The "Big 3" Multiple Comparison Procedures

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

P311, 2013

The "Big 3" Multiple Comparison Procedures

Introduction

The Omnibus F-Test

- Consider again a simple 1-way Analysis of Variance setup with $a = 4$ groups.
- One statistical question is "Are all the group means the same?"
- That question is addressed with the *omnibus (or overall) F-test.*
- This F-test addresses the question directly by testing the hypothesis

$$H_0 : \ \mu_1 = \mu_2 = \mu_3 = \mu_4 \tag{1}$$

- If the F-test rejects this null hypothesis, you can conclude that it is highly likely some means are different.

Introduction

The Omnibus F-Test

- Consider again a simple 1-way Analysis of Variance setup with $a = 4$ groups.
- One statistical question is "Are all the group means the same?"
- That question is addressed with the *omnibus (or overall) F-test.*
- This F-test addresses the question directly by testing the hypothesis

$$H_0: \ \mu_1 = \mu_2 = \mu_3 = \mu_4 \tag{1}$$

- If the F-test rejects this null hypothesis, you can conclude that it is highly likely some means are different.

Introduction
The Omnibus F-Test

- Consider again a simple 1-way Analysis of Variance setup with $a = 4$ groups.
- One statistical question is "Are all the group means the same?"
- That question is addressed with the *omnibus (or overall) F-test.*
- This F-test addresses the question directly by testing the hypothesis

$$H_0: \ \mu_1 = \mu_2 = \mu_3 = \mu_4 \tag{1}$$

- If the F-test rejects this null hypothesis, you can conclude that it is highly likely some means are different.

Introduction

The Omnibus F-Test

- Consider again a simple 1-way Analysis of Variance setup with $a = 4$ groups.
- One statistical question is "Are all the group means the same?"
- That question is addressed with the *omnibus (or overall) F-test.*
- This F-test addresses the question directly by testing the hypothesis

$$H_0: \ \mu_1 = \mu_2 = \mu_3 = \mu_4 \tag{1}$$

- If the F-test rejects this null hypothesis, you can conclude that it is highly likely some means are different.

Introduction
The Omnibus F-Test

- Consider again a simple 1-way Analysis of Variance setup with $a = 4$ groups.
- One statistical question is "Are all the group means the same?"
- That question is addressed with the *omnibus (or overall) F-test.*
- This F-test addresses the question directly by testing the hypothesis
$$H_0 : \ \mu_1 = \mu_2 = \mu_3 = \mu_4 \tag{1}$$
- If the F-test rejects this null hypothesis, you can conclude that it is highly likely some means are different.

Introduction

Alternative Hypotheses

- The omnibus hypothesis is a good starting place.
- However, you may enter the ANOVA situation with one or more other hypotheses that are of greater substantive interest.
- In that case, you may wish to perform other statistical tests.
- In previous lectures, we examined the *general* issues surrounding multiple hypothesis testing.
- We saw that there are several key problems that have to be dealt with when you perform additional hypothesis tests.
- Key among them are (1) the proliferation of Type-I errors when a significant number of tests are performed, and (2) the problem of *post hoc* inference, i.e., the fact that the probability model changes if you test a hypothesis after examining the data.

Introduction
Alternative Hypotheses

- The omnibus hypothesis is a good starting place.
- However, you may enter the ANOVA situation with one or more other hypotheses that are of greater substantive interest.
- In that case, you may wish to perform other statistical tests.
- In previous lectures, we examined the *general* issues surrounding multiple hypothesis testing.
- We saw that there are several key problems that have to be dealt with when you perform additional hypothesis tests.
- Key among them are (1) the proliferation of Type-I errors when a significant number of tests are performed, and (2) the problem of *post hoc* inference, i.e., the fact that the probability model changes if you test a hypothesis after examining the data.

Introduction
Alternative Hypotheses

- The omnibus hypothesis is a good starting place.
- However, you may enter the ANOVA situation with one or more other hypotheses that are of greater substantive interest.
- In that case, you may wish to perform other statistical tests.
- In previous lectures, we examined the *general* issues surrounding multiple hypothesis testing.
- We saw that there are several key problems that have to be dealt with when you perform additional hypothesis tests.
- Key among them are (1) the proliferation of Type-I errors when a significant number of tests are performed, and (2) the problem of *post hoc* inference, i.e., the fact that the probability model changes if you test a hypothesis after examining the data.

Introduction
Alternative Hypotheses

- The omnibus hypothesis is a good starting place.
- However, you may enter the ANOVA situation with one or more other hypotheses that are of greater substantive interest.
- In that case, you may wish to perform other statistical tests.
- In previous lectures, we examined the *general* issues surrounding multiple hypothesis testing.
- We saw that there are several key problems that have to be dealt with when you perform additional hypothesis tests.
- Key among them are (1) the proliferation of Type-I errors when a significant number of tests are performed, and (2) the problem of *post hoc* inference, i.e., the fact that the probability model changes if you test a hypothesis after examining the data.

Introduction

Alternative Hypotheses

- The omnibus hypothesis is a good starting place.
- However, you may enter the ANOVA situation with one or more other hypotheses that are of greater substantive interest.
- In that case, you may wish to perform other statistical tests.
- In previous lectures, we examined the *general* issues surrounding multiple hypothesis testing.
- We saw that there are several key problems that have to be dealt with when you perform additional hypothesis tests.
- Key among them are (1) the proliferation of Type-I errors when a significant number of tests are performed, and (2) the problem of *post hoc* inference, i.e., the fact that the probability model changes if you test a hypothesis after examining the data.

Introduction

Alternative Hypotheses

- The omnibus hypothesis is a good starting place.
- However, you may enter the ANOVA situation with one or more other hypotheses that are of greater substantive interest.
- In that case, you may wish to perform other statistical tests.
- In previous lectures, we examined the *general* issues surrounding multiple hypothesis testing.
- We saw that there are several key problems that have to be dealt with when you perform additional hypothesis tests.
- Key among them are (1) the proliferation of Type-I errors when a significant number of tests are performed, and (2) the problem of *post hoc* inference, i.e., the fact that the probability model changes if you test a hypothesis after examining the data.

