We have looked at 1-sample, 2-sample, and \( t \)-sample problems. We dealt with a treatment at \( t \) levels or with \( t \) treatments. A treatment with \( t \) levels could also be viewed as a **factor**. Now address experiments where several factors come into play. First we will do this for two such factors. How should we go about this?
### Insecticide Data

```r
> poison
```
y type delivery
1 3.1 I A
2 4.5 I A
3 4.6 I A
4 4.3 I A
5 3.6 II A
6 2.9 II A
7 4.0 II A
8 2.3 II A
9 2.2 III A
.....
44 3.8 II D
45 3.0 III D
46 3.6 III D
47 3.1 III D
48 3.3 III D```
We have 3 types of insecticides (I, II, and III) and 4 methods (A, B, C, D) of delivering the insecticide.

(I, II, and III) and (A, B, C, D) are the levels of the respective factors type of insecticide and insecticide delivery method.

The response $Y$ is the time to death in minutes.

Find the best insecticide and the best delivery method.

We have 48 experimental insects to experiment with.

Randomly divide the 48 insects into $12 = 3 \times 4$ groups of 4 insects each, assigning the respective groups to the 12 treatment combinations (I, A), (I, B), (I, C), (I, D), ..., (III, C), (III, D).

Randomize the order of all 48 runs to eliminate order biases.

This is a factorial design, specifically a $3 \times 4$ factor design with 4 replications.
Response Table

- Visualize responses in relation to the factor levels as follows:

<table>
<thead>
<tr>
<th>Insecticide Type</th>
<th>Delivery Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>I</td>
<td>$Y_{I,A}$</td>
</tr>
<tr>
<td>II</td>
<td>$Y_{II,A}$</td>
</tr>
<tr>
<td>III</td>
<td>$Y_{III,A}$</td>
</tr>
</tbody>
</table>

- $Y_{I,A}$ short for $(Y_{I,A,1}, \ldots, Y_{I,A,4})$ (replication depth), etc.
- More generically we would denote the $k^{th}$ response under level $i$ from factor 1 and under level $j$ from factor 2 by $Y_{ijk}$.
- Triplet notation $(i,j,k)$ is more useful than $\ell = 1,2,\ldots,48$.
- Useful in $\Sigma$ summation notation and also for identifying the factor levels, i.e., the factor level/replication coordinates.
First Look at Insecticide Boxplots

![Boxplot for type of poison](image1)

- Type I
- Type II
- Type III

![Boxplot for delivery method](image2)

- Method A
- Method B
- Method C
- Method D
First Impressions and Questions

- Insecticide type III and delivery method A seem to give the best combination.
- Is combination the right word here?
- Are the effects of delivery consistent across types, i.e., is the delivery effectiveness order (in terms of faster response time) the same from one insecticide type to another?
- It could be that delivery type A is not the fastest acting among all four when applied to insecticide type III.
- Method C could actually be better in combination with III.
Insecticide Responses by Delivery Method

delivery method A

delivery method B

delivery method C

delivery method D
Full Comparison of Insecticide Responses

time to death (minutes)

I.A I.B II.B III.B I.C II.C III.C I.D II.D III.D

type.delivery combinations
Insecticide type III seems to have lowest response with all 4 delivery methods.
The mean levels for each (III, delivery method) combination are \( \approx \) consistent.
The scatter within each (III, delivery method) combination is quite tight.
Delivery appears to have an effect on the response under type I and II, both in absolute terms and relative to each other.
It appears: scatter \( \uparrow \) as mean \( \uparrow \) across all combinations.
\( \Rightarrow \) variance stabilizing transformation. Deal with that first.
Linear Fit \[ \log(s_i) = a \times \log(\hat{\mu}_i) + b \]
Reciprocal Transform

- According to guideline: ⇒ \( \alpha = 2 \) or \( \lambda = 1 - \alpha = -1 \), i.e.,
  \( \tilde{Y}_{ijk} = Y_{ijk}^{\alpha} = 1 / Y_{ijk} \) a reciprocal transform for our response.

- A rationalization attempt: Suppose the absorption rate
  \( R = d / t \) (of dose \( d \) over time \( t \)) under any given combination
  is the most variable process aspect from insect to insect.

- Assume that this absorption variability (ingestion variability
  from insect to insect) is constant across all (insecticide type,
  delivery method) combinations.

- Assume lethal dose \( D \) is \( \approx \) constant for each type, \( \forall \) insects.

- Then the time to reach lethal dose is \( T = D / R \).

- The transformed \( 1 / T = R / D \) has \( \approx \) constant variability.

- A 1-term Taylor expansion around \( \mu_R \) with \( D \approx \) constant

\[
T = \frac{D}{R} \approx \frac{D}{\mu_R} - (R - \mu_R) \frac{D}{\mu_R^2} \Rightarrow \mu_T \approx \frac{D}{\mu_R}, \quad \text{var}(T) \approx \frac{D^2 \sigma_R^2}{\mu_R^4} \Rightarrow \sigma_T \propto \mu_T^2
\]
The variability is much more stable now.
Reciprocal Time by Delivery Method

delivery method A

reciprocal time to death (1/minutes)

I II III

0.0 0.1 0.2 0.3 0.4 0.5 0.6

●

●

●

●

●

●

●

●

reciprocal time to death (1/minutes)

delivery method B

I II III

0.0 0.1 0.2 0.3 0.4 0.5 0.6

●

●

●

●

●

●

●

●

reciprocal time to death (1/minutes)

delivery method C

I II III

0.0 0.1 0.2 0.3 0.4 0.5 0.6

●

●

●

●

●

●

●

●

reciprocal time to death (1/minutes)

delivery method D

I II III

0.0 0.1 0.2 0.3 0.4 0.5 0.6

●

●

●

●

●

●

●

●
Full Comparison of Reciprocal Times

Note the consistent variability across all 12 treatment combinations!
\[ \log(s_i) = a \times \log(\hat{\mu}_i) + b \] for Reciprocal Times

No strong linearity remains!
Assume `recip.time, type, delivery` are in the workspace. `type` and `delivery` are factors or write `as.factor(type)` in place of `type`!

```r
> anova(lm(recip.time ~ type))
```
or

```r
> anova(lm(recip.time ~ type, data=poison))
```

Analysis of Variance Table

```
Response: recip.time

             Df  Sum Sq Mean Sq  F value    Pr(>F)
type          2 0.34877 0.17439  25.6211 3.728e-08 ***
Residuals    45 0.30628 0.00681
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```
```r
> anova(lm(recip.time ~ delivery))

or

> anova(lm(recip.time ~ delivery, data=poison))

Analysis of Variance Table
Response: recip.time

<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>delivery</td>
<td>3</td>
<td>0.20414</td>
<td>0.06805</td>
<td>6.6401</td>
<td>0.0008496 ***</td>
</tr>
<tr>
<td>Residuals</td>
<td>44</td>
<td>0.45091</td>
<td>0.01025</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```
ANOVA for Reciprocal Times
vs All Type:Delivery Combinations

> anova(lm(recip.time ~ type:delivery))
or
> anova(lm(recip.time ~ type:delivery, data=poison))

Analysis of Variance Table
Response: recip.time

<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>type:delivery</td>
<td>11</td>
<td>0.56862</td>
<td>0.05169</td>
<td>21.531</td>
<td>1.289e-12 ***</td>
</tr>
<tr>
<td>Residuals</td>
<td>36</td>
<td>0.08643</td>
<td>0.00240</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- This is like a one-way ANOVA with 12 treatments.
- Legitimate but not enlightening.
### $\text{ANOVA.type}$ Analysis of Variance Table

**Response:** recip.time  

<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>type</td>
<td>2</td>
<td>0.34877</td>
<td>0.17439</td>
<td>25.621</td>
<td>3.728e-08 ***</td>
</tr>
<tr>
<td>Residuals</td>
<td>45</td>
<td>0.30628</td>
<td>0.00681</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### $\text{ANOVA.delivery}$ Analysis of Variance Table

**Response:** recip.time  

<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>delivery</td>
<td>3</td>
<td>0.20414</td>
<td>0.06805</td>
<td>6.6401</td>
<td>0.0008496 ***</td>
</tr>
<tr>
<td>Residuals</td>
<td>44</td>
<td>0.45091</td>
<td>0.01025</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### $\text{ANOVA.type.delivery}$ Analysis of Variance Table

**Response:** recip.time  

<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>type:delivery</td>
<td>11</td>
<td>0.56862</td>
<td>0.05169</td>
<td>21.531</td>
<td>1.289e-12 ***</td>
</tr>
<tr>
<td>Residuals</td>
<td>36</td>
<td>0.08643</td>
<td>0.00240</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Are these Analyses Appropriate?

- What does $MS_E$ represent in the first two ANOVAs?
- Compare these values to $MS_E$ in the third ANOVA.
- The $MS_E$’s in the first two ANOVAs are inflated because mean variation in the ignored factor is part of the error variation.
- On slide 14 note how the variation within each of the 4 delivery groups also reflects the response variation due to poison type (color).
- The variation within each of the three colors (poison type) also reflects the variation due to delivery method.
- Note that the p-value in the third ANOVA is very much smaller than in the other two ANOVAs.
- This is due to appropriately smaller $MS_E$ here.
The Third ANOVA

- The third ANOVA is technically correct in stating that the means change across factor level combinations.
- View the $3 \cdot 4$ combinations as $t = 12$ treatments/samples.
- Does the third ANOVA give any insight on the separate contributions of the type factor and the delivery factor?
  - No! It is insufficient in that regard.
- It is easy to conceive of situations where the 1$^{st}$ and 2$^{nd}$ ANOVA produce insignificant $F$-values but the 3$^{rd}$ ANOVA produces a highly significant $F$-value.
- This can result when the $MS_E$ in the first two ANOVAs are unduly inflated, i.e., $\gg MS_E$ in the third ANOVA.
- The next 5 slides illustrate this with doctored data.
The delivery variation is a good part of the “error” variation within type.
The type variation is a good part of the “error” variation within delivery.
3 type data sets seem well meshed when ignoring the 4 delivery methods.
4 delivery methods seem well meshed when ignoring the 3 poison types.
### Three ANOVAs for Modified Data

#### ANOVA.type 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>type</td>
<td>2</td>
<td>0.03592</td>
<td>0.01796</td>
<td>1.5582</td>
<td>0.2217</td>
</tr>
<tr>
<td>Residuals</td>
<td>45</td>
<td>0.51868</td>
<td>0.01153</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### ANOVA.delivery 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>delivery</td>
<td>3</td>
<td>0.08425</td>
<td>0.02808</td>
<td>2.6271</td>
<td>0.06208</td>
</tr>
<tr>
<td>Residuals</td>
<td>44</td>
<td>0.47035</td>
<td>0.01069</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### ANOVA.type.delivery 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>type:delivery</td>
<td>11</td>
<td>0.46817</td>
<td>0.04256</td>
<td>17.727</td>
<td>2.294e-11 ***</td>
</tr>
<tr>
<td>Residuals</td>
<td>36</td>
<td>0.08643</td>
<td>0.00240</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The first two ANOVAs confirm that neither factor alone (type or delivery) shows a significant effect (p-value \( \leq \alpha = .05 \)).

This confirms the meshing comments made below the last two plots.

By the data changes we aligned the means of all compared data sets more closely, but the variation from the ignored factor still inflates the \( MS_E \), leading to non-significant results.

The third ANOVA shows a highly significant effect of type:delivery combination. This is effected by the much reduced \( MS_E \) here (0.00240 as compared to 0.01153 or 0.01069).
For now we will deal only with the balanced model.

Same number $n$ of replications per factor level combination.

Unbalanced cases can get quite messy.\footnote{For an extensive treatment of the unbalanced case see S.R. Searle (1987), \textit{Linear Models for Unbalanced Data}, John Wiley & Sons, New York}

If time permits, may touch back on that later.

