Non-orthogonal Designs

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

P311, 2011
Non-orthogonal Designs

1. Introduction

2. A Non-Orthogonal $2 \times 2$ ANOVA
   - Assessing Gender Discrimination
   - Another Look at the Gender Effect

3. ANOVA Computations in R
   - Type III Sums of Squares

4. Which Method to Use?
   - Introduction
   - Comparisons of the Methods
   - Selection of a Method
Non-Orthogonal Designs
An Introduction

- So far, we’ve been examining 1-Way and 2-Way randomized designs in which the sample sizes are equal in each “cell” of the design. Such designs are said to be “orthogonal.”
- We’ve digressed to examine issues of robustness to violations of the normality and homogeneity of variances assumption.
- In this module, we tackle the problem of non-orthogonal designs — designs in which the “contrasts” used to test for main effects and interactions are no-longer uncorrelated because of unequal n’s in the different cells.
- Non-orthogonality poses some interesting (and still controversial) issues for analysis and interpretation.
- The following brief example will highlight some of the key issues, both conceptual and statistical.
Non-Orthogonal Designs
An Introduction

- So far, we’ve been examining 1-Way and 2-Way randomized designs in which the sample sizes are equal in each “cell” of the design. Such designs are said to be “orthogonal.”
- We’ve digressed to examine issues of robustness to violations of the normality and homogeneity of variances assumption.
- In this module, we tackle the problem of non-orthogonal designs — designs in which the “contrasts” used to test for main effects and interactions are no-longer uncorrelated because of unequal n’s in the different cells.
- Non-orthogonality poses some interesting (and still controversial) issues for analysis and interpretation.
- The following brief example will highlight some of the key issues, both conceptual and statistical.
So far, we’ve been examining 1-Way and 2-Way randomized designs in which the sample sizes are equal in each “cell” of the design. Such designs are said to be “orthogonal.”

We’ve digressed to examine issues of robustness to violations of the normality and homogeneity of variances assumption.

In this module, we tackle the problem of non-orthogonal designs — designs in which the “contrasts” used to test for main effects and interactions are no-longer uncorrelated because of unequal $n$’s in the different cells.

Non-orthogonality poses some interesting (and still controversial) issues for analysis and interpretation.

The following brief example will highlight some of the key issues, both conceptual and statistical.
Non-Orthogonal Designs
An Introduction

- So far, we’ve been examining 1-Way and 2-Way randomized designs in which the sample sizes are equal in each “cell” of the design. Such designs are said to be “orthogonal.”
- We’ve digressed to examine issues of robustness to violations of the normality and homogeneity of variances assumption.
- In this module, we tackle the problem of non-orthogonal designs — designs in which the “contrasts” used to test for main effects and interactions are no-longer uncorrelated because of unequal $n$’s in the different cells.
- Non-orthogonality poses some interesting (and still controversial) issues for analysis and interpretation.
- The following brief example will highlight some of the key issues, both conceptual and statistical.
So far, we’ve been examining 1-Way and 2-Way randomized designs in which the sample sizes are equal in each “cell” of the design. Such designs are said to be “orthogonal.”

We’ve digressed to examine issues of robustness to violations of the normality and homogeneity of variances assumption.

In this module, we tackle the problem of non-orthogonal designs — designs in which the “contrasts” used to test for main effects and interactions are no-longer uncorrelated because of unequal $n$’s in the different cells.

Non-orthogonality poses some interesting (and still controversial) issues for analysis and interpretation.

The following brief example will highlight some of the key issues, both conceptual and statistical.
A Non-Orthogonal $2 \times 2$ ANOVA

- All the $F$ tests for main effects and interactions in a $2 \times 2$ ANOVA can be accomplished as $t - tests$.
- Because of the simplicity of the design, we can see some important points that generalize to 2-Way ANOVA’s with more than 2 levels per factor.
- Scott Maxwell, in Chapter 7 of his classic textbook *Designing Experiments and Analyzing Data*, gives the following interesting example.
All the $F$ tests for main effects and interactions in a $2 \times 2$ ANOVA can be accomplished as $t - tests$.

Because of the simplicity of the design, we can see some important points that generalize to 2-Way ANOVA’s with more than 2 levels per factor.

Scott Maxwell, in Chapter 7 of his classic textbook *Designing Experiments and Analyzing Data*, gives the following interesting example.
A Non-Orthogonal 2 × 2 ANOVA

- All the \( F \) tests for main effects and interactions in a 2 × 2 ANOVA can be accomplished as \( t - \) tests.
- Because of the simplicity of the design, we can see some important points that generalize to 2-Way ANOVA’s with more than 2 levels per factor.
- Scott Maxwell, in Chapter 7 of his classic textbook *Designing Experiments and Analyzing Data*, gives the following interesting example.
Suppose that our data consists of 22 female and male employees of a large company. They are further divided into those with a college degree, and those without a degree.

The dependent variable is the employee’s annual salary, in thousands of dollars.

The raw data are shown below, and have been provided in the file Salary.csv.

### TABLE 7.15
**HYPOTHETICAL SALARY DATA (IN THOUSANDS) FOR FEMALE AND MALE EMPLOYEES**

<table>
<thead>
<tr>
<th></th>
<th>Females</th>
<th></th>
<th>Males</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>College Degree</td>
<td>No College Degree</td>
<td>College Degree</td>
<td>No College Degree</td>
</tr>
<tr>
<td>24</td>
<td>15</td>
<td>25</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>17</td>
<td>29</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>20</td>
<td>27</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>16</td>
<td></td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>25</td>
<td>17</td>
<td>27</td>
<td>20</td>
</tr>
</tbody>
</table>
Suppose that our data consists of 22 female and male employees of a large company. They are further divided into those with a college degree, and those without a degree.

The dependent variable is the employee’s annual salary, in thousands of dollars.

The raw data are shown below, and have been provided in the file Salary.csv.

**TABLE 7.15**

<table>
<thead>
<tr>
<th></th>
<th>Females</th>
<th>Males</th>
</tr>
</thead>
<tbody>
<tr>
<td>College Degree</td>
<td>No College Degree</td>
<td>College Degree</td>
</tr>
<tr>
<td>24</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td>26</td>
<td>17</td>
<td>29</td>
</tr>
<tr>
<td>25</td>
<td>20</td>
<td>27</td>
</tr>
<tr>
<td>24</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>27</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>25</td>
<td>17</td>
</tr>
</tbody>
</table>
Suppose that our data consists of 22 female and male employees of a large company. They are further divided into those with a college degree, and those without a degree.

The dependent variable is the employee’s annual salary, in thousands of dollars.

The raw data are shown below, and have been provided in the file `Salary.csv`.

<table>
<thead>
<tr>
<th>College Degree</th>
<th>No College Degree</th>
<th>College Degree</th>
<th>No College Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>15</td>
<td>25</td>
<td>19</td>
</tr>
<tr>
<td>26</td>
<td>17</td>
<td>29</td>
<td>18</td>
</tr>
<tr>
<td>25</td>
<td>20</td>
<td>27</td>
<td>21</td>
</tr>
<tr>
<td>24</td>
<td>16</td>
<td></td>
<td>20</td>
</tr>
<tr>
<td>27</td>
<td></td>
<td>21</td>
<td>24</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td>27</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
<td>19</td>
</tr>
</tbody>
</table>

Mean: 25, 17, 27, 20

**TABLE 7.15**
**HYPOTHETICAL SALARY DATA (IN THOUSANDS) FOR FEMALE AND MALE EMPLOYEES**
> salary <- read.csv("Salary.csv")
> salary

