Factorial Designs
Edpsy 580

Carolyn J. Anderson
Department of Educational Psychology

I L L I N O I S
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
Outline

- Motivation and Advantages
- Interactions: The *only* new concept.
- 2–Factor ANOVA.
- Factorial ANOVA as a GLM.
- Estimation
- Statistical Inference
- Examples
- Unbalanced designs
- Statistical Model and 3 or more factors...
Example 2-Way ANOVA


- **Response/dependent variable** = understanding as measured by 10 item inference verification test (IVT), $Y_i = \text{IVT}_i$.

- **Factors:**
  - Format (text or web)
  - Instructions participants received: write a Narrative (N), Summary (S), Explanation (E), Argument (A).
Some Data: Reading Scores

From Keppel:

- **Purpose**: studying things that effect reading scores.

- **Dependent variable**: reading scores.

- **Independent variables**:
  - Line length: Length of the line of text (3", 5", 7").
  - Contrast: The contrast between paper and print. There are three levels (low, medium, high).

---

Motivation/Advantages

- Example 2-Way ANOVA
- Some Data: Reading Scores
- Advantages of Factorial ANOVA
  - More efficient
  - Reduce $\sigma^2$
  - Experimental Control
  - Generality of Results
  - Generality of Results
  - Generality of Results

Interaction

2–Factor ANOVA

As a GLM

Estimation

Statistical Inference

Examples

Unbalanced Designs

Higher-way Designs
Advantages of Factorial ANOVA

- Economy
- Experimental control
- Generality of results

**Economy:**

Want to study the effect of two different factors on some behavior or performance measure.

Could to two 1-factor ANOVA’s.

- Suppose want \( n = 30 \) per level of a factor.
- 1-factor ANOVA for line length:
  
<table>
<thead>
<tr>
<th>3&quot;</th>
<th>5&quot;</th>
<th>7&quot;</th>
<th>( N = 90 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_1 = 30 )</td>
<td>( n_2 = 30 )</td>
<td>( n_3 = 30 )</td>
<td></td>
</tr>
</tbody>
</table>

- 1-factor ANOVA for Print/Paper contrast:
  
<table>
<thead>
<tr>
<th>Low</th>
<th>Med.</th>
<th>High</th>
<th>( N = 90 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_1 = 30 )</td>
<td>( n_2 = 30 )</td>
<td>( n_3 = 30 )</td>
<td></td>
</tr>
</tbody>
</table>

- Two separate experiments: need 180 subjects.
A single experiment, two factors with $n = 10$ subjects per “cell” (combination of the levels of the two factors).

<table>
<thead>
<tr>
<th>Line length</th>
<th>Print/paper contrast</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>3&quot;</td>
<td>$n_{11} = 10$</td>
<td>$n_{1+} = 30$</td>
</tr>
<tr>
<td></td>
<td>$n_{12} = 10$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$n_{13} = 10$</td>
<td></td>
</tr>
<tr>
<td>5&quot;</td>
<td>$n_{21} = 10$</td>
<td>$n_{2+} = 30$</td>
</tr>
<tr>
<td></td>
<td>$n_{22} = 10$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$n_{23} = 10$</td>
<td></td>
</tr>
<tr>
<td>7&quot;</td>
<td>$n_{31} = 10$</td>
<td>$n_{3+} = 30$</td>
</tr>
<tr>
<td></td>
<td>$n_{32} = 10$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$n_{33} = 10$</td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>$n_{+1} = 30$</td>
<td>$N = 90$</td>
</tr>
<tr>
<td></td>
<td>$n_{+2} = 30$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$n_{+3} = 30$</td>
<td></td>
</tr>
</tbody>
</table>

$N = 90$ & still $n_{jk} = 30$ subject per level of each factor.
Reduce $\sigma^2_e$

via Experimental Control

- In 1-factor ANOVA, there are uncontrolled variables that effect the dependent variable.
- Uncontrolled and unsystematic variables.
- Source of variance due to
  - Random assignment to groups (or random sampling).
  - Individual differences (subject variables).
- These contribute to $\sigma^2_e$. 

---

<table>
<thead>
<tr>
<th>Motivation/Advantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Example 2-Way ANOVA</td>
</tr>
<tr>
<td>● Some Data: Reading Scores</td>
</tr>
<tr>
<td>● Advantages of Factorial ANOVA</td>
</tr>
<tr>
<td>● More efficient</td>
</tr>
<tr>
<td>● Reduce $\sigma^2_e$</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-Factor ANOVA</td>
</tr>
<tr>
<td>As a GLM</td>
</tr>
<tr>
<td>Estimation</td>
</tr>
<tr>
<td>Statistical Inference</td>
</tr>
<tr>
<td>Examples</td>
</tr>
<tr>
<td>Unbalanced Designs</td>
</tr>
<tr>
<td>Higher-way Designs</td>
</tr>
</tbody>
</table>
Experimental Control

- Can reduce $\sigma^2_\epsilon$ and increase power by including variable(s) that effect the dependent variable even though you aren’t especially interested in the extra variable.

- For example... In reading experiment with contrast & line length as factors, reading level or achievement effects reading scores.
  - To reduce $\sigma^2_\epsilon$, use homogenous groups $\rightarrow$ subjects all same reading level.
  - **Drawback**: Results only generalize to others of the same reading level.

- Alternative: use reading level as a factor:
  - Randomly select individuals from each of the reading levels and randomly assign to levels of the other factors.
  - Advantages of adding another factor:
    - Gain reduction in $\sigma^2_\epsilon$ and therefore increase power.
    - Results more generalizable.
Generality of Results

- Results from single factor experiments lack generality to the extent to which specific values of other relevant variables are held constant.

- Factorial experiments allow effects of an independent variable to be averaged over levels of other relevant variable(s).

- Example: reading study...
Generality of Results

Example: reading study

- An overall effect of one factor is referred to as a **main effect** — effect obtained by combining scores across levels of other factor(s).

- Can collapse over print/paper contrast to look at main effect of text length on reading scores.

- Effect of text length represents a more general effect because average over 3 levels of paper/print contrast.

- Can collapse over text length to look at main effects of print/paper contrast.
Factorial designs provide the same information as a single factor design only up to a point.

Factorial designs provide the same type of information as single factor designs so long as there is no interaction between the two factors.
Interaction

- No interaction.

- Interaction.
  - Examples.
  - Definition(s).
  - Implications of interactions for theory.

- More examples of interactions and no interaction.
Interaction

- The only new concept for Factorial ANOVA.

- Best way to understand what “interaction” means is to first consider what it means for there be to no interaction.