Additive Effects Model:

$$Y_{ijk} = \mu + a_i + b_j + \epsilon_{ijk}, \quad i = 1, \ldots, t_1, \quad j = 1, \ldots, t_2, \quad k = 1, \ldots, n$$

with independent error terms $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$
The parameters $\mu, a_1, \ldots, a_{t_1}, b_1, \ldots, b_{t_2}$, are unidentifiable.

Adding a constant $c$ to $\mu$ and subtracting it from the $a_i$ ($b_j$) give the same means $\mu_{ij} = E(Y_{ijk})$ for all $(i, j)$.

$$\mu_{ij} = \mu + a_i + b_j = \mu + c + a_i - c + b_j = \mu' + a'_i + b_j$$

As in the 1-way ANOVA we only need $t_1 - 1$ parameters to distinguish between $t_1$ row levels, and similarly $t_2 - 1$ parameters to distinguish between $t_2$ column levels.

There are two customary ways of imposing side conditions that deal with this, i.e., to render all parameters as identifiable.

$$\sum_i a_i = 0 \quad \text{and} \quad \sum_j b_j = 0 \quad \text{sum-to-zero side conditions}$$

$$a_1 = 0 \quad \text{and} \quad b_1 = 0 \quad \text{set-to-zero side conditions}$$
Assume $\mu_{ij} = \mu + a_i + b_j$ with $\sum_i a_i = \sum_j b_j = 0$ or $\bar{a}_i = \bar{b}_j = 0$.

Here we identify $\mu$ with the average mean over all level combinations, because $\bar{\mu}_{..} = \mu + \bar{a}_i + \bar{b}_j = \mu$.

Since $\bar{\mu}_{i.} = \mu + a_i + \bar{b}_j = \mu + a_i$ we can interpret $a_i = \bar{\mu}_{i.} - \mu = \bar{\mu}_{i.} - \bar{\mu}_{..}$ as the average change from $\mu = \bar{\mu}_{..}$ due to level $i$ of factor one when averaged over all levels of factor two.

Similarly, $b_j = \bar{\mu}_{..} - \mu = \bar{\mu}_{..} - \bar{\mu}_{..}$.

The parameters $\mu, a_i, b_j$ with $\bar{a}_i = \bar{b}_j = 0$ define the means $\mu_{ij}$ with the above additive structure.

They are uniquely identified via the $\mu_{ij}$, as shown above.

Changes in both levels are additive $\mu_{ij} = \mu + a_i + b_j$
Set-to-zero Side Conditions

- Assume \( \mu_{ij} = \mu^* + a_i^* + b_j^* \) with \( a_1^* = b_1^* = 0 \).
- Here we identify the parameter \( \mu^* \) with the mean under the factor level combination \((1, 1)\): \( \mu_{11} = \mu^* + a_1^* + b_1^* = \mu^* \).
- We express each change from \( \mu^* \) due to other levels \( (\neq 1) \) in factor one via \( a_i^* \), i.e., \( \mu_{i1} = \mu^* + a_i^* + b_1^* = \mu^* + a_i^* \) and thus \( a_i^* = \mu_{i1} - \mu^* = \mu_{i1} - \mu_{11} \).
- Each change from \( \mu^* \) due to other levels \( (\neq 1) \) in factor two is expressed via \( b_j^* \), i.e., \( \mu_{1j} = \mu^* + a_1^* + b_j^* = \mu^* + b_j^* \) and thus \( b_j^* = \mu_{1j} - \mu^* = \mu_{1j} - \mu_{11} \).
- The parameters \( \mu^*, a_i^*, b_j^* \) with \( a_1^* = b_1^* = 0 \) define the means \( \mu_{ij} \) with the above additive structure.
- They are uniquely identified via the \( \mu_{ij} \), as shown above.
- Changes in both levels are additive \( \mu_{ij} = \mu^* + a_i^* + b_j^* \). 

Suppose you want to simulate such data.

How would you create such mean structures?

Pick any \( t_1 - 1 + t_2 - 1 + 1 = t_1 + t_2 - 1 \) numbers \( a_2, \ldots, a_{t_1}, b_2, \ldots, b_{t_2} \), and \( \mu \).

In the sum-to-zero case take

\[
a_1 = - \sum_{i=2}^{t_1} a_i \quad \text{and} \quad b_1 = - \sum_{j=2}^{t_2} b_j \quad \Rightarrow \quad \sum_{i=1}^{t_1} a_i = \sum_{j=1}^{t_2} b_j = 0
\]

In the set-to-zero case take \( a_1 = 0 \) and \( b_1 = 0 \).

In either case define

\[
\mu_{ij} = \mu + a_i + b_j
\]
The additive model is a **reduced model** since in the full model each factor level combination \((i,j)\) has its own mean \(\mu_{ij}\).

There are \(t_1 \cdot t_2\) such means \(\mu_{ij}\) which can vary freely.

In the additive model with identifiability restrictions we only have \(1 + (t_1 - 1) + (t_2 - 1) = t_1 + t_2 - 1\) **free parameters**.

Note that \(t_1 \cdot t_2 - [t_1 + t_2 - 1] = (t_1 - 1) \cdot (t_2 - 1)\) can be substantially greater than zero.

We get zero only when one of the factors has just one level.

In that case we are back in the 1-way (1-factor) ANOVA.

The second factor only has one level, i.e., does not change.
## A Tabular View of the Additive Model

### $t_1 \cdot t_2 = 3 \cdot 4$ Factorial Design

#### Factor 2

<table>
<thead>
<tr>
<th></th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$b_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_3$</td>
<td>$\mu + a_3 + b_1$</td>
<td>$\mu + a_3 + b_2$</td>
<td>$\mu + a_3 + b_3$</td>
<td>$\mu + a_3 + b_4$</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$\mu + a_2 + b_1$</td>
<td>$\mu + a_2 + b_2$</td>
<td>$\mu + a_2 + b_3$</td>
<td>$\mu + a_2 + b_4$</td>
</tr>
<tr>
<td>$a_1$</td>
<td>$\mu + a_1 + b_1$</td>
<td>$\mu + a_1 + b_2$</td>
<td>$\mu + a_1 + b_3$</td>
<td>$\mu + a_1 + b_4$</td>
</tr>
</tbody>
</table>

- **Rows** $i$ and $i'$ differ by $a_i - a_i' = \mu_{ij} - \mu_{i'j} = a_i^* - a_i^*$, the same amount across all columns $j = 1, \ldots, t_2$.

- **Columns** $j$ and $j'$ differ by $b_j - b_j' = \mu_{ij} - \mu_{ij'} = b_j^* - b_j^*$, the same amount across all rows $i = 1, \ldots, t_1$.

- Such differences are meaningful regardless of additive model parametrization, i.e., regardless of constraints (sum-to-zero or set-to-zero).
Additivity in Factor 1 $\iff$ Additivity in Factor 2

- If we take 5 numbers, say 5, 7, 9, 2, 3 in a row, and create four new rows by adding 2, or 4, or 5 we get the following tableau

| $a_1^*$ | 5 | 7 | 9 | 2 | 3 |
| $a_2^*$ | 7 | 9 | 11 | 4 | 5 |
| $a_3^*$ | 9 | 11 | 13 | 6 | 7 |
| $a_4^*$ | 10 | 12 | 14 | 7 | 8 |
| $b_j^*$ | 0 | 2 | 4 | $-3$ | $-2$ |

- The columns differ automatically by constant amounts $b_j^*$.
- $\iff$ additivity in factor 2.
- The column differences are set in the first row.
- They are not affected by translating that first row to various levels via the $a_i^*$ (additivity in factor 1).
Additive Model Decomposition

\[Y_{ijk} = \tilde{Y}_{...} + (\tilde{Y}_{i..} - \tilde{Y}_{...}) + (\tilde{Y}_{.j} - \tilde{Y}_{...}) + (Y_{ijk} - \tilde{Y}_{i..} - \tilde{Y}_{.j} + \tilde{Y}_{...}) = \hat{\mu} + \hat{a}_i + \hat{b}_j + \hat{\epsilon}_{ijk}\]

\[\sum_{i} \hat{a}_i = t_1 \frac{1}{t_1} \sum_{i} (\tilde{Y}_{i..} - \tilde{Y}_{...}) = t_1 (\tilde{Y}_{...} - \tilde{Y}_{...}) = 0\]

\[\sum_{j} \hat{b}_j = t_2 \frac{1}{t_2} \sum_{j} (\tilde{Y}_{.j} - \tilde{Y}_{...}) = t_2 (\tilde{Y}_{...} - \tilde{Y}_{...}) = 0\]

\[\sum_{ik} \hat{\epsilon}_{ijk} = nt_1 \frac{1}{nt_1} \sum_{ik} (Y_{ijk} - \tilde{Y}_{i..} - \tilde{Y}_{.j} + \tilde{Y}_{...}) = 0\]

\[\sum_{jk} \hat{\epsilon}_{ijk} = nt_2 \frac{1}{nt_2} \sum_{jk} (Y_{ijk} - \tilde{Y}_{i..} - \tilde{Y}_{.j} + \tilde{Y}_{...}) = 0\]

\[\sum_{ijk} \hat{\epsilon}_{ijk} = 0\]
Least Squares Estimates

\[ \hat{\mu} = \bar{Y}_{\ldots}, \quad \hat{a}_i = \bar{Y}_{i..} - \bar{Y}_{\ldots}, \quad \text{and} \quad \hat{b}_j = \bar{Y}_{.j.} - \bar{Y}_{\ldots} \]

are the least squares estimates of \( \mu, a_i, \) and \( b_j \)
subject to the conditions \( \sum_i a_i = 0 \) and \( \sum_j b_j = 0. \)

This follows without calculus from

\[
Q(\mu, a, b) = Q(\mu, a_1, \ldots, a_t, b_1, \ldots, b_t) = \sum_{ijk}(Y_{ijk} - \mu - a_i - b_j)^2
\]

\[ = \sum_{ijk} \{ \bar{Y}_{\ldots} - \mu + [(\bar{Y}_{i..} - \bar{Y}_{\ldots}) - a_i] \\
+ [(\bar{Y}_{.j.} - \bar{Y}_{\ldots}) - b_j] + (Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{\ldots}) \}^2 \\
= \sum_{ijk} \{(\bar{Y}_{\ldots} - \mu)^2 + [(\bar{Y}_{i..} - \bar{Y}_{\ldots}) - a_i]^2 \\
+ [(\bar{Y}_{.j.} - \bar{Y}_{\ldots}) - b_j]^2 + (Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{\ldots})^2 \}
\]

All cross product terms disappear (see next slide).