Introduction
Alternative Hypotheses

- We introduced the concepts of *Familywise Error Rate (FWER)* and *False Discovery Rate (FDR)* as two ways of assessing overall performance of a group of tests.
- In our ANOVA situation, we concentrate on methods that control FWER for various families of tests that have proven interesting in practice.
- These procedures have been described succinctly in the lecture slide handout *A Catalog of Multiple Comparison Procedures*.

Introduction

Alternative Hypotheses

- We introduced the concepts of *Familywise Error Rate (FWER)* and *False Discovery Rate (FDR)* as two ways of assessing overall performance of a group of tests.
- In our ANOVA situation, we concentrate on methods that control FWER for various families of tests that have proven interesting in practice.
- These procedures have been described succinctly in the lecture slide handout *A Catalog of Multiple Comparison Procedures.*

Introduction
Alternative Hypotheses

- We introduced the concepts of *Familywise Error Rate (FWER)* and *False Discovery Rate (FDR)* as two ways of assessing overall performance of a group of tests.
- In our ANOVA situation, we concentrate on methods that control FWER for various families of tests that have proven interesting in practice.
- These procedures have been described succinctly in the lecture slide handout *A Catalog of Multiple Comparison Procedures*.

Introduction

Alternative Hypotheses

- In this lecture, I review three of the most widely used procedures.
- These procedures are often the first three that are discussed in introductory texts.
- We might well call them the "Big 3" of multiple comparison testing for means.
- These 3 procedures are
  1. Planned Orthogonal Contrasts
  2. The Scheffé test
  3. The Tukey test

Introduction

Alternative Hypotheses

- In this lecture, I review three of the most widely used procedures.
- These procedures are often the first three that are discussed in introductory texts.
- We might well call them the "Big 3" of multiple comparison testing for means.
- These 3 procedures are
  1. Planned Orthogonal Contrasts
  2. The Scheffé test
  3. The Tukey test

Introduction

Alternative Hypotheses

- In this lecture, I review three of the most widely used procedures.
- These procedures are often the first three that are discussed in introductory texts.
- We might well call them the "Big 3" of multiple comparison testing for means.
- These 3 procedures are
  1. Planned Orthogonal Contrasts
  2. The Scheffé test
  3. The Tukey test

Introduction
Alternative Hypotheses

- In this lecture, I review three of the most widely used procedures.
- These procedures are often the first three that are discussed in introductory texts.
- We might well call them the "Big 3" of multiple comparison testing for means.
- These 3 procedures are
  1. Planned Orthogonal Contrasts
  2. The Scheffé test
  3. The Tukey test

Introduction

Alternative Hypotheses

- In this lecture, I review three of the most widely used procedures.
- These procedures are often the first three that are discussed in introductory texts.
- We might well call them the "Big 3" of multiple comparison testing for means.
- These 3 procedures are
    1. Planned Orthogonal Contrasts
    2. The Scheffé test
    3. The Tukey test

Introduction

Alternative Hypotheses

- In this lecture, I review three of the most widely used procedures.
- These procedures are often the first three that are discussed in introductory texts.
- We might well call them the "Big 3" of multiple comparison testing for means.
- These 3 procedures are
    1. Planned Orthogonal Contrasts
    2. The Scheffé test
    3. The Tukey test

Introduction

Alternative Hypotheses

- In this lecture, I review three of the most widely used procedures.
- These procedures are often the first three that are discussed in introductory texts.
- We might well call them the "Big 3" of multiple comparison testing for means.
- These 3 procedures are
  1. Planned Orthogonal Contrasts
  2. The Scheffé test
  3. The Tukey test

Planned Orthogonal Contrasts

Linear Combination Tests

- Planned Orthogonal Contrasts are linear combination hypotheses that represent experimental hypotheses. They are of the general form,

$$H_0: \ \Psi = \sum_j c_j \mu_j = 0 \tag{2}$$

- We saw how to phrase a substantive hypothesis as a linear combination of means in our earlier discussion of the *generalized t-statistic.*
- We also saw how to construct t-statistics to test such a hypothesis.
- Planned Orthogogonal Contrasts are linear combinations that have the following characteristics:
  1. They are *planned*, that is, are of interest prior to gathering or examining the data.
  2. They are *contrasts*, that is, the linear weights sum to zero.
  3. If there is more than one planned contrast, they are orthogonal to each other, that is, for contrasts $\Psi_1 = \sum_j c_{1j}\mu_j$ and $\Psi_2 = \sum_j c_{2j}\mu_j$, we have

$$\sum_j c_{1j} c_{2j} = 0 \tag{3}$$

Planned Orthogonal Contrasts

Linear Combination Tests

- Planned Orthogonal Contrasts are linear combination hypotheses that represent experimental hypotheses. They are of the general form,

$$H_0: \ \Psi = \sum_j c_j \mu_j = 0 \tag{2}$$

- We saw how to phrase a substantive hypothesis as a linear combination of means in our earlier discussion of the *generalized t-statistic*.
- We also saw how to construct t-statistics to test such a hypothesis.
- Planned Orthogogonal Contrasts are linear combinations that have the following characteristics:
  1. They are *planned*, that is, are of interest prior to gathering or examining the data.
  2. They are *contrasts*, that is, the linear weights sum to zero.
  3. If there is more than one planned contrast, they are orthogonal to each other, that is, for contrasts $\Psi_1 = \sum_j c_{1j}\mu_j$ and $\Psi_2 = \sum_j c_{2j}\mu_j$, we have

$$\sum_j c_{1j} c_{2j} = 0 \tag{3}$$

Planned Orthogonal Contrasts

Linear Combination Tests

- Planned Orthogonal Contrasts are linear combination hypotheses that represent experimental hypotheses. They are of the general form,