<table>
<thead>
<tr>
<th>Gender</th>
<th>Degree</th>
<th>Salary</th>
<th>Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>College</td>
<td>24</td>
<td>FemaleCollege</td>
</tr>
<tr>
<td>Female</td>
<td>College</td>
<td>26</td>
<td>FemaleCollege</td>
</tr>
<tr>
<td>Female</td>
<td>College</td>
<td>25</td>
<td>FemaleCollege</td>
</tr>
<tr>
<td>Female</td>
<td>College</td>
<td>24</td>
<td>FemaleCollege</td>
</tr>
<tr>
<td>Female</td>
<td>College</td>
<td>27</td>
<td>FemaleCollege</td>
</tr>
<tr>
<td>Female</td>
<td>College</td>
<td>24</td>
<td>FemaleCollege</td>
</tr>
<tr>
<td>Female</td>
<td>College</td>
<td>27</td>
<td>FemaleCollege</td>
</tr>
<tr>
<td>Female</td>
<td>College</td>
<td>23</td>
<td>FemaleCollege</td>
</tr>
<tr>
<td>Female No College</td>
<td></td>
<td>15</td>
<td>FemaleNoCollege</td>
</tr>
<tr>
<td>Female No College</td>
<td></td>
<td>17</td>
<td>FemaleNoCollege</td>
</tr>
<tr>
<td>Female No College</td>
<td></td>
<td>20</td>
<td>FemaleNoCollege</td>
</tr>
<tr>
<td>Female No College</td>
<td></td>
<td>16</td>
<td>FemaleNoCollege</td>
</tr>
<tr>
<td>Male</td>
<td>College</td>
<td>25</td>
<td>MaleCollege</td>
</tr>
<tr>
<td>Male</td>
<td>College</td>
<td>29</td>
<td>MaleCollege</td>
</tr>
<tr>
<td>Male</td>
<td>College</td>
<td>27</td>
<td>MaleCollege</td>
</tr>
<tr>
<td>Male No College</td>
<td></td>
<td>19</td>
<td>MaleNoCollege</td>
</tr>
<tr>
<td>Male No College</td>
<td></td>
<td>18</td>
<td>MaleNoCollege</td>
</tr>
<tr>
<td>Male No College</td>
<td></td>
<td>21</td>
<td>MaleNoCollege</td>
</tr>
<tr>
<td>Male No College</td>
<td></td>
<td>20</td>
<td>MaleNoCollege</td>
</tr>
<tr>
<td>Male No College</td>
<td></td>
<td>21</td>
<td>MaleNoCollege</td>
</tr>
<tr>
<td>Male No College</td>
<td></td>
<td>22</td>
<td>MaleNoCollege</td>
</tr>
<tr>
<td>Male No College</td>
<td></td>
<td>19</td>
<td>MaleNoCollege</td>
</tr>
</tbody>
</table>

> attach(salary)
Here is a table of cell means for the data.

<table>
<thead>
<tr>
<th>Gender</th>
<th>College</th>
<th>No College</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\bar{Y}_{11} = 25$</td>
<td>$\bar{Y}_{12} = 17$</td>
</tr>
<tr>
<td></td>
<td>$n_{11} = 8$</td>
<td>$n_{12} = 4$</td>
</tr>
<tr>
<td></td>
<td>$s_{11}^2 = 2.285714$</td>
<td>$s_{12}^2 = 4.666667$</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\bar{Y}_{21} = 27$</td>
<td>$\bar{Y}_{22} = 20$</td>
</tr>
<tr>
<td></td>
<td>$n_{21} = 3$</td>
<td>$n_{22} = 7$</td>
</tr>
<tr>
<td></td>
<td>$s_{21}^2 = 4$</td>
<td>$s_{22}^2 = 2$</td>
</tr>
</tbody>
</table>
Assessing Gender Discrimination

- Suppose that these data were gathered in an attempt to assess at this company, the following questions:
  1. Are men paid more than women?
  2. Are people with college degrees paid more than people without degrees?

- It turns out that there are different ways of viewing the data that lead to different answers to these questions.

- Let’s see why.
Assessing Gender Discrimination

Suppose that these data were gathered in an attempt to assess at this company, the following questions:

1. Are men paid more than women?
2. Are people with college degrees paid more than people without degrees?

It turns out that there are different ways of viewing the data that lead to different answers to these questions.

Let’s see why.
Suppose that these data were gathered in an attempt to assess at this company, the following questions:

1. Are men are paid more than women?
2. Are people with college degrees paid more than people without degrees?

It turns out that there are different ways of viewing the data that lead to different answers to these questions.

Let’s see why.
Assessing Gender Discrimination

Suppose that these data were gathered in an attempt to assess at this company, the following questions:

1. Are men paid more than women?
2. Are people with college degrees paid more than people without degrees?

It turns out that there are different ways of viewing the data that lead to different answers to these questions.

Let’s see why.
Assessing Gender Discrimination

Suppose that these data were gathered in an attempt to assess at this company, the following questions:

1. Are men paid more than women?
2. Are people with college degrees paid more than people without degrees?

It turns out that there are different ways of viewing the data that lead to different answers to these questions.

Let’s see why.
Suppose we look at the table of cell means and try to roughly assess the first question by examining the main effect for *Gender*.

- We compare the row means for *Gender = Male* and *Gender = Female*, averaging across college degree status.
- The average of the two *Female* means is \((25 + 17)/2 = 21\).
- The average of the two *Male* means is \((27 + 20)/2 = 23.5\).
- These row means are called the *unweighted row means*, because they are computed from the cell means without weighting them by the sample sizes for the cells.

<table>
<thead>
<tr>
<th>Degree</th>
<th>College</th>
<th>No College</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>(\bar{Y}_{11} = 25)</td>
<td>(\bar{Y}_{12} = 17)</td>
</tr>
<tr>
<td></td>
<td>(n_{11} = 8)</td>
<td>(n_{12} = 4)</td>
</tr>
<tr>
<td>Male</td>
<td>(\bar{Y}_{21} = 27)</td>
<td>(\bar{Y}_{22} = 20)</td>
</tr>
<tr>
<td></td>
<td>(n_{21} = 3)</td>
<td>(n_{22} = 7)</td>
</tr>
</tbody>
</table>
Suppose we look at the table of cell means and try to roughly assess the first question by examining the main effect for Gender.

We compare the row means for Gender = Male and Gender = Female, averaging across college degree status.

The average of the two Female means is \((25 + 17)/2 = 21\).

The average of the two Male means is \((27 + 20)/2 = 23.5\).

These row means are called the *unweighted row means*, because they are computed from the cell means without weighting them by the sample sizes for the cells.

---

<table>
<thead>
<tr>
<th>Gender</th>
<th>Degree</th>
<th>College</th>
<th>No College</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>(\bar{Y}_{11} = 25)</td>
<td>(n_{11} = 8)</td>
<td>(\bar{Y}_{12} = 17)</td>
</tr>
<tr>
<td>Male</td>
<td>(\bar{Y}_{21} = 27)</td>
<td>(n_{21} = 3)</td>
<td>(\bar{Y}_{22} = 20)</td>
</tr>
</tbody>
</table>

\[
\frac{25 + 17}{2} = 21
\]

\[
\frac{27 + 20}{2} = 23.5
\]
Assessing Gender Discrimination

- Suppose we look at the table of cell means and try to roughly assess the first question by examining the main effect for *Gender*.
- We compare the row means for *Gender* = *Male* and *Gender* = *Female*, averaging across college degree status.
- The average of the two *Female* means is \( \frac{25 + 17}{2} = 21 \).
- The average of the two *Male* means is \( \frac{27 + 20}{2} = 23.5 \).
- These row means are called the *unweighted row means*, because they are computed from the cell means without weighting them by the sample sizes for the cells.
Suppose we look at the table of cell means and try to roughly assess the first question by examining the main effect for Gender.

We compare the row means for Gender = Male and Gender = Female, averaging across college degree status.

The average of the two Female means is \( \frac{25 + 17}{2} = 21 \).

The average of the two Male means is \( \frac{27 + 20}{2} = 23.5 \).

These row means are called the unweighted row means, because they are computed from the cell means without weighting them by the sample sizes for the cells.
So the *Female − Male* difference is $21 − 23.5 = −2.5$. The average of the 4 overall cell means is $(25 + 17 + 27 + 20)/4 = 22.25$. In the classic ANOVA model, the main effect of being *Female* is $−1.25$ relative to the overall mean, while the main effect of being *Male* is $+1.25$ relative to the overall mean. The difference in the effects is $−2.5$.

This suggests that, in general, being *Female* is associated with a $\$2500$ negative salary differential.
So the *Female* – *Male* difference is $21 - 23.5 = -2.5$. The average of the 4 overall cell means is $(25 + 17 + 27 + 20)/4 = 22.25$. In the classic ANOVA model, the main effect of being *Female* is $-1.25$ relative to the overall mean, while the main effect of being *Male* is $+1.25$ relative to the overall mean. The difference in the effects is $-2.5$.

This suggests that, in general, being *Female* is associated with a $\$2500$ negative salary differential.
Assessing Gender Discrimination

- This agrees with well-established theory.
- Each of the sample cell means is an unbiased estimator of the corresponding population cell mean, and we learned in Psychology 310 that we can test the main effect of Gender by testing the null hypothesis

\[ H_0 : \frac{\mu_{11} + \mu_{12}}{2} = \frac{\mu_{21} + \mu_{22}}{2} \]  

or, equivalently

\[ H_0 : \psi_{Gender} = \mu_{11} + \mu_{12} - \mu_{21} - \mu_{22} = 0 \]  

- An unbiased estimator of the quantity of interest is

\[ \hat{\psi}_{Gender} = \bar{Y}_{11} + \bar{Y}_{12} - \bar{Y}_{21} - \bar{Y}_{22} \]
Assessing Gender Discrimination

- This agrees with well-established theory.
- Each of the sample cell means is an unbiased estimator of the corresponding population cell mean, and we learned in Psychology 310 that we can test the main effect of Gender by testing the null hypothesis

\[
H_0 : \frac{\mu_{11} + \mu_{12}}{2} = \frac{\mu_{21} + \mu_{22}}{2} \quad (1)
\]

or, equivalently

\[
H_0 : \psi_{Gender} = \mu_{11} + \mu_{12} - \mu_{21} - \mu_{22} = 0 \quad (2)
\]

- An unbiased estimator of the quantity of interest is

\[
\hat{\psi}_{Gender} = \bar{Y}_{11} + \bar{Y}_{12} - \bar{Y}_{21} - \bar{Y}_{22} \quad (3)
\]
This agrees with well-established theory.