Note that \( \sum_i \hat{a}_i = \sum_j \hat{b}_j = 0, \) solutions satisfy constraints.
\[
\sum_{ijk}(\bar{Y}_{...} - \mu)[(\bar{Y}_{i..} - \bar{Y}_{...}) - a_i] = (\bar{Y}_{...} - \mu)t_2 n \left\{ \sum_i (\bar{Y}_{i..} - \bar{Y}_{...}) - \sum_i a_i \right\} = 0
\]

\[
\sum_{ijk}(\bar{Y}_{...} - \mu)[(\bar{Y}_{.j.} - \bar{Y}_{...}) - b_j] = (\bar{Y}_{...} - \mu)t_1 n \left\{ \sum_j (\bar{Y}_{.j.} - \bar{Y}_{...}) - \sum_j b_j \right\} = 0
\]

\[
Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...} = \hat{\epsilon}_{ijk}
\]

\[
\sum_{ijk}(\bar{Y}_{...} - \mu)\hat{\epsilon}_{ijk} = (\bar{Y}_{...} - \mu) \sum_{ijk} \hat{\epsilon}_{ijk} = 0
\]

\[
\sum_{ijk}[(\bar{Y}_{i..} - \bar{Y}_{...}) - a_i][(\bar{Y}_{.j.} - \bar{Y}_{...}) - b_j] = n \sum_i [\hat{a}_i - a_i] \sum_j [\hat{b}_j - b_j] = 0
\]

\[
\sum_{ijk}[(\bar{Y}_{i..} - \bar{Y}_{...}) - a_i]\hat{\epsilon}_{ijk} = \sum_i [(\bar{Y}_{i..} - \bar{Y}_{...}) - a_i] \sum_{jk} \hat{\epsilon}_{ijk} = 0
\]

\[
\sum_{ijk}[(\bar{Y}_{.j.} - \bar{Y}_{...}) - b_j]\hat{\epsilon}_{ijk} = \sum_j [(\bar{Y}_{.j.} - \bar{Y}_{...}) - b_j] \sum_{ik} \hat{\epsilon}_{ijk} = 0
\]
As sum-to-zero estimates we have:

\[ \hat{\mu} = \bar{Y}_{..}, \quad \hat{a}_i = \bar{Y}_{i..} - \bar{Y}_{..}, \quad \text{and} \quad \hat{b}_j = \bar{Y}_{.j} - \bar{Y}_{..}, \]

\[ \implies \quad \hat{\mu}_{ij} = \hat{\mu} + \hat{a}_i + \hat{b}_j. \]

Sum-to-zero estimates \( \rightarrow \) set-to-zero estimates via

\[ \hat{\mu}^* = \hat{\mu}_{11} = \hat{\mu} + \hat{a}_1 + \hat{b}_1, \quad \hat{a}_i^* = \hat{\mu}_{i1} - \hat{\mu}_{11} = \hat{a}_i - \hat{a}_1, \quad \hat{b}_j^* = \hat{\mu}_{1j} - \hat{\mu}_{11} = \hat{b}_j - \hat{b}_1 \]

Note that \( \hat{a}_1^* = \hat{b}_1^* = 0. \)

Furthermore, \( \hat{\mu}_{ij} = \hat{\mu}^* + \hat{a}_i^* + \hat{b}_j^* = \hat{\mu} + \hat{a}_i + \hat{b}_j \) with

\[ \hat{\mu}^* = \bar{Y}_{1..} + \bar{Y}_{.1} - \bar{Y}_{..} \]
\[ \hat{a}_i^* = \bar{Y}_{i..} - \bar{Y}_{1..} \]
\[ \hat{b}_j^* = \bar{Y}_{.j} - \bar{Y}_{.1}. \]
The set-to-zero estimates $\hat{\mu}^*, \hat{a}_i^*, \hat{b}_j^*$ with $\hat{a}_1^* = 0$, $\hat{b}_1^* = 0$, define $\hat{\mu}_{ij} = \hat{\mu}^* + \hat{a}_i^* + \hat{b}_j^*$.

With these we get the sum-to-zero equivalent representation

\[
\begin{align*}
\hat{\mu} &= \hat{\mu}_{..} = \hat{\mu}^* + \hat{a}^* + \hat{b}^* \\
\hat{a}_i &= \hat{\mu}_{i..} - \hat{\mu}_{..} = \hat{a}_i^* - \hat{a}^* \\
\hat{b}_j &= \hat{\mu}_{.j} - \hat{\mu}_{..} = \hat{b}_j^* - \hat{b}^* \\
\end{align*}
\]

$\implies$ $\hat{\mu}_{ij} = \hat{\mu}^* + \hat{a}_i^* + \hat{b}_j^* = \hat{\mu} + \hat{a}_i + \hat{b}_j$

$\implies$ $\hat{\mu} = \bar{Y}_{..}$, $\hat{a}_i = \bar{Y}_{i..} - \bar{Y}_{..}$, and $\hat{b}_j = \bar{Y}_{.j} - \bar{Y}_{..}$

Set-to-zero is what `lm` in R gives as coefficients.
The fitted mean per treatment combination under either parametrization (sum-to-zero or set-to-zero) are the same, i.e.,

$$\hat{\mu}_{ij} = \hat{\mu} + \hat{a}_i + \hat{b}_j = \hat{\mu}^* + \hat{a}_i^* + \hat{b}_j^*$$

Components of the fitted values have different interpretations. This is completely analogous to the previous parameter version

$$\mu_{ij} = \mu + a_i + b_j = \mu^* + a_i^* + b_j^*.$$ 

Explicitly, in terms of the data

$$\hat{\mu}_{ij} = \bar{Y}_{..} + (\bar{Y}_{i..} - \bar{Y}_{..}) + (\bar{Y}_{.j} - \bar{Y}_{..}) = \bar{Y}_{i..} + \bar{Y}_{.j} - \bar{Y}_{..}.$$ 

Note

$$\hat{\epsilon}_{ijk} = Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j} + \bar{Y}_{..} = Y_{ijk} - \hat{\mu}_{ij}.$$
Orthogonal Decomposition of the Data Vector

\[ Y = \begin{pmatrix} Y_{111} \\ \vdots \\ Y_{11n} \\ \vdots \\ Y_{1t21} \\ \vdots \\ Y_{1t2n} \\ \vdots \\ Y_{t111} \\ \vdots \\ Y_{t11n} \\ \vdots \\ Y_{t1t21} \\ \vdots \\ Y_{t1t2n} \end{pmatrix} = \begin{pmatrix} \hat{\mu} \\ \vdots \\ \hat{\mu} \\ \vdots \\ \hat{\mu} \\ \vdots \\ \hat{\mu} \\ \vdots \\ \hat{\mu} \end{pmatrix} + \begin{pmatrix} \hat{a}_1 \\ \vdots \\ \hat{a}_1 \\ \vdots \\ \hat{a}_1 \\ \vdots \\ \hat{a}_1 \end{pmatrix} + \begin{pmatrix} \hat{b}_1 \\ \vdots \\ \hat{b}_1 \\ \vdots \\ \hat{b}_1 \\ \vdots \\ \hat{b}_1 \end{pmatrix} + \begin{pmatrix} \hat{\epsilon}_{111} \\ \vdots \\ \hat{\epsilon}_{11n} \\ \vdots \\ \hat{\epsilon}_{1t21} \\ \vdots \\ \hat{\epsilon}_{1t2n} \\ \vdots \\ \hat{\epsilon}_{t111} \\ \vdots \\ \hat{\epsilon}_{t11n} \\ \vdots \\ \hat{\epsilon}_{t1t21} \\ \vdots \\ \hat{\epsilon}_{t1t2n} \end{pmatrix} = \hat{\mu} 1 + \hat{a} + \hat{b} + \hat{\epsilon} \]
Orthogonalities

\[
\sum_{ijk} \mu \, \hat{a}_i = \mu \, t_2 n \sum_i \hat{a}_i = 0
\]

\[
\sum_{ijk} \mu \, \hat{b}_j = \mu \, t_1 n \sum_j \hat{b}_j = 0
\]

\[
\sum_{ijk} \mu \, \hat{\epsilon}_{ijk} = \mu \, \sum_{ijk} \hat{\epsilon}_{ijk} = 0
\]

\[
\sum_{ijk} \hat{a}_i \, \hat{b}_j = n \sum_i \hat{a}_i \sum_j \hat{b}_j = 0
\]

\[
\sum_{ijk} \hat{a}_i \, \hat{\epsilon}_{ijk} = \sum_i \left( \hat{a}_i \sum_{jk} \hat{\epsilon}_{ijk} \right) = 0
\]

\[
\sum_{ijk} \hat{b}_j \, \hat{\epsilon}_{ijk} = \sum_j \left( \hat{b}_j \sum_{ik} \hat{\epsilon}_{ijk} \right) = 0
\]
Sum of Squares (SS) Decomposition

Orthogonality $\implies$ following SS decomposition (Pythagoras again)

$$
\sum_{ijk} Y_{ijk}^2 = \sum_{ijk} \bar{Y}_{...}^2 + \sum_{ijk} (\bar{Y}_{i..} - \bar{Y}_{...})^2 + \sum_{ijk} (\bar{Y}_{..j} - \bar{Y}_{...})^2 \\
+ \sum_{ijk} (Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{..j} + \bar{Y}_{...})^2 \\
= \sum_{ijk} \hat{\mu}^2 + \sum_{ijk} \hat{a}_i^2 + \sum_{ijk} \hat{b}_j^2 + \sum_{ijk} \hat{\epsilon}_{ijk}^2
$$

$$
\implies \sum_{ijk} (Y_{ijk} - \bar{Y}_{...})^2 = \sum_{ijk} Y_{ijk}^2 - \sum_{ijk} \bar{Y}_{...}^2 = \sum_{ijk} Y_{ijk}^2 - \sum_{ijk} \hat{\mu}^2 \\
= \sum_{ijk} \hat{a}_i^2 + \sum_{ijk} \hat{b}_j^2 + \sum_{ijk} \hat{\epsilon}_{ijk}^2 \\
= \sum_{ijk} (\bar{Y}_{i..} - \bar{Y}_{...})^2 + \sum_{ijk} (\bar{Y}_{..j} - \bar{Y}_{...})^2 \\
+ \sum_{ijk} (Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{..j} + \bar{Y}_{...})^2
$$

$$
SS_T = SS_A + SS_B + SS_E
$$
Interpretation of SS Decomposition

\[ SS_T = SS_A + SS_B + SS_E \]

- \( SS_T = \sum_{ijk} (Y_{ijk} - \bar{Y}_{...})^2 \):
  Total variation of data around the grand or overall mean
- \( SS_A = \sum_{ijk} (\bar{Y}_{i..} - \bar{Y}_{...})^2 \):
  Variation of means around the grand mean
  (by factor 1 level, averaged over the levels of factor 2)
- \( SS_B = \sum_{ijk} (\bar{Y}_{.j.} - \bar{Y}_{...})^2 \):
  Variation of means around the grand mean
  (by factor 2 level, averaged over the levels of factor 1)
- \( SS_E = \sum_{ijk} (Y_{ijk} - \hat{\mu}_{ij})^2 \):
  Variation of data around the fitted additive model value.

\[
SS_E = \sum_{ijk} (Y_{ijk} - \left( \bar{Y}_{...} + (\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{.j.} - \bar{Y}_{...}) \right))^2
\]

\[
= \sum_{ijk} (Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{...})^2 = \sum_{ijk} \epsilon_{ijk}^2
\]
In $\mathbf{Y} = \hat{\mu} \mathbf{1} + \hat{\mathbf{a}} + \hat{\mathbf{b}} + \hat{\mathbf{e}}$ the component vectors are orthogonal to each other.

There is 1 degree of freedom in $\hat{\mu} \mathbf{1}$ and thus there are $N - 1 = t_1 t_2 n - 1$ degrees of freedom (df) in $\mathbf{Y} - \hat{\mu} \mathbf{1} \perp \hat{\mu} \mathbf{1}$ and thus in $SS_T$.

Although the vector $\hat{\mathbf{a}}$ contains $t_1$ distinct values, only $t_1 - 1$ can vary freely, due to sum-to-zero or set-to-zero constraints.

There are $t_1 - 1$ df in that vector and thus in $SS_A$.

Similarly, there are $t_2 - 1$ df in $\hat{\mathbf{b}}$ and thus in $SS_B$.

