University of Washington

STATISTICS

Applied Statistics and Experimental Design
Factorial Design

Fritz Scholz

Fall Quarter 2008

Factorial Design

So far we have looked at 1-sample, 2-sample, and t-sample problems.

In the latter 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 we will 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

```
> poison=read.csv("poison.csv",header=T)
> 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
```

Insecticide Example

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.

We want to 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), $\ldots$, (III,C), (III,D).
Randomize the order of all 48 runs to eliminate order biases.

This is a factorial design, specifically a 3×4 factor design with 4 replications.

Response Table

It is useful to visualize the responses in relation to the factor levels as follows:

		Delivery Method			
		A	B	C	D
Insecticide Type	I	$\mathbf{Y}_{\text{I,A}}$	$\mathbf{Y}_{\text{I,B}}$	$\mathbf{Y}_{\text{I,C}}$	$\mathbf{Y}_{\text{I,D}}$
	II	$\mathbf{Y}_{\text{II,A}}$	$\mathbf{Y}_{\text{II,B}}$	$\mathbf{Y}_{\text{II,C}}$	$\mathbf{Y}_{\text{II,D}}$
	III	$\mathbf{Y}_{\text{III,A}}$	$\mathbf{Y}_{\text{III,B}}$	$\mathbf{Y}_{\text{III,C}}$	$\mathbf{Y}_{\text{III,D}}$

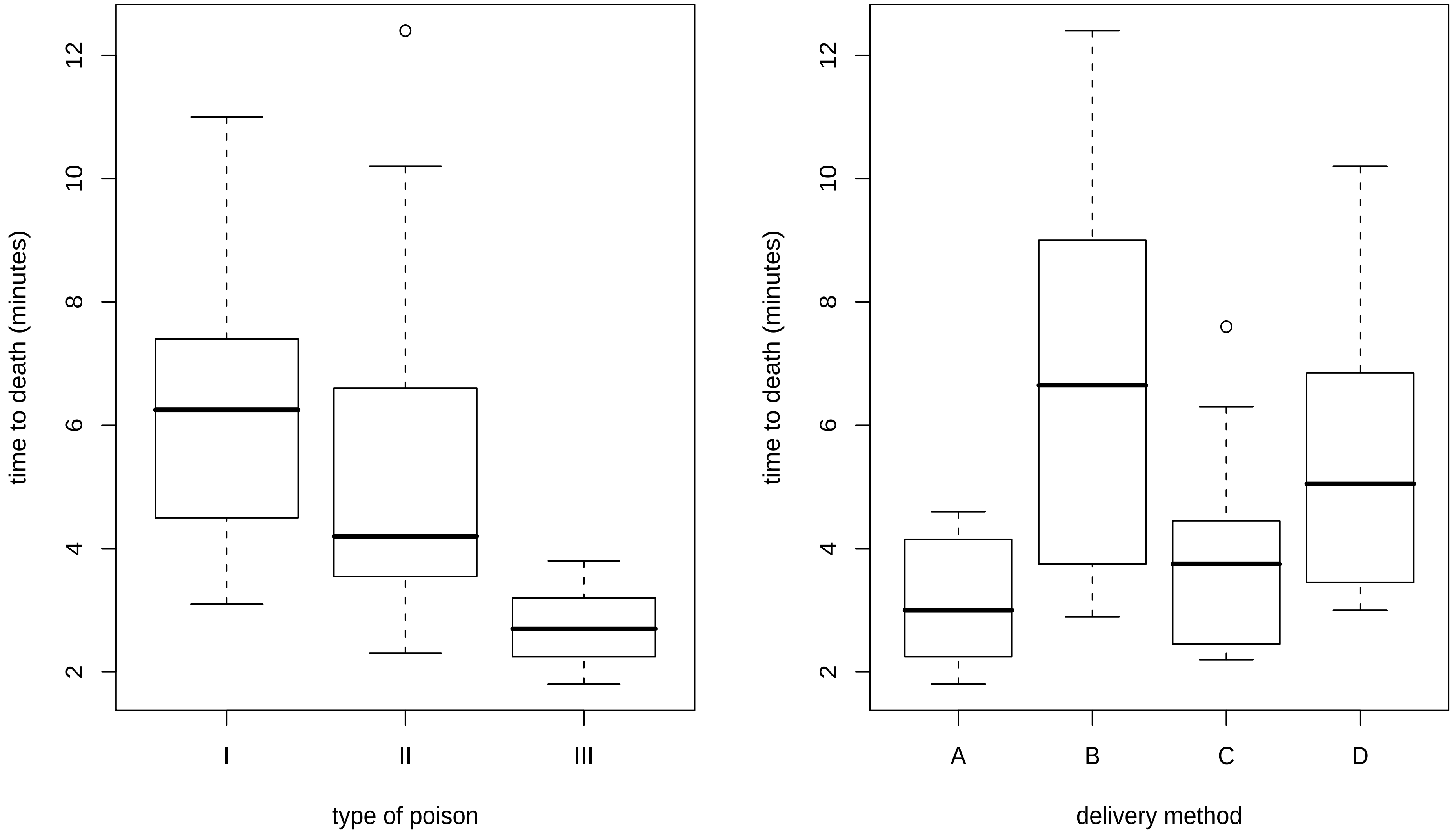
where $\mathbf{Y}_{\text{I,A}}$ stands short for $(Y_{\text{I,A},1},\ldots,Y_{\text{I,A},4})$ (replication depth), and so on.

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}.

This triplet notation (i,j,k) is more useful than a single index $\ell = 1,2,\ldots,48$.

Useful in Σ summation notation and also for identifying the factor levels, i.e., the factor level/replication coordinates.

First Look at Insecticide Boxplots



First Impressions and Questions

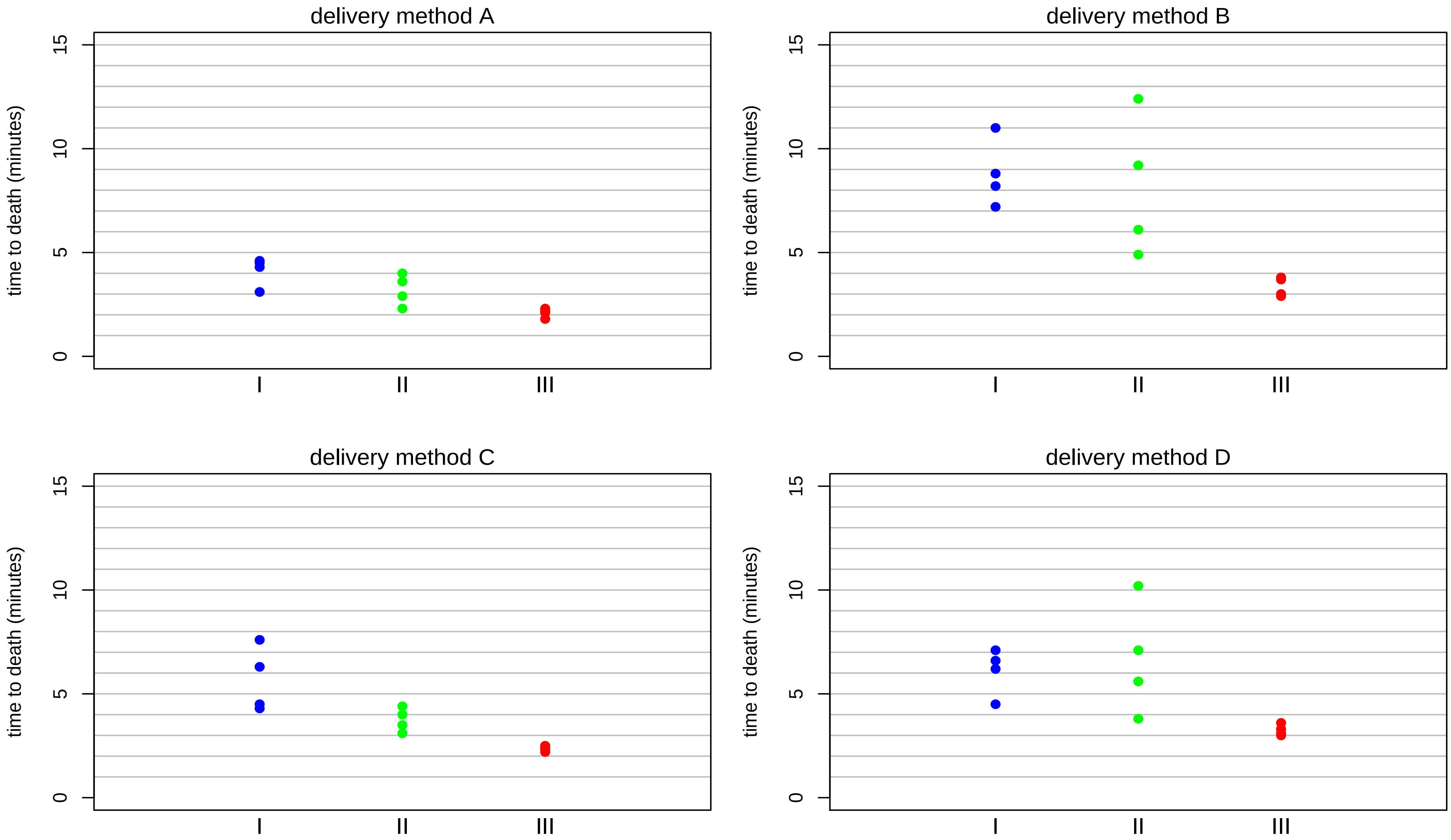
Insecticide type III and delivery method A seem to give the best combination.
Is combination the right word here?

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?

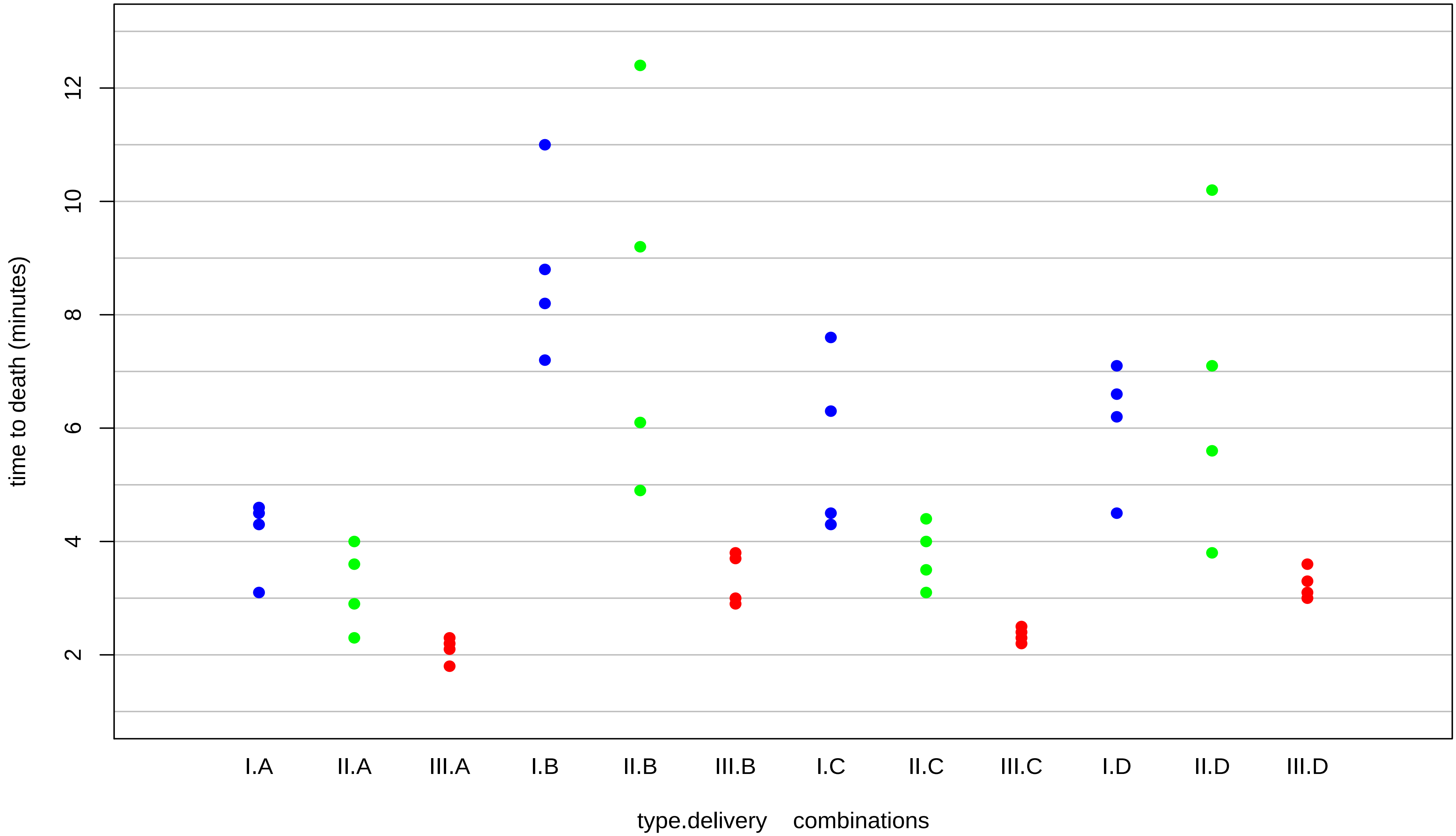
It could be that delivery type A is not the fastest acting among all four when applied to insecticide type III.