Each of the sample cell means is an unbiased estimator of the corresponding population cell mean, and we learned in Psychology 310 that we can test the main effect of Gender by testing the null hypothesis

\[
H_0 : \frac{\mu_{11} + \mu_{12}}{2} = \frac{\mu_{21} + \mu_{22}}{2}
\]

or, equivalently

\[
H_0 : \psi_{\text{Gender}} = \mu_{11} + \mu_{12} - \mu_{21} - \mu_{22} = 0
\]

An unbiased estimator of the quantity of interest is

\[
\hat{\psi}_{\text{Gender}} = \bar{Y}_{11} + \bar{Y}_{12} - \bar{Y}_{21} - \bar{Y}_{22}
\]
Assessing Gender Discrimination

- The standard $t$ statistic for testing the null hypothesis of Equation 3 is

\[
t = \frac{\hat{\psi}}{\sqrt{(1/n_{11} + 1/n_{12} + 1/n_{21} + 1/n_{22})\hat{\sigma}^2}}
\]

(4)

where $\hat{\sigma}^2$, called $MS_{error}$, $MS_{residuals}$, or $MS_{S|cells}$ in various ANOVA texts, is given (for $a$ levels of factor $A$ and $b$ levels of factor $B$) by

\[
\hat{\sigma}^2 = \frac{\sum(n_{ij} - 1)s_{ij}^2}{\sum n_{ij} - ab}
\]

(5)

- It is well known that, if you square a $t$ statistic, you get an $F$ statistic with 1 numerator degree of freedom, and denominator degrees of freedom equal to those of the $t$ statistic.
The standard $t$ statistic for testing the null hypothesis of Equation 3 is

$$t = \frac{\hat{\psi}}{\sqrt{(1/n_{11} + 1/n_{12} + 1/n_{21} + 1/n_{22})\hat{\sigma}^2}}$$

(4)

where $\hat{\sigma}^2$, called $MS_{error}$, $MS_{residuals}$, or $MS_{S|cells}$ in various ANOVA texts, is given (for $a$ levels of factor $A$ and $b$ levels of factor $B$) by

$$\hat{\sigma}^2 = \frac{\sum(n_{ij} - 1)s_{ij}^2}{\sum n_{ij} - ab}$$

(5)

It is well known that, if you square a $t$ statistic, you get an $F$ statistic with 1 numerator degree of freedom, and denominator degrees of freedom equal to those of the $t$ statistic.
Squaring our \( t \) statistic, and inverting part of the denominator and moving it to the numerator, we get

\[
F = \frac{\tilde{n} \hat{\psi}^2}{\hat{\sigma}^2}
\]  

(6)

where \( \tilde{n} \) is the **harmonic mean** of the \( n_{ij} \), given by

\[
\tilde{n} = \frac{1}{\sum_{ij} \frac{1}{n_{ij}}}
\]  

(7)

We can show that \( \hat{\psi}^2 \) is equal to \( b \) times the variance of the **unweighted** row means shown in our previous calculations, and so we may write

\[
F_{gender} = \frac{b\tilde{n}s^2_{Y_{ij\bullet}}}{\hat{\sigma}^2} = \frac{s^2_{Y_{ij\bullet}}}{\hat{\sigma}^2/b\tilde{n}}
\]  

(8)

So, once again, we see that the main effect for factor A is assessed by comparing the sample variance of the (unweighted) row means with an estimate of the variance of the row means, given by an estimate of \( \sigma^2 \) divided the “effective \( n \)” for the row means.
Squaring our $t$ statistic, and inverting part of the denominator and moving it to the numerator, we get

$$F = \frac{\tilde{n} \hat{\psi}^2}{\hat{\sigma}^2} \quad (6)$$

where $\tilde{n}$ is the harmonic mean of the $n_{ij}$, given by

$$\tilde{n} = \frac{1}{\sum_{ij} \frac{1}{n_{ij}}} \quad (7)$$

We can show that $\hat{\psi}^2$ is equal to $b$ times the variance of the unweighted row means shown in our previous calculations, and so we may write

$$F_{gender} = \frac{b\tilde{n}s_{\tilde{Y}_{j\cdot}}^2}{\hat{\sigma}^2} = \frac{s_{\tilde{Y}_{j\cdot}}^2}{\hat{\sigma}^2/b\tilde{n}} \quad (8)$$

So, once again, we see that the main effect for factor A is assessed by comparing the sample variance of the (unweighted) row means with an estimate of the variance of the row means, given by an estimate of $\sigma^2$ divided the “effective n” for the row means.
Assessing Gender Discrimination

- Squaring our $t$ statistic, and inverting part of the denominator and moving it to the numerator, we get

$$F = \frac{\tilde{n}\hat{\psi}^2}{\hat{\sigma}^2}$$

(6)

where $\tilde{n}$ is the harmonic mean of the $n_{ij}$, given by

$$\tilde{n} = \frac{1}{\sum_{ij} \frac{1}{n_{ij}}}$$

(7)

- We can show that $\hat{\psi}^2$ is equal to $b$ times the variance of the unweighted row means shown in our previous calculations, and so we may write

$$F_{gender} = \frac{b\tilde{n}s_{Y_j\cdot}^2}{\hat{\sigma}^2} = \frac{s_{Y_j\cdot}^2}{\hat{\sigma}^2/b\tilde{n}}$$

(8)

- So, once again, we see that the main effect for factor A is assessed by comparing the sample variance of the (unweighted) row means with an estimate of the variance of the row means, given by an estimate of $\sigma^2$ divided the “effective n” for the row means.
We can calculate $\hat{\sigma}^2$ quickly in R as follows:

```r
> var11 <- var(Salary[Group=="FemaleCollege"])
> var12 <- var(Salary[Group=="FemaleNoCollege"])
> n11 <- length(Salary[Group=="FemaleCollege"])
> n12 <- length(Salary[Group=="FemaleNoCollege"])
> var21 <- var(Salary[Group=="MaleCollege"])
> var22 <- var(Salary[Group=="MaleNoCollege"])
> n21 <- length(Salary[Group=="MaleCollege"])
> n22 <- length(Salary[Group=="MaleNoCollege"])
> sigma.hat.squared <- ((n11-1)*var11 + (n12-1)*var12 +
                        (n21-1)*var21 + (n22-1)*var22)/(n11+n12+n21+n22-4)
> n.tilde <- 4/(1/n11 + 1/n12 + 1/n21 + 1/n22)
> F.stat <- 2*n.tilde*var(c(21,23.5))/sigma.hat.squared
> F.stat
[1] 10.57343
> df <- n11+n12+n21+n22-4
> df
[1] 18
> pvalue <- 1 - pf(F.stat,1,df)
> pvalue
[1] 0.004428981
```

Our $F$ statistic for the main effect of Gender is therefore 10.57, and is significant beyond the 0.01 level.
Seeking to verify our calculation with R, we set up the standard ANOVA.

```r
> fit.1 <- lm(Salary ~ Gender * Degree)
> anova(fit.1)
```