By orthogonal complement there are

$$t_1 t_2 n - 1 - (t_1 - 1) - (t_2 - 1) = (t_1 - 1)(t_2 - 1) + t_1 t_2 (n - 1) = df_E$$

df in the residual error vector $\hat{\mathbf{e}}$ and thus in $SS_E$. 
### ANOVA Table for the Additive Model

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$SS_A$</td>
<td>$t_1 - 1$</td>
<td>$MS_A = SS_A/(t_1 - 1)$</td>
<td>$MS_A/MS_E$</td>
</tr>
<tr>
<td>B</td>
<td>$SS_B$</td>
<td>$t_2 - 1$</td>
<td>$MS_B = SS_B/(t_2 - 1)$</td>
<td>$MS_B/MS_E$</td>
</tr>
<tr>
<td>Error</td>
<td>$SS_E$</td>
<td>df$_E$</td>
<td>$MS_E = SS_E/df_E$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$SS_T$</td>
<td>$t_1 t_2 n - 1$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$df_E = t_1 t_2 n - 1 - (t_1 - 1) - (t_2 - 1) = (t_1 - 1)(t_2 - 1) + t_1 t_2 (n - 1)$

- View $(t_1 - 1)(t_2 - 1) = t_1 t_2 - [1 + (t_1 - 1) + (t_2 - 1)]$ as the number of means $\mu_{ij}$ left unexplained by the additive model.
- View $t_1 t_2 (n - 1)$ as the degrees of freedom of within cell variation ($n - 1$ per cell) totaled over all $t_1 t_2$ cells.
out <- lm(recip.time ~ type + delivery)

out$coeff

Coefficients:
(Intercept) typeII typeIII deliveryB deliveryC deliveryD
 0.26977  0.04686  0.19964 -0.16574  -0.05721  -0.13583

\hat{\mu}^*, \hat{a}_2^*, \hat{a}_3^*, \hat{b}_2^*, \hat{b}_3^*, \hat{b}_4^*

- Note the implicit set-to-zero form of the parameter estimates in out$coeff! intercept = \hat{\mu}^* with \hat{a}_1^* = \hat{b}_1^* = 0.
- \hat{\mu}^* represents the mean under the treatment combination (typeI, deliveryA)
- \hat{a}_i^*, \hat{b}_j^* represent additive mean deviation effects from this baseline \hat{\mu}^*.
Below are the sum-to-zero estimates corresponding to the previous slide, using the conversion formulas from slide 41.

$\hat{\mu}$

\[
\hat{\mu} = 0.2622376
\]

$\hat{a}_1, \hat{a}_2, \hat{a}_3$

\[
\begin{array}{ccc}
\text{typeI} & \text{typeII} & \text{typeIII} \\
-0.08216887 & -0.03530475 & 0.11747362 \\
\end{array}
\]

$\hat{b}_1, \hat{b}_2, \hat{b}_3, \hat{b}_4$

\[
\begin{array}{cccc}
\text{deliveryA} & \text{deliveryB} & \text{deliveryC} & \text{deliveryD} \\
0.08969690 & -0.07604334 & 0.03248336 & -0.04613693 \\
\end{array}
\]
```r
anova(lm(recip.time ~ type + delivery))
```

### Analysis of Variance Table

**Response: recip.time**

<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>type</td>
<td>2</td>
<td>0.34877</td>
<td>0.17439</td>
<td>71.708</td>
<td>2.865e-14 ***</td>
</tr>
<tr>
<td>delivery</td>
<td>3</td>
<td>0.20414</td>
<td>0.06805</td>
<td>27.982</td>
<td>4.192e-10 ***</td>
</tr>
<tr>
<td>Residuals</td>
<td>42</td>
<td>0.10214</td>
<td>0.00243</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
The ANOVA table states p-values based on null-distributions derived from the model and $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$.

Under the additive model

\[ \Rightarrow SS_E \sim \sigma^2 \chi^2_{df_E} \text{ and } E(MS_E) = E(SS_E/df_E) = \sigma^2 \]

\[ SS_A \sim \sigma^2 \chi^2_{t_1-1, \lambda_1} \text{ with } ncp \]

\[ \lambda_1 = \sum_{ijk} (\bar{\mu}_{i \cdot} - \bar{\mu}_{..})^2 / \sigma^2 = \sum_{ijk} a_i^2 / \sigma^2 = n t_2 \sum_i a_i^2 / \sigma^2 \]

\[ E(MS_A) = E \left( \frac{SS_A}{t_1 - 1} \right) = \sigma^2 + \sigma^2 \frac{\lambda_1}{t_1 - 1} = \sigma^2 + \frac{nt_2}{t_1 - 1} \sum_i a_i^2 \]

\[ SS_B \sim \sigma^2 \chi^2_{t_2-1, \lambda_2} \text{ with } ncp \]

\[ \lambda_2 = \sum_{ijk} (\bar{\mu}_{\cdot j} - \bar{\mu}_{..})^2 / \sigma^2 = \sum_{ijk} b_j^2 / \sigma^2 = n t_1 \sum_j b_j^2 / \sigma^2 \]

\[ E(MS_B) = E \left( \frac{SS_B}{t_2 - 1} \right) = \sigma^2 + \sigma^2 \frac{\lambda_2}{t_2 - 1} = \sigma^2 + \frac{nt_1}{t_2 - 1} \sum_j b_j^2 \]

\[ nt_2 \text{ and } nt_1 \text{ act as multipliers in the noncentrality parameters!} \]

Looking at both factors jointly, we benefit from the common $\sigma^2$ assumption.
SS_A, SS_B and SS_E are independent due to orthogonality of component vectors.

Under $H_A : a_1 = \ldots = a_{t_1} = 0$ we have $\lambda_1 = 0$ and thus

$$F_A = \frac{MS_A}{MS_E} \sim F_{t_1-1,df_E}.$$ 

Under $H_B : b_1 = \ldots = b_{t_2} = 0$ we have $\lambda_2 = 0$ and thus

$$F_B = \frac{MS_B}{MS_E} \sim F_{t_2-1,df_E}.$$ 

These $F$-distributions are the basis for the p-values in the previous ANOVA table.

These p-values correspond to testing $H_A$ and $H_B$, respectively.
In the additive model we have:  \( \mu_{ij} = \mu + a_i + b_j \)

Could compare the full model estimate of \( \mu_{ij} \), namely the average \( \bar{Y}_{ij} \), over the \( n \) observations in cell \((i, j)\), with the additive model fitted value for that same cell, i.e., with

\[
\hat{\mu}_{ij} = \hat{\mu} + \hat{a}_i + \hat{b}_j = \bar{Y}_{...} + (\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{.j} - \bar{Y}_{...}) = \bar{Y}_{i..} + \bar{Y}_{.j} - \bar{Y}_{...}.
\]

\( \bar{Y}_{ij} \) depends only on data from cell \((i, j)\), averaging \( n \) values.

\( \hat{\mu}_{ij} \) depends on data from cells other than cell \((i, j)\) and is more strongly averaged.
The full model $Y_{ijk} = \mu_{ij} + \epsilon_{ijk}$ can be written in the following equivalent form:

$$Y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + \epsilon_{ijk} = \mu + a_i + b_j + c_{ij} + \epsilon_{ijk}$$

Here it is assumed that $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma^2)$.

The equivalent full model form decomposes the mean structure in $\mu_{ij}$ into two components, namely the previously considered additive model $\mu + a_i + b_j$ and the extent $c_{ij} = (ab)_{ij}$ to which this additive model does not explain $\mu_{ij}$, i.e., $c_{ij} = \mu_{ij} - (\mu + a_i + b_j)$.

These parameters $c_{ij} = (ab)_{ij}$ are called interaction terms.

The use of the notational device $(ab)$ is just a mnemonic to indicate the inseparable or joint action of the factors $A$ and $B$, i.e., their interaction.
While there are $t_1 t_2$ mean parameters $\mu_{ij}$, the alternate parametrization has $1 + t_1 + t_2 + t_1 t_2$ parameters

$$\mu, a_1, \ldots, a_{t_1}, b_1, \ldots, b_{t_2}, c_{11}, \ldots, c_{t_1 t_2}$$

To make these latter parameters identifiable we need to impose again certain side conditions.

There are two customary ways which parallel the previous identifiability resolution in the case of the additive model.

1) $a_1 = b_1 = c_{1j} = c_{ij} = 0 \forall i, j$
   set-to-zero side condition ($\texttt{lm} \text{ output in R}$)

2) $\sum_i a_i = \sum_j b_j = \sum_i c_{ij} = \sum_j c_{ij} = 0 \forall i, j$
   sum-to-zero side condition
Define

\[ \mu = \bar{\mu}_{..} = \frac{1}{t_1 t_2} \sum_i \sum_j \mu_{ij}, \quad \bar{\mu}_{i.} = \frac{1}{t_2} \sum_j \mu_{ij}, \quad \bar{\mu}.j = \frac{1}{t_1} \sum_i \mu_{ij} \]

and then all parameters \( \mu, a_i, b_j, c_{ij} \) are determined from the \( \mu_{ij} \) via

\[ a_i = \bar{\mu}_{i.} - \bar{\mu}_{..}, \quad b_j = \bar{\mu}.j - \bar{\mu}_{..} \]

\[ c_{ij} = \mu_{ij} - \bar{\mu}_{i.} - \bar{\mu}.j + \bar{\mu}_{..} \]

\[ = \mu_{ij} - (\bar{\mu}_{i.} - \bar{\mu}_{..}) - (\bar{\mu}.j - \bar{\mu}_{..}) - \bar{\mu}_{..} = \mu_{ij} - a_i - b_j - \mu \]

\[ \implies \mu_{ij} = \mu + a_i + b_j + c_{ij} \quad \text{with} \quad \sum_i a_i = \sum_j b_j = \sum_i c_{ij} = \sum_j c_{ij} = 0 \]

satisfying the sum-to-zero side conditions.
Define all parameters $\mu^*, a^*_i, b^*_j, c^*_{ij}$ from the $\mu_{ij}$ via

\[ \begin{align*}
\mu^* &= \mu_{11}, \\
a^*_i &= \mu_{i1} - \mu_{11} = \mu_{i1} - \mu^*, \\
b^*_j &= \mu_{1j} - \mu_{11} = \mu_{1j} - \mu^*, \\
c^*_{ij} &= \mu_{ij} - \mu_{i1} - \mu_{1j} + \mu_{11} \\
&= \mu_{ij} - (\mu_{i1} - \mu_{11}) - (\mu_{1j} - \mu_{11}) - \mu_{11} \\
&= \mu_{ij} - a^*_i - b^*_j - \mu^*.
\end{align*} \]

\[ \mu_{ij} = \mu^* + a^*_i + b^*_j + c^*_{ij} \] with $a^*_1 = 0, b^*_1 = 0, c^*_{i1} = c^*_{1j} = 0 \ \forall i, j$

satisfying the set-to-zero side conditions.

Whatever the parametrization, we can easily go from one to the other via the previous definitions in terms of the $\mu_{ij}$. 
Decomposition and Least Squares Estimation

Extend decomposition as follows (with orthogonal components)

\[ Y_{ijk} = \bar{Y}_{..} + (\bar{Y}_{i..} - \bar{Y}_{..}) + (\bar{Y}_{.j} - \bar{Y}_{..}) \]
\[ + (\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{.j} + \bar{Y}_{..}) + (Y_{ijk} - \bar{Y}_{ij}) \]
\[ = \mu + \hat{a}_i + \hat{b}_j + \hat{c}_{ij} + \hat{\epsilon}_{ijk} \]

\[ \sum_{ijk} (Y_{ijk} - \mu - a_i - b_j - c_{ij})^2 \]
\[ = \sum_{ijk} [(\bar{Y}_{..} - \mu) + (\bar{Y}_{i..} - \bar{Y}_{..} - a_i) + (\bar{Y}_{.j} - \bar{Y}_{..} - b_j) \]
\[ + (\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{.j} + \bar{Y}_{..} - c_{ij}) + (Y_{ijk} - \bar{Y}_{ij})]^2 \]
\[ = \sum_{ijk} [(\bar{Y}_{..} - \mu)^2 + (\bar{Y}_{i..} - \bar{Y}_{..} - a_i)^2 + (\bar{Y}_{.j} - \bar{Y}_{..} - b_j)^2 \]
\[ + (\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{.j} + \bar{Y}_{..} - c_{ij})^2 + (Y_{ijk} - \bar{Y}_{ij})^2] \]
Full Model Least Squares Estimates (LSEs)

- Cross product terms = 0 in the previous quadratic expansion, because of the component orthogonality in the decomposition.