Delivery method C could actually be better in combination with III.

Insecticide Responses by Delivery Method



Full Comparison of Insecticide Responses



Comments

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.

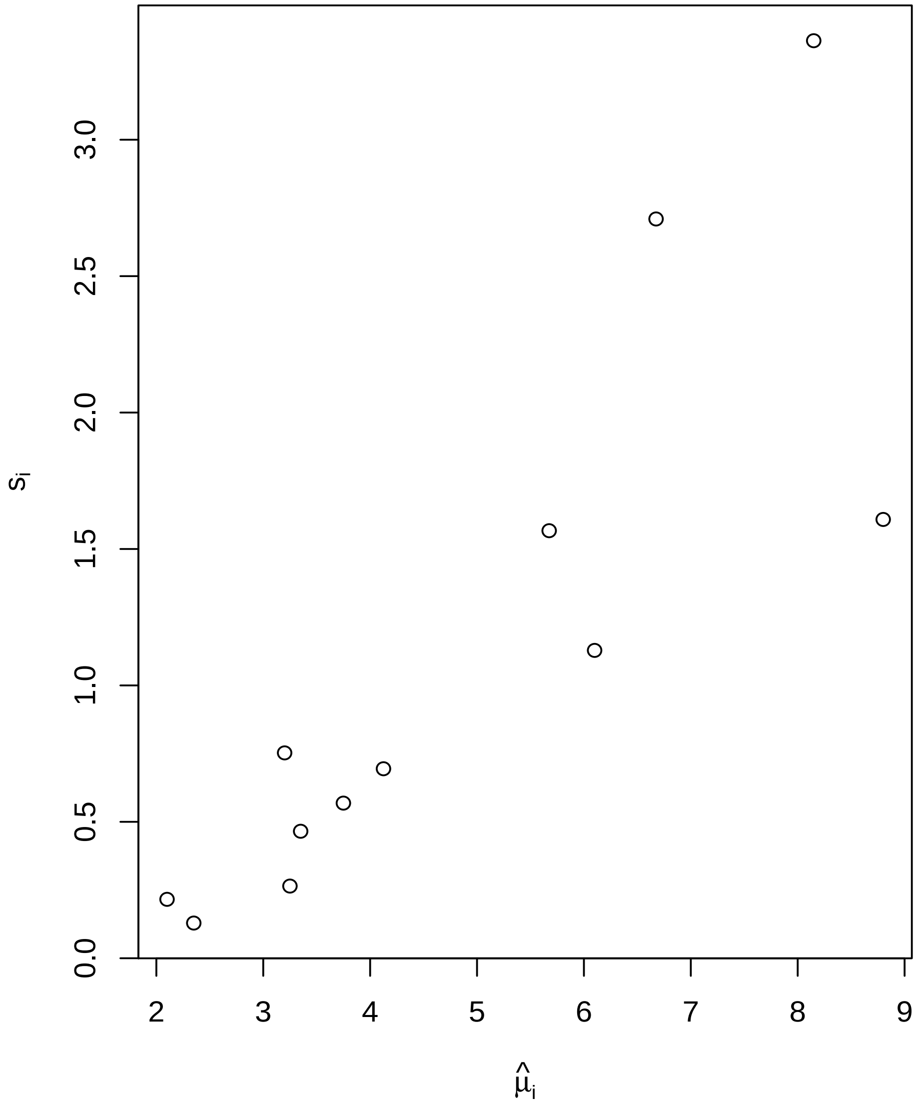
Delivery appears to have an effect on the response under type I and II,

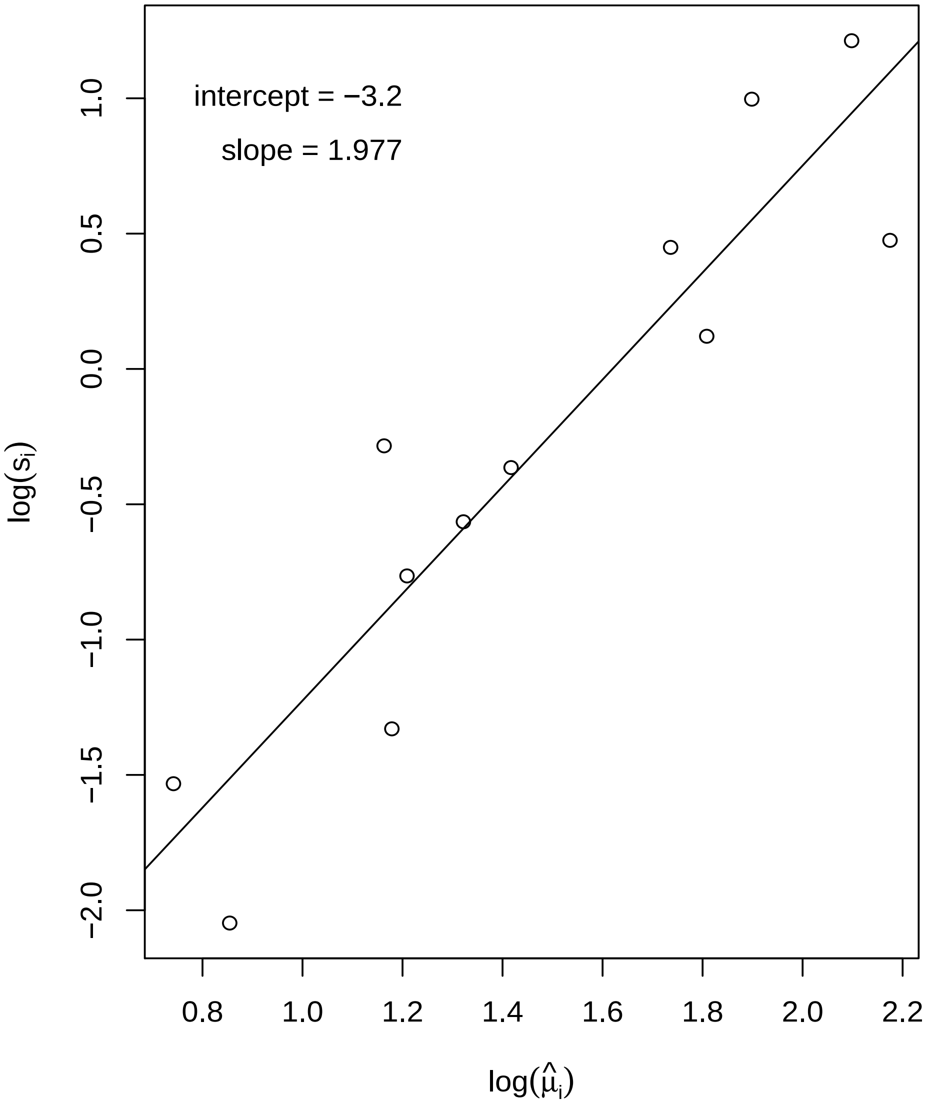
both in absolute terms and relative to each other.

It appears that scatter $\nearrow$ as mean $\nearrow$ across all combinations.

$\Longrightarrow$ variance stabilizing transformation. Deal with that first.

Linear Fit $\log(s_i) = a \times \log(\hat{\mu}_i) + b$





Reciprocal Transform

According to our guidelines this suggests $\alpha = 2$ or $\lambda = 1 - \alpha = -1$,

i.e., $\tilde{Y}_{ijk} = Y_{ijk}^{-1} = 1/Y_{ijk}$ a reciprocal transform for our response times.

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.

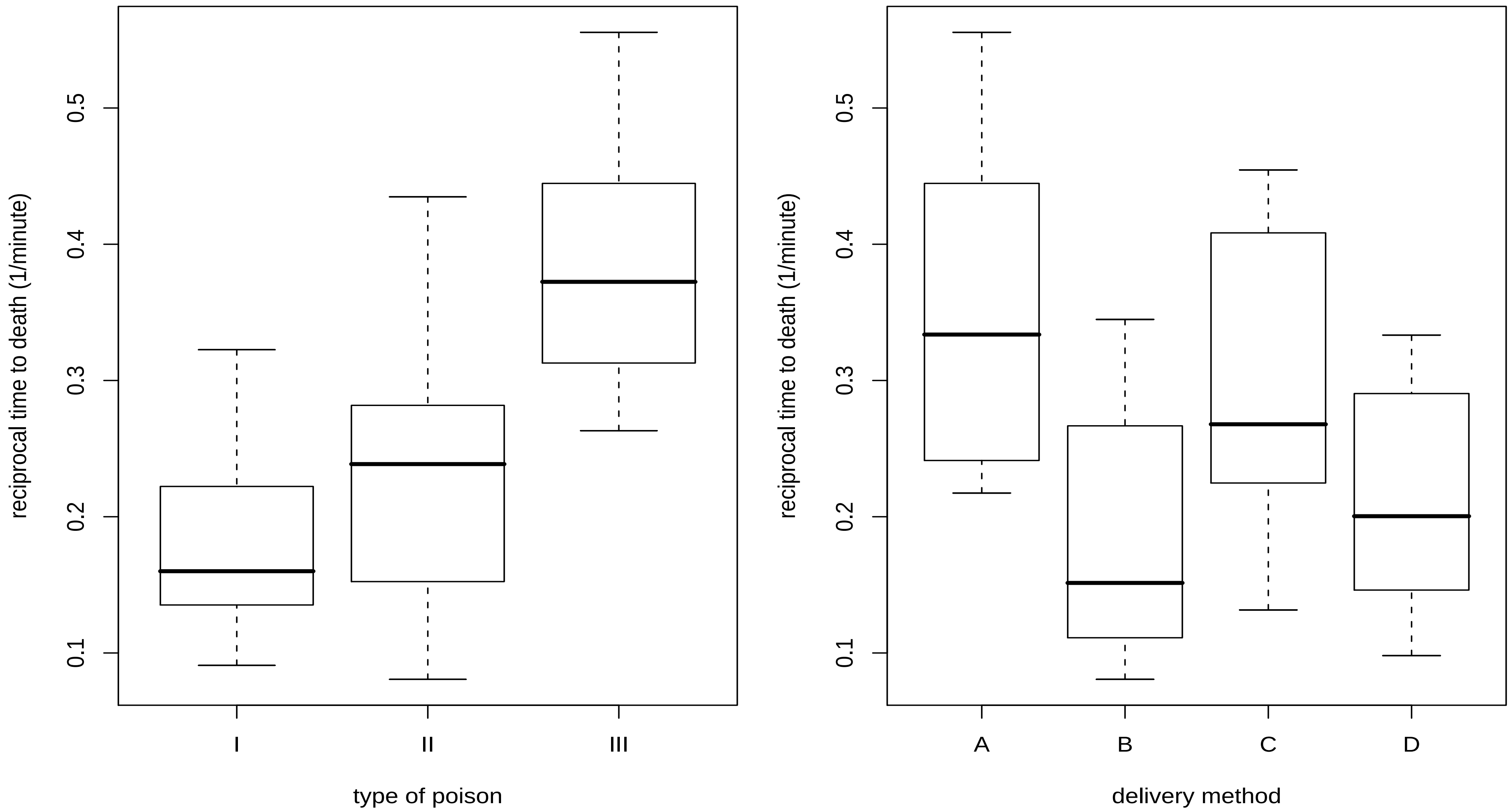
Assume further, that the lethal dose D is $\approx$ constant for each type for all insects.