Analysis of Variance Table

```plaintext
Response: Salary

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>1</td>
<td>0.297</td>
<td>0.297</td>
<td>0.1069</td>
<td>0.7475</td>
</tr>
<tr>
<td>Degree</td>
<td>1</td>
<td>272.392</td>
<td>272.392</td>
<td>98.0611</td>
<td>1.038e-08 ***</td>
</tr>
<tr>
<td>Gender:Degree</td>
<td>1</td>
<td>1.175</td>
<td>1.175</td>
<td>0.4229</td>
<td>0.5237</td>
</tr>
<tr>
<td>Residuals</td>
<td>18</td>
<td>50.000</td>
<td>2.778</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

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

What? Instead of a highly significant $F$, we obtain a value of 0.1069, with a $p$-value of 0.7475. What did we do wrong?

James H. Steiger

Non-orthogonal Designs
Another Look at the Gender Effect

- The answer is — in a sense, nothing.
- Welcome to the world of unbalanced designs and Types I, II, and III (not to mention IV) Sums of Squares!
- R did not provide the answer we expected because, by default, R computes its ANOVA using *Type I Sums of Squares*. Other programs, such as SAS and SPSS, analyze unbalanced ANOVA designs using *Type III Sums of Squares* by default.
- We can force R to override its default in several ways, and in a subsequent section I shall demonstrate two of them.
- But first, let’s examine another way to think of the salary data analysis.

James H. Steiger  Non-orthogonal Designs
Another Look at the Gender Effect

- The answer is — in a sense, nothing.
- Welcome to the world of unbalanced designs and Types I, II, and III (not to mention IV) Sums of Squares!
- R did not provide the answer we expected because, by default, R computes its ANOVA using *Type I Sums of Squares*. Other programs, such as SAS and SPSS, analyze unbalanced ANOVA designs using *Type III Sums of Squares* by default.
- We can force R to override its default in several ways, and in a subsequent section I shall demonstrate two of them.
- But first, let’s examine another way to think of the salary data analysis.
Another Look at the Gender Effect

- The answer is — in a sense, nothing.
- Welcome to the world of unbalanced designs and Types I, II, and III (not to mention IV) Sums of Squares!
- R did not provide the answer we expected because, by default, R computes its ANOVA using *Type I Sums of Squares*. Other programs, such as SAS and SPSS, analyze unbalanced ANOVA designs using *Type III Sums of Squares* by default.
- We can force R to override its default in several ways, and in a subsequent section I shall demonstrate two of them.
- But first, let’s examine another way to think of the salary data analysis.
Another Look at the Gender Effect

- The answer is — in a sense, nothing.
- Welcome to the world of unbalanced designs and Types I, II, and III (not to mention IV) Sums of Squares!
- R did not provide the answer we expected because, by default, R computes its ANOVA using *Type I Sums of Squares*. Other programs, such as SAS and SPSS, analyze unbalanced ANOVA designs using *Type III Sums of Squares* by default.
- We can force R to override its default in several ways, and in a subsequent section I shall demonstrate two of them.
- But first, let’s examine another way to think of the salary data analysis.
Another Look at the Gender Effect

- The answer is — in a sense, nothing.
- Welcome to the world of unbalanced designs and Types I, II, and III (not to mention IV) Sums of Squares!
- R did not provide the answer we expected because, by default, R computes its ANOVA using *Type I Sums of Squares*. Other programs, such as SAS and SPSS, analyze unbalanced ANOVA designs using *Type III Sums of Squares* by default.
- We can force R to override its default in several ways, and in a subsequent section I shall demonstrate two of them.
- But first, let’s examine another way to think of the salary data analysis.
Another Look at the Gender Effect
The Weighted Means Approach

- Our previous analysis did not *weight* the individual cell means in the two columns when estimating the two row means.
- For example, when estimating the mean salary for females, we simply averaged the two row means of 25 and 17, obtaining an estimate of 21.
- But suppose the sample sizes in the two cells (i.e., 8 and 4) actually represented the relative proportions of women who have college degrees and do not have college degrees. Then, in order to properly estimate the overall average salary for women in the population, we would have to weight the two sample means.
Another Look at the Gender Effect
The Weighted Means Approach

- Our previous analysis did not weight the individual cell means in the two columns when estimating the two row means.
- For example, when estimating the mean salary for females, we simply averaged the two row means of 25 and 17, obtaining an estimate of 21.
- But suppose the sample sizes in the two cells (i.e., 8 and 4) actually represented the relative proportions of women who have college degrees and do not have college degrees. Then, in order to properly estimate the overall average salary for women in the population, we would have to weight the two sample means.
Another Look at the Gender Effect  
The Weighted Means Approach

- Our previous analysis did not *weight* the individual cell means in the two columns when estimating the two row means.
- For example, when estimating the mean salary for females, we simply averaged the two row means of 25 and 17, obtaining an estimate of 21.
- But suppose the sample sizes in the two cells (i.e., 8 and 4) actually represented the relative proportions of women who have college degrees and do not have college degrees. Then, in order to properly estimate the overall average salary for women in the population, we would have to weight the two sample means.
Another Look at the Gender Effect
The Weighted Means Approach

- In that case, our weighted estimate of the first row mean would be

\[
\frac{n_{11} \bar{Y}_{11} + n_{12} \bar{Y}_{12}}{n_{11} + n_{12}} = \frac{2}{3} 25 + \frac{1}{3} 17 = 22.3333 \quad (9)
\]

- Correspondingly, the weighted estimate of the second row mean is

\[
\frac{n_{21} \bar{Y}_{21} + n_{22} \bar{Y}_{22}}{n_{21} + n_{22}} = \frac{3}{10} 27 + \frac{7}{20} 20 = 22.1 \quad (10)
\]

- Something truly interesting has been revealed. Note that, although college educated women earn less than their male counterparts, and non-college educated women earn less than their male counterparts, in the data as a whole, women average higher income than men!

- This is an example of Simpson’s Paradox.
Another Look at the Gender Effect

The Weighted Means Approach

- In that case, our weighted estimate of the first row mean would be
  \[ \frac{n_{11} \overline{Y}_{11} + n_{12} \overline{Y}_{12}}{n_{11} + n_{12}} = \frac{2}{3} 25 + \frac{1}{3} 17 = 22.3333 \quad (9) \]

- Correspondingly, the weighted estimate of the second row mean is
  \[ \frac{n_{21} \overline{Y}_{21} + n_{22} \overline{Y}_{22}}{n_{21} + n_{22}} = \frac{3}{10} 27 + \frac{7}{20} 20 = 22.1 \quad (10) \]

- Something truly interesting has been revealed. Note that, although college educated women earn less than their male counterparts, and non-college educated women earn less than their male counterparts, in the data as a whole, women average higher income than men!
- This is an example of Simpson’s Paradox.
Another Look at the Gender Effect
The Weighted Means Approach

- In that case, our weighted estimate of the first row mean would be

\[
\frac{n_{11}\bar{Y}_{11} + n_{12}\bar{Y}_{12}}{n_{11} + n_{12}} = \frac{2}{3}25 + \frac{1}{3}17 = 22.3333 \quad (9)
\]

- Correspondingly, the weighted estimate of the second row mean is

\[
\frac{n_{21}\bar{Y}_{21} + n_{22}\bar{Y}_{22}}{n_{21} + n_{22}} = \frac{3}{10}27 + \frac{7}{20}20 = 22.1 \quad (10)
\]

- Something truly interesting has been revealed. Note that, although college educated women earn less than their male counterparts, and non-college educated women earn less than their male counterparts, in the data as a whole, women average higher income than men!

- This is an example of *Simpson’s Paradox.*
Another Look at the Gender Effect
The Weighted Means Approach

- In that case, our weighted estimate of the first row mean would be

\[
\frac{n_{11} \bar{Y}_{11} + n_{12} \bar{Y}_{12}}{n_{11} + n_{12}} = \frac{2}{3} 25 + \frac{1}{3} 17 = 22.3333
\] (9)

- Correspondingly, the weighted estimate of the second row mean is

\[
\frac{n_{21} \bar{Y}_{21} + n_{22} \bar{Y}_{22}}{n_{21} + n_{22}} = \frac{3}{10} 27 + \frac{7}{20} 20 = 22.1
\] (10)

- Something truly interesting has been revealed. Note that, although college educated women earn less than their male counterparts, and non-college educated women earn less than their male counterparts, in the data as a whole, women average higher income than men!

- This is an example of *Simpson’s Paradox.*
Another Look at the Gender Effect
The Weighted Means Approach

- Suppose we were to test for the main effect of *Gender* as before, except this time using weights.
- We first construct the $t$ statistic. But this time, the null hypothesis is

$$\frac{n_{11}}{n_{1\bullet}} \mu_{11} + \frac{n_{12}}{n_{1\bullet}} \mu_{12} = \frac{n_{21}}{n_{2\bullet}} \mu_{21} + \frac{n_{22}}{n_{2\bullet}} \mu_{22}$$  (11)

- We’ll load in a brief function to compute the $t$ statistic, and feed it our data for analysis.
Another Look at the Gender Effect
The Weighted Means Approach

- Suppose we were to test for the main effect of Gender as before, except this time using weights.
- We first construct the $t$ statistic. But this time, the null hypothesis is

$$\frac{n_{11}}{n_{1\bullet}} \mu_{11} + \frac{n_{12}}{n_{1\bullet}} \mu_{12} = \frac{n_{21}}{n_{2\bullet}} \mu_{21} + \frac{n_{22}}{n_{2\bullet}} \mu_{22}$$

(11)

- We’ll load in a brief function to compute the $t$ statistic, and feed it our data for analysis.
Another Look at the Gender Effect

The Weighted Means Approach

- Suppose we were to test for the main effect of Gender as before, except this time using weights.
- We first construct the $t$ statistic. But this time, the null hypothesis is

$$\frac{n_{11}}{n_{1*}} \mu_{11} + \frac{n_{12}}{n_{1*}} \mu_{12} = \frac{n_{21}}{n_{2*}} \mu_{21} + \frac{n_{22}}{n_{2*}} \mu_{22}$$

(11)

- We’ll load in a brief function to compute the $t$ statistic, and feed it our data for analysis.
Another Look at the Gender Effect
The Weighted Means Approach

```r
> GeneralizedT <- function(means, sds, ns, wts, k0 = 0) {
+   {  
+     J <- length(means) 
+     df <- sum(ns) - J 
+     VarEstimate <- sum((ns - 1) * sds^2) / df 
+     num <- sum(wts * means) - k0 
+     den <- sqrt(sum(wts^2 / ns) * VarEstimate) 
+     t <- num / den 
+     return(c(t, df)) 
+   }
> GeneralizedT(c(25, 17, 27, 20), sqrt(c(2.285714, 4.666667, 4, 2)), 
+   c(8, 4, 3, 7), c(2/3, 1/3, -3/10, -7/10))
[1] 0.3269696 18.0000000
```

If we square the resulting $t$ statistic, we obtain the $F$ statistic given by the ANOVA analysis.