- The least squares estimates (LSEs, in sum-to-zero form) are

\[ \hat{\mu} = \bar{Y}_{\ldots} \]  
\[ \hat{a}_i = \bar{Y}_{i\ldots} - \bar{Y}_{\ldots} \]  
\[ \hat{b}_j = \bar{Y}_{.j.} - \bar{Y}_{\ldots} \]  
\[ \hat{c}_{ij} = \bar{Y}_{ij.} - \bar{Y}_{i\ldots} - \bar{Y}_{.j.} + \bar{Y}_{\ldots} \]  

with \( E(\hat{\mu}) = \mu \), \( E(\hat{a}_i) = a_i \), \( E(\hat{b}_j) = b_j \) and \( E(\hat{c}_{ij}) = c_{ij} \),

- The LSEs are unbiased.

- The fitted values for the \( \mu_{ij} \) are

\[ \hat{\mu}_{ij} = \hat{\mu} + \hat{a}_i + \hat{b}_j + \hat{c}_{ij} = \hat{\mu}_{ij} + \hat{c}_{ij} \]  
\[ = \bar{Y}_{\ldots} + (\bar{Y}_{i\ldots} - \bar{Y}_{\ldots}) + (\bar{Y}_{.j.} - \bar{Y}_{\ldots}) + (\bar{Y}_{ij.} - \bar{Y}_{i\ldots} - \bar{Y}_{.j.} + \bar{Y}_{\ldots}) \]  
\[ = \bar{Y}_{ij.} \quad \text{i.e., the cell means.} \]

- with residuals \( \hat{\epsilon}_{ijk} = Y_{ijk} - \bar{Y}_{ij.} \).
Sum of Squares Decomposition

Using $\mu = a_i = b_j = c_{ij} = 0$ in previous decomposition

$$
\sum_{ijk} Y_{ijk}^2 = \sum_{ijk} \left[ \bar{Y}^2 + (\bar{Y}_{i..} - \bar{Y}_{...})^2 + (\bar{Y}_{.j} - \bar{Y}_{...})^2 \\
+ (\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{.j} + \bar{Y}_{...})^2 + (Y_{ijk} - \bar{Y}_{ij})^2 \right]
$$

$$
\sum_{ijk} (Y_{ijk} - \bar{Y}_{...})^2 = \sum_{ijk} \left[ (\bar{Y}_{i..} - \bar{Y}_{...})^2 + (\bar{Y}_{.j} - \bar{Y}_{...})^2 \\
+ (\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{.j} + \bar{Y}_{...})^2 + (Y_{ijk} - \bar{Y}_{ij})^2 \right]
$$

$$
SS_T = SS_A + SS_B + SS_{AB} + SS_E
$$

$$
SS_T = \sum_{ijk} (Y_{ijk} - \bar{Y}_{...})^2
$$

$$
SS_A = \sum_{ijk} (\bar{Y}_{i..} - \bar{Y}_{...})^2 = \sum_{ijk} \hat{a}_{i}^2, \quad SS_B = \sum_{ijk} (\bar{Y}_{.j} - \bar{Y}_{...})^2 = \sum_{ijk} \hat{b}_j^2
$$

$$
SS_{AB} = \sum_{ijk} (\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{.j} + \bar{Y}_{...})^2 = \sum_{ijk} \hat{c}_{ij}^2, \quad SS_E = \sum_{ijk} (Y_{ijk} - \bar{Y}_{ij})^2
$$
ANOVA Table for the Full Model

<table>
<thead>
<tr>
<th>Source</th>
<th>$SS$</th>
<th>df</th>
<th>$MS$</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$SS_A$</td>
<td>$t_1 - 1$</td>
<td>$MS_A = SS_A/(t_1 - 1)$</td>
<td>$MS_A/MS_E$</td>
</tr>
<tr>
<td>B</td>
<td>$SS_B$</td>
<td>$t_2 - 1$</td>
<td>$MS_B = SS_B/(t_2 - 1)$</td>
<td>$MS_B/MS_E$</td>
</tr>
<tr>
<td>AB</td>
<td>$SS_{AB}$</td>
<td>$(t_1 - 1)(t_2 - 1)$</td>
<td>$MS_{AB} = SS_{AB}/[(t_1 - 1)(t_2 - 1)]$</td>
<td>$MS_{AB}/MS_E$</td>
</tr>
<tr>
<td>Error</td>
<td>$SS_E$</td>
<td>$t_1 t_2 (n - 1)$</td>
<td>$MS_E = SS_E/[t_1 t_2 (n - 1)]$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$SS_T$</td>
<td>$t_1 t_2 n - 1$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Distributional Facts for the Full Model

\[
\frac{SS_A}{\sigma^2} \sim \chi^2_{t_1-1}, \lambda_A, \quad \lambda_A = \frac{\sum_{ijk} a^2_i}{\sigma^2}, \quad \frac{SS_B}{\sigma^2} \sim \chi^2_{t_2-1}, \lambda_B, \quad \lambda_B = \frac{\sum_{ijk} b^2_j}{\sigma^2}
\]

\[
\frac{SS_{AB}}{\sigma^2} \sim \chi^2_{(t_1-1)(t_2-1)}, \lambda_{AB}, \quad \lambda_{AB} = \frac{\sum_{ijk} c^2_{ij}}{\sigma^2}, \quad \frac{SS_E}{\sigma^2} \sim \chi^2_{t_1t_2(n-1)}
\]

\[SS_A, SS_B, SS_{AB}, SS_E\] are statistically independent (orthogonality).

\[
F_A = MS_A / MS_E \sim F_{t_1-1, t_1t_2(n-1)}, \lambda_A
\]

\[
F_B = MS_B / MS_E \sim F_{t_2-1, t_1t_2(n-1)}, \lambda_B
\]

\[
F_{AB} = MS_{AB} / MS_E \sim F_{(t_1-1)(t_2-1), t_1t_2(n-1)}, \lambda_{AB}
\]
Expected MS for the Full Model

\[ E(\text{MS}_A) = E \left( \frac{SS_A}{t_1 - 1} \right) = \sigma^2 + \frac{\sum_{ijk} a^2_i}{t_1 - 1} = \sigma^2 \left( 1 + \frac{\lambda_A}{t_1 - 1} \right) \]

\[ E(\text{MS}_B) = E \left( \frac{SS_B}{t_2 - 1} \right) = \sigma^2 + \frac{\sum_{ijk} b^2_j}{t_2 - 1} = \sigma^2 \left( 1 + \frac{\lambda_B}{t_2 - 1} \right) \]

\[ E(\text{MS}_{AB}) = E \left( \frac{SS_{AB}}{(t_1 - 1)(t_2 - 1)} \right) = \sigma^2 + \frac{\sum_{ijk} c^2_{ij}}{(t_1 - 1)(t_2 - 1)} = \sigma^2 \left( 1 + \frac{\lambda_{AB}}{(t_1 - 1)(t_2 - 1)} \right) \]

\[ E(\text{MS}_E) = \sigma^2. \]
Reject $H_{0A}: a_1 = \ldots = a_{t_1} = 0$ whenever $F_A$ is too large.

For a level $\alpha$ test reject $H_{0A}$ whenever

$$F_A \geq q_f(1 - \alpha, t_1 - 1, t_1 t_2 (n - 1)).$$

Reject $H_{0B}: b_1 = \ldots = b_{t_2} = 0$ whenever $F_B$ is too large.

For a level $\alpha$ test reject $H_{0B}$ whenever

$$F_B \geq q_f(1 - \alpha, t_2 - 1, t_1 t_2 (n - 1)).$$

Reject $H_{0AB}: c_{ij} = 0 \forall i, j$ whenever $F_{AB}$ is too large.

For a level $\alpha$ test reject $H_{0AB}$ whenever

$$F_{AB} \geq q_f(1 - \alpha, (t_1 - 1)(t_2 - 1), t_1 t_2 (n - 1)).$$

$H_{0AB}: c_{ij} = 0 \forall i, j$ means that the additive model $\mu_{ij} = \mu + a_i + b_j$ is sufficient to explain the mean structure.

Rejecting $H_{0AB} \implies$ the additive model will not provide a sufficient explanation.
Comments on the Full Model ANOVA Table

\[ SS_{E}^{\text{additive model}} = \sum_{ijk} (Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j} + \bar{Y}_{...})^2 \neq SS_{E}^{\text{full model}} = \sum_{ijk} (Y_{ijk} - \bar{Y}_{ij})^2 \]

\[ \sum_{ijk} (Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j} + \bar{Y}_{...})^2 = \sum_{ijk} (Y_{ijk} - \bar{Y}_{ij})^2 + \sum_{ijk} (\bar{Y}_{ij} - \bar{Y}_{i..} - \bar{Y}_{.j} + \bar{Y}_{...})^2 \]

\[ MS_{E}^{\text{additive model}} = \frac{\sum_{ijk} (Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j} + \bar{Y}_{...})^2}{(t_1 - 1)(t_2 - 1) + t_1 t_2 (n - 1)} \neq MS_{E}^{\text{full model}} = \frac{\sum_{ijk} (Y_{ijk} - \bar{Y}_{ij})^2}{t_1 t_2 (n - 1)} \]

\[ F_{A}^{\text{additive model}} = \frac{MS_{A}}{MS_{E}^{\text{additive model}}} \neq F_{A}^{\text{full model}} = \frac{MS_{A}}{MS_{E}^{\text{full model}}} \]

\[ F_{B}^{\text{additive model}} = \frac{MS_{B}}{MS_{E}^{\text{additive model}}} \neq F_{B}^{\text{full model}} = \frac{MS_{B}}{MS_{E}^{\text{full model}}} \]
Reciprocal Time to Death (Insecticide Data)

> round(recip.time,5)
  [1] 0.32258 0.22222 0.21739 0.23256 0.27778 0.34483 0.25000
  [8] 0.43478 0.45455 0.47619 0.55556 0.43478 0.12195 0.09091
 [15] 0.11364 0.13889 0.10870 0.16393 0.20408 0.08065 0.33333
 [22] 0.27027 0.26316 0.34483 0.23256 0.22222 0.15873 0.13158
 [29] 0.22727 0.28571 0.32258 0.25000 0.43478 0.40000 0.41667
 [36] 0.45455 0.22222 0.14085 0.15152 0.16129 0.17857 0.09804
 [43] 0.14085 0.26316 0.33333 0.27778 0.32258 0.30303
Both type and delivery are in factor form.

Don’t need to invoke as.factor(type) and as.factor(delivery) in the call of lm.
out.lmFULL <- lm(recip.time ~ type*delivery)

recip.time ~ type*delivery

\[ \text{compare} \quad \longleftrightarrow \quad \text{recip.time} \sim \text{type}+\text{delivery} \]

> out.lmFULL
Call:
lm(formula = recip.time ~ type * delivery)

Coefficients:

<table>
<thead>
<tr>
<th>(Intercept)</th>
<th>typeII</th>
<th>typeIII</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.248688</td>
<td>0.078159</td>
<td>0.231580</td>
</tr>
<tr>
<td>deliveryB</td>
<td>deliveryC</td>
<td>deliveryD</td>
</tr>
<tr>
<td>-0.132342</td>
<td>-0.062416</td>
<td>-0.079720</td>
</tr>
<tr>
<td>typeII:deliveryB</td>
<td>typeIII:deliveryB</td>
<td>typeII:deliveryC</td>
</tr>
<tr>
<td>-0.055166</td>
<td>-0.045030</td>
<td>0.006961</td>
</tr>
<tr>
<td>typeIII:deliveryC</td>
<td>typeII:deliveryD</td>
<td>typeIII:deliveryD</td>
</tr>
<tr>
<td>0.008646</td>
<td>-0.076974</td>
<td>-0.091368</td>
</tr>
</tbody>
</table>

Note the set-to-zero form of parameter estimates.
What else is in `out.lmFULL`?

```r
> names(out.lmFULL)
[1] "coefficients"  "residuals"   "effects"
[5] "fitted.values" "assign"      "qr"
[9] "contrasts"     "xlevels"     "call"
[13] "model"
```
Thus it appears that the additive model is quite acceptable.