Then the time to reach lethal dose is $T = D/R$. If we took $1/T = R/D$ as transformed response we would have constant variability in $1/T$.

Linearizing by a 1-term Taylor expansion around μ_R and treating D as a constant

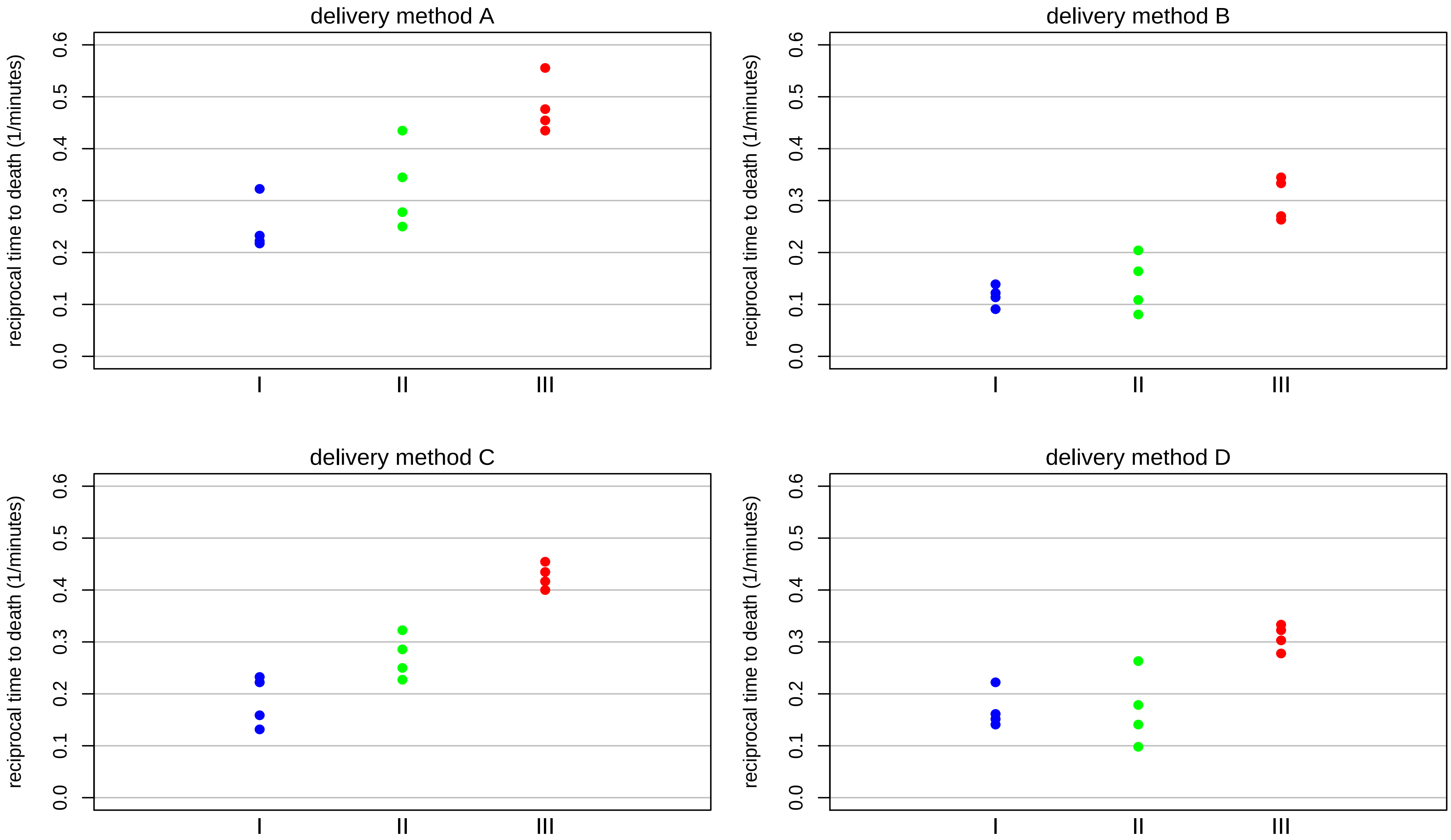
$$T = \frac{D}{R} \approx \frac{D}{\mu_R} - (R - \mu_R)\frac{1}{\mu_R^2} \;\Rightarrow\; \mu_T \approx \frac{D}{\mu_R}\,,\;\; \mathrm{var}(T) \approx \frac{\sigma_R^2}{\mu_R^4} \;\Rightarrow\; \sigma_T \propto \mu_T^2$$

Reciprocal Time Boxplots

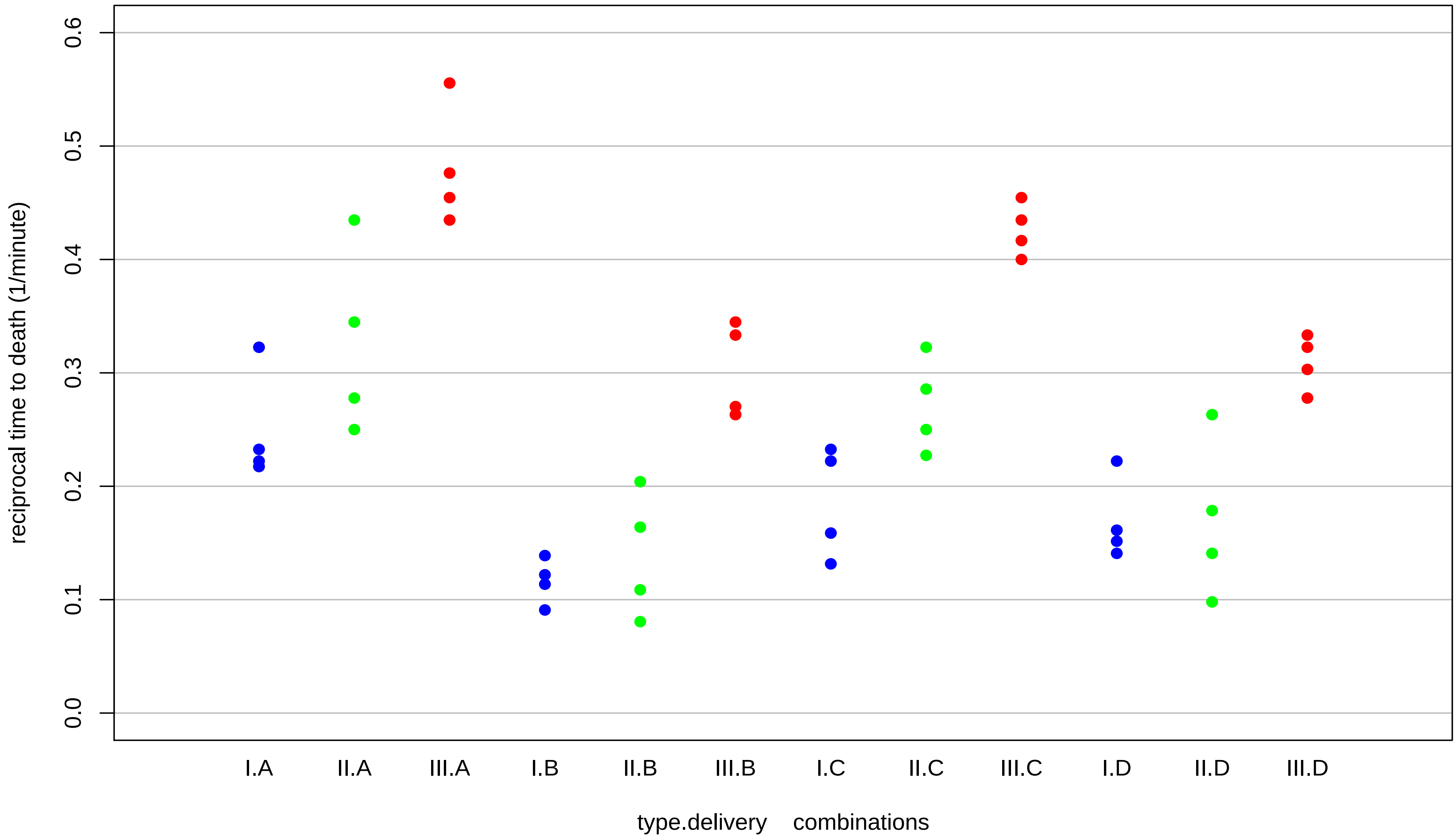


The variability is much more stable now.

Reciprocal Time by Delivery Method



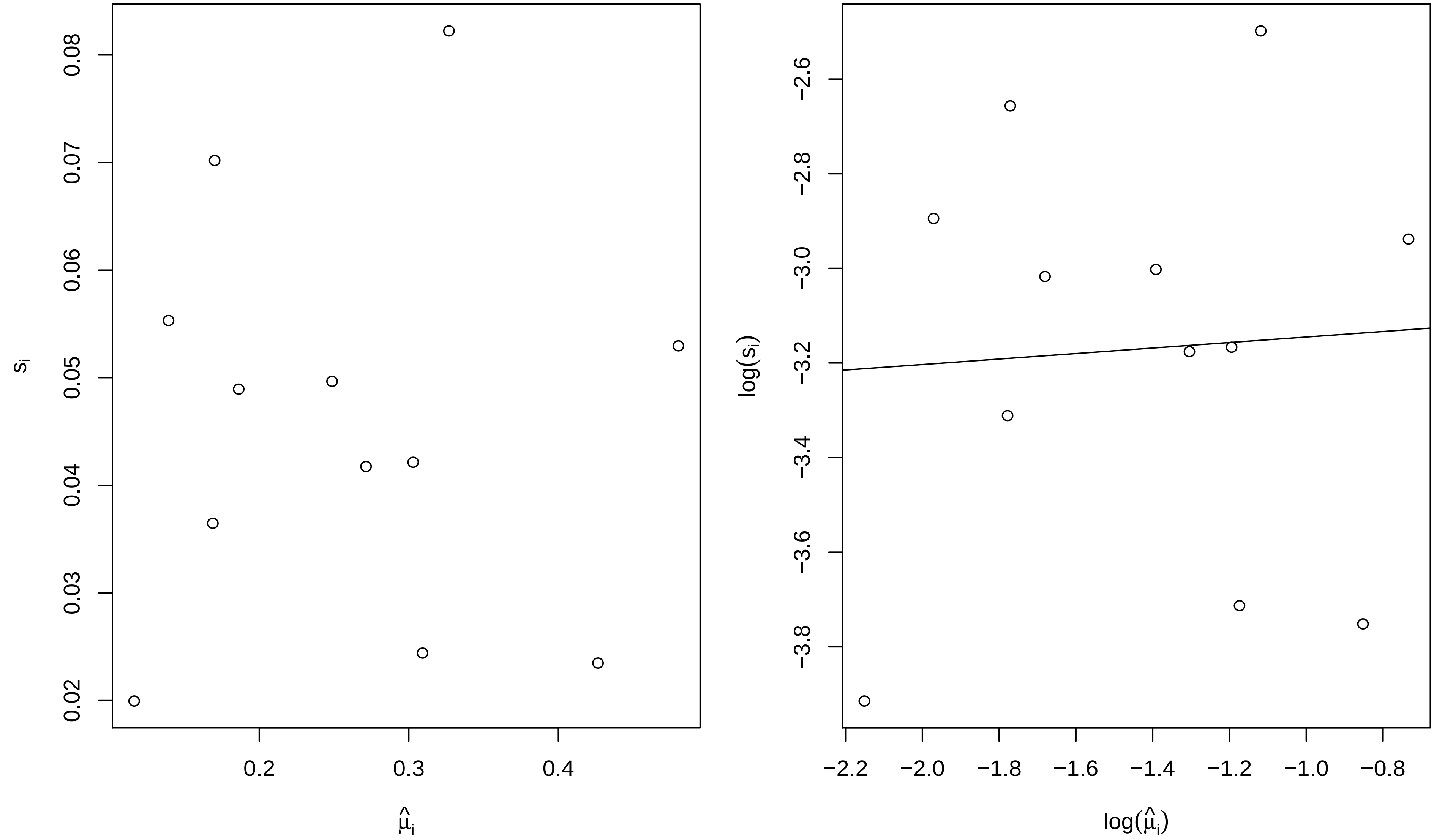
Full Comparison of Reciprocal Times



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



ANOVA of Reciprocal Times vs Type

Assuming `recip.time, type, delivery` are variables in the workspace.