Hypothetical example from Keppel:

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Line Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Factor A)</td>
</tr>
<tr>
<td></td>
<td>3 in. 5 in. 7 in.</td>
</tr>
<tr>
<td>Contrast</td>
<td>(a₁) (a₂) (a₃)</td>
</tr>
<tr>
<td>Low (b₁)</td>
<td>0.89 2.22 2.89</td>
</tr>
<tr>
<td>Med. (b₂)</td>
<td>3.89 5.22 5.89</td>
</tr>
<tr>
<td>High (b₃)</td>
<td>4.22 5.55 6.22</td>
</tr>
<tr>
<td>Mean</td>
<td>3.00 4.33 5.00</td>
</tr>
</tbody>
</table>
No Interaction: Effect of Length

Main Effects

Simple Effects

- Interaction: Effect of Length
- No Interaction: Effect of Contrast
- No Interaction & Interaction

- Interaction
- Interaction: Effect of Line Length
- Interaction: Effect of Contrast
- Interactions
- Definitions of Interactions
- Wiley-Voss: Interaction?
- Main Effects and Interactions

- ANOVA Model: (No)
- Implications of Interaction
- Classic Example

2-Factor ANOVA

Factorial Designs
Motivation/Advantages

Interaction
- Interaction
- Interaction
- No Interaction: Effect of Length
- No Interaction: Effect of Contrast
- Interaction: Effect of Line Length
- Interaction: Effect of Contrast
- Interactions
- Definitions of Interactions
- Wiley-Voss: Interaction?
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- ANOVA Model: (No) Interaction
- Implications of Interaction
- Classic Example

2–Factor ANOVA

Factorial Designs

As a GLM

No Interaction: Effect of Contrast

Main Effects

Simple Effects
No Interaction

- Marginal means are representative of the pattern of means in the cells.

- The figures on the right are plots of cell means with a separate line for levels of the other factor.

- Each line can be considered a single factor experiment and the whole set of them as simple effects of contrast or line length.

- Parallel lines for simple effects → the exact pattern of differences obtained with one factor is the same at each level of the other factor.
No Interaction & Interaction

- **No interaction** means that the effect of an independent variable (factor) is exactly the same for each level of the other factor.

- **Interaction** means that the effect of an independent variable (factor) is different depending on the level of the other factor.
Interaction

- Hypothetical example from Keppel:

<table>
<thead>
<tr>
<th>Contrast (Factor B)</th>
<th>Line Length (Factor A)</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3 in.</td>
<td>5 in.</td>
</tr>
<tr>
<td>Low (b₁)</td>
<td>(a₁)</td>
<td>(a₂)</td>
</tr>
<tr>
<td>Med. (b₂)</td>
<td>1.00</td>
<td>2.00</td>
</tr>
<tr>
<td>High (b₃)</td>
<td>3.00</td>
<td>5.00</td>
</tr>
<tr>
<td></td>
<td>5.00</td>
<td>5.00</td>
</tr>
<tr>
<td>Mean</td>
<td>3.00</td>
<td>4.33</td>
</tr>
</tbody>
</table>

- Note: The marginal means are identical to those from the example of no interaction
Interaction: Effect of Line Length

Main Effects

Interaction Effects

Motivation/Advantages

Interaction

- Interaction
- Interaction
- No Interaction: Effect of Length
- No Interaction: Effect of Contrast
- No Interaction & Interaction
- Interaction: Effect of Line Length
- Interaction: Effect of Contrast
- Interactions
- Definitions of Interactions
- Wiley-Voss: Interaction?
- Main Effects and Interactions
- ANOVA Model: (No)
- Implications of Interaction
- Classic Example

2–Factor ANOVA

Factorial Designs

As a GLM
Interaction: Effect of Contrast

Main Effects

Interaction Effects

Motivation/Advantages

Interaction
- Interaction
  - No Interaction: Effect of Length
  - No Interaction: Effect of Contrast
  - No Interaction & Interaction
  - Interaction: Effect of Line Length
- Interaction: Effect of Contrast
- Interactions
- Definitions of Interactions
- Wiley-Voss: Interaction?
- Main Effects and Interactions
- ANOVA Model: (No) Interaction
- Implications of Interaction
- Classic Example

2–Factor ANOVA

Factorial Designs

As a GLM
Interactions

- In the interaction example, the marginal means are the same as those from the no interaction example. The cell means differ.

- The simple effects for the factors are not the same as the main effects.

- The lines on the right figures (cell mean plots) are not parallel.

- A quick glance at the plot of cell means shows that the form or nature of the interaction (relationship) between the two factors.
Definitions of Interactions

(From Keppel)

Definition 1: “An interaction is present when the effects of one independent variable on behavior changes at different levels of the second independent variable.”

- The focus is on the effect of factors on behavior (the response or dependent variable) and **not** that the two factors influence each other.

- Interaction refers to the way the two factor combine to influence behavior.
Definitions of Interactions

(From Keppel)

- **Definition 2:** “An interaction is present when the simple effects of one factor are not the same at different levels of the second factor.”

- **Definition 3:** “An interaction is present when the main effects of a factor are not representative of the simple effects of that factor.”

If an interaction is present, the conclusions based only on main effects do not fully or accurately describe the data.
Wiley-Voss: Interaction?

- Interaction
- No Interaction: Effect of Length
- No Interaction: Effect of Contrast
- No Interaction & Interaction
- Interaction: Effect of Line Length
- Interaction: Effect of Contrast
- Interactions
- Definitions of Interactions
- Wiley-Voss: Interaction?
- Main Effects and Interactions
- Main Effects and Interactions
- ANOVA Model: (No)
- Interaction
- Implications of Interaction
- Classic Example

2–Factor ANOVA

Factorial Designs

Slide 24 of 92
Main Effects and Interactions

- **Motivation/Advantages**
- **Interaction**
- Interaction
  - No Interaction: Effect of Length
  - No Interaction: Effect of Contrast
- Interaction & Interaction
- Interaction: Effect of Line Length
- Interaction: Effect of Contrast
- Interactions
- Definitions of Interactions
- Wiley-Voss: Interaction?
- **Main Effects and Interactions**
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- ANOVA Model: (No)
- Interaction
- Implications of Interaction
- Classic Example

![Graph](image-url)
Main Effects and Interactions

Main Effect A?  Main Effect B?  Interaction?

Motivation/Advantages

Interaction
- Interaction
- Interaction
- No Interaction: Effect of Length
- No Interaction: Effect of Contrast
- No Interaction
- No Interaction & Interaction
- Interaction
- Interaction: Effect of Line Length
- Interaction: Effect of Contrast
- Interactions
- Definitions of Interactions
- Definitions of Interactions
- Wiley-Voss: Interaction?
- Main Effects and Interactions

Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- ANOVA Model: (No) Interaction
- Implications of Interaction
- Classic Example

2–Factor ANOVA

As a GLM
Main Effects and Interactions

[Graph showing main effects and interactions with factors A and B]

Motivation/Advantages

Interaction

- Interaction
- No Interaction: Effect of Length
- No Interaction: Effect of Contrast
- Interaction
- Interaction: Effect of Line Length
- Interaction: Effect of Contrast
- Interactions
- Definitions of Interactions
- Wiley-Voss: Interaction?
- Main Effects and Interactions
- ANOVA Model: (No Interaction)

2-Factor ANOVA
Main Effects and Interactions

Motivation/Advantages

Interaction
- Interaction
- No Interaction: Effect of Length
- No Interaction: Effect of Contrast
- No Interaction & Interaction
- Interaction
- Interaction: Effect of Line Length
- Interaction: Effect of Contrast
- Interactions
- Definitions of Interactions
- Wiley-Voss: Interaction?
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- ANOVA Model: (No)
- Interaction
- Implications of Interaction
- Classic Example
Main Effects and Interactions

Motivation/Advantages

Interaction
- Interaction
- No Interaction: Effect of Length
- No Interaction: Effect of Contrast
- No Interaction & Interaction
- Interaction: Effect of Line Length
- Interaction: Effect of Contrast
- Interactions
- Definitions of Interactions
- Wiley-Voss: Interaction?
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- ANOVA Model: (No)
- Implications of Interaction
- Classic Example