$$H_0: \ \Psi = \sum_j c_j \mu_j = 0 \tag{2}$$

- We saw how to phrase a substantive hypothesis as a linear combination of means in our earlier discussion of the *generalized t-statistic*.
- We also saw how to construct t-statistics to test such a hypothesis.
- Planned Orthogogonal Contrasts are linear combinations that have the following characteristics:
  1. They are *planned*, that is, are of interest prior to gathering or examining the data.
  2. They are *contrasts*, that is, the linear weights sum to zero.
  3. If there is more than one planned contrast, they are orthogonal to each other, that is, for contrasts $\Psi_1 = \sum_j c_{1j}\mu_j$ and $\Psi_2 = \sum_j c_{2j}\mu_j$, we have

$$\sum_j c_{1j} c_{2j} = 0 \tag{3}$$

Planned Orthogonal Contrasts

Linear Combination Tests

- Planned Orthogonal Contrasts are linear combination hypotheses that represent experimental hypotheses. They are of the general form,

$$H_0: \ \Psi = \sum_j c_j \mu_j = 0 \tag{2}$$

- We saw how to phrase a substantive hypothesis as a linear combination of means in our earlier discussion of the *generalized t-statistic*.
- We also saw how to construct t-statistics to test such a hypothesis.
- Planned Orthogogonal Contrasts are linear combinations that have the following characteristics:
  1. They are *planned*, that is, are of interest prior to gathering or examining the data.
  2. They are *contrasts*, that is, the linear weights sum to zero.
  3. If there is more than one planned contrast, they are orthogonal to each other, that is, for contrasts $\Psi_1 = \sum_j c_{1j}\mu_j$ and $\Psi_2 = \sum_j c_{2j}\mu_j$, we have

$$\sum_j c_{1j} c_{2j} = 0 \tag{3}$$

Planned Orthogonal Contrasts

Linear Combination Tests

- Planned Orthogonal Contrasts are linear combination hypotheses that represent experimental hypotheses. They are of the general form,

$$H_0: \; \Psi = \sum_j c_j \mu_j = 0 \tag{2}$$

- We saw how to phrase a substantive hypothesis as a linear combination of means in our earlier discussion of the *generalized t-statistic*.
- We also saw how to construct t-statistics to test such a hypothesis.
- Planned Orthogogonal Contrasts are linear combinations that have the following characteristics:
  1. They are *planned*, that is, are of interest prior to gathering or examining the data.
  2. They are *contrasts*, that is, the linear weights sum to zero.
  3. If there is more than one planned contrast, they are orthogonal to each other, that is, for contrasts $\Psi_1 = \sum_j c_{1j}\mu_j$ and $\Psi_2 = \sum_j c_{2j}\mu_j$, we have

$$\sum_j c_{1j} c_{2j} = 0 \tag{3}$$

Planned Orthogonal Contrasts

Linear Combination Tests

- Planned Orthogonal Contrasts are linear combination hypotheses that represent experimental hypotheses. They are of the general form,

$$H_0: \ \Psi = \sum_j c_j \mu_j = 0 \tag{2}$$

- We saw how to phrase a substantive hypothesis as a linear combination of means in our earlier discussion of the *generalized t-statistic*.
- We also saw how to construct t-statistics to test such a hypothesis.
- Planned Orthogogonal Contrasts are linear combinations that have the following characteristics:
  1. They are *planned*, that is, are of interest prior to gathering or examining the data.
  2. They are *contrasts*, that is, the linear weights sum to zero.
  3. If there is more than one planned contrast, they are orthogonal to each other, that is, for contrasts $\Psi_1 = \sum_j c_{1j}\mu_j$ and $\Psi_2 = \sum_j c_{2j}\mu_j$, we have

$$\sum_j c_{1j} c_{2j} = 0 \tag{3}$$

Planned Orthogonal Contrasts

Linear Combination Tests

- Planned Orthogonal Contrasts are linear combination hypotheses that represent experimental hypotheses. They are of the general form,

$$H_0: \ \Psi = \sum_j c_j \mu_j = 0 \tag{2}$$

- We saw how to phrase a substantive hypothesis as a linear combination of means in our earlier discussion of the *generalized t-statistic*.
- We also saw how to construct t-statistics to test such a hypothesis.
- Planned Orthogogonal Contrasts are linear combinations that have the following characteristics:
  1. They are *planned*, that is, are of interest prior to gathering or examining the data.
  2. They are *contrasts*, that is, the linear weights sum to zero.
  3. If there is more than one planned contrast, they are orthogonal to each other, that is, for contrasts $\Psi_1 = \sum_j c_{1j}\mu_j$ and $\Psi_2 = \sum_j c_{2j}\mu_j$, we have

$$\sum_j c_{1j} c_{2j} = 0 \tag{3}$$

Planned Orthogonal Contrasts

An Example

- Suppose you are performing a social psychology experiment examining the effects of violent movies on willingness to be aggressive during an experimental test.
- Subjects are randomly divided into 4 groups. Groups 1, 2, and 3 view violent movies, while Group 4 views a neutral control movie.
- Movies 1 and 2 involve sexually explicit violence, while Movie 3 depicts violence of a non-sexual nature.
- You realize before the experiment is ever performed that the omnibus hypothesis concerning whether any of the movies is different really doesn't matter to you.
- You are more interested in the following questions:
  1. Is the average of the 3 violent movies the same as that of the control movie?
  2. Is the average of the two sexually explicit violent movies different from the third violent movie?
- Let's express these two experimental hypotheses as contrasts and see if they satisfy the definition of *orthogonal contrasts*

Planned Orthogonal Contrasts

An Example

- Suppose you are performing a social psychology experiment examining the effects of violent movies on willingness to be aggressive during an experimental test.
- Subjects are randomly divided into 4 groups. Groups 1, 2, and 3 view violent movies, while Group 4 views a neutral control movie.
- Movies 1 and 2 involve sexually explicit violence, while Movie 3 depicts violence of a non-sexual nature.
- You realize before the experiment is ever performed that the omnibus hypothesis concerning whether any of the movies is different really doesn't matter to you.
- You are more interested in the following questions:
  1. Is the average of the 3 violent movies the same as that of the control movie?
  2. Is the average of the two sexually explicit violent movies different from the third violent movie?
- Let's express these two experimental hypotheses as contrasts and see if they satisfy the definition of *orthogonal contrasts*