`type` and `delivery` are factors or write `as.factor(type)` in place of `type`!

```
> anova(lm(recip.time ∼ type))
```

or

```
> 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  25.621 3.728e-08 ***
Residuals 45 0.30628 0.00681
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

ANOVA of Reciprocal Times vs Delivery

```
> anova(lm(recip.time ~ delivery))
```

or

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

```
Analysis of Variance Table
Response: recip.time
          Df  Sum Sq Mean Sq F value    Pr(>F)
delivery   3 0.20414 0.06805  6.6401 0.0008496 ***
Residuals 44 0.45091 0.01025
```

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
              Df  Sum Sq Mean Sq F value    Pr(>F)
type:delivery 11 0.56862 0.05169  21.531 1.289e-12 ***
Residuals     36 0.08643 0.00240
```

This is like a one-way ANOVA with 12 treatments. Legitimate but not enlightening.

All Three ANOVAs for Reciprocal Times

```
$ANOVA.type Analysis of Variance Table
Response: recip.time
          Df  Sum Sq Mean Sq F value    Pr(>F)
type       2 0.34877 0.17439  25.621 3.728e-08 ***
Residuals 45 0.30628 0.00681

$ANOVA.delivery Analysis of Variance Table
Response: recip.time
          Df  Sum Sq Mean Sq F value    Pr(>F)
delivery   3 0.20414 0.06805  6.6401 0.0008496 ***
Residuals 44 0.45091 0.01025

$ANOVA.type.delivery Analysis of Variance Table
Response: recip.time
              Df  Sum Sq Mean Sq F value    Pr(>F)
type:delivery 11 0.56862 0.05169  21.531 1.289e-12 ***
Residuals     36 0.08643 0.00240
```

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} in the first two ANOVAs are inflated because mean variation in the ignored factor is absorbed as 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).

Similarly, the variation within each of the three colors (poison type) 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.   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.

We view the $3 \times 4 = 12$ combinations as $t = 12$ treatments or 12 samples.

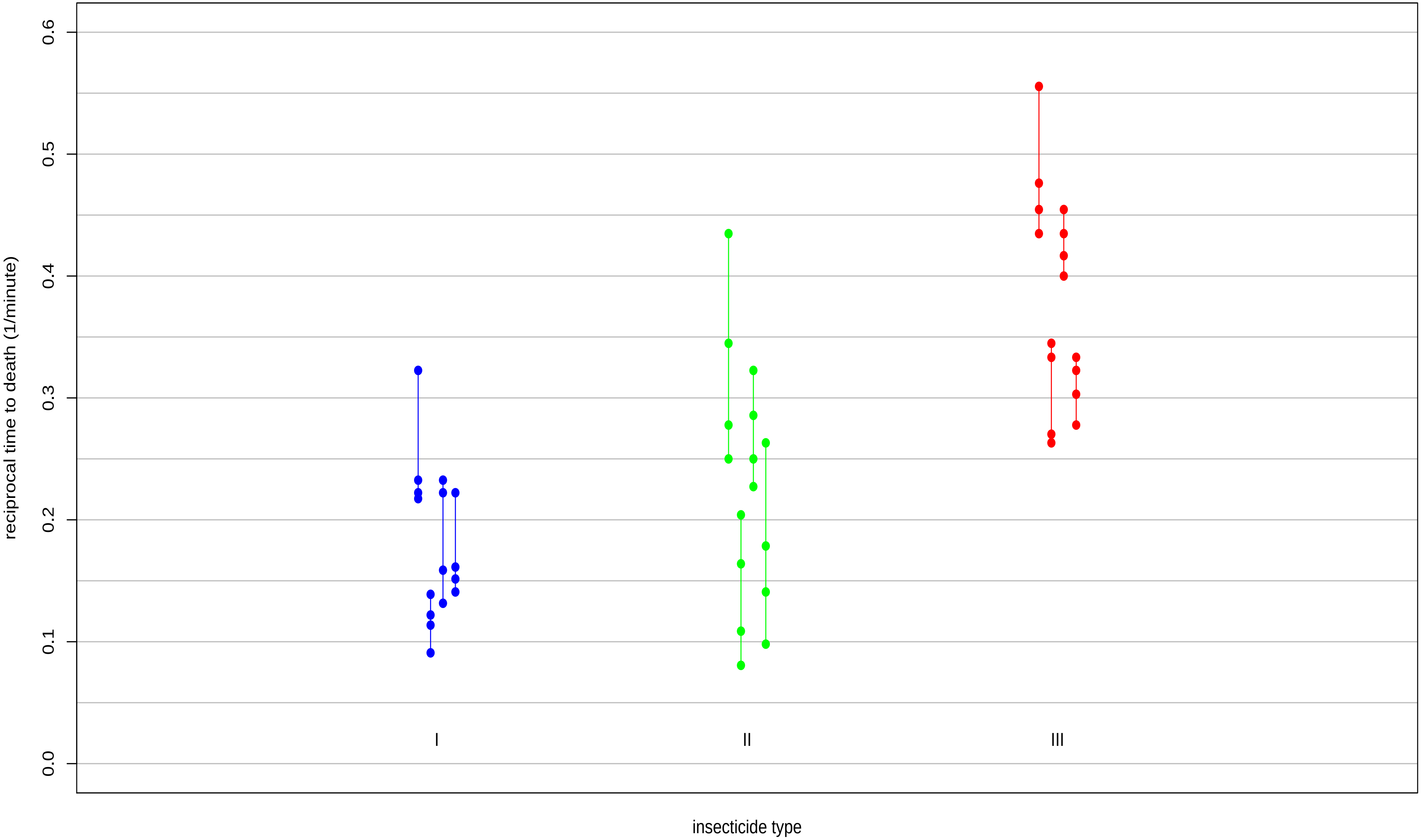
Does the third ANOVA give any insight on the separate contributions of the type factor and the delivery factor? No! Hence it is insufficient.

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 could come about when the MS_{E} in the first two ANOVAs are unduly inflated compared to a correct MS_{E} in the third ANOVA.

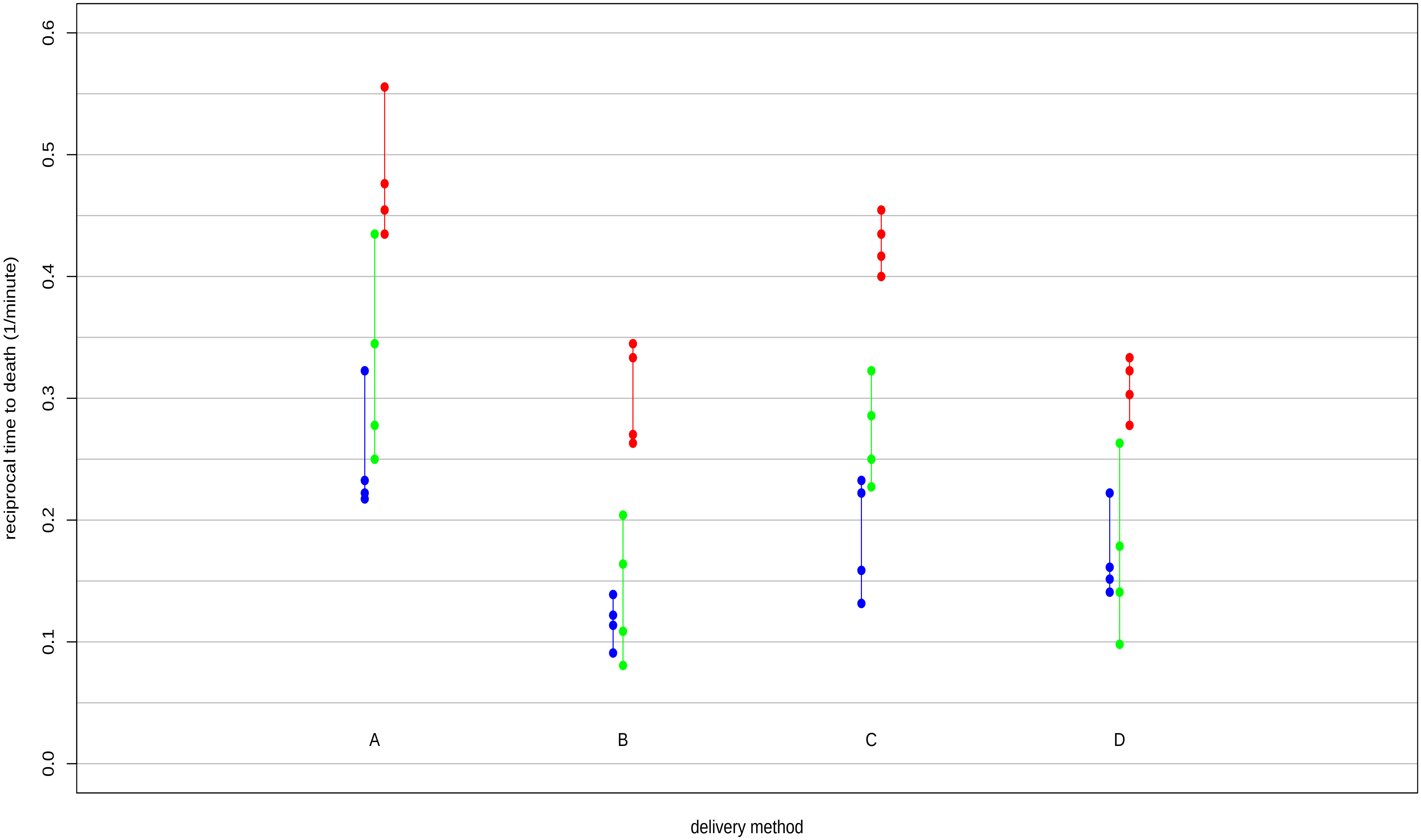
The next 5 slides illustrate this by doctoring the data appropriately.

View by Type



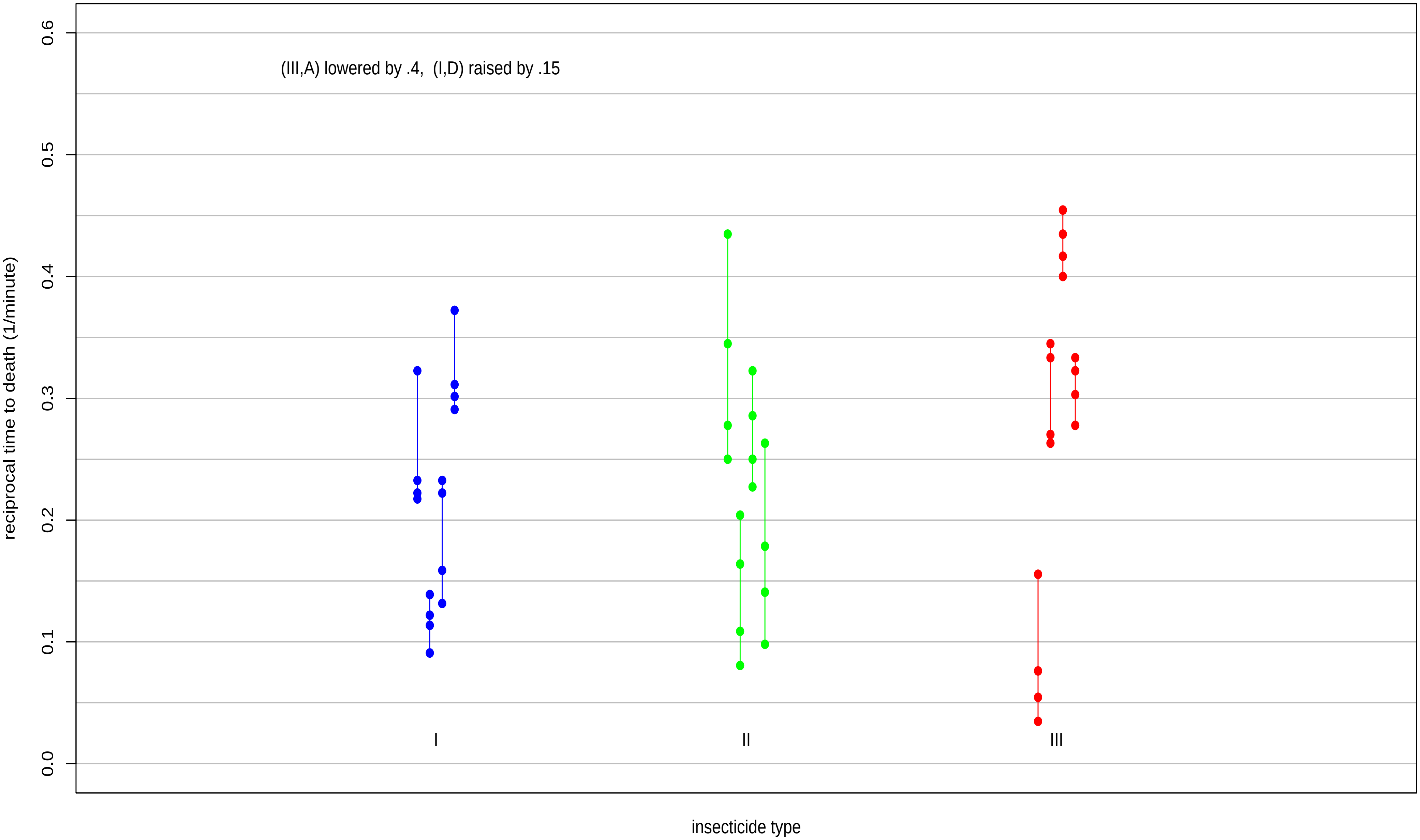
The delivery variation is a good part of the “error” variation within type.