```r
> GeneralizedT(c(25, 17, 27, 20), sqrt(c(2.285714, 4.666667, 4, 2)),  
+   c(8, 4, 3, 7), c(2/3, 1/3, -3/10, -7/10))^2
[1] 0.1069091
```
Another Look at the Gender Effect
The Degree Effect

- Suppose we now use the same approach to assess the effect of *Degree* on *Salary*.
  
  ```
  > GeneralizedT(c(25,17,27,20),sqrt(c(2.285714,4.666667, + c(8,4,3,7),c(8/11,-4/11,3/11,-7/11)))[1]^2
  [1] 87.20182
  ```

- This $F$ is highly significant, but it *does not agree* with the $F$ value of 98.0611 shown in our ANOVA table. What happened?

- Type I sums of squares, it turns out, are order dependent, and effects are processed in the order given in the model statement.
So, if we reverse the order of *Degree* and *Gender* in the model specification, we obtain an $F$ of 87.2, agreeing with our $t$ test-based analysis.

```r
> fit.2 <- lm(Salary ~ Degree * Gender)
> anova(fit.2)
```

Analysis of Variance Table

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</td>
<td>1</td>
<td>242.227</td>
<td>242.227</td>
<td>87.2018</td>
<td>2.534e-08 ***</td>
</tr>
<tr>
<td>Gender</td>
<td>1</td>
<td>30.462</td>
<td>30.462</td>
<td>10.9662</td>
<td>0.003881 **</td>
</tr>
<tr>
<td>Degree:Gender</td>
<td>1</td>
<td>1.175</td>
<td>1.175</td>
<td>0.4229</td>
<td>0.523690</td>
</tr>
<tr>
<td>Residuals</td>
<td>18</td>
<td>50.000</td>
<td>2.778</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

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

Why does order matter in the case of non-orthogonal designs, while it doesn’t change the results when sample sizes are equal?
Another Look at the Gender Effect

The Degree Effect

- Type I sums of squares are based on a sequential modeling approach, using multiple regression.
- When predictors in multiple regression are uncorrelated, or orthogonal, then the effect of a predictor does not depend on other predictors in the equation, so order makes no difference.
- However, if they are not uncorrelated, the result of a significance test for a particular predictor depends on the order in which it was entered into the model.
- We can show that, in this case, however, the contrasts are not orthogonal. If we assume that each of our means has variance $\sigma^2/n_{ij}$, then we can compute the covariance between the contrasts computed for the Gender and Degree main effects.
Recall from our early discussion of linear combinations that, if group means are independent, we can compute the covariance between linear combinations without worrying covariances between different means.

In this case, the sample based estimates of the two linear combinations are

\[
\hat{\psi}_{\text{Gender}} = \frac{2}{3} \bar{Y}_{11} + \frac{1}{3} \bar{Y}_{12} - \frac{3}{10} \bar{Y}_{21} - \frac{7}{10} \bar{Y}_{22}
\]

\[
\hat{\psi}_{\text{Degree}} = \frac{8}{11} \bar{Y}_{11} - \frac{4}{11} \bar{Y}_{12} + \frac{3}{11} \bar{Y}_{21} - \frac{7}{11} \bar{Y}_{22}
\]

Using the “heuristic rule,” we can calculate the covariance between the two linear combinations by taking products and applying the conversion rule.

\[
\text{Cov} \left( \hat{\psi}_{\text{Degree}}, \hat{\psi}_{\text{Gender}} \right) = \frac{8 \times 2 \sigma^2}{3 \times 11 n_{11}} - \frac{4 \times 1 \sigma^2}{3 \times 11 n_{12}} - \frac{3 \times 3 \sigma^2}{11 \times 11 n_{21}} + \frac{8 \times 7 \sigma^2}{11 \times 11 n_{22}}
\]

\[
= \frac{8 \times 2}{3 \times 11} - \frac{4}{3 \times 11} - \frac{3 \times 3}{11 \times 11} + \frac{8 \times 7}{11 \times 11}
\]

\[
= \left( \frac{2}{33} - \frac{1}{33} - \frac{3}{121} + \frac{8}{121} \right) \sigma^2
\]

\[
= 26/363
\]
Covariance between Contrasts

Recall from our early discussion of linear combinations that, if group means are independent, we can compute the covariance between linear combinations without worrying covariances between different means.

In this case, the sample based estimates of the two linear combinations are

\[
\hat{\psi}_{\text{Gender}} = \frac{2}{3} \bar{Y}_{11} + \frac{1}{3} \bar{Y}_{12} - \frac{3}{10} \bar{Y}_{21} - \frac{7}{10} \bar{Y}_{22}
\]

\[
\hat{\psi}_{\text{Degree}} = \frac{8}{11} \bar{Y}_{11} - \frac{4}{11} \bar{Y}_{12} + \frac{3}{11} \bar{Y}_{21} - \frac{7}{11} \bar{Y}_{22}
\]

Using the “heuristic rule,” we can calculate the covariance between the two linear combinations by taking products and applying the conversion rule.

\[
\text{Cov} \left( \hat{\psi}_{\text{Degree}}, \hat{\psi}_{\text{Gender}} \right) = \frac{8 \times 2}{3 \times 11} \frac{\sigma^2}{n_{11}} - \frac{4 \times 1}{3 \times 11} \frac{\sigma^2}{n_{12}} - \frac{3 \times 3}{11 \times 11} \frac{\sigma^2}{n_{21}} + \frac{8 \times 7}{11 \times 11} \frac{\sigma^2}{n_{22}}
\]

\[
= \frac{8 \times 2}{3 \times 11} \frac{\sigma^2}{8} - \frac{4 \times 1}{3 \times 11} \frac{\sigma^2}{4} - \frac{3 \times 3}{11 \times 11} \frac{\sigma^2}{3} + \frac{8 \times 7}{11 \times 11} \frac{\sigma^2}{7}
\]

\[
= \left( \frac{2}{33} - \frac{1}{33} - \frac{3}{121} + \frac{8}{121} \right) \frac{\sigma^2}{8}
\]

\[
= \frac{26}{363} \frac{\sigma^2}{8}
\]
Covariance between Contrasts

- Recall from our early discussion of linear combinations that, if group means are independent, we can compute the covariance between linear combinations without worrying covariances between different means.
- In this case, the sample based estimates of the two linear combinations are

\[
\hat{\psi}_{\text{Gender}} = \frac{2}{3}\bar{Y}_{11} + \frac{1}{3}\bar{Y}_{12} - \frac{3}{10}\bar{Y}_{21} - \frac{7}{10}\bar{Y}_{22}
\]

\[
\hat{\psi}_{\text{Degree}} = \frac{8}{11}\bar{Y}_{11} - \frac{4}{11}\bar{Y}_{12} + \frac{3}{11}\bar{Y}_{21} - \frac{7}{11}\bar{Y}_{22}
\]

- Using the “heuristic rule,” we can calculate the covariance between the two linear combinations by taking products and applying the conversion rule.

\[
\text{Cov} \left( \hat{\psi}_{\text{Degree}}, \hat{\psi}_{\text{Gender}} \right) = \frac{8 \times 2}{3 \times 11} \sigma^2 - \frac{4 \times 1}{3 \times 11} \sigma^2 - \frac{3 \times 3}{11 \times 11} \sigma^2 + \frac{8 \times 7}{11 \times 11} \sigma^2
\]

\[
= \frac{8 \times 2}{3 \times 11} \sigma^2 - \frac{4 \times 1}{3 \times 11} \sigma^2 - \frac{3 \times 3}{11 \times 11} \sigma^2 + \frac{8 \times 7}{11 \times 11} \sigma^2
\]

\[
= \left( \frac{2}{33} - \frac{1}{33} - \frac{3}{121} - \frac{8}{121} \right) \sigma^2
\]

\[
= \frac{26}{363}
\]
Type III Sums of Squares

- It is easy to verify that, with equal sample sizes, the corresponding contrasts have zero covariance and are orthogonal.
- Note that the latter depends on equality of variances as well as equality of sample sizes.
- We saw in the preceding section how the default `anova` procedure in R generates a hierarchical, weighted-means-based analysis. Results depend on the order in which terms are entered into the model equation.
- The sums of squares for this type of analysis are called *Type I Sums of Squares*, a nomenclature that is believed to have originated with SAS.
- The alternative approach, based on unweighted-means, is non-hierarchical, and is referred to as *Type III Sums of Squares*. In this approach, sums of squares are based on the dropping of only one term at a time *from the full model*, rather than dropping a sequence of terms. Consequently, the results of the analysis do not depend on the order terms were entered into the model.
Here, we demonstrate with our analyses of Gender and Degree main effects, how to obtain the classic Type III Sums of Squares with unbalanced data.