2–Factor ANOVA

As a GLM
Main Effects and Interactions

2-Factor ANOVA

Factorial Designs

As a GLM

Motivation/Advantages

Interaction
- Interaction
- Interaction
- No Interaction: Effect of Length
- No Interaction: Effect of Contrast
- No Interaction
- No Interaction & Interaction
- Interaction
- Interaction: Effect of Line Length
- Interaction: Effect of Contrast
- Interactions
- Definitions of Interactions
- Definitions of Interactions
- Wiley-Voss: Interaction?
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- ANOVA Model: (No) Interaction
- Implications of Interaction
- Classic Example

Main Effect A? Main Effect B? Interaction?
Main Effects and Interactions

Motivation/Advantages

Interaction
- Interaction
- No Interaction: Effect of Length
- No Interaction: Effect of Contrast
- No Interaction
- Interaction
- Interaction: Effect of Line Length

Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- ANOVA Model: (No)
- Implications of Interaction
- Classic Example

Factor A

Main Effect A? Main Effect B? Interaction?

Cell Mean

2–Factor ANOVA

Factorial Designs

As a GLM
Main Effects and Interactions

Motivation/Advantages

Interaction

- Interaction
- Interaction
- No Interaction: Effect of Length
- No Interaction: Effect of Contrast
- No Interaction
- No Interaction & Interaction
- Interaction
- Interaction: Effect of Line Length
- Interaction: Effect of Contrast
- Interactions
- Definitions of Interactions
- Wiley-Voss: Interaction?
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- Main Effects and Interactions
- ANOVA Model? (No)
- Interaction
- Implications of Interaction
- Classic Example

2–Factor ANOVA

Factor A

B

b1

b2

b3
ANOVA Model: (No) Interaction

When there is **no interaction**, the term **Additive Effects** describe the joint or combined influence or effect of the factors on the dependent variable,

\[ Y_{ijk} = \mu + \alpha_j + \beta_k + \epsilon_{ijk} \]

When there is an **interaction**, the term **nonadditive effects** is used. There is an additional effect that must be added to specify the joint effect of the two factors,

\[ Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha \beta)_{jk} + \epsilon_{ijk} \]
Implications of Interaction

For theory.

- If there’s an interaction, need a more complex theory to explain the phenomenon than when there is no interaction.

- Often theory’s predict the presence of an interaction and interactions are often what’s interesting.

- Finding interactions can lead to greater understanding of behavior.
Classic Example

- Experiment by Lashley on the effect of brain damage on maze learning by rats. Lashley varied the amount damage of cortical tissue. He destroyed 1% to 10% (small) to over 50% (large). He tested rats on 3 mazes (3 difficulty levels).

- Two factors:
  - Amount of brain damage
  - Difficulty of the maze.

- Results: There was little difference among rats due to amount of brain damage when they ran the easiest mazes, but very dramatic differences on the most difficult mazes.

- The destruction of cortical tissue primarily effects the acquisition of complex learning. There is no uniform, overall learning deficit.

- If Lashley had only used 1 maze, he would have never discovered this.
2–Factor ANOVA

- Design & Notation.
- 2-Factor ANOVA Model.
- Estimation.
- Statistical Hypotheses.
- Sums of squares.
- Degrees of freedom.
- Mean squares.
- F-ratios.
Design

- **Completely Crossed**: Every level of one factor is combined with every other level of the other factor(s). All combinations of levels and factors occur.

- **Balanced**: Each combination of factors (cell in the design) has the same number of subjects (observations); i.e.,

\[ n_{ij} = n. \]

- **Factorial design**: Two or more factors that are completely crossed.

- **Two-Way ANOVA**: A factorial design involving 2 crossed factors and balanced design.

- **Replicated Experiments**: More than one observation in each cell of the design.

- **Orthogonal design**: Can estimate each treatment effect separately.

- **Factorial designs are orthogonal if**
  - Random and independent samples within each cell.
  - Balanced (or proportional, i.e., \( n_{jk} = n_j n_k / N \))
Motivation/Advantages

Interaction

2–Factor ANOVA

● 2–Factor ANOVA
● Design
● Notation
● Notation: Various Sample Means
● 2-Factor ANOVA Model

As a GLM

Estimation

Statistical Inference

Examples

Unbalanced Designs

Higher-way Designs

<table>
<thead>
<tr>
<th>Notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>■ $Y_{ijk}$ is the observation on the $i$th subject (individual, etc) on the $j$th level of “row” factor (Factor $A$) and the $k$th level of the “column” factor (Factor $B$).</td>
</tr>
<tr>
<td>■ $j$ is the index for the levels Factor $A$.</td>
</tr>
<tr>
<td>■ $J$ is the total number of levels of Factor $A$, so $j = 1, \ldots, J$.</td>
</tr>
<tr>
<td>■ $k$ is the index for the levels Factor $B$.</td>
</tr>
<tr>
<td>■ $K$ is the total number of levels of Factor $B$, so $k = 1, \ldots, K$.</td>
</tr>
<tr>
<td>■ $i$ is the index for subjects within levels of the two factors.</td>
</tr>
<tr>
<td>■ $n_{jk}$ is number of subjects in the combination of levels $j$ and $k$ on factors $A$ and $B$, respectively, so $i = 1, \ldots, n_{jk}$.</td>
</tr>
<tr>
<td>■ $(j, k)$ cell is a combination of levels $j$ and $k$ on factors $A$ &amp; $B$.</td>
</tr>
<tr>
<td>■ $n_j = \sum_k n_{jk}$ is the number of subjects who are in the $j$th level of Factor $A$.</td>
</tr>
<tr>
<td>■ $n_k = \sum_j n_{jk}$ is the number of subjects who are in the $k$th level of Factor $B$.</td>
</tr>
<tr>
<td>■ $N = n_{..} = \sum_j \sum_k n_{jk}$ is the total number of subjects.</td>
</tr>
</tbody>
</table>
Notation: Various Sample Means

- $\bar{Y}_{jk} = (\sum_i Y_{ijk})/n_{jk}$ is the mean over subjects within the $(j, k)$th cell.

- $\bar{Y}_j = \sum_k \sum_i Y_{ijk}/n_{jk} = \sum_k \bar{Y}_{jk}/K$, the mean over subjects and levels of factor $B$. It’s the (marginal) mean for level $j$ of factor $A$.

- $\bar{Y}_k = \sum_j \sum_i Y_{ijk}/n_{jk} = \sum_j \bar{Y}_{jk}/J$, the mean over individuals and levels of factor $A$. It’s the (marginal) mean for level $k$ of factor $B$.

- $\bar{Y}. = \sum_j \sum_k \sum_i Y_{ijk}/N$, the grand mean
The model,

\[ Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk} \]

where,

- \( \mu \) is the grand mean (overall).
- \( \alpha_j \) is the treatment effect for level \( j \) of factor \( A \).
- \( \beta_k \) is the treatment effect for level \( k \) of factor \( B \).
- \( (\alpha\beta)_{jk} \) is the interaction effect.
- \( \epsilon_{jk} \) is the error (residual).