Both factors play strongly in the additive model.
Graphical View of No Interaction Effect

If one view looks parallel so will the other.
If One View Looks Parallel so Will the Other.

\[
\begin{array}{cccc}
\mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} \\
\mu_{11} + d_2 & \mu_{12} + d_2 & \mu_{13} + d_2 & \mu_{14} + d_2 \\
\mu_{11} + d_3 & \mu_{12} + d_3 & \mu_{13} + d_3 & \mu_{14} + d_3 \\
\end{array}
\]

Row differences are constant!

\[
\begin{align*}
\Delta_2 &= \mu_{12} - \mu_{11} \text{ is the difference between column 2 and column 1} \\
\Delta_3 &= \mu_{13} - \mu_{11} \text{ is the difference between column 3 and column 1} \\
\Delta_4 &= \mu_{14} - \mu_{11} \text{ is the difference between column 4 and column 1.}
\end{align*}
\]

\[
\begin{array}{cccc}
\mu_{11} & \mu_{11} + \Delta_2 & \mu_{11} + \Delta_3 & \mu_{11} + \Delta_4 \\
\mu_{21} & \mu_{21} + \Delta_2 & \mu_{21} + \Delta_3 & \mu_{21} + \Delta_4 \\
\mu_{31} & \mu_{31} + \Delta_2 & \mu_{31} + \Delta_3 & \mu_{31} + \Delta_4 \\
\end{array}
\]

Used \( \mu_{11} = \mu_{11}, \quad \mu_{21} = \mu_{11} + d_2, \) \, and \, \mu_{31} = \mu_{11} + d_3.

Similarly one argues going the other direction.
Strong evidence of factor level differences.
Examine them individually to see which differences matter.
A naive approach: Perform a 2-sample $t$-test.
Or look at the corresponding confidence intervals.
Comparing type I with type II means we get

```r
t.test(recip.time[type == "I"], recip.time[type == "II"], var.equal=T)
```

Two Sample t-test

data:  recip.time[type == "I"] and recip.time[type == "II"]
t = -1.6383, df = 30, p-value = 0.1118
alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:
-0.10528550 0.01155725
sample estimates:
mean of x mean of y
0.1800688 0.2269329
Comparing Types I and II of Insecticide

points are jittered vertically and horizontally
Getting $MS_E$ in Types I and II Comparison

```r
> anova(lm(recip.time[type=="I" | type=="II"] ~ + type[type=="I" | type=="II"]))
Analysis of Variance Table

Response: recip.time[type == "I" | type == "II"]

<table>
<thead>
<tr>
<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>type[type == &quot;I&quot;</td>
<td>type == &quot;II&quot;]</td>
<td>1</td>
<td>0.017570</td>
<td>0.017570</td>
</tr>
<tr>
<td>Residuals</td>
<td>30</td>
<td>0.196394</td>
<td>0.006546</td>
<td></td>
</tr>
</tbody>
</table>
```

- Note the same p-value 0.1118 as in previous $t$-test.
- We are doing the same test, since $t_f^2 = F_{1,f}$.
- From this table we get $s = \sqrt{MS_E} = \sqrt{0.006546} = .08091$.
- This could also have been backed out from previous $t$-based confidence interval.
- `type == "I" | type == "II"` = logical OR.
What is Wrong?

- In the previous 2-sample \( t \)-test/interval we treated the observations as i.i.d. from two populations.
- Ignored the known variations due to the delivery method.
- The pooled sample standard deviation from these 2 “samples” confounds variation between delivery method means with variation (\( \sigma \)) of within (delivery,type) combination.
- Our “reference distribution” will thus be too dispersed.
- Our test will be less discriminating.
- Our confidence interval will be too wide.
Closer Look in Comparing Types I and II

Note the reduced variability within same color clusters.
According to our (accepted) additive model we have

\[ Y_{1jk} = \mu + a_1 + b_j + \epsilon_{1jk} \quad \text{and} \]
\[ Y_{2jk} = \mu + a_2 + b_j + \epsilon_{2jk} \]

with \( \epsilon_{ijk} \sim \text{i.i.d. } \mathcal{N}(0, \sigma^2) \)

The difference due to type I and type II is captured by \( a_1 - a_2 \).

This can be interpreted as the difference between the mean response under type I and the mean response under type II:

\[
\bar{\mu}_1. - \bar{\mu}_2. = \frac{\sum_j \mu_{1j}}{4} - \frac{\sum_j \mu_{2j}}{4} = \sum_j (\mu + a_1 + b_j) - \sum_j (\mu + a_2 + b_j) = \frac{4}{4} (a_1 - a_2)
\]

The effect of type I vs. type II can be interpreted as a contrast in cell means.

Think of it as contrasting the effects of interest while canceling/averaging out other effects.
Even in the full model with interactions the previous contrast in means stays the same since

$$\bar{\mu}_1 - \bar{\mu}_2 = \frac{\sum_j \mu_{1j}}{4} - \frac{\sum_j \mu_{2j}}{4}$$

$$= \frac{\sum_j (\mu + a_1 + b_j + c_{1j}) - \sum_j (\mu + a_2 + b_j + c_{2j})}{4}$$

$$= \mu + a_1 + \frac{\sum_j (b_j + c_{1j})}{4} - \mu - a_2 - \frac{\sum_j (b_j + c_{2j})}{4}$$

$$= a_1 - a_2$$

$\mu$, $b_j$, $c_{1j}$ and $c_{2j}$ are canceled out.
Estimated Contrast

- The natural estimate of $\bar{\mu}_1. - \bar{\mu}_2. = a_1 - a_2$ is
  $\hat{\mu}_1. - \hat{\mu}_2. = \hat{a}_1 - \hat{a}_2 = (\bar{Y}_1.. - \bar{Y}...) - (\bar{Y}_2.. - \bar{Y}...) = \bar{Y}_1.. - \bar{Y}_2..$

- Same as in previous (naive) 2-sample $t$-test/interval.
- Can also be viewed as the contrast of estimated cell averages

$$\hat{a}_1 - \hat{a}_2 = \bar{Y}_1.. - \bar{Y}_2.. = \frac{\bar{Y}_{11.} + \bar{Y}_{12.} + \bar{Y}_{13.} + \bar{Y}_{14.}}{4} - \frac{\bar{Y}_{21.} + \bar{Y}_{22.} + \bar{Y}_{23.} + \bar{Y}_{24.}}{4}$$

$$= \frac{\hat{\mu}_{11} + \hat{\mu}_{12} + \hat{\mu}_{13} + \hat{\mu}_{14}}{4} - \frac{\hat{\mu}_{21} + \hat{\mu}_{22} + \hat{\mu}_{23} + \hat{\mu}_{24}}{4}$$

because $\hat{c}_{i1} + \hat{c}_{i2} + \hat{c}_{i3} + \hat{c}_{i4} = 0$ in $\sum_j \hat{\mu}_{ij} = \sum_j (\hat{\mu}_{ij} + \hat{c}_{ij})$. 
However, \( \text{var}(\bar{Y}_{1..} - \bar{Y}_{2..}) = \text{var}(\bar{Y}_{1..}) + \text{var}(\bar{Y}_{2..}) = \frac{\sigma^2}{nt_2} + \frac{\sigma^2}{nt_2} = \frac{2\sigma^2}{nt_2} \)

\[
\frac{(\bar{Y}_{1..} - \bar{Y}_{2..} - (a_1 - a_2))}{s/\sigma} \quad = \quad \frac{(\bar{Y}_{1..} - \bar{Y}_{2..} - (a_1 - a_2))}{s\sqrt{2/(nt_2)}} \quad = \quad \frac{\hat{a}_1 - \hat{a}_2 - (a_1 - a_2)}{s\sqrt{2/(nt_2)}} \sim t_f .
\]

We have two options in choosing \( s \) and the corresponding \( f \):

\[
s^2 = s^2_x = MS_{\text{full model}} = \frac{\sum_{ijk}(Y_{ijk} - \bar{Y}_{ij.})^2}{t_1t_2(n-1)} \quad \text{and thus} \quad f = t_1 t_2(n-1) .
\]

or \( s^2 = s^2_+ = MS_{\text{additive model}} = \frac{\sum_{ijk}(Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.j.} + \bar{Y}_{..})^2}{t_1t_2(n-1) + (t_1-1)(t_2-1)} \)

and thus \( f = t_1 t_2(n-1) + (t_1 - 1)(t_2 - 1) = t_1 t_2 n - t_1 - t_2 + 1 .\)
\textit{t-Test for } H_0: a_1 = a_2 \\

- Reject \( H_0: a_1 = a_2 \) when

\[
\left| \frac{\hat{a}_1 - \hat{a}_2}{\sqrt{2MS_E/(nt_2)}} \right| > t_{1-\alpha/2,f}
\]

- or \( |\hat{a}_1 - \hat{a}_2| > \sqrt{2MS_E/(nt_2)}t_{1-\alpha/2,f} \)

- or \( |\hat{a}_1 - \hat{a}_2| > SE(\hat{a}_1 - \hat{a}_2)t_{1-\alpha/2,f} = LSD_A \)

- \( LSD_A \) = least significant difference with which any estimated differences \( \hat{a}_i - \hat{a}_{i'} \) (\( i \neq i' \)) in factor A levels can be compared.

- Parallel with our previous use of \( LSD \) in the ANOVA situation.
t-Test for $H_0: b_1 = b_2$

- Reject $H_0: b_1 = b_2$ when

$$\left| \frac{\hat{b}_1 - \hat{b}_2}{\sqrt{2MS_E/(nt_1)}} \right| > t_{1-\alpha/2,f}$$

or $$|\hat{b}_1 - \hat{b}_2| > \sqrt{2MS_E/(nt_1)}t_{1-\alpha/2,f}$$

or $$|\hat{b}_1 - \hat{b}_2| > SE(\hat{b}_1 - \hat{b}_2)t_{1-\alpha/2,f} = LSD_B,$$

$LSD_B = \text{least significant difference with which any estimated differences } \hat{b}_j - \hat{b}_{j'} \text{ (} j \neq j' \text{)} \text{ in factor } B \text{ levels can be compared.}$

- Note the change ($t_2 \leftrightarrow t_1$) between $LSD_A$ and $LSD_B$. 

---
Factor Level Means

```r
> mean(recip.time[type=="I"])
[1] 0.1800688
> mean(recip.time[type=="II"])
[1] 0.2269329
> mean(recip.time[type=="III"])
[1] 0.3797112
> mean(recip.time[delivery=="A"])
[1] 0.3519345
> mean(recip.time[delivery=="B"])
[1] 0.1861943
> mean(recip.time[delivery=="C"])
[1] 0.294721
> mean(recip.time[delivery=="D"])
[1] 0.2161007
```
**LSD_A and LSD_B**

For the Type Factor

\[
LSD_A = t_{.975,f} SE(\hat{a}_1 - \hat{a}_2) = 2.028094 \sqrt{.00240 \frac{2}{4.4}} = 0.03513
\]
for the full model with \( f = 36 \).

\[
LSD_A = t_{.975,f} SE(\hat{a}_1 - \hat{a}_2) = 2.018082 \sqrt{.00243 \frac{2}{4.4}} = 0.03517
\]
for the additive model with \( f = 42 \).

For the Delivery Factor

\[
LSD_B = t_{.975,f} SE(\hat{b}_1 - \hat{b}_2) = 2.028094 \sqrt{.00240 \frac{2}{3.4}} = 0.04056
\]
for the full model with \( f = 36 \).

\[
LSD_B = t_{.975,f} SE(\hat{b}_1 - \hat{b}_2) = 2.018082 \sqrt{.00243 \frac{2}{3.4}} = 0.04061
\]
for the additive model with \( f = 42 \).
### LSD Groupings

**Type**

\[ LSD_A = 0.0352 \]

<table>
<thead>
<tr>
<th>Poison Type</th>
<th>Mean ( \hat{\mu}_i )</th>
<th>LSD Grouping</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0.180</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>0.227</td>
<td>2</td>
</tr>
<tr>
<td>III</td>
<td>0.380</td>
<td>3</td>
</tr>
</tbody>
</table>

Note difference to naive 2-sample \( t \)-test.