View by Delivery



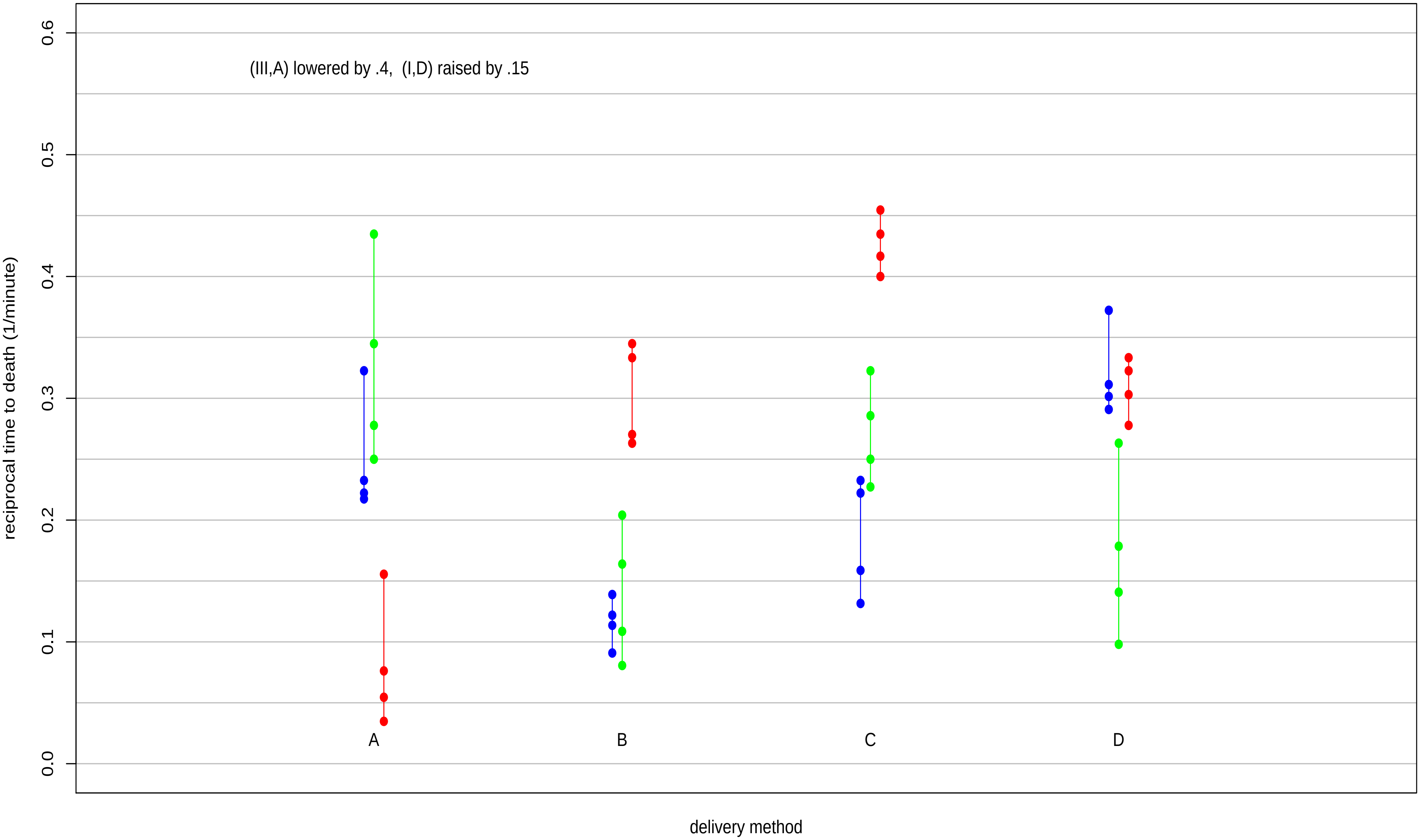
The type variation is a good part of the “error” variation within delivery.

View by Type



The 3 type data sets seem well meshed when ignoring the 4 delivery methods.

View by Delivery



The 4 delivery methods seem well meshed when ignoring the 3 poison types.

Three ANOVAs for Modified Data

```
$ANOVA.type  Analysis of Variance Table
Response: recip.time
          Df  Sum Sq Mean Sq F value Pr(>F)
type       2 0.03592 0.01796  1.5582 0.2217
Residuals 45 0.51868 0.01153

$ANOVA.delivery  Analysis of Variance Table
Response: recip.time
          Df  Sum Sq Mean Sq F value  Pr(>F)
delivery   3 0.08425 0.02808  2.6271 0.06208 .
Residuals 44 0.47035 0.01069

$ANOVA.type.delivery  Analysis of Variance Table
Response: recip.time
              Df  Sum Sq Mean Sq F value    Pr(>F)
type:delivery 11 0.46817 0.04256  17.727 2.294e-11 ***
Residuals     36 0.08643 0.00240
```

Three ANOVAs for Modified Data

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 (0.08643 as compared to 0.51868 or 0.47035).

Additive Effects Model

We will deal only with the balanced model, same number n of replications per factor level combination. Unbalanced cases can get quite messy.[*]

Additive Effects Model:

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

with the error terms $\varepsilon_{ijk}\,, \ i = 1,\ldots,t_1\,, \ j = 1,\ldots,t_2\,, \ k = 1,\ldots,n$ assumed to be independent with mean zero and common variance σ^2.

[*] And extensive treatment of the unbalanced case can be found in S.R. Searle (1987), *Linear Models for Unbalanced Data*, John Wiley & Sons, New York

Dealing with Identifiability Issues

As it is, the model parameters, μ, $a_1,\ldots,a_{t_1}$, and $b_1,\ldots,b_{t_2}$, are unidentifiable, since adding a constant c to μ and subtracting it from the a_i (or the b_j) would give the same means $\mu_{ij} = E(Y_{ijk})$ for all factor level combinations (i,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.

1) Require: $\sum_i a_i = 0$ and $\sum_j b_j = 0$ sum-to-zero side conditions,

2) Require: $a_1 = 0$ and $b_1 = 0$ set-to-zero side conditions.

Sum-to-zero Side Conditions

Assume $\mu_{ij} = \mu + a_i + b_j$ with $\sum_i a_i = \sum_j b_j = 0$ or $\bar{a}_\bullet = \bar{b}_\bullet = 0$.

Here we identify μ with the average mean over all level combinations,

because $\bar{\mu}_{\bullet\bullet} = \mu + \bar{a}_\bullet + \bar{b}_\bullet = \mu$.

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

The parameters μ, a_i, b_j with $\bar{a}_\bullet = \bar{b}_\bullet = 0$ define the means μ_{ij} with the above additive structure and in turn are uniquely identified via the μ_{ij}, as shown above.

Changes in both levels are additive $\mu_{ij} = \mu + a_i + b_j$

Set-to-zero Side Conditions

Assume $\mu_{ij} = \mu^\star + a_i^\star + b_j^\star$ with $a_1^\star = b_1^\star = 0$.

Here we identify the parameter $\mu^\star$ with the mean under the factor level combination $(1,1)$, i.e. with level 1 for each factor, since $\mu_{11} = \mu^\star + a_1^\star + b_1^\star = \mu^\star$.

We express each change from $\mu^\star$ due to other levels ($\neq 1$) in factor one via $a_i^\star$, i.e., $\mu_{i1} = \mu^\star + a_i^\star + b_1^\star = \mu^\star + a_i^\star$ and thus $a_i^\star = \mu_{i1} - \mu^\star = \mu_{i1} - \mu_{11}$.

Similarly, each change from $\mu^\star$ due to other levels ($\neq 1$) in factor two is expressed via $b_j^\star$, i.e., $\mu_{1j} = \mu^\star + a_1^\star + b_j^\star = \mu^\star + b_j^\star$ and thus $b_j^\star = \mu_{1j} - \mu^\star = \mu_{1j} - \mu_{11}$.

The parameters $\mu^\star, a_i^\star, b_j^\star$ with $a_1^\star = b_1^\star = 0$ define the means μ_{ij} with the above additive structure and in turn are uniquely identified via the μ_{ij}, as shown above.

Changes in both levels are additive $\mu_{ij} = \mu^\star + a_i^\star + b_j^\star$

How to Create Such Mean Structures

If you were 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 μ.

In the sum-to-zero case take

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

In the set-to-zero case take $a_1 = 0$ and $b_1 = 0$

In either case define

$$\mu_{ij} = \mu + a_i + b_j$$

Additive Model as Reduced Model

The additive model is a reduced model since in the full model each factor level combination (i,j) has its own mean μ_{ij}. There are $t_1 \times t_2$ such means μ_{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 (see previous slide).

Note that $t_1 \times t_2 - [t_1+t_2-1] = (t_1-1)\times(t_2-1)$ 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 situation, because then the second factor only has one level, i.e., does not change.

A Tabular View of the Additive Model

$t_1 \times t_2 = 3 \times 4$ Factorial Design

Factor 2

		b_1	b_2	b_3	b_4
	a_3	$\mu + a_3 + b_1$	$\mu + a_3 + b_2$	$\mu + a_3 + b_3$	$\mu + a_3 + b_4$
Factor 1	a_2	$\mu + a_2 + b_1$	$\mu + a_2 + b_2$	$\mu + a_2 + b_3$	$\mu + a_2 + b_4$
	a_1	$\mu + a_1 + b_1$	$\mu + a_1 + b_2$	$\mu + a_1 + b_3$	$\mu + a_1 + b_4$

Rows i and i' differ by $a_i - a_{i'} = \mu_{ij} - \mu_{i'j} = a_i^\star - a_{i'}^\star$ across all columns.

Columns j and j' differ by $b_j - b_{j'} = \mu_{i,j} - \mu_{ij'} = b_j^\star - b_{j'}^\star$ across all rows.

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 $\Longleftrightarrow$ 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

	5	7	9	2	3
$a_2^\star = 2$	7	9	11	4	5
$a_3^\star = 4$	9	11	13	6	7
$a_4^\star = 5$	10	12	14	7	8
		$b_2^\star = 2$	$b_3^\star = 4$	$b_4^\star = -3$	$b_5^\star = -2$

The columns differ automatically by constant amounts $b_j^\star$ $\Rightarrow$ additivity in factor 2.

The column differences are set in the first row and are not affected by translating that first row to various levels via the $a_i^\star$ (additivity in factor 1).