Planned Orthogonal Contrasts

An Example

- Suppose you are performing a social psychology experiment examining the effects of violent movies on willingness to be aggressive during an experimental test.
- Subjects are randomly divided into 4 groups. Groups 1, 2, and 3 view violent movies, while Group 4 views a neutral control movie.
- Movies 1 and 2 involve sexually explicit violence, while Movie 3 depicts violence of a non-sexual nature.
- You realize before the experiment is ever performed that the omnibus hypothesis concerning whether any of the movies is different really doesn't matter to you.
- You are more interested in the following questions:
  1. Is the average of the 3 violent movies the same as that of the control movie?
  2. Is the average of the two sexually explicit violent movies different from the third violent movie?
- Let's express these two experimental hypotheses as contrasts and see if they satisfy the definition of *orthogonal contrasts*

Planned Orthogonal Contrasts
An Example

- Suppose you are performing a social psychology experiment examining the effects of violent movies on willingness to be aggressive during an experimental test.
- Subjects are randomly divided into 4 groups. Groups 1, 2, and 3 view violent movies, while Group 4 views a neutral control movie.
- Movies 1 and 2 involve sexually explicit violence, while Movie 3 depicts violence of a non-sexual nature.
- You realize before the experiment is ever performed that the omnibus hypothesis concerning whether any of the movies is different really doesn't matter to you.
- You are more interested in the following questions:
  1. Is the average of the 3 violent movies the same as that of the control movie?
  2. Is the average of the two sexually explicit violent movies different from the third violent movie?
- Let's express these two experimental hypotheses as contrasts and see if they satisfy the definition of *orthogonal contrasts*

Planned Orthogonal Contrasts
An Example

- Suppose you are performing a social psychology experiment examining the effects of violent movies on willingness to be aggressive during an experimental test.
- Subjects are randomly divided into 4 groups. Groups 1, 2, and 3 view violent movies, while Group 4 views a neutral control movie.
- Movies 1 and 2 involve sexually explicit violence, while Movie 3 depicts violence of a non-sexual nature.
- You realize before the experiment is ever performed that the omnibus hypothesis concerning whether any of the movies is different really doesn't matter to you.
- You are more interested in the following questions:
  1. Is the average of the 3 violent movies the same as that of the control movie?
  2. Is the average of the two sexually explicit violent movies different from the third violent movie?
- Let's express these two experimental hypotheses as contrasts and see if they satisfy the definition of *orthogonal contrasts*

Planned Orthogonal Contrasts

An Example

- Suppose you are performing a social psychology experiment examining the effects of violent movies on willingness to be aggressive during an experimental test.
- Subjects are randomly divided into 4 groups. Groups 1, 2, and 3 view violent movies, while Group 4 views a neutral control movie.
- Movies 1 and 2 involve sexually explicit violence, while Movie 3 depicts violence of a non-sexual nature.
- You realize before the experiment is ever performed that the omnibus hypothesis concerning whether any of the movies is different really doesn't matter to you.
- You are more interested in the following questions:
  1. Is the average of the 3 violent movies the same as that of the control movie?
  2. Is the average of the two sexually explicit violent movies different from the third violent movie?
- Let's express these two experimental hypotheses as contrasts and see if they satisfy the definition of *orthogonal contrasts*

Planned Orthogonal Contrasts
An Example

- Suppose you are performing a social psychology experiment examining the effects of violent movies on willingness to be aggressive during an experimental test.
- Subjects are randomly divided into 4 groups. Groups 1, 2, and 3 view violent movies, while Group 4 views a neutral control movie.
- Movies 1 and 2 involve sexually explicit violence, while Movie 3 depicts violence of a non-sexual nature.
- You realize before the experiment is ever performed that the omnibus hypothesis concerning whether any of the movies is different really doesn't matter to you.
- You are more interested in the following questions:
  1. Is the average of the 3 violent movies the same as that of the control movie?
  2. Is the average of the two sexually explicit violent movies different from the third violent movie?
- Let's express these two experimental hypotheses as contrasts and see if they satisfy the definition of *orthogonal contrasts*

Planned Orthogonal Contrasts
An Example

- Suppose you are performing a social psychology experiment examining the effects of violent movies on willingness to be aggressive during an experimental test.
- Subjects are randomly divided into 4 groups. Groups 1, 2, and 3 view violent movies, while Group 4 views a neutral control movie.
- Movies 1 and 2 involve sexually explicit violence, while Movie 3 depicts violence of a non-sexual nature.
- You realize before the experiment is ever performed that the omnibus hypothesis concerning whether any of the movies is different really doesn't matter to you.
- You are more interested in the following questions:
  1. Is the average of the 3 violent movies the same as that of the control movie?
  2. Is the average of the two sexually explicit violent movies different from the third violent movie?
- Let's express these two experimental hypotheses as contrasts and see if they satisfy the definition of *orthogonal contrasts*

Planned Orthogonal Contrasts
An Example

- Our first hypothesis asks: Is the average of the 3 violent movies the same as that of the control movie?
- This might be written as

$$\frac{1}{3}(\mu_1 + \mu_2 + \mu_3) = \mu_4 \tag{4}$$

  or, equivalently
- This might be written as

$$\Psi_1 = \frac{1}{3}\mu_1 + \frac{1}{3}\mu_2 + \frac{1}{3}\mu_3 - \mu_4 = 0 \tag{5}$$

- Is the hypothesis a contrast?
- Yes, it is, because the linear weights are $1/3, 1/3, 1/3$, and -1, and sum to zero.