**Delivery**

\[ LSD_B = 0.0406 \]

<table>
<thead>
<tr>
<th>Poison Delivery</th>
<th>Mean ( \hat{\mu}_j )</th>
<th>LSD Grouping</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>0.186</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>0.216</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>0.295</td>
<td>2</td>
</tr>
<tr>
<td>A</td>
<td>0.352</td>
<td>3</td>
</tr>
</tbody>
</table>
The plots on the previous slide were produced by:

```
> par(mfrow=c(2,1),mar=c(4,4,1,1)+.1)
> interaction.plot(type,delivery,recip.time, 
  col=c("green","red","blue","cyan"))
> interaction.plot(delivery,type,recip.time,  
  col=c("green","red","blue"))
```

Here `mar=c(4,4,1,1)+.1` inside `par` sets plot margins.

`mfrow=c(2,1)` sets up plotting for two plots per page, one above the other.
The full model is \( Y_{ijk} = \mu_{ij} + \epsilon_{ijk} \).

We can reparametrize it as

\[
Y_{ijk} = \mu + a_i + b_j + c_{ij} + \epsilon_{ijk}
\]

with replications \( k = 1, \ldots, n \), where

1) \( \mu = \frac{\sum_{ij} \mu_{ij}}{(t_1 t_2)} = \bar{\mu}_{..} \) is the grand mean.
2) \( a_i = \frac{\sum_j (\mu_{ij} - \bar{\mu}_{..})}{t_2} = \bar{\mu}_i - \bar{\mu}_{..} \) with \( \sum_i a_i = 0 \).
3) \( b_j = \frac{\sum_i (\mu_{ij} - \bar{\mu}_{..})}{t_2} = \bar{\mu}_j - \bar{\mu}_{..} \) with \( \sum_j b_j = 0 \).

\( a_i \) and \( b_j \) are also referred to as main effects.

4) \( c_{ij} = \mu_{ij} - \bar{\mu}_i - \bar{\mu}_j + \bar{\mu}_{..} = \mu_{ij} - (\mu + a_i + b_j) \) with \( \sum_i c_{ij} = \sum_j c_{ij} = 0 \).

In the additive model the only change is 4) \( c_{ij} = 0 \ \forall \ i, j \).
The Interpretation of $a_i - a_{i'}$

- In the additive or main effects model we have
  \[ a_i - a_{i'} = \mu_{ij} - \mu_{i'j} = \mu + a_i + b_j - (\mu + a_{i'} + b_j) \quad \forall j = 1, \ldots, t_2, \]

- $a_i - a_{i'}$ = difference in mean response between levels $i$ and $i'$ of factor $A$, and it is the same for each level $j$ of factor $B$.

- In the full or interaction model (since $b_\cdot = \bar{c}_i = \bar{c}_{i'} = 0$)
  \[ a_i - a_{i'} = \bar{\mu}_i - \bar{\mu}_{i'} = \mu + a_i + b_\cdot + \bar{c}_i - (\mu + a_{i'} + b_\cdot + \bar{c}_{i'}) \quad \forall j = 1, \ldots, t_2, \]

- $a_i - a_{i'}$ = difference in mean response between levels $i$ and $i'$ of factor $A$, when averaged over all levels $j$ of factor $B$.

- Corresponding interpretations hold for $b_j - b_{j'}$. 
Main effects \( a_i \) and \( b_j \) are called that way because their additive effects on the mean \( \mu_{ij} \) are easily understood.

Interactions can be more complicated and may need more scrutiny in order to develop some understanding.

We will just give a few example situations that illustrate some distinct and very different situations.

There are certainly many other possibilities.

For simplicity we consider a 2 × 4 two-factor experiment.
Interaction Pattern 1
Interaction Pattern 2


2  4  6  8
Interaction Pattern 3

- A1.B1
- A1.B3
- A1.B4
- A2.B1
- A2.B2
- A2.B3
- A2.B4

Box plots for different combinations of A1, B1, A2, B2, A1, B3, A2, B3, A1, B4, and A2, B4.
Comments on Interaction Patterns

- Pattern 1 shows a linear interaction trend along the levels of factor B.
- Factor B seems to have no additive or main effect.
- Pattern 2 seems to have no additive effect from factor B and almost no additive effect from factor A.
- Only for \( B_4 \) of factor B is there a clear change in factor A.
- Hence \( B_4 \) acts as an interaction switch.
- Pattern 3 shows no additive or main effect due to factor B.
- If the colors of the last two boxes were switched we would have a clear additive or main effect due to factor A (±2.5).
- This would give a far simpler data explanation and it suggests the possibility of a labeling error.
When we did our ANOVA to examine the effects of a factor A of interest, we saw that the power of the $F$-test is an increasing function of the noncentrality parameter 

$$\lambda = \sum_i n_i (\mu_i - \bar{\mu})^2 / \sigma^2.$$ 

Cannot influence the size of treatment effects, $|\mu_i - \bar{\mu}|$, 

We can influence the sample sizes $n_i$ and possibly $\sigma$. 

Note $\sigma \rightarrow \sigma / 2 \iff n_i \rightarrow 4 \cdot n_i$ !! 

How can we influence $\sigma$ itself? 

We need to understand what may affect $\sigma$. 

Often $\sigma$ is due to variability of hidden or ignored factors. 

Consider the delivery factor in our insecticide experiment. 

It had a definite effect on the measured response times. 

Had we ignored it or left it to happenstance which delivery was used for each experimental unit (insect), we would have confounded the variability due to delivery with the remaining variability within (delivery,type).
Ignoring a Factor

- Assuming an additive model (similar reasoning under full model):

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>SS_A</td>
<td>(t_1 - 1)</td>
<td>(SS_A/(t_1 - 1))</td>
<td>(MS_A/MS_E)</td>
</tr>
<tr>
<td>B</td>
<td>SS_B</td>
<td>(t_2 - 1)</td>
<td>(SS_B/(t_2 - 1))</td>
<td>(MS_B/MS_E)</td>
</tr>
<tr>
<td>Error</td>
<td>SS_E</td>
<td>(df_E)</td>
<td>(SS_E/df_E)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>(SS_T)</td>
<td>(t_1 t_2 n - 1)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- If we ignored factor B, we would treat \(SS'_E = SS_B + SS_E\) as our error sum of squares with \(df'_E = df_E + t_2 - 1\).

- \(SS'_E/df'_E = MS'_E\) would be a legitimate estimate of \(\sigma^2\) if \(b_1 = \ldots = b_{t_2} = 0\).

- If not, \(MS'_E\) would be inflated, it would estimate \(\sigma'^2 > \sigma^2 \Rightarrow\) loss of power.
Blocking consists of stratifying experimental units into groups that will have more homogeneous responses within groups.

Possibly quite inhomogeneous responses between groups.

Such grouping/blocking can be accomplished by an appropriately chosen factor, where the levels of that factor define the different groups.

Blocking will be beneficial if the factor used for blocking causes variation in the response as the levels of that factor change.

Beneficial in the sense that this variation can be eliminated.

Thus experimental units within a level of that factor (i.e., within a block) will not experience that change and will thus be more homogeneous in their response.

Treatment comparisons within blocks, block by block.
Typical Blocking Criteria

- **Location**: Experiments are conducted over varying locations and location is judged to have an effect on the response.
- **Time**: If time of day, month, or year are likely to affect response. Results of the study are to stand regardless of time.
- **Litters**: If animals in the same litter are likely to produce more homogeneous responses. Useful in medical experiments.
- **Batches of Material**: Variations in the process for creating experimental material are likely to show up in the responses.
- **Any variation inducing aspect of an experiment** (education, income, ...) that is not considered a treatment of interest.
The statistician Sir Ronald Aylmer Fisher (1890-1962) developed experimental design in an agricultural setting at Rothamstead Experimental Station.

See [http://www.bookrags.com/Ronald_Fisher](http://www.bookrags.com/Ronald_Fisher) on his great influence as one of the founding fathers of statistics.

To some he is more famous for his work in Genetics.

“He bred poultry, mice, snails, and other creatures and published his findings in several papers that contributed to scientists’ overall understanding of genetic dominance.”

Apparently he had much to do with the fact that the significance level of .05 is so entrenched until today, see: [http://www.jerrydallal.com/LHSP/p05.htm](http://www.jerrydallal.com/LHSP/p05.htm)

Thus it is only fitting to consider an agricultural experiment to illustrate blocking.
Nitrogen Fertilizer Timing

- A nitrogen fertilizer can be administered according to 6 different timing schedules (treatments).
- The response is the nitrogen uptake ($\text{ppm} \times 10^{-2}$).
- The experimental material: One irrigated field.
- Subdividing the field into different experimental units for use with different treatments could be affected by soil moisture variation, caused by a sloping field gradient.
- This moisture gradient is mainly across the width of the field.
- This suggest rows along the field length as blocks.
Field Moisture Gradient

Dry

Soil wetness

Wet
Field subdivided into 4 rows (or blocks) with 6 plots each. The 6 treatments were randomly assigned to each row. Remaining moisture variability within rows affects $MS_E$.

<table>
<thead>
<tr>
<th>row</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40.89</td>
<td>37.99</td>
<td>37.18</td>
<td>34.98</td>
<td>34.89</td>
<td>42.07</td>
</tr>
<tr>
<td>2</td>
<td>41.22</td>
<td>49.42</td>
<td>45.85</td>
<td>50.15</td>
<td>41.99</td>
<td>46.69</td>
</tr>
<tr>
<td>3</td>
<td>44.57</td>
<td>52.68</td>
<td>37.61</td>
<td>36.94</td>
<td>46.65</td>
<td>40.23</td>
</tr>
<tr>
<td>4</td>
<td>41.90</td>
<td>39.20</td>
<td>43.29</td>
<td>40.45</td>
<td>42.91</td>
<td>39.97</td>
</tr>
</tbody>
</table>

Also available as `fertilizerdata.csv` on web site.
Randomized Complete Block (RCB) Design

- Experimental units are blocked into presumably more homogeneous groups.
- The blocks are complete, i.e., each treatment in each block.
- The blocks are balanced
  - \( t_1 = 4 \) observations for each treatment level.
  - \( t_2 = 6 \) treatments for each block level (row).
  - \( n = 1 \) observation per (block, treatment) combination.
# ANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Block</td>
<td>$SS_{Block}$</td>
<td>$t_1 - 1$</td>
<td>$SS_{Block}/(t_1 - 1)$</td>
<td>$MS_{Block}/MS_E$</td>
</tr>
<tr>
<td>Treatment</td>
<td>$SS_{Treat}$</td>
<td>$t_2 - 1$</td>
<td>$SS_{Treat}/(t_2 - 1)$</td>
<td>$MS_{Treat}/MS_E$</td>
</tr>
<tr>
<td>Error</td>
<td>$SS_E$</td>
<td>df$_E$</td>
<td>$SS_E/df_E$</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$SS_T$</td>
<td>$t_1 t_2 n - 1$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Note that here $n = 1$, thus $t_1 t_2 n - 1 = t_1 t_2 - 1$, and $df_E = (t_1 - 1)(t_2 - 1) + t_1 t_2 (n - 1) = (t_1 - 1)(t_2 - 1)$.
- The table is the same as in the additive 2-factor ANOVA.
- Just relabel factors $A$ and $B$ to Block and Treatment.
Nitrogen Fertilizer Results

> anova(lm(response~as.factor(row)+as.factor(treatment),
          data=fertilizerdata))

Analysis of Variance Table

Response: response

Df  Sum Sq Mean Sq  F value  Pr(>F)
as.factor(row)   3 197.004  65.668  9.1198 0.001116 **
as.factor(treatment) 5 201.316  40.263  5.5917 0.004191 **
Residuals         15 108.008   7.201
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Both treatment and blocking factor are significant at .005.
Nitrogen Fertilizer Results (Ignoring the Blocking)

> anova(lm(response~as.factor(treatment),data=fertilizerdata))

Analysis of Variance Table

Response: response

<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>as.factor(treatment)</td>
<td>5</td>
<td>201.316</td>
<td>40.263</td>
<td>2.3761</td>
<td>0.08024</td>
</tr>
<tr>
<td>Residuals</td>
<td>18</td>
<td>305.012</td>
<td>16.945</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

- Not significant at .05.
- Affected by extra variation induced by wetness gradient.
- Less discrimination power to see the fertilizer timing effect.
Compare the treatment $F$-statistics for the last two ANOVAs.

\[
F_{\text{Treat}}^{(1)} = \frac{\sum_{ij}(\bar{Y}_j - \bar{Y}_{..})^2/5}{\sum_{ij}(Y_{ij} - \bar{Y}_j - [\bar{Y}_i. - \bar{Y}_{..}])^2/15}
\]

and

\[
F_{\text{Treat}}^{(2)} = \frac{\sum_{ij}(\bar{Y}_j - \bar{Y}_{..})^2/5}{\sum_{ij}(Y_{ij} - \bar{Y}_j)^2/18}
\]

In $F_{\text{Treat}}^{(1)}$ the variation of the $Y_{ij}$ around $\bar{Y}_j$ is corrected for the variation due to $\bar{Y}_i. - \bar{Y}_{..}$, i.e., the row variation.