Assumption: \( \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2) \) i.i.d. So

\[ Y_{ijk} \sim \mathcal{N}(\mu_{jk}, \sigma^2) \text{ where } \mu_{jk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}. \]
2-Factor ANOVA Model as GLM

- Define variables denoting levels of factors A:
  \[
  X_{iA(j)} = \begin{cases} 
  1 & \text{if level } j \text{ of } A \\
  0 & \text{otherwise}
  \end{cases} \quad \text{for } j = 1, \ldots, J
  \]

- Define variables denoting levels of factors B:
  \[
  X_{iB(k)} = \begin{cases} 
  1 & \text{if level } k \text{ of } B \\
  0 & \text{otherwise}
  \end{cases} \quad \text{for } k = 1, \ldots, K
  \]

- For the interaction, \( X_{AB(jk)} = X_{A(j)}X_{B(k)} \)

- The general linear model,
  \[
  Y_{ijk} = \beta_0^* + \beta_1^* X_{iA(1)} + \ldots + \beta_J^* X_{iA(J)} + \beta_{J+1}^* X_{iB(1)} + \ldots + \beta_{J+K}^* X_{iB(K)} \\
  + \beta_{(J+K+1)}^* X_{iAB(11)} + \ldots + \beta_{J+K+JK}^* X_{iAB(JK)} + \epsilon_{ijk}
  \]
Define $X_{io} = 1$ and

$$X_{i1} = \begin{cases} 
1 & \text{if Narrative} \\
0 & \text{otherwise}
\end{cases} \quad \quad X_{i2} = \begin{cases} 
1 & \text{if Summary} \\
0 & \text{otherwise}
\end{cases}$$

$$X_{i3} = \begin{cases} 
1 & \text{if Explanation} \\
0 & \text{otherwise}
\end{cases} \quad \quad X_{i4} = \begin{cases} 
1 & \text{if Argument} \\
0 & \text{otherwise}
\end{cases}$$

Define dummy variables for format

$$X_{i5} = \begin{cases} 
1 & \text{if Text} \\
0 & \text{otherwise}
\end{cases} \quad \quad X_{i6} = \begin{cases} 
1 & \text{if web} \\
0 & \text{otherwise}
\end{cases}$$

Interactions:

$$X_7 = X_{i1}X_{i5}$$

$$X_8 = X_{i1}X_{i6}$$

$$\vdots$$

$$X_{i14} = X_{i4}X_{i6}$$
2-Factor ANOVA Model as GLM

\[ Y_{ijk} = \beta_o^* + \sum_{m=1}^{14} \beta_m^* X_{m11} + \epsilon_{im} \]

\[ Y_{ijk} = \begin{cases} 
\beta_o^* + \beta_1^* + \beta_5^* + \beta_7^* + \epsilon_{11} & \text{if Nar. & text} \\
\beta_o^* + \beta_1^* + \beta_6^* + \beta_8^* + \epsilon_{12} & \text{if Nar. & web} \\
\beta_o^* + \beta_2^* + \beta_5^* + \beta_9^* + \epsilon_{21} & \text{if Sum. & text} \\
\beta_o^* + \beta_2^* + \beta_6^* + \beta_{10}^* + \epsilon_{22} & \text{if Sum. & web} \\
\beta_o^* + \beta_3^* + \beta_5^* + \beta_{11}^* + \epsilon_{31} & \text{if Exp. & text} \\
\beta_o^* + \beta_3^* + \beta_6^* + \beta_{12}^* + \epsilon_{32} & \text{if Exp. & web} \\
\beta_o^* + \beta_4^* + \beta_5^* + \beta_{13}^* + \epsilon_{41} & \text{if Arg. & text} \\
\beta_o^* + \beta_4^* + \beta_6^* + \beta_{14}^* + \epsilon_{42} & \text{if Arg. & web} 
\end{cases} \]
2-Factor ANOVA Model as GLM

The correspondence between the two forms of the model

<table>
<thead>
<tr>
<th>Intercept</th>
<th>Factor A</th>
<th>Factor B</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu = \beta_0^*$</td>
<td>$\alpha_1 = \beta_1^*$</td>
<td>$\beta_1 = \beta_5^*$</td>
</tr>
<tr>
<td>$\alpha_2 = \beta_2^*$</td>
<td>$\beta_2 = \beta_6^*$</td>
<td></td>
</tr>
<tr>
<td>$\alpha_3 = \beta_3^*$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_4 = \beta_4^*$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

And for the interactions:

$$(\alpha \beta)_{11} = \beta_7^*, \quad (\alpha \beta)_{12} = \beta_8^*, \quad ..., \quad (\alpha \beta)_{42} = \beta_{14}^*$$

SO...
2-Factor ANOVA Model

Traditional ANOVA form:

\[ Y_{ijk} = \mu + \alpha_j + \beta_j + (\alpha\beta)_{jk} + \epsilon_{ijk} \]

\[
Y_{ijk} = \begin{cases} 
\mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11} + \epsilon_{i11} & \text{if Narrative & text} \\
\mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12} + \epsilon_{i12} & \text{if Narrative & web} \\
\mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21} + \epsilon_{i21} & \text{if Summary & text} \\
\mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22} + \epsilon_{i22} & \text{if Summary & web} \\
\mu + \alpha_3 + \beta_1 + (\alpha\beta)_{31} + \epsilon_{i31} & \text{if Explantion & text} \\
\mu + \alpha_3 + \beta_2 + (\alpha\beta)_{32} + \epsilon_{i32} & \text{if Explantion & web} \\
\mu + \alpha_4 + \beta_1 + (\alpha\beta)_{41} + \epsilon_{i41} & \text{if Arguement & text} \\
\mu + \alpha_4 + \beta_2 + (\alpha\beta)_{42} + \epsilon_{i42} & \text{if Arguement & web}
\end{cases}
\]
Estimation

Least Squares: values that minimize

$$
\sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \epsilon_{ijk}^2 = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (Y_{ijk} - (\text{guess}_{ijk}))^2
$$

$$
= \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (Y_{ijk} - (\mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}))^2
$$

$$
= \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} (Y_{ijk} - \mu_{jk})^2
$$
Estimation

Least Squares: values that minimize

\[ \sum \sum \sum (Y_{ijk} - (\mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}))^2 \]

Solution:

\[ \hat{\mu} = \frac{1}{N} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{ijk} = \bar{Y}. \]

\[ \hat{\alpha}_j = \frac{1}{n.k} \sum_i \sum_k Y_{ijk} - \bar{Y}. = \bar{Y}_j - \bar{Y}. \]

\[ \hat{\beta}_k = \frac{1}{n.j} \sum_i \sum_j Y_{ijk} - \bar{Y}. = \bar{Y}_k - \bar{Y}. \]

\[ (\hat{\alpha}\hat{\beta})_{jk} = \bar{Y}_{jk} - \bar{Y}_j - \bar{Y}_k + \bar{Y}. \]

\[ \hat{\epsilon}_{ijk} = Y_{ijk} - \bar{Y}_{jk} \]
Quick checks...