Planned Orthogonal Contrasts

An Example

- Our first hypothesis asks: Is the average of the 3 violent movies the same as that of the control movie?
- This might be written as

$$\frac{1}{3}(\mu_1 + \mu_2 + \mu_3) = \mu_4 \tag{4}$$

or, equivalently

- This might be written as

$$\Psi_1 = \frac{1}{3}\mu_1 + \frac{1}{3}\mu_2 + \frac{1}{3}\mu_3 - \mu_4 = 0 \tag{5}$$

- Is the hypothesis a contrast?
- Yes, it is, because the linear weights are $1/3, 1/3, 1/3$, and -1, and sum to zero.

Planned Orthogonal Contrasts

An Example

- Our first hypothesis asks: Is the average of the 3 violent movies the same as that of the control movie?
- This might be written as

$$\frac{1}{3}(\mu_1 + \mu_2 + \mu_3) = \mu_4 \tag{4}$$

or, equivalently

- This might be written as

$$\Psi_1 = \frac{1}{3}\mu_1 + \frac{1}{3}\mu_2 + \frac{1}{3}\mu_3 - \mu_4 = 0 \tag{5}$$

- Is the hypothesis a contrast?
- Yes, it is, because the linear weights are $1/3, 1/3, 1/3$, and -1, and sum to zero.

Planned Orthogonal Contrasts
An Example

- Our first hypothesis asks: Is the average of the 3 violent movies the same as that of the control movie?
- This might be written as

$$\frac{1}{3}(\mu_1 + \mu_2 + \mu_3) = \mu_4 \tag{4}$$

or, equivalently

- This might be written as

$$\Psi_1 = \frac{1}{3}\mu_1 + \frac{1}{3}\mu_2 + \frac{1}{3}\mu_3 - \mu_4 = 0 \tag{5}$$

- Is the hypothesis a contrast?
- Yes, it is, because the linear weights are $1/3, 1/3, 1/3$, and -1, and sum to zero.

Planned Orthogonal Contrasts
An Example

- Our first hypothesis asks: Is the average of the 3 violent movies the same as that of the control movie?
- This might be written as

$$\frac{1}{3}(\mu_1 + \mu_2 + \mu_3) = \mu_4 \tag{4}$$

or, equivalently
- This might be written as

$$\Psi_1 = \frac{1}{3}\mu_1 + \frac{1}{3}\mu_2 + \frac{1}{3}\mu_3 - \mu_4 = 0 \tag{5}$$

- Is the hypothesis a contrast?
- Yes, it is, because the linear weights are $1/3, 1/3, 1/3$, and -1, and sum to zero.

Planned Orthogonal Contrasts

An Example

- Our second hypothesis asks: Is the average of the two sexually explicit violent movies different from the third violent movie?
- This can be written as

$$\Psi_2 = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \mu_3 = 0 \tag{6}$$

- Is Ψ_2 a contrast?
- Yes, it is, because the linear weights sum to zero.
- But are Ψ_1 and Ψ_2 orthogonal?
- To test whether they are orthogonal, we “line up” the linear weights and see if their sum of cross-products is equal to zero.

Planned Orthogonal Contrasts

An Example

- Our second hypothesis asks: Is the average of the two sexually explicit violent movies different from the third violent movie?
- This can be written as

$$\Psi_2 = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \mu_3 = 0 \tag{6}$$

- Is Ψ_2 a contrast?
- Yes, it is, because the linear weights sum to zero.
- But are Ψ_1 and Ψ_2 orthogonal?
- To test whether they are orthogonal, we "line up" the linear weights and see if their sum of cross-products is equal to zero.

Planned Orthogonal Contrasts

An Example

- Our second hypothesis asks: Is the average of the two sexually explicit violent movies different from the third violent movie?
- This can be written as

$$\Psi_2 = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \mu_3 = 0 \tag{6}$$

- Is Ψ_2 a contrast?
- Yes, it is, because the linear weights sum to zero.
- But are Ψ_1 and Ψ_2 orthogonal?
- To test whether they are orthogonal, we "line up" the linear weights and see if their sum of cross-products is equal to zero.

Planned Orthogonal Contrasts

An Example

- Our second hypothesis asks: Is the average of the two sexually explicit violent movies different from the third violent movie?
- This can be written as

$$\Psi_2 = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \mu_3 = 0 \tag{6}$$

- Is Ψ_2 a contrast?
- Yes, it is, because the linear weights sum to zero.
- But are Ψ_1 and Ψ_2 orthogonal?
- To test whether they are orthogonal, we "line up" the linear weights and see if their sum of cross-products is equal to zero.

Planned Orthogonal Contrasts
An Example

- Our second hypothesis asks: Is the average of the two sexually explicit violent movies different from the third violent movie?
- This can be written as

$$\Psi_2 = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \mu_3 = 0 \tag{6}$$

- Is Ψ_2 a contrast?
- Yes, it is, because the linear weights sum to zero.
- But are Ψ_1 and Ψ_2 orthogonal?
- To test whether they are orthogonal, we "line up" the linear weights and see if their sum of cross-products is equal to zero.

Planned Orthogonal Contrasts

An Example

- Our second hypothesis asks: Is the average of the two sexually explicit violent movies different from the third violent movie?
- This can be written as

$$\Psi_2 = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2 - \mu_3 = 0 \tag{6}$$

- Is Ψ_2 a contrast?
- Yes, it is, because the linear weights sum to zero.
- But are Ψ_1 and Ψ_2 orthogonal?
- To test whether they are orthogonal, we "line up" the linear weights and see if their sum of cross-products is equal to zero.

Planned Orthogonal Contrasts
An Example

- The table summarizes the linear weights for the two contrasts.

Contrast	μ_1	μ_2	μ_3	μ_4
Ψ_1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	-1
Ψ_2	$\frac{1}{2}$	$\frac{1}{2}$	-1	0

- The two contrasts *are* orthogonal, since

$$\begin{aligned}\left(\frac{1}{3}\right)\left(\frac{1}{2}\right)+\left(\frac{1}{3}\right)\left(\frac{1}{2}\right)+\left(\frac{1}{3}\right)(-1)+(-1)(0) &= \frac{1}{6}+\frac{1}{6}-\frac{1}{3}\\ &= 0\end{aligned}$$

Planned Orthogonal Contrasts

An Example

- The table summarizes the linear weights for the two contrasts.