```r
> options(contrasts = c("contr.sum","contr.poly"))
> fit <- lm(Salary ~ Gender * Degree)
> drop1(fit,~.,test="F")

Single term deletions

Model:
Salary ~ Gender * Degree

Df Sum of Sq RSS AIC F value Pr(>F)
<none> 50.000 26.062
Gender 1 29.371 79.371 34.228 10.5734 0.004429 **
Degree 1 264.336 314.336 64.507 95.1608 1.306e-08 ***
Gender:Degree 1 1.175 51.175 24.573 0.4229 0.523690
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> fit2 <- lm(Salary ~ Degree * Gender)
> drop1(fit2,~.,test="F")

Single term deletions

Model:
Salary ~ Degree * Gender

Df Sum of Sq RSS AIC F value Pr(>F)
<none> 50.000 26.062
Degree 1 264.336 314.336 64.507 95.1608 1.306e-08 ***
Gender 1 29.371 79.371 34.228 10.5734 0.004429 **
Degree:Gender 1 1.175 51.175 24.573 0.4229 0.523690
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
Here, we demonstrate with our analyses of *Gender* and *Degree* main effects, how to obtain the Type II Sums of Squares with unbalanced data. First, the `car` library must be loaded. Then, the `Anova` command must be invoked, with the `Type=2` (or, alternatively, `Type="II"`) option.

```r
> options(contrasts = c("contr.sum","contr.poly"))
> fit <- lm(Salary ~ Gender * Degree)
> Anova(fit,type=2)
```