In $F_{\text{Treat}}^{(2)}$ such a correction is omitted, i.e., the row variation is ignored and absorbed as part of the $MS_E$:

\[
\sum_{ij}(Y_{ij} - \bar{Y}_j)^2 = \sum_{ij}(Y_{ij} - \bar{Y}_j - [\bar{Y}_i. - \bar{Y}_{..}])^2 + \sum_{ij}(\bar{Y}_i. - \bar{Y}_{..})^2
\]
The previous decomposition could use some elaboration:

\[
\sum_{ij} (Y_{ij} - \bar{Y}_{.j} - [\bar{Y}_{i.} - \bar{Y}_{..}])^2 = \sum_{ij} (Y_{ij} - \bar{Y}_{.j})^2 + \sum_{ij} (\bar{Y}_{i.} - \bar{Y}_{..})^2 - 2 \sum_{ij} (Y_{ij} - \bar{Y}_{.j})(\bar{Y}_{i.} - \bar{Y}_{..})
\]

\[
= \sum_{ij} (Y_{ij} - \bar{Y}_{.j})^2 + \sum_{ij} (\bar{Y}_{i.} - \bar{Y}_{..})^2 - 2 \sum_{ij} (\bar{Y}_{i.} - \bar{Y}_{..})^2
\]

\[
= \sum_{ij} (Y_{ij} - \bar{Y}_{.j})^2 - \sum_{ij} (\bar{Y}_{i.} - \bar{Y}_{..})^2
\]

\[\implies \sum_{ij} (Y_{ij} - \bar{Y}_{.j})^2 = \sum_{ij} (Y_{ij} - \bar{Y}_{.j} - [\bar{Y}_{i.} - \bar{Y}_{..}])^2 + \sum_{ij} (\bar{Y}_{i.} - \bar{Y}_{..})^2
\]
A Look at the Block Adjustment

- Model: \( Y_{ij} = \mu + b_i + \tau_j + \epsilon_{ij} \)
- Estimated block effect \( \hat{b}_i = \bar{Y}_i - \bar{Y} \)
- Consider the block adjusted observations
  \[ Z_{ij} = Y_{ij} - \hat{b}_i = Y_{ij} - (\bar{Y}_i - \bar{Y}_.) \]
- These \( Z_{ij} \), when plotted on separate levels for each block identifier \( i \), will look more aligned, more like “replicates.”
- Thus it makes sense to look at the average \( \bar{Z}.j - \bar{Z}. \) over those blocks as estimate for the \( j^{th} \) treatment effect.

\[
\begin{align*}
\bar{Z}.j - \bar{Z}. & = \bar{Y}_j - 0 - (\bar{Y}_. - 0) = \bar{Y}_j - \bar{Y}_. \\
Z_{ij} - \bar{Z}.j & = Y_{ij} - (\bar{Y}_i - \bar{Y}_.) - (\bar{Y}_j - 0) = Y_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y}_.
\end{align*}
\]

- Compare the treatment effect dispersion
  \[ \sum_{ij}(\bar{Z}.j - \bar{Z}.)^2 = \sum_{ij}(\bar{Y}_j - \bar{Y}_.)^2 \]
  against the within “replication” dispersion:
  \[ \sum_{ij}(Z_{ij} - \bar{Z}.j)^2 = \sum_{ij}(Y_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y}_.)^2 \]
  which leads back to our \( F^{(1)}_{\text{Treat}} \).
Visual Block Adjustment

response

response adjusted for block effect
Clear block effect.

Compare variability between and within blocks.

After the blocks are shifted to remove the individual block effects the data sets appear aligned on top of each other.

Pattern relationship within each block is undisturbed.

We are just using a magnified scale.

Easy to see differences between symbol groups (treatments) compared to variation within symbol groups.

Not easy to see without prior alignment.
Randomization Tests Revisited

- Assign 6 treatments randomly within each row.
- The number of possible assignments is
  \[(6!)^4 = 720^4 = 268,738,560,000 \approx 2.7 \cdot 10^{11}\]
- This is roughly \(1/10,000\) the number of patterns assigning 6 treatments in groups of 4 without row blocking restriction
  \[\left(\begin{array}{c}24 \\ 4\end{array}\right)\left(\begin{array}{c}20 \\ 4\end{array}\right)\left(\begin{array}{c}16 \\ 4\end{array}\right)\left(\begin{array}{c}12 \\ 4\end{array}\right)\left(\begin{array}{c}8 \\ 4\end{array}\right)\left(\begin{array}{c}4 \\ 4\end{array}\right) = 3.246671 \cdot 10^{15}\]
- If the treatment effects were identical, then the treatment assignment would not have affected the responses.
- Any other of the \(2.7 \cdot 10^{11}\) treatment assignments would have given us the same results \(Y_{ij}\).
- Test \(H_0 : \tau_1 = \ldots = \tau_6 (= 0)\) using as test statistics \(F_{Treat}\) and its resulting randomization reference distribution

\[
F_{Treat} = \frac{\sum_{ij}(\bar{Y}_{.j} - \bar{Y}_{..})^2/(6 - 1)}{\sum_{ij}(Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2/((6 - 1) \times (4 - 1))}
\]
According to the premise of no treatment effect, when calculating $F_{Treat}$ for all these $2.7 \cdot 10^{11}$ treatment assignments, we get $2.7 \cdot 10^{11}$ equally likely values of $F_{Treat}$, i.e., the reference distribution of $F_{Treat}$.

This distribution by theory $\approx F_{5,15}$-distribution.

$F_{Treat}^{obs}$, would then be like one of these equally likely values.

If $F_{Treat}^{obs}$ is in the far upper tail of the reference distribution, judge the result significant evidence against $H_0$.

Strength of conviction $\iff$ p-value $= P(F_{Treat} \geq F_{Treat}^{obs})$
Randomization Test for Treatment Effect

Randomization Distribution (F-test for Treatment Effect)

with $F_{5,15}$ density superimposed

$N_{sim} = 1e+05$

$p$-value = 0.00127
Comparing Results

- Randomization reference distribution
  \[ p\text{-value} = 0.00127 = 127/100000. \]

- Normal theory test for the same hypothesis and using the same test statistic came up with a \( p\)-value = 0.004191 from \( F_{5,15} \).

- Question: Is this explainable by statistical variation alone?
  Such statistical variation would come from the fact that we approximated the true randomization reference distribution by simulation from \( N_{\text{sim}} = 100,000 \) treatment allocations.

- A 95% upper confidence bound for the \( p\)-value resulting from the full true reference distribution can be obtained from R via
  \[ \text{qgamma}(0.95, 127 + 1)/100000 = 0.001471603 \] or via
  \[ \text{qbeta}(0.95, 127 + 1, 100000 - 127) = 0.001471455. \]

- That still leaves a wide gap to 0.004191.

- \( F_{5,15} \) approximation is not good enough out in the far tail.

- The superimposed \( F \)-density looks mostly like a good fit.
Deliberately undo the balanced design in poison/delivery data.

```r
> recip.time0 <- recip.time[c(-1,-5,-15)]
> delivery0 <- delivery[c(-1,-5,-15)]
> type0 <- type[c(-1,-5,-15)]

> table(type,delivery)
  delivery
 type  A  B  C  D
  I   4  4  4  4
  II  4  4  4  4
  III 4  4  4  4

> table(type0,delivery0)
  delivery0
 type0  A  B  C  D
  I   3  3  4  4
  II  3  4  4  4
  III 4  4  4  4
```
ANOVA in Unbalanced Designs

delivery0type0 <- lm(recip.time0 ~ delivery0*type0)
> coef(delivery0type0)

(Intercept) delivery0B delivery0C delivery0D
0.22405722 -0.10680749 -0.03778486 -0.05508903

type0II type0III delivery0B:type0II delivery0C:type0II delivery0D:type0II
0.11914618 0.25621130 -0.09705669 -0.03402663

delivery0D:type0II delivery0B:type0III delivery0C:type0III delivery0D:type0III
-0.11796097 -0.07056376 -0.01598499 -0.11599898

> anova(delivery0type0)

Analysis of Variance Table

Response: recip.time0

<table>
<thead>
<tr>
<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>delivery0</td>
<td>3</td>
<td>0.19124</td>
<td>0.063745</td>
<td>27.7042</td>
</tr>
<tr>
<td>type0</td>
<td>2</td>
<td>0.34039</td>
<td>0.170193</td>
<td>73.9673</td>
</tr>
<tr>
<td>delivery0:type0</td>
<td>6</td>
<td>0.02142</td>
<td>0.003570</td>
<td>1.5514</td>
</tr>
<tr>
<td>Residuals</td>
<td>33</td>
<td>0.07593</td>
<td>0.002301</td>
<td></td>
</tr>
</tbody>
</table>
ANOVA in Unbalanced Design (order reversed)

```r
> type0delivery0 <- lm(recip.time0 ~ type0*delivery0)
> coef(type0delivery0)
  (Intercept) type0II type0III delivery0B
delivery0C delivery0D type0II:delivery0B type0III:delivery0B
-0.03778486 -0.05508903 -0.09705669 -0.07056376
  type0II:delivery0C type0III:delivery0C type0II:delivery0D type0III:delivery0D
-0.03402663 -0.01598499 -0.11796097 -0.11599898

> anova(type0delivery0)

Analysis of Variance Table

Response: recip.time0

<table>
<thead>
<tr>
<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>type0</td>
<td>2</td>
<td>0.35058</td>
<td>0.175291</td>
<td>76.1830</td>
</tr>
<tr>
<td>delivery0</td>
<td>3</td>
<td>0.18104</td>
<td>0.060347</td>
<td>26.2271</td>
</tr>
<tr>
<td>type0:delivery0</td>
<td>6</td>
<td>0.02142</td>
<td>0.003570</td>
<td>1.5514</td>
</tr>
<tr>
<td>Residuals</td>
<td>33</td>
<td>0.07593</td>
<td>0.002301</td>
<td></td>
</tr>
</tbody>
</table>
```
The type and delivery data subspaces are no longer orthogonal.

Each \( x \in \text{delivery subspace} \) can be decomposed into
\[
x = y + z \text{ with } y \in \text{type subspace}.
\]

The projection onto the type subspace will reflect a component within the delivery subspace.

The subsequent projection onto the delivery subspace will only account for the part not yet explained by the previous projection.

Reversing type and delivery has a corresponding effect.

The order of projection matters for the distances (or SS’s).

The parameter estimates are the same (same model).