Contrast	μ_1	μ_2	μ_3	μ_4
Ψ_1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	-1
Ψ_2	$\frac{1}{2}$	$\frac{1}{2}$	-1	0

- The two contrasts *are* orthogonal, since

$$\begin{aligned}\left(\frac{1}{3}\right)\left(\frac{1}{2}\right)+\left(\frac{1}{3}\right)\left(\frac{1}{2}\right)+\left(\frac{1}{3}\right)(-1)+(-1)(0) &= \frac{1}{6}+\frac{1}{6}-\frac{1}{3} \\ &= 0\end{aligned}$$

Planned Orthogonal Contrasts

Calculating the Test Statistic

- The test statistic for a Planned Orthogonal Contrast is calculated the same way as the generalized t statistic discussed in the first weeks of the course. That is

$$\Psi = \sum_{j=1}^{a} c_j \mu_j = 0 \tag{7}$$

$$t_{n_\bullet - a} = \frac{\hat{\Psi}}{\sqrt{\hat{\sigma}^2_{\hat{\Psi}}}} = \frac{\sum_{j=1}^{a} c_j \bar{X}_{\bullet j}}{\sqrt{\left(\sum_{j=1}^{a} \frac{c_j^2}{n_j}\right) MS_{S|A}}} \tag{8}$$

- In evaluating significance, if you are in fact testing more than one planned orthogonal contrast, you can control FWER by using either a Bonferroni or a Hochberg procedure.
- The Bonferroni procedure uses a critical value at the $FWER/k$ significance level to control the FWER at the desired level. This critical value can also be used to construct a confidence interval on the linear combination Ψ.
- The Hochberg procedure takes the two-sided p-values from the t-statistic and subjects them to the Hochberg sequential testing method discussed in the lecture notes on *Multiple Hypothesis Tests*.

Planned Orthogonal Contrasts

Calculating the Test Statistic

- The test statistic for a Planned Orthogonal Contrast is calculated the same way as the generalized t statistic discussed in the first weeks of the course. That is

$$\Psi = \sum_{j=1}^{a} c_j \mu_j = 0 \tag{7}$$

$$t_{n_\bullet - a} = \frac{\hat{\Psi}}{\sqrt{\hat{\sigma}^2_{\hat{\Psi}}}} = \frac{\sum_{j=1}^{a} c_j \bar{X}_{\bullet j}}{\sqrt{\left(\sum_{j=1}^{a} \frac{c_j^2}{n_j}\right) MS_{S|A}}} \tag{8}$$

- In evaluating significance, if you are in fact testing more than one planned orthogonal contrast, you can control FWER by using either a Bonferroni or a Hochberg procedure.
- The Bonferroni procedure uses a critical value at the $FWER/k$ significance level to control the FWER at the desired level. This critical value can also be used to construct a confidence interval on the linear combination Ψ.
- The Hochberg procedure takes the two-sided p-values from the t-statistic and subjects them to the Hochberg sequential testing method discussed in the lecture notes on *Multiple Hypothesis Tests*.

Planned Orthogonal Contrasts

Calculating the Test Statistic

- The test statistic for a Planned Orthogonal Contrast is calculated the same way as the generalized t statistic discussed in the first weeks of the course. That is

$$\Psi = \sum_{j=1}^{a} c_j \mu_j = 0 \tag{7}$$

$$t_{n_\bullet - a} = \frac{\hat{\Psi}}{\sqrt{\hat{\sigma}^2_{\hat{\Psi}}}} = \frac{\sum_{j=1}^{a} c_j \bar{X}_{\bullet j}}{\sqrt{\left(\sum_{j=1}^{a} \frac{c_j^2}{n_j}\right) MS_{S|A}}} \tag{8}$$

- In evaluating significance, if you are in fact testing more than one planned orthogonal contrast, you can control FWER by using either a Bonferroni or a Hochberg procedure.
- The Bonferroni procedure uses a critical value at the $FWER/k$ significance level to control the FWER at the desired level. This critical value can also be used to construct a confidence interval on the linear combination Ψ.
- The Hochberg procedure takes the two-sided p-values from the t-statistic and subjects them to the Hochberg sequential testing method discussed in the lecture notes on *Multiple Hypothesis Tests.*

Planned Orthogonal Contrasts

Calculating the Test Statistic

- The test statistic for a Planned Orthogonal Contrast is calculated the same way as the generalized t statistic discussed in the first weeks of the course. That is

$$\Psi = \sum_{j=1}^{a} c_j \mu_j = 0 \tag{7}$$

$$t_{n_\bullet - a} = \frac{\hat{\Psi}}{\sqrt{\hat{\sigma}^2_{\hat{\Psi}}}} = \frac{\sum_{j=1}^{a} c_j \bar{X}_{\bullet j}}{\sqrt{\left(\sum_{j=1}^{a} \frac{c_j^2}{n_j}\right) MS_{S|A}}} \tag{8}$$

- In evaluating significance, if you are in fact testing more than one planned orthogonal contrast, you can control FWER by using either a Bonferroni or a Hochberg procedure.
- The Bonferroni procedure uses a critical value at the $FWER/k$ significance level to control the FWER at the desired level. This critical value can also be used to construct a confidence interval on the linear combination Ψ.
- The Hochberg procedure takes the two-sided p-values from the t-statistic and subjects them to the Hochberg sequential testing method discussed in the lecture notes on *Multiple Hypothesis Tests*.

The Scheffé Test

- The Scheffé test is designed to control FWER at α for *any number* of *post hoc* contrast tests after observing a significant F statistic in the omnibus ANOVA hypothesis test.
- The method also allows simultaneous confidence intervals to be constructed for the entire family of tests.
- The tremendous flexibility and generality of the procedure means that, in order to control FWER at α, it must be rather conservative to provide the desired level of protection.