\[ Y_{ijk} = \hat{\mu} + \hat{\alpha}_j + \hat{\beta}_k + (\hat{\alpha}\hat{\beta})_{jk} + \hat{\epsilon}_{ijk} \]

\[ = \bar{Y}_{..} + (\bar{Y}_{j.} - \bar{Y}_{..}) + (\bar{Y}_{.k} - \bar{Y}_{..}) + (\bar{Y}_{jk} - \bar{Y}_{j.} - \bar{Y}_{.k} + \bar{Y}_{..}) + (Y_{ijk} - \bar{Y}_{jk}) \]

\[ = Y_{ijk} \]

\[ \hat{\mu}_{jk} = \hat{\mu} + \hat{\alpha}_j + \hat{\beta}_k + (\hat{\alpha}\hat{\beta})_{jk} \]

\[ = \bar{Y}_{..} + (\bar{Y}_{j.} - \bar{Y}_{..}) + (\bar{Y}_{.k} - \bar{Y}_{..}) + \bar{Y}_{jk} - \bar{Y}_{j.} - \bar{Y}_{.k} + \bar{Y}_{..} \]

\[ = \bar{Y}_{jk} \]
Statistical Hypotheses

- **Main effects Factor A:**
  \[ H_0(A) : \mu_1 = \mu_2 = \ldots = \mu_J. \]
  or  \[ H_0(A) : \alpha_1 = \alpha_2 = \ldots = \alpha_J \]

- **Main effects Factor B:**
  \[ H_0(B) : \mu_1 = \mu_2 = \ldots = \mu_K \]
  or  \[ H_0(B) : \beta_1 = \beta_2 = \ldots = \beta_K \]

- **Interaction Effects:**
  \[ H_0(AB) : (\alpha_1\beta_1) = (\alpha_1\beta_2) = \ldots = (\alpha_1\beta_K) = \ldots = (\alpha_1\beta_J K) \]

- **Why not** \[ H_0(AB) : \mu_{11} = \mu_{12} = \ldots = \mu_{JK} \]?
For **Balanced designs**

- In 2-Factor ANOVA, the sums of squared deviations from means partitions as

\[ SS_{total} = SS_{model} + SS_{residual} \]

- Now we further partition \( SS_{model} (= SS_{between}) \) into parts due to main effects for Factor \( A \), main effects due to Factor \( B \) and interaction effects between Factors \( A \) and \( B \),

\[ SS_{total} = (SS_A + SS_B + SS_{AB}) + SS_{residual} \]

where...
Sums of Squares and Factorial ANOVA

\[ SS_{total} = \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{i=1}^{n_{jk}} (Y_{ijk} - \bar{Y}.)^2 = (nJK - 1)s^2 \]

\[ SS_A = \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{i=1}^{n_{jk}} (\bar{Y}_{j.} - \bar{Y}.)^2 = nK \sum_j \hat{\alpha}_j^2 \]

\[ SS_B = \sum_{j=1}^{J} \sum_{k=1}^{K} \sum_{i=1}^{n_{jk}} (\bar{Y}_{.k} - \bar{Y}.)^2 = nJ \sum_k \hat{\beta}_k^2 \]

\[ SS_{AB} = \sum_j \sum_k \sum_i (\bar{Y}_{jk} - \bar{Y}_{j.} - \bar{Y}_{.k} + \bar{Y}.)^2 = n \sum_j \sum_k (\hat{\alpha}_j \hat{\beta}_k)^2 \]

\[ SS_{residual} = \sum_j \sum_k \sum_i (Y_{ijk} - \bar{Y}_{jk})^2 = (n - 1) \sum_j \sum_k s_{jk}^2 \]

Qualification for unbalanced designs.
Partitioning the Variance

Just as in 1-Factor ANOVA,

\[
SS_{total} = SS_{model} + SS_{residual}
\]

\[
= (SS_A + SS_B + SS_{AB}) + SS_{residual}
\]

\[
\frac{SS_{total}}{N - 1} = \frac{SS_A}{N - 1} + \frac{SS_B}{N - 1} + \frac{SS_{AB}}{N - 1} + \frac{SS_{residual}}{N - 1}
\]

\[\text{var}(Y_{ijk}) = (\text{due to A}) + (\text{due to B}) + (\text{due to AB}) + (\text{due to error})\]
Variance Explained

Just as in 1-Factor ANOVA,

$$\frac{SS_{total}}{SS_{total}} = \frac{SS_A}{SS_{total}} + \frac{SS_B}{SS_{total}} + \frac{SS_{AB}}{SS_{total}} + \frac{SS_{residual}}{SS_{total}}$$

$$1 = \frac{SS_{model}}{SS_{total}} + \frac{SS_{residual}}{SS_{total}}$$

$$1 = R^2 + (1 - R^2)$$

where $R^2 = \text{squared correlation between } Y_{ijk} \text{ and } \hat{Y}_{ijk} (= \bar{Y}_{jk}).$
Degrees of Freedom

For **Balanced designs**

- $\nu_{total} = N - 1$.
- $\nu_A = J - 1$ because $\sum_{j=1}^{J}(\bar{Y}_{.j} - \bar{Y}..) = 0$.
- $\nu_B = K - 1$ because $\sum_{k=1}^{K}(\bar{Y}_{.k} - \bar{Y}..) = 0$.
- $\nu_{AB} = (J - 1)(K - 1)$ because $\sum_{j}(\hat{\alpha}\hat{\beta})_{jk} = \sum_{k}(\hat{\alpha}\hat{\beta})_{jk} = 0$
- $\nu_{residual} = \nu_{error} = \nu_{\epsilon} = J\cdot K\cdot (n - 1) = N - J\cdot K$

- Like the sums of squares, $\nu_{total} = \nu_A + \nu_B + \nu_{AB} + \nu_{residual}$
Mean Squares

- General:

\[ MS = \frac{\text{(Sum of Square)}}{\text{(Degrees of freedom)}}. \]

- Specific ones for 2-way ANOVA:

\[
MS_A = \frac{SS_A}{\nu_A} \\
MS_B = \frac{SS_B}{\nu_B} \\
MS_{AB} = \frac{SS_{AB}}{\nu_{AB}} \\
MS_{\text{error}} = MS_{\text{within}} = \frac{SS_{\text{error}}}{\nu_{\text{error}}} = \frac{SS_{\text{residual}}}{\nu_{\text{residual}}}
\]
Expected Mean Squares

- Regardless of whether any of the null hypotheses are true,
  \[ E(MS_{error}) = \sigma^2_\epsilon \]

- Consider \( H_{o(A)} : \mu_1 = \mu_2 = \ldots = \mu_J \) or equivalently \( H_{o(A)} : \alpha_1 = \alpha_2 = \ldots = \alpha_J \).
  - If \( H_{o(A)} \) is \textbf{True}, then
    \[ E(MS_A) = \sigma^2_\epsilon \]
  - If \( H_{o(A)} \) is \textbf{False}, then
    \[ E(MS_A) = \sigma^2_\epsilon + \frac{nK \sum_{j=1}^{J} \alpha^2_j}{J - 1} \]
Expected Mean Squares (continued)

- Consider $H_{o(B)} : \mu_1 = \mu_2 = \ldots = \mu_K$ or equivalently $H_{o(B)} : \beta_1 = \beta_2 = \ldots = \beta_K$.

