critical coefficients and choose one of contrasts to manually verify its confidence intervals. [**Hint**: you need fist find the contrast coefficients for all pairwise treatment comparisons for the reagent in terms of 12 treatment combinations which are arranged by the lexical order. For example, the contrast coefficients for reagent 1 – reagent 2 in terms of 12 treatment combinations are:

$$[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, -\frac{1}{3}, 0, 0, 0, 0, 0, 0]$$

Solution: R program for this exercise is provided.