The Scheffé Test

- The Scheffé test is designed to control FWER at α for *any number* of *post hoc* contrast tests after observing a significant F statistic in the omnibus ANOVA hypothesis test.
- The method also allows simultaneous confidence intervals to be constructed for the entire family of tests.
- The tremendous flexibility and generality of the procedure means that, in order to control FWER at α, it must be rather conservative to provide the desired level of protection.

The Scheffé Test

- The Scheffé test is designed to control FWER at α for *any number* of *post hoc* contrast tests after observing a significant F statistic in the omnibus ANOVA hypothesis test.
- The method also allows simultaneous confidence intervals to be constructed for the entire family of tests.
- The tremendous flexibility and generality of the procedure means that, in order to control FWER at α, it must be rather conservative to provide the desired level of protection.

The Scheffé Test

- One performs the Scheffé procedure exactly the same as the generalized t procedure, either when constructing the t-statistic for a hypothesis test or when constructing a confidence interval on the linear combination Ψ.
- For example, the test statistic for the procedure is calculated using Equation 8. The only difference is that, instead of using a critical value from the t distribution, one instead uses the following critical value.
- Let F^* be the critical value used for the ANOVA F-test, i.e.,

$$F^* = F_{1-\alpha, a-1, n_\bullet - a}$$

- Then the critical value used in the Scheffé test is

$$S = \sqrt{(a-1)F^*} \tag{9}$$

The Scheffé Test

- One performs the Scheffé procedure exactly the same as the generalized t procedure, either when constructing the t-statistic for a hypothesis test or when constructing a confidence interval on the linear combination Ψ.
- For example, the test statistic for the procedure is calculated using Equation 8. The only difference is that, instead of using a critical value from the t distribution, one instead uses the following critical value.
- Let F^* be the critical value used for the ANOVA F-test, i.e.,

$$F^* = F_{1-\alpha, a-1, n_\bullet - a}$$

- Then the critical value used in the Scheffé test is

$$S = \sqrt{(a-1)F^*} \tag{9}$$

The Scheffé Test

- One performs the Scheffé procedure exactly the same as the generalized t procedure, either when constructing the t-statistic for a hypothesis test or when constructing a confidence interval on the linear combination Ψ.
- For example, the test statistic for the procedure is calculated using Equation 8. The only difference is that, instead of using a critical value from the t distribution, one instead uses the following critical value.
- Let F^* be the critical value used for the ANOVA F-test, i.e.,

$$F^* = F_{1-\alpha, a-1, n_\bullet - a}$$

- Then the critical value used in the Scheffé test is

$$S = \sqrt{(a-1)F^*} \tag{9}$$

The Scheffé Test

- One performs the Scheffé procedure exactly the same as the generalized t procedure, either when constructing the t-statistic for a hypothesis test or when constructing a confidence interval on the linear combination Ψ.
- For example, the test statistic for the procedure is calculated using Equation 8. The only difference is that, instead of using a critical value from the t distribution, one instead uses the following critical value.
- Let F^* be the critical value used for the ANOVA F-test, i.e.,

$$F^* = F_{1-\alpha, a-1, n_\bullet - a}$$

- Then the critical value used in the Scheffé test is

$$S = \sqrt{(a-1)F^*} \tag{9}$$

The Tukey Test

- The Tukey procedure allows one to conduct all possible pairwise comparisons between pairs of means, after looking at the data, while controlling FWER at α.
- One reason for the great popularity of the method is its simplicity.
- To perform the Tukey test, one calculates a single value called the Honestly Significant Difference (HSD). If any two means are farther apart than HSD, they are declared statistically significant.

The Tukey Test

- The Tukey procedure allows one to conduct all possible pairwise comparisons between pairs of means, after looking at the data, while controlling FWER at α.
- One reason for the great popularity of the method is its simplicity.
- To perform the Tukey test, one calculates a single value called the Honestly Significant Difference (HSD). If any two means are farther apart than HSD, they are declared statistically significant.

The Tukey Test

- The Tukey procedure allows one to conduct all possible pairwise comparisons between pairs of means, after looking at the data, while controlling FWER at α.
- One reason for the great popularity of the method is its simplicity.
- To perform the Tukey test, one calculates a single value called the Honestly Significant Difference (HSD). If any two means are farther apart than HSD, they are declared statistically significant.

The Tukey Test

- To calculate the HSD, one needs a critical value q^* from the *Studentized Range Distribution*.
- The critical value q^* can be calulated in R using the function `qtukey`, and is

$$q^* = q_{1-\alpha,a,n_\bullet - a} \tag{10}$$

- The HSD value is calculated as

$$HSD = q^* \sqrt{\frac{MS_{S|A}}{n}} \tag{11}$$

where n is the sample size per group.

Simultaneous Confidence Intervals from the Tukey Test

- To construct confidence intervals on any pairwise mean difference $\mu_i - \mu_j$, simply use the HSD as follows.

$$\overline{X}_{\bullet i} - \overline{X}_{\bullet j} \pm HSD \tag{12}$$

Displaying the Results of a Tukey Test

- The results of a set of Tukey tests may displayed in a variety of ways.
- Some popular methods are the *line plot* and the *letter plot*, along with various tabular presentations

Some Artificial Data

- You can load a small artificial data set with $a = 4$ groups with $n = 3$ per group from the course website.

```
> data <- read.csv(
+   "http://www.statpower.net/Content/311/Lecture Notes/TukeyData.csv")
> data
   group   x
1      1 1.0
2      1 2.0
3      1 3.0
4      2 4.0
5      2 5.0
6      2 6.0
7      3 7.0
8      3 8.0
9      3 9.0
10     4 7.5
11     4 8.5
12     4 9.5
```

Some Artificial Data

- We can load the code for the generalized t statistic as follows:

```
> source(
+   "http://www.statpower.net/Content/311/Handout/GT2/GeneralizedTCode.r"
+       )
```