**Anova Table (Type II tests)**

Response: Salary

<table>
<thead>
<tr>
<th></th>
<th>Sum Sq</th>
<th>Df</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>30.462</td>
<td>1</td>
<td>10.9662</td>
<td>0.003881 **</td>
</tr>
<tr>
<td>Degree</td>
<td>272.392</td>
<td>1</td>
<td>98.0611</td>
<td>1.038e-08 ***</td>
</tr>
<tr>
<td>Gender:Degree</td>
<td>1.175</td>
<td>1</td>
<td>0.4229</td>
<td>0.523690</td>
</tr>
<tr>
<td>Residuals</td>
<td>50.000</td>
<td>18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
In a similar manner, one may analyze *Gender* and *Degree* main effects, with the Type III Sums of Squares with unbalanced data. First, the *car* library must be loaded. Then, the *Anova* command must be invoked, with the `Type=3` (or, alternatively, `Type="III"`) option.

```r
> options(contrasts = c("contr.sum","contr.poly"))
> fit <- lm(Salary ~ Gender * Degree)
> Anova(fit,type=3)

Anova Table (Type III tests)

Response: Salary

<table>
<thead>
<tr>
<th></th>
<th>Sum Sq</th>
<th>Df</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>9305.8</td>
<td>1</td>
<td>3350.0845</td>
<td>&lt; 2.2e-16 ***</td>
</tr>
<tr>
<td>Gender</td>
<td>29.4</td>
<td>1</td>
<td>10.5734</td>
<td>0.004429 **</td>
</tr>
<tr>
<td>Degree</td>
<td>264.3</td>
<td>1</td>
<td>95.1608</td>
<td>1.306e-08 ***</td>
</tr>
<tr>
<td>Gender:Degree</td>
<td>1.2</td>
<td>1</td>
<td>0.4229</td>
<td>0.523690</td>
</tr>
<tr>
<td>Residuals</td>
<td>50.0</td>
<td>18</td>
<td>---</td>
<td></td>
</tr>
</tbody>
</table>

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
Clearly, when analyzing unbalanced designs, which method (Type I or Type III) is employed can have a substantial effect on the results.

Since we haven’t yet gotten to the later chapters in the textbook that discuss multiple regression and its relationship to ANOVA, we should defer some of the more technical discussion comparing the two methods.

At that time, we’ll discuss in detail that there are actually 4 methods (Types I, II, III, and IV).

Type IV sums of squares are an extension of Type III designed to specifically handle the case in which entire cells in the factorial design have no observations, either by happenstance or because they are technically or ethically unfeasible.

Type II sums of squares are an adaptation of Type I that eliminates the dependence on the order terms are listed in the model.
Clearly, when analyzing unbalanced designs, which method (Type I or Type III) is employed can have a substantial effect on the results.

Since we haven’t yet gotten to the later chapters in the textbook that discuss multiple regression and its relationship to ANOVA, we should defer some of the more technical discussion comparing the two methods.

At that time, we’ll discuss in detail that there are actually 4 methods (Types I, II, III, and IV).

Type IV sums of squares are an extension of Type III designed to specifically handle the case in which entire cells in the factorial design have no observations, either by happenstance or because they are technically or ethically unfeasible.

Type II sums of squares are an adaptation of Type I that eliminates the dependence on the order terms are listed in the model.
Clearly, when analyzing unbalanced designs, which method (Type I or Type III) is employed can have a substantial effect on the results.

Since we haven’t yet gotten to the later chapters in the textbook that discuss multiple regression and its relationship to ANOVA, we should defer some of the more technical discussion comparing the two methods.

At that time, we’ll discuss in detail that there are actually 4 methods (Types I, II, III, and IV).

Type IV sums of squares are an extension of Type III designed to specifically handle the case in which entire cells in the factorial design have no observations, either by happenstance or because they are technically or ethically unfeasible.

Type II sums of squares are an adaptation of Type I that eliminates the dependence on the order terms are listed in the model.
Clearly, when analyzing unbalanced designs, which method (Type I or Type III) is employed can have a substantial effect on the results.

Since we haven’t yet gotten to the later chapters in the textbook that discuss multiple regression and its relationship to ANOVA, we should defer some of the more technical discussion comparing the two methods.

At that time, we’ll discuss in detail that there are actually 4 methods (Types I, II, III, and IV).

Type IV sums of squares are an extension of Type III designed to specifically handle the case in which entire cells in the factorial design have no observations, either by happenstance or because they are technically or ethically unfeasible.

Type II sums of squares are an adaptation of Type I that eliminates the dependence on the order terms are listed in the model.
Introduction

Clearly, when analyzing unbalanced designs, which method (Type I or Type III) is employed can have a substantial effect on the results.

Since we haven’t yet gotten to the later chapters in the textbook that discuss multiple regression and its relationship to ANOVA, we should defer some of the more technical discussion comparing the two methods.

At that time, we’ll discuss in detail that there are actually 4 methods (Types I, II, III, and IV).

Type IV sums of squares are an extension of Type III designed to specifically handle the case in which entire cells in the factorial design have no observations, either by happenstance or because they are technically or ethically unfeasible.

Type II sums of squares are an adaptation of Type I that eliminates the dependence on the order terms are listed in the model.
Comparisons of the methods
Equivalence Relationships

- A number of resources on the internet list situations in which the various methods yield identical results. The first point is that, if cell sample sizes are equal, all methods are equivalent.
- In the case of unequal cell sizes, the following relationships hold in a 2-way factorial design (assuming that the terms in the model are entered in the order $A, B, AB$).

<table>
<thead>
<tr>
<th>Effect</th>
<th>Equivalence Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>III=IV</td>
</tr>
<tr>
<td>$B$</td>
<td>I=II, III=IV</td>
</tr>
<tr>
<td>$AB$</td>
<td>I=II=III=IV</td>
</tr>
</tbody>
</table>
Comparisons of the methods
Equivalence Relationships

- A number of resources on the internet list situations in which the various methods yield identical results. The first point is that, if cell sample sizes are equal, all methods are equivalent.
- In the case of unequal cell sizes, the following relationships hold in a 2-way factorial design (assuming that the terms in the model are entered in the order $A, B, AB$).

<table>
<thead>
<tr>
<th>Effect</th>
<th>Equivalence Relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>III=IV</td>
</tr>
<tr>
<td>$B$</td>
<td>I=II, III=IV</td>
</tr>
<tr>
<td>$AB$</td>
<td>I=II=III=IV</td>
</tr>
</tbody>
</table>
Comparisons of the methods
Calculation Procedures

- ANOVA calculations can be accomplished in a multiple regression context, and sums of squares are actually differences in the sum of squared residuals obtained from fitting two nested models in multiple regression.
- When a qualitative variable (i.e., a “factor”) is involved, the predictors are specially coded variables.
- The table below shows how sums of squares are calculated for the different terms in a 2-factor ANOVA.
Comparisons of the methods
Calculation Procedures

- ANOVA calculations can be accomplished in a multiple regression context, and sums of squares are actually differences in the sum of squared residuals obtained from fitting two nested models in multiple regression.
- When a qualitative variable (i.e., a “factor”) is involved, the predictors are specially coded variables.
- The table below shows how sums of squares are calculated for the different terms in a 2-factor ANOVA.
Comparisons of the methods
Calculation Procedures

- ANOVA calculations can be accomplished in a multiple regression context, and sums of squares are actually differences in the sum of squared residuals obtained from fitting two nested models in multiple regression.
- When a qualitative variable (i.e., a “factor”) is involved, the predictors are specially coded variables.
- The table below shows how sums of squares are calculated for the different terms in a 2-factor ANOVA.
## Comparisons of the methods

### Calculation Procedures

<table>
<thead>
<tr>
<th>Term</th>
<th>Type I SS</th>
<th>Type II SS</th>
<th>Type III SS</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$SS(A) = R() - R(A)$</td>
<td>$SS(A</td>
<td>B) = R(B) - R(A,B)$</td>
</tr>
<tr>
<td>$B$</td>
<td>$SS(B</td>
<td>A) = R(A) - R(A,B)$</td>
<td>$SS(B</td>
</tr>
<tr>
<td>$AB$</td>
<td>$SS(AB</td>
<td>A,B) = R(A,B) - R(A,B,AB)$</td>
<td>$SS(AB</td>
</tr>
</tbody>
</table>
Comparisons of the methods
Calculation Procedures

- The Venn diagram on the next page represents the fact that the overall sources of variation in the 2-Way factorial design are not generally orthogonal with unequal $n$. Following Rudolf Cardinal’s excellent webpage, we can gain some insight into the meaning of the various $SS$ methods as follows.

- Again, assuming $A$ is first in the model specification, Type I $SS$ (“sequential”) has
  - $SS_A = t + u + v + w$
  - $SS_B = x + y$
  - $SS_{AB} = z$

- This asks the questions: what’s the whole effect of $A$ (ignoring $B$)? What’s the effect of $B$, over and above the effect of $A$? What’s the effect of the $AB$ interaction, over and above the effects of $A$ and $B$? These could be written as tests of $A$, and $B|A$, and $AB|A, B$. 

James H. Steiger | Non-orthogonal Designs
Comparisons of the methods
Calculation Procedures

- The Venn diagram on the next page represents the fact that the overall sources of variation in the 2-Way factorial design are not generally orthogonal with unequal $n$. Following Rudolf Cardinal’s excellent webpage, we can gain some insight into the meaning of the various $SS$ methods as follows.

- Again, assuming $A$ is first in the model specification, Type I $SS$ (“sequential”) has

  1. $SS_A = t + u + v + w$
  2. $SS_B = x + y$
  3. $SS_{AB} = z$

- This asks the questions: what’s the whole effect of $A$ (ignoring $B$)? What’s the effect of $B$, over and above the effect of $A$? What’s the effect of the $AB$ interaction, over and above the effects of $A$ and $B$? These could be written as tests of $A$, and $B|A$, and $AB|A, B$. 

James H. Steiger
Non-orthogonal Designs
The Venn diagram on the next page represents the fact that the overall sources of variation in the 2-Way factorial design are not generally orthogonal with unequal $n$. Following Rudolf Cardinal’s excellent webpage, we can gain some insight into the meaning of the various $SS$ methods as follows.

Again, assuming $A$ is first in the model specification, Type I $SS$ (“sequential”) has

1. $SS_A = t + u + v + w$
2. $SS_B = x + y$
3. $SS_{AB} = z$

This asks the questions: what’s the whole effect of $A$ (ignoring $B$)? What’s the effect of $B$, over and above the effect of $A$? What’s the effect of the $AB$ interaction, over and above the effects of $A$ and $B$? These could be written as tests of $A$, and $B|A$, and $AB|A,B$. 

James H. Steiger  
Non-orthogonal Designs
The Venn diagram on the next page represents the fact that the overall sources of variation in the 2-Way factorial design are not generally orthogonal with unequal $n$. Following Rudolf Cardinal’s excellent webpage, we can gain some insight into the meaning of the various $SS$ methods as follows.

Again, assuming $A$ is first in the model specification, Type I $SS$ (“sequential”) has

1. $SS_A = t + u + v + w$
2. $SS_B = x + y$
3. $SS_{AB} = z$

This asks the questions: what’s the whole effect of $A$ (ignoring $B$)? What’s the effect of $B$, over and above the effect of $A$? What’s the effect of the $AB$ interaction, over and above the effects of $A$ and $B$? These could be written as tests of $A$, and $B|A$, and $AB|A,B$. 

James H. Steiger
Non-orthogonal Designs
The Venn diagram on the next page represents the fact that the overall sources of variation in the 2-Way factorial design are not generally orthogonal with unequal $n$. Following Rudolf Cardinal’s excellent webpage, we can gain some insight into the meaning of the various $SS$ methods as follows.

Again, assuming $A$ is first in the model specification, Type I $SS$ (“sequential”) has

1. $SS_A = t + u + v + w$
2. $SS_B = x + y$
3. $SS_{AB} = z$

This asks the questions: what’s the whole effect of $A$ (ignoring $B$)? What’s the effect of $B$, over and above the effect of $A$? What’s the effect of the $AB$ interaction, over and above the effects of $A$ and $B$? These could be written as tests of $A$, and $B|A$, and $AB|A,B$. 

James H. Steiger  
Non-orthogonal Designs
Comparisons of the methods

Calculation Procedures

- The Venn diagram on the next page represents the fact that the overall sources of variation in the 2-Way factorial design are not generally orthogonal with unequal $n$. Following Rudolf Cardinal’s excellent webpage, we can gain some insight into the meaning of the various $SS$ methods as follows.