  - If $H_{o(B)}$ is True, then
    \[ E(MS_B) = \sigma^2 \]
  
  - If $H_{o(B)}$ is False, then
    \[ E(MS_B) = \sigma^2 + \frac{nJ \sum_{k=1}^{K} \beta^2_k}{K - 1} \]
Expected Mean Squares (continued)

- Consider interaction null hypothesis
  \[ H_{o(AB)} : (\alpha \beta)_{11} = (\alpha \beta)_{12} = \ldots = (\alpha \beta)_{jk} = \ldots = (\alpha \beta)_{JK} \]

- If \( H_{o(AB)} \) is True, then
  \[ E(MS_{AB}) = \sigma^2 \]

- If \( H_{o(AB)} \) is False, then
  \[ E(MS_{AB}) = \sigma^2 + \frac{n \sum_{j=1}^{J} \sum_{k=1}^{K} (\alpha \beta)_{jk}^2}{(J - 1)(K - 1)} \]
\( F \)-Ratios

- To test a null hypothesis, we form ratios of mean squares such that if an \( H_0 \) is true, then the ratio is expected to equal one,

\[
    F = \frac{M S_{hypothesis}}{M S_{error}} = \frac{\hat{\sigma}^2}{\hat{\sigma}^2}
\]

- If an \( H_0 \) is false, then the ratio is expected to greater than 1,

\[
    F = \frac{M S_{hypothesis}}{M S_{error}} = \frac{\hat{\sigma}^2 + (\text{treatment effects})}{\hat{\sigma}^2}
\]
Sampling Distribution of an $F$-Ratio

- Regardless of whether any null hypothesis is true,

\[ \frac{SS_{\text{error}}}{\sigma^2_\epsilon} \sim \chi^2_{\nu_{\text{error}}} \quad \text{so} \quad \frac{MS_{\text{error}}}{\sigma^2_\epsilon} \sim \chi^2_{\nu_{\text{error}}}/\nu_{\text{error}} \]

- As an example, Main Effect for Factor A: If $H_{o(A)}$ is true,

\[ \frac{SS_A}{\sigma^2_\epsilon} \sim \chi^2_{\nu_A} \quad \text{so} \quad \frac{SS_A/\nu_A}{\sigma^2_\epsilon} = \frac{MS_A}{\sigma^2_\epsilon} \sim \chi^2_{\nu_A}/\nu_A \]

- $F$-ratio for Factor A if $H_{o(A)}$ is true,

\[ F_{\nu_A,\nu_{\text{error}}} = \frac{MS_A}{\sigma^2_\epsilon} / MS_{\text{error}} \sim \frac{\chi^2_{\nu_A}/\nu_A}{\chi^2_{\nu_{\text{error}}}/\nu_{\text{error}}} \sim F_{\nu_A,\nu_{\text{error}}} \]

- Expected value of $F$ ratio when $H_{o(A)}$ is False,

\[ F_{\nu_A,\nu_{\text{error}}} = \frac{MS_A}{MS_{\text{error}}} = \frac{\sigma^2_\epsilon + \frac{nK \sum_{j=1}^{J} \alpha_j^2}{J-1}}{\sigma^2_\epsilon} \]
Sampling Distribution of an $F$-Ratio

- For Factor $B$, if $H_{o(B)}$ is true, then

$$F_{\nu_B,\nu_{error}} = \frac{MS_B}{MS_{error}} \sim F_{\nu_B,\nu_{error}}$$

- For the interaction, if $H_{o(AB)}$ is true, then

$$F_{\nu_{AB},\nu_{error}} = \frac{MS_{AB}}{MS_{error}} \sim F_{\nu_{AB},\nu_{error}}$$
### Summary Table for 2-Factor ANOVA

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor A</td>
<td>( J - 1 )</td>
<td>( nK \sum_j (\bar{Y}<em>{j.} - \bar{Y}</em>{..})^2 )</td>
<td>( \frac{SS_A}{\nu_A} )</td>
<td>( \frac{MS_A}{MSError} )</td>
</tr>
<tr>
<td>Factor B</td>
<td>( K - 1 )</td>
<td>( nJ \sum_k (\bar{Y}<em>{.k} - \bar{Y}</em>{..})^2 )</td>
<td>( \frac{SS_B}{\nu_B} )</td>
<td>( \frac{MS_B}{MSError} )</td>
</tr>
<tr>
<td>Interaction</td>
<td>((J - 1)(K - 1))</td>
<td>( n \sum_j \sum_k (\bar{Y}<em>{jk} - \bar{Y}</em>{..})^2 )</td>
<td>( \frac{SS_{AB}}{\nu_{AB}} )</td>
<td>( \frac{MS_{AB}}{MSError} )</td>
</tr>
<tr>
<td>Error</td>
<td>( JK(n - 1) )</td>
<td>( \sum_j \sum_k \sum_i (Y_{ijk} - \bar{Y}_{jk})^2 )</td>
<td>( \frac{SS_{error}}{\nu_{error}} )</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>( nJK - 1 )</td>
<td>( \sum_j \sum_k \sum_i (Y_{ijk} - \bar{Y}_{..})^2 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Assumptions are that \( Y_{ijk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} + \epsilon_{ijk} \) and \( \epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2_{\epsilon}) \) & independent.
Example 2-Way ANOVA


- **Response/dependent variable** = understanding as measured by 10 item inference verification test (IVT), $Y_i = IVT_i$.

- **Factors**:
  - Format (text or web)
  - Instructions participants received: write a Narrative (N), Summary (S), Explanation (E), Argument (A).
# Example 2-Way ANOVA

Summary statistics: $\bar{Y}$ and $s^2$, where $n_{cell} = 8$.

<table>
<thead>
<tr>
<th>Instructions</th>
<th>Format</th>
<th>N</th>
<th>S</th>
<th>E</th>
<th>A</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Text</td>
<td>71.25</td>
<td>72.5</td>
<td>68.75</td>
<td>73.75</td>
<td>71.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>126.79</td>
<td>164.29</td>
<td>69.64</td>
<td>141.07</td>
<td>116.83</td>
</tr>
<tr>
<td></td>
<td>Web</td>
<td>76.25</td>
<td>73.75</td>
<td>72.5</td>
<td>90.0</td>
<td>78.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>255.36</td>
<td>255.36</td>
<td>107.14</td>
<td>114.29</td>
<td>215.73</td>
</tr>
<tr>
<td></td>
<td>Totals</td>
<td>73.75</td>
<td>73.13</td>
<td>70.63</td>
<td>81.88</td>
<td>74.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>185.00</td>
<td>196.25</td>
<td>86.25</td>
<td>189.58</td>
<td>174.58</td>
</tr>
</tbody>
</table>

$n_{j} = 4(8) = 32$, $n_{.k} = 2(8) = 16$ & $N = 2(4)(8) = 96$.