- The header of the function shows the form for inputting the data.

```
> GeneralizedT<-function(means,sds,ns,wts,k0=0,CI=FALSE,conf=0.95,null=0){}
> # means => a vector of group means
> #          example.. means <- c(1.12,3.51)
> # sds => a vector of corresponding standard deviations
> # ns  => a vector of sample sizes
> # wts => a vector of linear weights to be applied
> # k0  => the constant that the linear combination is, by hypothesis, equal to
> #        default value is 0
> # CI  => set this equal to TRUE if you want a confidence interval
> # conf => the confidence level, by default 0.95 for a 95% interval
> # null => an indicator as to where the null hypothesis region is relative to k0
> #         0 indicates equal to k0, i.e., a 2-sided test
> #         -1 indicates that the null hypothesis is of the form H0: kappa <= k0
> #         1 indicates that the null hypothesis is of the form H0: kappa >= k0
> # NOTE!  Entering null incorrectly will result in the
> #        p-value being reported incorrectly!
```

- Note that, to operate, the function needs vectors of means, sds, ns, and linear weights for the 4 groups.

Calculating Summary Statistics

- Here is a simple function to calculate summary statistics:

```
> summary.stats <- function (x,group) {
+   means <- tapply(x, group, mean, na.rm=TRUE)
+   sds  <- tapply(x,group, sd, na.rm = TRUE)
+   valid <- function (x) {return(sum(!is.na(x)) )}
+   ns <- tapply(x,group, valid )
+   output <- list(means = means, sds=sds, ns=ns)
+   return(output)
+   }
> output <- summary.stats(data$x,data$group)
> means <- output$means
> sds <- output$sds
> ns <- output$ns
> means

  1   2   3   4
2.0 5.0 8.0 8.5

> sds

1 2 3 4
1 1 1 1

> ns

1 2 3 4
3 3 3 3
```

Planned Contrasts

- We recall from a previous slide that we have two planned orthogonal contrasts.
- Let's compute the t statistics. The contrasts were defined by

```
> wts.1 <- c(1/3,1/3,1/3,-1)
> wts.2 <- c(1/2,1/2,-1,0)
```

- The t values are

```
> t.1 <- GeneralizedT(means,sds,ns,wts.1)
> t.2 <- GeneralizedT(means,sds,ns,wts.2)
> t.1
[1] -5.2500000000  8.0000000000  0.0007738347
> t.2
[1] -6.3639610307  8.0000000000  0.0002173371
```

- With a Bonferroni correction, and FWER rate of 0.05, each test is performed at the 0.025 significance levels. Both t statistics have p-values way below this significant level, so both hypotheses are rejected.

The Scheffé Test

- Imagine now that the two contrast hypotheses that we tested in the preceding section were actually only thought of after the experimenter examined the data.
- In this case, the Scheffé test procedure and will control $FWER$ at the α level.
- We need to compute the S critical value. Since there are 4 groups with $n = 3$ observations per group, our degrees of freedom for the F-test are 3 and 8.
- The Scheffé critical value may then be computed as

```
> a <- 4
> n <- 3
> df1 <- a - 1
> df2 <- a * (n - 1)
> F.crit <- qf(0.95, df1, df2)
> S <- sqrt((a-1)*F.crit)
> S

[1] 3.492641
```

- Both t statistics exceed this critical value by a wide margin, and so these orthogonal contrast hypotheses can be rejected even when performed *post hoc*.

The Tukey Test

- Suppose we wished to perform all possible pairwise comparisons among the 4 means.
- There are a number of ways to do this in R.
- The first approach uses the `HST.test` in the `agricolae` library to perform the calculations.

```
> library(agricolae)
> data$group <- factor(data$group)
> fit <- aov(x ~ group, data=data)
> HSD.test(fit,"group",group=TRUE)
Study:

HSD Test for x

Mean Square Error:  1

group,  means

    x   std.err r Min. Max.
1 2.0 0.5773503 3  1.0  3.0
2 5.0 0.5773503 3  4.0  6.0
3 8.0 0.5773503 3  7.0  9.0
4 8.5 0.5773503 3  7.5  9.5

alpha: 0.05 ; Df Error: 8
Critical Value of Studentized Range: 4.52881

Honestly Significant Difference: 2.614709

Means with the same letter are not significantly different.

Groups, Treatments and means
a          4          8.5
a          3          8
b          2          5
c          1          2
```

The Tukey Test

- Note that besides containing the HSD, the table displays the results of the Tukey test on the ordered means by means of a letter plot on the ordered groups.
- The two groups with the largest means, groups 4 and 3, are not significantly different, and they both have the letter "a" next to them.
- On the other hand, groups 2 and 1 have different letters. This indicates that these groups are significantly different from groups 4 and 3, and significantly different from each other.
- The *HSD* value given in the output agrees with our calculation in R.

```
> MS.S.A <- 1.0
> a <- 4
> n <- 3
> df1 <- a
> df2 <- a * (n - 1)
> HSD <- qtukey(0.95,df1,df2) * sqrt(MS.S.A / n)
> HSD
[1] 2.614709
```

The Tukey Test

- An alternative approach, called the *line plot*, draws lines either under or alongside the list of means, with each solid line corresponding to a letter in the letter plot.
- That approach was demonstrated in class.

The Tukey Test

- Still another approach uses the `TukeyHSD` function, which displays the mean differences pairwise, along with a confidence interval on the mean difference and an "adjusted p-value," which is less than 0.05 if the result is significant with $FWER$ set at 0.05.

```
> TukeyHSD(fit)

  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = x ~ group, data = data)

$group
    diff        lwr      upr     p adj
2-1  3.0  0.3852905 5.614709 0.0259193
3-1  6.0  3.3852905 8.614709 0.0003667
4-1  6.5  3.8852905 9.114709 0.0002084
3-2  3.0  0.3852905 5.614709 0.0259193
4-2  3.5  0.8852905 6.114709 0.0113928
4-3  0.5 -2.1147095 3.114709 0.9252929
```