Additive Model Decomposition

$$\begin{aligned}
Y_{ijk} &= \bar{Y}_{\bullet\bullet\bullet} + (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}) + (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet}) + (Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet} - [\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}] - [\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet}]) \\
&= \bar{Y}_{\bullet\bullet\bullet} + (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}) + (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet}) + (Y_{ijk} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet}) \\
&= \hat{\mu} \quad + \qquad \hat{a}_i \qquad + \qquad \hat{b}_j \qquad + \qquad\qquad \hat{\varepsilon}_{ijk}
\end{aligned}$$

Note that

$$\begin{aligned}
\sum_i \hat{a}_i &= t_1 \frac{1}{t_1} \sum_i (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}) = t_1(\bar{Y}_{\bullet\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}) = 0 \qquad \text{sum to zero} \\
\sum_j \hat{b}_j &= t_2 \frac{1}{t_2} \sum_j (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet}) = t_2(\bar{Y}_{\bullet\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}) = 0 \qquad \text{sum to zero} \\
\sum_{ik} \hat{\varepsilon}_{ijk} &= nt_1 \frac{1}{nt_1} \sum_{ik} (Y_{ijk} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet}) = nt_1(\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet}) = 0 \\
\sum_{jk} \hat{\varepsilon}_{ijk} &= nt_2 \frac{1}{nt_2} \sum_{jk} (Y_{ijk} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet}) = nt_2(\bar{Y}_{i\bullet\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet} + \bar{Y}_{\bullet\bullet\bullet}) = 0 \\
\sum_{ijk} \hat{\varepsilon}_{ijk} &= 0 \qquad \text{all residuals sum to zero.}
\end{aligned}$$

Least Squares Estimates

$\hat{\mu} = \bar{Y}_{\bullet\bullet\bullet}$, $\hat{a}_i = \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}$, and $\hat{b}_j = \bar{Y}_{\bullet j \bullet} - \bar{Y}_{\bullet\bullet\bullet}$ are the least squares estimates of μ, a_i, and b_j subject to the conditions $\sum_i a_i = 0$ and $\sum_j b_j = 0$.

Using the above decomposition identity this follows without calculus from

$$Q(\mu, a_1, \ldots, a_{t_1}, b_1, \ldots, b_{t_2}) = \sum_{ijk} (Y_{ijk} - \mu - a_i - b_j)^2 =$$

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

$$\sum_{ijk} \left\{(\bar{Y}_{\bullet\bullet\bullet} - \mu)^2 + [(\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}) - a_i]^2 + [(\bar{Y}_{\bullet j \bullet} - \bar{Y}_{\bullet\bullet\bullet}) - b_j]^2 + (Y_{ijk} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j \bullet} + \bar{Y}_{\bullet\bullet\bullet})^2\right\} .$$

All cross product terms disappear (see next slide). It is now obvious that the minimization of $Q(\mu, a_1, \ldots, a_{t_1}, b_1, \ldots, b_{t_2})$ is accomplished by taking $\mu = \hat{\mu} = \bar{Y}_{\bullet\bullet\bullet}$,

$a_i = \hat{a}_i = \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}$ and $b_j = \hat{b}_j = \bar{Y}_{\bullet j \bullet} - \bar{Y}_{\bullet\bullet\bullet}$.

Note that $\sum_i \hat{a}_i = \sum_j \hat{b}_j = 0$, i.e., the solutions satisfy the constraints.

Cross Product Terms $= 0$

$$\sum_{ijk}(\bar{Y}_{\bullet\bullet\bullet}-\mu)[(\bar{Y}_{i\bullet\bullet}-\bar{Y}_{\bullet\bullet\bullet})-a_i] = (\bar{Y}_{\bullet\bullet\bullet}-\mu)t_2 n\left\{\sum_i(\bar{Y}_{i\bullet\bullet}-\bar{Y}_{\bullet\bullet\bullet})-\sum_i a_i\right\}=0$$

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

$$\sum_{ijk}(\bar{Y}_{\bullet\bullet\bullet}-\mu)(Y_{ijk}-\bar{Y}_{i\bullet\bullet}-\bar{Y}_{\bullet j\bullet}+\bar{Y}_{\bullet\bullet\bullet}) = (\bar{Y}_{\bullet\bullet\bullet}-\mu)\sum_{ijk}\hat{\varepsilon}_{ijk}=0$$

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

$$\sum_{ijk}[(\bar{Y}_{i\bullet\bullet}-\bar{Y}_{\bullet\bullet\bullet})-a_i](Y_{ijk}-\bar{Y}_{i\bullet\bullet}-\bar{Y}_{\bullet j\bullet}+\bar{Y}_{\bullet\bullet\bullet}) = \sum_i\left\{[(\bar{Y}_{i\bullet\bullet}-\bar{Y}_{\bullet\bullet\bullet})-a_i]\sum_{jk}\hat{\varepsilon}_{ijk}\right\}=0$$

$$\sum_{ijk}[(\bar{Y}_{\bullet j\bullet}-\bar{Y}_{\bullet\bullet\bullet})-b_j](Y_{ijk}-\bar{Y}_{i\bullet\bullet}-\bar{Y}_{\bullet j\bullet}+\bar{Y}_{\bullet\bullet\bullet}) = \sum_j\left\{[(\bar{Y}_{\bullet j\bullet}-\bar{Y}_{\bullet\bullet\bullet})-b_j]\sum_{ik}\hat{\varepsilon}_{ijk}\right\}=0$$

Estimates: Sum-to-Zero $\longrightarrow$ Set-to-Zero

As sum-to-zero estimates we have: $\hat{\mu} = \bar{Y}_{\bullet\bullet\bullet}$, $\hat{a}_i = \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}$, and $\hat{b}_j = \bar{Y}_{\bullet j \bullet} - \bar{Y}_{\bullet\bullet\bullet}$,

$\Longrightarrow \quad \hat{\mu}_{ij} = \hat{\mu} + \hat{a}_i + \hat{b}_j$.

These sum-to-zero estimates are converted to their set-to-zero counterparts via

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

Note that $\hat{a}_1^\star = \hat{b}_1^\star = 0$.

Furthermore, $\hat{\mu}_{ij} = \hat{\mu}^\star + \hat{a}_i^\star + \hat{b}_j^\star = \hat{\mu} + \hat{a}_i + \hat{b}_j$ and

$$\hat{\mu}^\star = \bar{Y}_{1\bullet\bullet} + \bar{Y}_{\bullet 1 \bullet} - \bar{Y}_{\bullet\bullet\bullet}, \quad \hat{a}_i^\star = \bar{Y}_{i\bullet\bullet} - \bar{Y}_{1\bullet\bullet}, \quad \hat{b}_j^\star = \bar{Y}_{\bullet j \bullet} - \bar{Y}_{\bullet 1 \bullet}$$

Estimates: Set-to-Zero $\longrightarrow$ Sum-to-Zero

We can go from the set-to-zero estimates $\hat{\mu}^\star, \hat{a}_i^\star, \hat{b}_j^\star$ with $\hat{a}_1^\star = 0$ and $\hat{b}_1^\star = 0$,

which define $\hat{\mu}_{ij} = \hat{\mu}^\star + \hat{a}_i^\star + \hat{b}_j^\star$,

to the sum-to-zero equivalent representation

$$\hat{\mu} = \hat{\mu}_{\bullet\bullet} = \hat{\mu}^\star + \hat{a}_\bullet^\star + \hat{b}_\bullet^\star\,, \quad \hat{a}_i = \hat{\mu}_{i\bullet} - \hat{\mu}_{\bullet\bullet} = \hat{a}_i^\star - \hat{a}_\bullet^\star\,, \quad \hat{b}_j = \hat{\mu}_{\bullet j} - \hat{\mu}_{\bullet\bullet} = \hat{b}_j^\star - \hat{b}_\bullet^\star$$

$$\Longrightarrow \qquad \hat{\mu}_{ij} = \hat{\mu}^\star + \hat{a}_i^\star + \hat{b}_j^\star = \hat{\mu} + \hat{a}_i + \hat{b}_j$$

$$\Longrightarrow \quad \hat{\mu} = \bar{Y}_{\bullet\bullet\bullet}\,, \quad \hat{a}_i = \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}\,, \quad \text{and} \quad \hat{b}_j = \bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet}$$

Set-to-zero is what `lm` in R gives as coefficients, see example later.

Fitted Models

The fitted mean per treatment combination under either parametrization (sum-to-zero or set-to-zero) are the same, i.e.,

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

Only the components of these fitted values have different interpretations.

This is completely analogous to the previous parameter version

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

Explicitly, in terms of the data

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

Note

$$\hat{\varepsilon}_{ijk} = Y_{ijk} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet} = Y_{ijk} - \hat{\mu}_{ij} \ .$$

Orthogonal Decomposition of the Data Vector

$$\mathbf{Y} = \begin{pmatrix} Y_{111} \\ \vdots \\ Y_{11n} \\ \vdots \\ Y_{1t_21} \\ \vdots \\ Y_{1t_2n} \\ \vdots \\ \vdots \\ Y_{t_111} \\ \vdots \\ Y_{t_11n} \\ \vdots \\ Y_{t_1t_21} \\ \vdots \\ Y_{t_1t_2n} \end{pmatrix} = \begin{pmatrix} \hat{\mu} \\ \vdots \\ \hat{\mu} \\ \vdots \\ \hat{\mu} \\ \vdots \\ \hat{\mu} \\ \vdots \\ \vdots \\ \hat{\mu} \\ \vdots \\ \hat{\mu} \\ \vdots \\ \hat{\mu} \\ \vdots \\ \hat{\mu} \end{pmatrix} \overset{\perp}{+} \begin{pmatrix} \hat{a}_1 \\ \vdots \\ \hat{a}_1 \\ \vdots \\ \hat{a}_1 \\ \vdots \\ \hat{a}_1 \\ \vdots \\ \vdots \\ \hat{a}_{t_1} \\ \vdots \\ \hat{a}_{t_1} \\ \vdots \\ \hat{a}_{t_1} \\ \vdots \\ \hat{a}_{t_1} \end{pmatrix} \overset{\perp}{+} \begin{pmatrix} \hat{b}_1 \\ \vdots \\ \hat{b}_1 \\ \vdots \\ \hat{b}_{t_2} \\ \vdots \\ \hat{b}_{t_2} \\ \vdots \\ \vdots \\ \hat{b}_1 \\ \vdots \\ \hat{b}_1 \\ \vdots \\ \hat{b}_{t_2} \\ \vdots \\ \hat{b}_{t_2} \end{pmatrix} \overset{\perp}{+} \begin{pmatrix} \hat{\varepsilon}_{111} \\ \vdots \\ \hat{\varepsilon}_{11n} \\ \vdots \\ \hat{\varepsilon}_{1t_21} \\ \vdots \\ \hat{\varepsilon}_{1t_2n} \\ \vdots \\ \vdots \\ \hat{\varepsilon}_{t_111} \\ \vdots \\ \hat{\varepsilon}_{t_11n} \\ \vdots \\ \hat{\varepsilon}_{t_1t_21} \\ \vdots \\ \hat{\varepsilon}_{t_1t_2n} \end{pmatrix} = \hat{\mu}\mathbf{1} + \hat{\mathbf{a}} + \hat{\mathbf{b}} + \hat{\varepsilon}$$

Orthogonalities

$$
\begin{aligned}
\sum_{ijk} \hat{\mu}\, \hat{a}_i &= \hat{\mu}\, t_2 n \sum_i \hat{a}_i = 0 \\
\sum_{ijk} \hat{\mu}\, \hat{b}_j &= \hat{\mu}\, t_1 n \sum_j \hat{b}_j = 0 \\
\sum_{ijk} \hat{\mu}\, \hat{\varepsilon}_{ijk} &= \hat{\mu} \sum_{ijk} \hat{\varepsilon}_{ijk} = 0 \\
\sum_{ijk} \hat{a}_i\, \hat{b}_j &= n \sum_i \hat{a}_i \sum_j \hat{b}_j = 0 \\
\sum_{ijk} \hat{a}_i\, \hat{\varepsilon}_{ijk} &= \sum_i (\hat{a}_i \sum_{jk} \hat{\varepsilon}_{ijk}) = 0 \\
\sum_{ijk} \hat{b}_j\, \hat{\varepsilon}_{ijk} &= \sum_j (\hat{b}_j \sum_{ik} \hat{\varepsilon}_{ijk}) = 0
\end{aligned}
$$