Mean Plot

Wiley & Voss Data: Mean Plot

Percent Correction Inferences

Writing task (1—nar, 2—sum, 3—expl, 4—arg)

Presentation format (1—chapter, 2—web)
Mean Plot

Presentation format (1—chapter, 2—web)
Estimate the effects

- Grand Mean: $\hat{\mu} = \bar{Y}.. = 74.84$

- Main Effects Format (F):
  
  Chapter: $\hat{\alpha}_1 = \bar{Y}_1. - \bar{Y}.. = 71.56 - 74.84 = -3.28$
  Web: $\hat{\alpha}_2 = \bar{Y}_2. - \bar{Y}.. = 78.13 - 74.84 = 3.28$

- Main Effects Instruction (I):
  
  Narrative: $\hat{\beta}_1 = \bar{Y}_1. - \bar{Y}.. = 73.75 - 74.84 = -1.09$
  Summary: $\hat{\beta}_2 = \bar{Y}_2. - \bar{Y}.. = 73.84 - 74.84 = -1.72$
  Explanation: $\hat{\beta}_3 = \bar{Y}_3. - \bar{Y}.. = 70.63 - 74.84 = -4.22$
  Argument: $\hat{\beta}_4 = \bar{Y}_4. - \bar{Y}.. = 81.88 - 74.84 = 7.03$
Estimate the effects

Interaction Effects (FI):

\[
(\hat{\alpha\beta})_{11} = \bar{Y}_{11} - \bar{Y}_1 - \bar{Y}_1 + \bar{Y}_. = 71.25 - 71.56 - 73.75 + 74.84 = 0.78
\]

\[
(\hat{\alpha\beta})_{12} = \bar{Y}_{12} - \bar{Y}_1 - \bar{Y}_2 + \bar{Y}_. = 72.50 - 71.56 - 73.13 + 74.84 = 2.66
\]

\[
(\hat{\alpha\beta})_{13} = \bar{Y}_{13} - \bar{Y}_1 - \bar{Y}_3 + \bar{Y}_. = 68.75 - 71.56 - 70.63 + 74.84 = 1.40
\]

\[
(\hat{\alpha\beta})_{14} = \bar{Y}_{14} - \bar{Y}_1 - \bar{Y}_3 + \bar{Y}_. = 73.75 - 71.56 - 81.88 + 74.84 = -4.84
\]
Entries in the cells are \((\hat{\alpha}\beta)_{jk}\)

<table>
<thead>
<tr>
<th>Format</th>
<th>N</th>
<th>S</th>
<th>E</th>
<th>A</th>
<th>(\hat{\alpha}_j)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>.78</td>
<td>2.66</td>
<td>1.41</td>
<td>-4.84</td>
<td>-3.28</td>
</tr>
<tr>
<td>Web</td>
<td>- .78</td>
<td>-2.66</td>
<td>-1.41</td>
<td>4.84</td>
<td>3.28</td>
</tr>
<tr>
<td>(\hat{\beta}_k)</td>
<td>-1.09</td>
<td>-1.72</td>
<td>-4.22</td>
<td>7.03</td>
<td>(\hat{\mu} = 74.84)</td>
</tr>
</tbody>
</table>

\[ \hat{Y}_{ijk} = \hat{\mu} + \hat{\alpha}_j + \hat{\beta}_k + (\hat{\alpha}\beta)_{jk} = \bar{Y}_{jk} \]
Computing Sums of Squares

- **Total corrected for grand mean,**
  \[
  SS_{total} = \sum_{j} \sum_{k} \sum_{i} (Y_{ijk} - \bar{Y})^2
  \]
  \[
  = (N - 1)s^2 = (64 - 1)(174.58) = 19,998.44
  \]

- **Sum of squares for error (residual, within),**
  \[
  SS_{error} = \sum_{j} \sum_{k} \sum_{i} (Y_{ijk} - \bar{Y}_{jk})^2 = \sum_{j} \sum_{k} (n - 1)s^2_{jk}
  \]
  \[
  = (8 - 1)(126.78 + 164.29 + 69.64 + 141.07 + 255.36 + 255.36 + 107.14 + 114.29)
  \]
  \[
  = \frac{7(1233.93)}{7} = 8,637.50
  \]
Computing Sums of Squares

Main Effects Format,

\[ SS_F = nK \sum_j (\bar{Y}_j - \bar{Y}.)^2 = nK \sum_j \hat{\alpha}_j^2 \]

= \((8)(4)((-3.28)^2 + (3.28)^2) = 689.06\)

Main Effect Instruction,

\[ SS_I = nJ \sum_k (\bar{Y}_k - \bar{Y}.)^2 = nJ \sum_k \hat{\beta}_k^2 \]

= \((8)(2)((-1.09)^2 + (-1.72)^2 + (-4.22)^2 + (7.03)^2) = 1,142.19\)
Computing Sums of Squares

- Interaction between Format and Instruction,

\[ SS_{IF} = n \sum_j \sum_k (\bar{Y}_{jk} - \bar{Y}_j - \bar{Y}_k + \bar{Y}_{..})^2 \]

\[ = n \sum_j \sum_k (\hat{\alpha}\hat{\beta}_{jk})^2 \]

- Or if we’re pretty sure we did our computations correctly,

\[ SS_{IF} = SS_{total} - SS_F - SS_I - SS_{error} \]

\[ = 10,998.44 - 1,142.19 - 689.06 - 8,637.50 \]

\[ = 529.69 \]
Degrees of Freedom

- Main effect Format: $\nu_F = (2 - 1) = 1$
- Main effect Instruction: $\nu_I = (4 - 1) = 3$
- Interaction between Format & Instruction: $\nu_{FI} = (2 - 1)(4 - 1) = 3$
- Error: $\nu_{error} = 2(4)(8 - 1) = 56$
- Total (corrected): $\nu_{total} = nJK - 1 = 64 - 1 = 63$
## ANOVA Summary Table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructions</td>
<td>1</td>
<td>1,142.19</td>
<td>380.73</td>
<td>2.47</td>
<td>.07</td>
</tr>
<tr>
<td>Format</td>
<td>3</td>
<td>689.06</td>
<td>689.06</td>
<td>4.47</td>
<td>.04</td>
</tr>
<tr>
<td>Interaction (FI)</td>
<td>3</td>
<td>529.69</td>
<td>176.56</td>
<td>1.14</td>
<td>.34</td>
</tr>
<tr>
<td>Error</td>
<td>56</td>
<td>8,637.44</td>
<td>154.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total (corrected)</strong></td>
<td>63</td>
<td>10,998.45</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:**

From 2-way ANOVA: \[ \hat{\sigma}_e^2 = 154.24 \]

(Instructions & Format)

From 1-way ANOVA: \[ \hat{\sigma}_e^2 = 164.27 \]

(only Instructions)
SAS and Factorial ANOVA

- **Commands:**

  ```
  PROC GLM DATA=ivt;
  CLASS format instruct;
  MODEL ivt = format instruct format*instruct ;
  MEANS format instruct format*instruct;
  TITLE 'Wiley Voss Data';
  RUN;
  ```

- **ANALYST demonstration.**
Unbalanced Designs

- Not a problem for 1-Factor ANOVA.
- Have to give careful thought to how to deal with.
- If $n_{ij} \neq (n_j + n_k)/N$, then the sum of squares total does not (uniquely) partition into three (independent) parts.
- Example: I deleted some observations from the Wiley and Voss data so that we have unbalanced data and we get...
Unbalanced 2-Way ANOVA

Summary statistics: $\bar{Y}$ and $s$, where $n_{cell} = 8$, except $n_{11} = 4$, $n_{21} = 5$, and $n_{24} = 3$.