  - Again, assuming $A$ is first in the model specification, Type I $SS$ ("sequential") has:
    1. $SS_A = t + u + v + w$
    2. $SS_B = x + y$
    3. $SS_{AB} = z$

  - This asks the questions: what’s the whole effect of $A$ (ignoring $B$)? What’s the effect of $B$, over and above the effect of $A$? What’s the effect of the $AB$ interaction, over and above the effects of $A$ and $B$? These could be written as tests of $A$, and $B|A$, and $AB|A,B$. 

James H. Steiger
Non-orthogonal Designs
Comparisons of the methods
Calculation Procedures

Fig. 9.6 The partitioning of variability in a two-factor design.
Comparisons of the methods
Calculation Procedures

- **Type II SS** ("hierarchical"): 
  1. $SS_A = t + w$
  2. $SS_B = x + y$
  3. $SS_{AB} = z$

This adjusts terms for all other terms except higher-order terms including the same predictors (in this example, adjusting the main effects of $A$ and $B$ for each other, but not for the interaction).

By "adjust for", we mean "not include any portion of the variance that overlaps with". These could be written as tests of $A|B$, and $B|A$, and $AB|A, B$. 

James H. Steiger  
Non-orthogonal Designs
Comparisons of the methods
Calculation Procedures

- **Type II $SS$ (“hierarchical”):**
  1. $SS_A = t + w$
  2. $SS_B = x + y$
  3. $SS_{AB} = z$

- This adjusts terms for all other terms except higher-order terms including the same predictors (in this example, adjusting the main effects of $A$ and $B$ for each other, but not for the interaction).

- By “adjust for”, we mean “not include any portion of the variance that overlaps with”. These could be written as tests of $A|B$, and $B|A$, and $AB|A, B$. 
Comparisons of the methods
Calculation Procedures

- **Type II SS** ("hierarchical"):
  1. $SS_A = t + w$
  2. $SS_B = x + y$
  3. $SS_{AB} = z$

- This adjusts terms for all other terms except higher-order terms including the same predictors (in this example, adjusting the main effects of $A$ and $B$ for each other, but not for the interaction).
- By "adjust for", we mean "not include any portion of the variance that overlaps with". These could be written as tests of $A|B$, and $B|A$, and $AB|A, B$. 
Comparisons of the methods
Calculation Procedures

- Type II $SS$ (“hierarchical”):
  1. $SS_A = t + w$
  2. $SS_B = x + y$
  3. $SS_{AB} = z$

  This adjusts terms for all other terms except higher-order terms including the same predictors (in this example, adjusting the main effects of $A$ and $B$ for each other, but not for the interaction).

  By “adjust for”, we mean “not include any portion of the variance that overlaps with”. These could be written as tests of $A|B$, and $B|A$, and $AB|A, B$. 
Comparisons of the methods
Calculation Procedures

- **Type II SS** (“hierarchical”):
  1. $SS_A = t + w$
  2. $SS_B = x + y$
  3. $SS_{AB} = z$

- This adjusts terms for all other terms except higher-order terms including the same predictors (in this example, adjusting the main effects of $A$ and $B$ for each other, but not for the interaction).

- By “adjust for”, we mean “not include any portion of the variance that overlaps with”. These could be written as tests of $A|B$, and $B|A$, and $AB|A, B$. 

Comparisons of the methods
Calculation Procedures

- **Type II SS** ("hierarchical"):
  1. $SS_A = t + w$
  2. $SS_B = x + y$
  3. $SS_{AB} = z$

  This adjusts terms for all other terms except higher-order terms including the same predictors (in this example, adjusting the main effects of $A$ and $B$ for each other, but not for the interaction).

  By "adjust for", we mean "not include any portion of the variance that overlaps with". These could be written as tests of $A|B$, and $B|A$, and $AB|A, B$. 
Comparisons of the methods
Calculation Procedures

Fig. 9.6 The partitioning of variability in a two-factor design.
Comparisons of the methods
Calculation Procedures

- **Type III SS** ("marginal", "orthogonal")
  1. $SS_A = t$
  2. $SS_B = x$
  3. $SS_{AB} = z$

  This assesses the contribution of each predictor over and above all others. These could be written as tests of $A|B$, $AB$, and $B|A$, $AB$, and $AB|A$, $B$. 
Comparisons of the methods
Calculation Procedures

- **Type III $SS$ ("marginal", "orthogonal")**
  1. $SS_A = t$
  2. $SS_B = x$
  3. $SS_{AB} = z$

  This assesses the contribution of each predictor over and above all others. These could be written as tests of $A|B, AB, B|A, AB$, and $AB|A, B$. 

James H. Steiger
Non-orthogonal Designs
Comparisons of the methods
Calculation Procedures

- **Type III SS** ("marginal", "orthogonal")
  1. $SS_A = t$
  2. $SS_B = x$
  3. $SS_{AB} = z$

This assesses the contribution of each predictor over and above all others. These could be written as tests of $A|B$, $AB$, and $B|A$, $AB$, and $AB|A, B$. 

James H. Steiger

Non-orthogonal Designs
Comparisons of the methods
Calculation Procedures

- **Type III SS** ("marginal", "orthogonal")
  1. $SS_A = t$
  2. $SS_B = x$
  3. $SS_{AB} = z$

  This assesses the contribution of each predictor over and above all others. These could be written as tests of $A|B$, $AB$, and $B|A$, $AB$, and $AB|A$, $B$. 
Comparisons of the methods
Calculation Procedures

• Type III $SS$ ("marginal", "orthogonal")
  
  1. $SS_A = t$
  2. $SS_B = x$
  3. $SS_{AB} = z$

• This assesses the contribution of each predictor over and above all others. These could be written as tests of $A|B, AB$, and $B|A, AB$, and $AB|A, B$. 
Comparisons of the methods
Calculation Procedures

Fig. 9.6 The partitioning of variability in a two-factor design.
As we have seen, Type I and Type III sums of squares test different hypotheses. This fact, for most users, probably takes precedence over other arguments advanced in favor of one method or another.

However, main effects only make sense when there is no interaction. Moreover, the *marginality principle* in multiple regression states that interactions should only be considered in models that include all effects subordinate to the interaction. Type III sums of squares test main effects for factor A by removing A from a model that still included the AB interaction. This is seen as a violation of the marginality principle.

Type II sums of squares are designed to test A and B main effects with maximum power *when there is no interaction between A and B*. 

James H. Steiger

Non-orthogonal Designs
Some other Relevant Points

- As we have seen, Type I and Type III sums of squares test different hypotheses. This fact, for most users, probably takes precedence over other arguments advanced in favor of one method or another.

- However, main effects only make sense when there is no interaction. Moreover, the *marginality principle* in multiple regression states that interactions should only be considered in models that include all effects subordinate to the interaction. Type III sums of squares test main effects for factor A by removing A from a model that still included the AB interaction. This is seen as a violation of the marginality principle.

- Type II sums of squares are designed to test A and B main effects with maximum power *when there is no interaction between A and B*. 
As we have seen, Type I and Type III sums of squares test different hypotheses. This fact, for most users, probably takes precedence over other arguments advanced in favor of one method or another.

However, main effects only make sense when there is no interaction. Moreover, the *marginality principle* in multiple regression states that interactions should only be considered in models that include all effects subordinate to the interaction. Type III sums of squares test main effects for factor A by removing A from a model that still included the AB interaction. This is seen as a violation of the marginality principle.

Type II sums of squares are designed to test A and B main effects with maximum power *when there is no interaction between A and B*.
Selection of a Method

- If there is any consensus among statisticians, it is probably in favor of Type III SS in most situations.
- However, different major “authorities” have come down strongly in favor of one method or another.
- Which method you choose depends on your scientific priorities.
- Clearly, if you have no priorities, or if you don’t understand the methods, it is impossible to make an intelligent choice.
- SPSS and SAS use Type III SS by default. Some editors will not give you a choice, but will demand that you use Type III for “compatibility” with SPSS and SAS.
- If there is no interaction effect, Type II SS will probably give you more power than Type III SS, while maintaining the marginality principle. However, there is always the chance that you are committing a Type II error by failing to reject the hypothesis of no interaction, in which case the additional power is illusory.
Selection of a Method

- If there is any consensus among statisticians, it is probably in favor of Type III SS in most situations.
- However, different major “authorities” have come down strongly in favor of one method or another.
- Which method you choose depends on your scientific priorities.
- Clearly, if you have no priorities, or if you don’t understand the methods, it is impossible to make an intelligent choice.
- SPSS and SAS use Type III SS by default. Some editors will not give you a choice, but will demand that you use Type III for “compatibility” with SPSS and SAS.
- If there is no interaction effect, Type II SS will probably give you more power than Type III SS, while maintaining the marginality principle. However, there is always the chance that you are committing a Type II error by failing to reject the hypothesis of no interaction, in which case the additional power is illusory.
If there is any consensus among statisticians, it is probably in favor of Type III SS in most situations.

However, different major “authorities” have come down strongly in favor of one method or another.

Which method you choose depends on your scientific priorities.

Clearly, if you have no priorities, or if you don’t understand the methods, it is impossible to make an intelligent choice.

SPSS and SAS use Type III SS by default. Some editors will not give you a choice, but will demand that you use Type III for “compatibility” with SPSS and SAS.

If there is no interaction effect, Type II SS will probably give you more power than Type III SS, while maintaining the marginality principle. However, there is always the chance that you are committing a Type II error by failing to reject the hypothesis of no interaction, in which case the additional power is illusory.
Selection of a Method

- If there is any consensus among statisticians, it is probably in favor of Type III SS in most situations.
- However, different major “authorities” have come down strongly in favor of one method or another.
- Which method you choose depends on your scientific priorities.
- Clearly, if you have no priorities, or if you don’t understand the methods, it is impossible to make an intelligent choice.
- SPSS and SAS use Type III SS by default. Some editors will not give you a choice, but will demand that you use Type III for “compatibility” with SPSS and SAS.
- If there is no interaction effect, Type II SS will probably give you more power than Type III SS, while maintaining the marginality principle. However, there is always the chance that you are committing a Type II error by failing to reject the hypothesis of no interaction, in which case the additional power is illusory.
Selection of a Method

- If there is any consensus among statisticians, it is probably in favor of Type III SS in most situations.
- However, different major “authorities” have come down strongly in favor of one method or another.
- Which method you choose depends on your scientific priorities.
- Clearly, if you have no priorities, or if you don’t understand the methods, it is impossible to make an intelligent choice.
- SPSS and SAS use Type III SS by default. Some editors will not give you a choice, but will demand that you use Type III for “compatibility” with SPSS and SAS.
- If there is no interaction effect, Type II SS will probably give you more power than Type III SS, while maintaining the marginality principle. However, there is always the chance that you are committing a Type II error by failing to reject the hypothesis of no interaction, in which case the additional power is illusory.
If there is any consensus among statisticians, it is probably in favor of Type III SS in most situations.

However, different major “authorities” have come down strongly in favor of one method or another.

Which method you choose depends on your scientific priorities.

Clearly, if you have no priorities, or if you don’t understand the methods, it is impossible to make an intelligent choice.

SPSS and SAS use Type III SS by default. Some editors will not give you a choice, but will demand that you use Type III for “compatibility” with SPSS and SAS.

If there is no interaction effect, Type II SS will probably give you more power than Type III SS, while maintaining the marginality principle. However, there is always the chance that you are committing a Type II error by failing to reject the hypothesis of no interaction, in which case the additional power is illusory.