Sum of Squares (SS) Decomposition

From this orthogonality obtain the following SS decomposition (Pythagoras again)

$$\begin{aligned}\sum_{ijk} Y_{ijk}^2 &= \sum_{ijk} \bar{Y}_{\bullet\bullet\bullet}^2 + \sum_{ijk} (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 + \sum_{ijk} (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 + \sum_{ijk} (Y_{ijk} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2 \\ &= \sum_{ijk} \hat{\mu}^2 + \sum_{ijk} \hat{a}_i^2 + \sum_{ijk} \hat{b}_j^2 + \sum_{ijk} \hat{\varepsilon}_{ijk}^2\end{aligned}$$

$$\begin{aligned}\Longrightarrow \sum_{ijk} (Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet})^2 &= \sum_{ijk} Y_{ijk}^2 - \sum_{ijk} \bar{Y}_{\bullet\bullet\bullet}^2 = \sum_{ijk} Y_{ijk}^2 - \sum_{ijk} \hat{\mu}^2 \\ &= \sum_{ijk} \hat{a}_i^2 + \sum_{ijk} \hat{b}_j^2 + \sum_{ijk} \hat{\varepsilon}_{ijk}^2 \\ &= \sum_{ijk} (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 + \sum_{ijk} (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 + \sum_{ijk} (Y_{ijk} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2\end{aligned}$$

$$SS_{\mathrm{T}} = SS_{\mathrm{A}} + SS_{\mathrm{B}} + SS_{\mathrm{E}}$$

Interpretation of SS Decomposition

$$SS_{\mathrm{T}} \quad = \quad SS_{\mathrm{A}} \;+\; SS_{\mathrm{B}} \;+\; SS_{\mathrm{E}}$$

$SS_{\mathrm{T}} = \sum_{ijk}(Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet})^2$: Total variation of data around the grand or overall mean

$SS_{\mathrm{A}} = \sum_{ijk}(\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2$: Variation of means around the grand mean
(by factor 1 level, averaged over the levels of factor 2)

$SS_{\mathrm{B}} = \sum_{ijk}(\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2$:Variation of means around the grand mean
(by factor 2 level, averaged over the levels of factor 1)

$SS_{\mathrm{E}} = \sum_{ijk}(Y_{ijk} - \hat{\mu}_{ij})^2$: Variation of data around the fitted additive model value.

$$\begin{aligned} SS_{\mathrm{E}} &= \sum_{ijk}(Y_{ijk} - \overbrace{[\bar{Y}_{\bullet\bullet\bullet} + (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}) + (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})]}^{\hat{\mu}_{ij}})^2 \\ &= \sum_{ijk}(Y_{ijk} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2 = \sum_{ijk}\hat{\varepsilon}_{ijk}^2 \end{aligned}$$

Degrees of Freedom

In $\mathbf{Y} = \hat{\mu}\mathbf{1} + \hat{\mathbf{a}} + \hat{\mathbf{b}} + \hat{\varepsilon}$ the component vectors are orthogonal to each other.

There is 1 degree of freedom in $\hat{\mu}\mathbf{1}$ and thus there are 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 the 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 the second factor vector $\hat{\mathbf{b}}$ and thus in SS_{B}.

By orthogonal complement there are

$$t_1 t_2 n - 1 - (t_1 - 1) - (t_2 - 1) = (t_1 - 1)(t_2 - 1) + t_1 t_2 (n-1) = \mathrm{df}_{\mathrm{E}}$$

df in the residual error vector $\hat{\varepsilon}$.

ANOVA Table for the Additive Model

Source	SS	df	MS	F
A	SS_{A}	$t_1 - 1$	$MS_A = SS_{\mathrm{A}}/(t_1 - 1)$	$MS_{\mathrm{A}}/MS_{\mathrm{E}}$
B	SS_{B}	$t_2 - 1$	$MS_B = SS_{\mathrm{B}}/(t_2 - 1)$	$MS_{\mathrm{B}}/MS_{\mathrm{E}}$
Error	SS_{E}	df_{E}	$MS_E = SS_{\mathrm{E}}/\mathrm{df}_{\mathrm{E}}$	
Total	SS_{T}	$t_1 t_2 n - 1$		

where

$$\mathrm{df}_{\mathrm{E}} = t_1 t_2 n - 1 - (t_1 - 1) - (t_2 - 1) = (t_1 - 1)(t_2 - 1) + t_1 t_2 (n - 1)$$

Here view $(t_1 - 1)(t_2 - 1) = t_1 t_2 - [1 + (t_1 - 1) + (t_2 - 1)]$

as the number of means μ_{ij} possibly left unexplained by the additive model

and $t_1 t_2 (n - 1)$ as the degrees of freedom of within cell variation ($n - 1$ per cell)

totaled over all $t_1 t_2$ cells.

`lm` on Reciprocal Time to Death

```
out=lm(recip.time ~ type + delivery)
```

```
out$coef
```

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

$\hat{\mu}^\star$ $\quad \hat{a}_2^\star$ $\quad \hat{a}_3^\star$ $\quad \hat{b}_2^\star$ $\quad \hat{b}_3^\star$ $\quad \hat{b}_4^\star$

Note the implicit set-to-zero form of the parameter estimates in `out$coef`!

`intercept` $= \hat{\mu}^\star$ with $\hat{a}_1^\star = \hat{b}_1^\star = 0$.

$\hat{\mu}^\star$ represents the mean under the treatment combination `(typeI,deliveryA)`

and $\hat{a}_i^\star$, $\hat{b}_j^\star$ represent additive mean deviation effects from this baseline $\hat{\mu}^\star$.

Sum-to-Zero Estimates

Below are the sum-to-zero estimates corresponding to the previous slide, using the conversion formulas from slide 40.

```
$mu.hat
  0.2622376
```

$\hat{\mu}$

```
$a.hat
      typeI      typeII     typeIII
-0.08216887 -0.03530475  0.11747362
```

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

```
$b.hat
  deliveryA   deliveryB   deliveryC   deliveryD
 0.08969690 -0.07604334  0.03248336 -0.04613693
```