<table>
<thead>
<tr>
<th>Instructions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Format</strong></td>
</tr>
<tr>
<td><strong>Text</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Web</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Totals</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Unbalanced 2-Way ANOVA: SS’s

**Sum of Squares: Format,**

\[
SS_{\text{format}} = 28(72.14 - 75.00)^2 + 24(78.33 - 75.00)^2
\]
\[
= 28(-2.86)^2 + 24(3.33)^2
\]
\[
= 495.1624
\]

**Sum of Squares: Instructions,**

\[
SS_{\text{instruc}} = 9(80.00 - 75.00)^2 + 16(73.13 - 75.00)^2
\]
\[
+ 16(70.63 - 75.00)^2 + 11(80.00 - 75.00)^2
\]
\[
= 9(5)^2 + 16(-1.87)^2 + 16(-4.37)^2 + 11(5)^2
\]
\[
= 861.5008
\]

**Note:** there’s quite a bit of rounding error because I only retained 2 decimal places.
Unbalanced 2-Way ANOVA: SS’s

**Sum of Squares: Interaction between Format & Instructions,**

\[
SS_{FI} = 4(75.00 - 72.14 - 80.00 + 75.00)^2 + 8(72.50 - 72.14 - 73.13 + 75.00)^2 \\
+ 8(68.75 - 72.14 - 70.63 + 75.00)^2 + 8(73.75 - 72.14 - 80.00 + 75.00)^2 \\
+ 5(84.00 - 78.33 + 80.00 + 75.00)^2 + 8(73.75 - 78.33 + 73.13 + 75.00)^2 \\
+ 8(72.50 - 78.33 - 70.63 + 75.00)^2 + 3(96.67 - 78.33 - 80.00 + 75.00)^2 \\
= 4(-2.14)^2 + 8(2.48)^2 + 8(0.98)^2 + 8(-3.39)^2 \\
+ 5(0.67)^2 + 8(-2.71)^2 + 8(-1.46)^2 + 3(13.34)^2 \\
= 779.0585
\]
## Unbalanced 2-Way ANOVA

### ANOVA Summary Table

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instructions</td>
<td>1</td>
<td>495.61</td>
<td>967.70</td>
<td>6.81</td>
<td>.01</td>
</tr>
<tr>
<td>Format</td>
<td>3</td>
<td>861.50</td>
<td>478.30</td>
<td>3.37</td>
<td>.03</td>
</tr>
<tr>
<td>Interaction (FI)</td>
<td>3</td>
<td>779.06</td>
<td>245.11</td>
<td>1.73</td>
<td>.18</td>
</tr>
<tr>
<td>Error</td>
<td>44</td>
<td>6249.17</td>
<td>142.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total (corrected)</strong></td>
<td><strong>51</strong></td>
<td><strong>8,500.00</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

However,

\[
495.61 + 861.50 + 779.06 + 6249.17 = 8,385.34 \neq 8,500
\]
The overlay represents covariances (correlations, negative or positive) between the effects that results from non-orthogonal design.
Unbalanced Designs

There are 4 possibilities:

- **Type I** $SS$.
- **Type II** $SS$ (briefly).
- **Type III** $SS$.
- **Type IV** $SS$ (won’t talk about this one).
Type I Sums of Squares

Hierarchical Method

- First compute $SS_A$ (for example) in the usual way
  $$SS_A = t + u + v + w$$

- Next compute $SS_B$ adjusting for factor $A$,
  $$SS_B|A = x + y$$

- Compute $SS_{AB}$ adjusting for factors $A$ and $B$
  $$SS_{AB}|A,B = z$$

- Order matters!
### Type I Sums of Squares: Hierarchical

#### Model
```
model ivt = format instruct format*instruct;
```

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>FORMAT</td>
<td>1</td>
<td>495.238095</td>
<td>495.238095</td>
<td>3.49</td>
<td>.0685</td>
</tr>
<tr>
<td>INSTRUCT</td>
<td>3</td>
<td>1020.274934</td>
<td>340.091645</td>
<td>2.39</td>
<td>.0811</td>
</tr>
<tr>
<td>FORMAT*INSTRUCT</td>
<td>3</td>
<td>735.320304</td>
<td>245.106768</td>
<td>1.73</td>
<td>.1755</td>
</tr>
</tbody>
</table>

#### Model
```
model ivt = instruct format format*instruct;
```

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type I SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>INSTRUCT</td>
<td>3</td>
<td>862.50000000</td>
<td>287.5000000</td>
<td>2.02</td>
<td>.1243</td>
</tr>
<tr>
<td>FORMAT</td>
<td>1</td>
<td>653.0130293</td>
<td>653.0130293</td>
<td>4.60</td>
<td>.0376</td>
</tr>
<tr>
<td>FORMAT*INSTRUCT</td>
<td>3</td>
<td>735.3203040</td>
<td>245.1067680</td>
<td>1.73</td>
<td>.1755</td>
</tr>
</tbody>
</table>

In both cases, $S_I + SS_F + SS_{IF} = 2250.82$, which is $SS_{model}$. ```
Type III Sums of Squares

“Regression Strategy”: Compute each sums of squares adjusting for everything else in the model.

\[ SS_{A|B,AB} = t. \]
\[ SS_{B|A,AB} = x. \]
\[ SS_{AB|A,B} = z. \]

- Can make decisions for the hypothesis regardless of order, etc.
- This will make more sense after studying regression.
### Type III Sums of Squares

**Regression Strategy:**

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>INSTRUCT</td>
<td>3</td>
<td>1434.913138</td>
<td>478.304379</td>
<td>3.37</td>
<td>.0268</td>
</tr>
<tr>
<td>FORMAT</td>
<td>1</td>
<td>967.697239</td>
<td>967.697239</td>
<td>6.81</td>
<td>.0123</td>
</tr>
<tr>
<td>FORMAT*INSTRUCT</td>
<td>3</td>
<td>735.320304</td>
<td>245.106768</td>
<td>1.73</td>
<td>.1755</td>
</tr>
</tbody>
</table>
Type II Sums of Squares

Appelbaum & Cramer main advocates of this.

- Underlying assumption is the interaction effect is zero in the population.
- Adjust $SS$ for all others of the same or lower order.
- $SS_{A|B} = t + w$.
- $SS_{B|A} = x + y$.
- $SS_{AB|A,B} = z$. 
The same as for 1-Factor ANOVA

- Normality
- Homogeneity of variance
- Independence
3 and Higher Factor ANOVA

- Adding a third factor → adds more terms to our model.

- ANOVA model for 3 factors:

\[ Y_{ijkl} = \mu_{jkl} + \epsilon_{ijkl} \]

where \( \epsilon \sim \mathcal{N}(0, \sigma^2) \).

- We specify a model for \( \mu_{jkl} \).

- Some Possible models for \( \mu_{jkl} \):
  - Main effects or additive
  - 2-way interaction
  - 3-way interaction
Main effects or additive:

\[ \mu_{jkl} = \mu + \alpha_j + \beta_k + \gamma_l \]

2-way interactions:

\[ \mu_{jkl} = \mu + \alpha_j + \beta_k + \gamma_l + (\alpha\beta)_{jk} + (\alpha\gamma)_{jl} + (\beta\gamma)_{kl} \]

3-way interaction:

\[ \mu_{jkl} = \mu + \alpha_j + \beta_k + \gamma_l + (\alpha\beta)_{jk} + (\alpha\gamma)_{jl} + (\beta\gamma)_{kl} + (\alpha\beta\gamma)_{jkl} \]
3-Factor Interaction

The 2-way means are not representative of the pattern in the cells of the table.

Example of a 3-Factor Interaction
Higher-Way Designs

Some issues with higher-way designs:

- If cell size is $n = 1$ the,
  - You can’t test the highest way interaction because $\nu_{error} = 0$
  - The interaction sums of squares is used as an estimate of “error”, which means you have to assume that you do not have an interaction.

- To pool or not to pool?