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

`anova(lm(recip.time ~ type + delivery))`

```
Analysis of Variance Table

Response: recip.time
          Df  Sum Sq Mean Sq F value    Pr(>F)
type       2 0.34877 0.17439  71.708 2.865e-14 ***
delivery   3 0.20414 0.06805  27.982 4.192e-10 ***
Residuals 42 0.10214 0.00243
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Distributional Facts

Obviously the previous ANOVA table states p-values based on certain null-distributions.

These are based on the following facts derived from the normality assumptions:

SS_{A}, SS_{B} and SS_{E} are independent due to orthogonality of component vectors.

Under the additive model $SS_{\mathrm{E}} \sim \sigma^2\chi^2_{\mathrm{df}_{\mathrm{E}}}$ and $E(MS_{\mathrm{E}}) = E(SS_{\mathrm{E}}/\mathrm{df}_{\mathrm{E}}) = \sigma^2$.

$SS_{\mathrm{A}} \sim \sigma^2\chi^2_{t_1-1,\lambda_1}$ with `ncp` $\lambda_1 = \sum_{ijk}(\bar{\mu}_{i\bullet} - \bar{\mu}_{\bullet\bullet})^2/\sigma^2 = \sum_{ijk} a_i^2/\sigma^2 = n\,t_2 \sum_i a_i^2/\sigma^2$.

$E(MS_{\mathrm{A}}) = E(SS_{\mathrm{A}}/(t_1-1)) = \sigma^2 + \sigma^2\lambda_1/(t_1-1) = \sigma^2 + n\,t_2\sum_i a_i^2/(t_1-1)$.

$SS_{\mathrm{B}} \sim \sigma^2\chi^2_{t_2-1,\lambda_2}$ with `ncp` $\lambda_2 = \sum_{ijk}(\bar{\mu}_{\bullet j} - \bar{\mu}_{\bullet\bullet})^2/\sigma^2 = \sum_{ijk} b_j^2/\sigma^2 = n\,t_1 \sum_j b_j^2/\sigma^2$.

$E(MS_{\mathrm{B}}) = E(SS_{\mathrm{B}}/(t_2-1)) = \sigma^2 + \sigma^2\lambda_2/(t_2-1) = \sigma^2 + n\,t_1\sum_j b_j^2/(t_2-1)$.

Note how $n\,t_2$ and $n\,t_1$ act as multipliers in the noncentrality parameters!

Looking at both factors jointly, we benefit from the common σ^2 assumption.

Distributional Facts (continued)

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

$F_\mathrm{A} = MS_\mathrm{A}/MS_\mathrm{E} \sim F_{t_1-1,\mathrm{df}_\mathrm{E}}$.

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

$F_\mathrm{B} = MS_\mathrm{B}/MS_\mathrm{E} \sim F_{t_2-1,\mathrm{df}_\mathrm{E}}$.

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.

How Well Does the Additive Model Fit?

In the additive model we have: $\mu_{ij} = \mu + a_i + b_j$

We could compare the natural full model estimate of μ_{ij}, namely the average

$\bar{Y}_{ij\bullet}$ over all n observations under the factor level combination (i,j),

with the additive model fitted value for that same cell (i,j), i.e., with

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

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

Note that $\hat{\mu}_{ij}$ depends on data from cells other than cell (i,j)

and is more strongly averaged:

$$i^{\text{th}} \text{ row average} + j^{\text{th}} \text{ column average} - \text{grand average} \quad = \quad \bar{Y}_{i\bullet\bullet} + \bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet}$$

Full Model Representation

The full model $Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$ can be written in the following equivalent form:

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

Here it is assumed that $\varepsilon_{ijk} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$.

The equivalent full model form decomposes the mean structure in μ_{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 μ_{ij}, i.e., $c_{ij} = \mu_{ij} - (\mu + a_i + b_j)$.

These parameters $c_{ij} = (ab)_{ij}$ are also referred to as 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.

Identifiability Issues

While there are $t_1 \times t_2$ mean parameters μ_{ij} there are $1+t_1+t_2+t_1 \times t_2$ parameters $\mu, a_1, \ldots, a_{t_1}, b_1, \ldots, b_{t_2}, c_{11}, \ldots, c_{t_1 t_2}$ in the alternate parametrization.

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) $a_1 = b_1 = c_{1j} = c_{i1} = 0$ for all i, j set-to-zero side condition `lm` output in R

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

$\mu_{ij} \Longrightarrow$ Sum-to-Zero Parametrization

Define

$$\mu = \bar{\mu}_{\bullet\bullet} = \frac{1}{t_1 t_2}\sum_i \sum_j \mu_{ij}, \qquad \bar{\mu}_{i\bullet} = \frac{1}{t_2}\sum_j \mu_{ij}, \qquad \bar{\mu}_{\bullet j} = \frac{1}{t_1}\sum_i \mu_{ij}$$

and then all parameters μ, a_i, b_j, c_{ij} are determined from the μ_{ij} via

$$\begin{aligned} a_i &= \bar{\mu}_{i\bullet} - \bar{\mu}_{\bullet\bullet}, \qquad b_j = \bar{\mu}_{\bullet j} - \bar{\mu}_{\bullet\bullet} \\ c_{ij} &= \mu_{ij} - \bar{\mu}_{i\bullet} - \bar{\mu}_{\bullet j} + \bar{\mu}_{\bullet\bullet} \\ &= \mu_{ij} - (\bar{\mu}_{i\bullet} - \bar{\mu}_{\bullet\bullet}) - (\bar{\mu}_{\bullet j} - \bar{\mu}_{\bullet\bullet}) - \bar{\mu}_{\bullet\bullet} \quad = \quad \mu_{ij} - a_i - b_j - \mu \,. \end{aligned}$$

$$\Longrightarrow \quad \mu_{ij} = \mu + a_i + b_j + c_{ij} \qquad \text{with} \qquad \sum_i a_i = \sum_j b_j = \sum_i c_{ij} = \sum_j c_{ij} = 0$$

satisfying the sum-to-zero side conditions.

$\mu_{ij} \quad \Longrightarrow \quad$ Set-to-Zero Parametrization

Define all parameters $\mu^\star, a_i^\star, b_j^\star, c_{ij}^\star$ from the μ_{ij} via

$$\mu^\star = \mu_{11}, \qquad a_i^\star = \mu_{i1} - \mu_{11} = \mu_{i1} - \mu^\star, \qquad b_j^\star = \mu_{1j} - \mu_{11} = \mu_{1j} - \mu^\star,$$

$$\begin{aligned} c_{ij}^\star &= \mu_{ij} - \mu_{i1} - \mu_{1j} + \mu_{11} \\ &= \mu_{ij} - (\mu_{i1} - \mu_{11}) - (\mu_{1j} - \mu_{11}) - \mu_{11} \\ &= \mu_{ij} - a_i^\star - b_j^\star - \mu^\star \,. \end{aligned}$$

$\Longrightarrow \quad \mu_{ij} = \mu^\star + a_i^\star + b_j^\star + c_{ij}^\star$ with $a_1^\star = 0, \quad b_1^\star = 0, \quad c_{i1}^\star = c_{1j}^\star = 0 \quad \forall i, j$

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 μ_{ij}.

Decomposition and Least Squares Estimation

We extend our decomposition as follows (again with orthogonal components)

$$\begin{aligned} Y_{ijk} &= \bar{Y}_{\bullet\bullet\bullet} + (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}) + (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet}) + (\bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet}) + (Y_{ijk} - \bar{Y}_{ij\bullet}) \\ &= \hat{\mu} + \hat{a}_i + \hat{b}_j + \hat{c}_{ij} + \hat{\varepsilon}_{ijk} \end{aligned}$$

$$\sum_{ijk} (Y_{ijk} - \mu - a_i - b_j - c_{ij})^2$$

$$= \sum_{ijk} [(\bar{Y}_{\bullet\bullet\bullet} - \mu) + (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet} - a_i) + (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet} - b_j)$$

$$+ (\bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet} - c_{ij}) + (Y_{ijk} - \bar{Y}_{ij\bullet})]^2$$

$$= \sum_{ijk} \Big[(\bar{Y}_{\bullet\bullet\bullet} - \mu)^2 + (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet} - a_i)^2 + (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet} - b_j)^2$$

$$+ (\bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet} - c_{ij})^2 + (Y_{ijk} - \bar{Y}_{ij\bullet})^2 \Big]$$

Full Model Least Squares Estimates (LSEs)

The cross product terms in the previous quadratic expansion again are zero because of the component orthogonality in the decomposition.

One sees directly that the least squares estimates (LSEs, in sum-to-zero form) are

$$\hat{\mu} = \bar{Y}_{\bullet\bullet\bullet}\ , \quad \hat{a}_i = \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}\ , \quad \hat{b}_j = \bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet}\ , \quad \hat{c}_{ij} = \bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet}$$

with $E(\hat{\mu}) = \mu$, $E(\hat{a}_i) = a_i$, $E(\hat{b}_j) = b_j$ and $E(\hat{c}_{ij}) = c_{ij}$, i.e., the LSEs are unbiased.

The fitted values for the μ_{ij} are

$$\begin{aligned}
\hat{\hat{\mu}}_{ij} &= \hat{\mu} + \hat{a}_i + \hat{b}_j + \hat{c}_{ij} = \hat{\mu}_{ij} + \hat{c}_{ij} \\
&= \bar{Y}_{\bullet\bullet\bullet} + (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}) + (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet}) + (\bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet}) \\
&= \bar{Y}_{ij\bullet} \qquad \text{i.e., the cell means.}
\end{aligned}$$

with residuals $\hat{\varepsilon}_{ijk} = Y_{ijk} - \bar{Y}_{ij\bullet}$.

Sum of Squares Decomposition

Using the previous least squares decomposition with $\mu = a_i = b_j = c_{ij} = 0$ we get

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

or

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

or $SS_{\mathrm{T}} = SS_{\mathrm{A}} + SS_{\mathrm{B}} + SS_{\mathrm{AB}} + SS_{\mathrm{E}}$ with $SS_{\mathrm{T}} = \sum_{ijk} (Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet})^2$

$$SS_{\mathrm{A}} = \sum_{ijk} (\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = \sum_{ijk} \hat{a}_i^2 \,, \qquad SS_{\mathrm{B}} = \sum_{ijk} (\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 = \sum_{ijk} \hat{b}_j^2$$

$$SS_{\mathrm{AB}} = \sum_{ijk} (\bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2 = \sum_{ijk} \hat{c}_{ij}^2 \quad \text{and} \quad SS_{\mathrm{E}} = \sum_{ijk} (Y_{ijk} - \bar{Y}_{ij\bullet})^2 = \sum_{ijk} \hat{\varepsilon}_{ijk}^2$$

ANOVA Table for the Full Model

Source	SS	df	MS	F
A	SS_{A}	$t_1 - 1$	$MS_{\text{A}} = SS_{\text{A}}/(t_1 - 1)$	$MS_{\text{A}}/MS_{\text{E}}$
B	SS_{B}	$t_2 - 1$	$MS_{\text{B}} = SS_{\text{B}}/(t_2 - 1)$	$MS_{\text{B}}/MS_{\text{E}}$
AB	SS_{AB}	$(t_1 - 1)(t_2 - 1)$	$MS_{\text{AB}} = SS_{\text{AB}}/[(t_1 - 1)(t_2 - 1)]$	$MS_{\text{AB}}/MS_{\text{E}}$
Error	SS_{E}	$t_1 t_2 (n - 1)$	$MS_{\text{E}} = SS_{\text{E}}/[t_1 t_2 (n - 1)]$	
Total	SS_{T}	$t_1 t_2 n - 1$		

Distributional Facts for the Full Model

$$\frac{SS_{\mathrm{A}}}{\sigma^2} \sim \chi^2_{t_1-1,\lambda_{\mathrm{A}}} \text{ with } \lambda_{\mathrm{A}} = \frac{\sum_{ijk} a_i^2}{\sigma^2} \quad \text{and} \quad \frac{SS_{\mathrm{B}}}{\sigma^2} \sim \chi^2_{t_2-1,\lambda_{\mathrm{B}}} \text{ with } \lambda_{\mathrm{B}} = \frac{\sum_{ijk} b_j^2}{\sigma^2}$$

$$\frac{SS_{\mathrm{AB}}}{\sigma^2} \sim \chi^2_{(t_1-1)(t_2-1),\lambda_{\mathrm{AB}}} \text{ with } \lambda_{\mathrm{AB}} = \frac{\sum_{ijk} c_{ij}^2}{\sigma^2} \quad \text{and} \quad \frac{SS_{\mathrm{E}}}{\sigma^2} \sim \chi^2_{t_1t_2(n-1)}$$

SS_{A}, SS_{B}, SS_{AB}, and SS_{E} are statistically independent (orthogonality).

$$F_{\mathrm{A}} = MS_{\mathrm{A}}/MS_{\mathrm{E}} \sim F_{t_1-1,\, t_1t_2(n-1),\lambda_{\mathrm{A}}}, \quad F_{\mathrm{B}} = MS_{\mathrm{B}}/MS_{\mathrm{E}} \sim F_{t_2-1,\, t_1t_2(n-1),\lambda_{\mathrm{B}}}$$

$$F_{\mathrm{AB}} = MS_{\mathrm{AB}}/MS_{\mathrm{E}} \sim F_{(t_1-1)(t_2-1),\, t_1t_2(n-1),\lambda_{\mathrm{AB}}}$$

Expected MS for the Full Model

$$
\begin{aligned}
E(MS_{\mathrm{A}}) &= E\left(\frac{SS_{\mathrm{A}}}{t_1-1}\right) = \sigma^2 + \frac{\sum_{ijk} a_i^2}{t_1-1} = \sigma^2\left(1+\frac{\lambda_{\mathrm{A}}}{t_1-1}\right) \\
E(MS_{\mathrm{B}}) &= E\left(\frac{SS_{\mathrm{B}}}{t_2-1}\right) = \sigma^2 + \frac{\sum_{ijk} b_j^2}{t_2-1} = \sigma^2\left(1+\frac{\lambda_{\mathrm{B}}}{t_2-1}\right) \\
E(MS_{\mathrm{AB}}) &= E\left(\frac{SS_{\mathrm{AB}}}{(t_1-1)(t_2-1)}\right) = \sigma^2 + \frac{\sum_{ijk} c_{ij}^2}{(t_1-1)(t_2-1)} \\
&= \sigma^2\left(1+\frac{\lambda_{\mathrm{AB}}}{(t_1-1)(t_2-1)}\right) \\
E(MS_{\mathrm{E}}) &= \sigma^2.
\end{aligned}
$$

F-Tests for Full Model

Reject $H_{0\mathrm{A}} : a_1 = \ldots = a_{t_1} = 0$ whenever F_{A} is too large.

For a level α test reject $H_{0\mathrm{A}}$ whenever $F_{\mathrm{A}} \geq \mathtt{qf}(1-\alpha, \mathtt{t}_1 - 1, \mathtt{t}_1\mathtt{t}_2(\mathtt{n}-1))$.

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

For a level α test reject $H_{0\mathrm{B}}$ whenever $F_{\mathrm{B}} \geq \mathtt{qf}(1-\alpha, \mathtt{t}_2 - 1, \mathtt{t}_1\mathtt{t}_2(\mathtt{n}-1))$.

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

For a level α test reject $H_{0\mathrm{AB}}$ whenever $F_{\mathrm{AB}} \geq \mathtt{qf}(1-\alpha, (\mathtt{t}_1 - 1)(\mathtt{t}_2 - 1), \mathtt{t}_1\mathtt{t}_2(\mathtt{n}-1))$.

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

Rejecting $H_{0\mathrm{AB}} \implies$ the additive model will not provide a sufficient explanation.

Comments on the Full Model ANOVA Table

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

$$\sum_{ijk}(Y_{ijk} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2 = \sum_{ijk}(Y_{ijk} - \bar{Y}_{ij\bullet})^2 + \sum_{ijk}(\bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2$$

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

$$F_{\text{A}}^{\text{additive model}} = MS_{\text{A}} / MS_{\text{E}}^{\text{additive model}} \neq F_{\text{A}}^{\text{full model}} = MS_{\text{A}} / MS_{\text{E}}^{\text{full model}}$$

$$F_{\text{B}}^{\text{additive model}} = MS_{\text{B}} / MS_{\text{E}}^{\text{additive model}} \neq F_{\text{B}}^{\text{full model}} = MS_{\text{B}} / MS_{\text{E}}^{\text{full model}}$$

Reciprocal Time to Death (Insecticide Data)

```
> recip.time
 [1] 0.32258065 0.22222222 0.21739130 0.23255814 0.27777778 0.34482759
 [7] 0.25000000 0.43478261 0.45454545 0.47619048 0.55555556 0.43478261
[13] 0.12195122 0.09090909 0.11363636 0.13888889 0.10869565 0.16393443
[19] 0.20408163 0.08064516 0.33333333 0.27027027 0.26315789 0.34482759
[25] 0.23255814 0.22222222 0.15873016 0.13157895 0.22727273 0.28571429
[31] 0.32258065 0.25000000 0.43478261 0.40000000 0.41666667 0.45454545
[37] 0.22222222 0.14084507 0.15151515 0.16129032 0.17857143 0.09803922
[43] 0.14084507 0.26315789 0.33333333 0.27777778 0.32258065 0.30303030
```

Factors of Insecticide Data

```
> type
 [1] I   I   I   I   II  II  II  II  III III III III
[13] I   I   I   I   II  II  II  II  III III III III
[25] I   I   I   I   II  II  II  II  III III III III
[37] I   I   I   I   II  II  II  II  III III III III
Levels: I II III

> delivery
 [1] A A A A A A A A A A A A B B B B B B B B B B B B
[25] C C C C C C C C C C C C D D D D D D D D D D D D
Levels: A B C D
```

Note that both `type` and `delivery` are in factor form. Thus we don't have to invoke `as.factor(type)` and `as.factor(delivery)` in the call of `lm`.

`out.lmFULL=lm(recip.time ~ type*delivery)`

$\texttt{recip.time} \sim \texttt{type*delivery} \overset{\text{compare}}{\longleftrightarrow} \texttt{recip.time} \sim \texttt{type+delivery}$

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

Coefficients:
      (Intercept)             typeII             typeIII
         0.248688           0.078159            0.231580
        deliveryB          deliveryC           deliveryD
        -0.132342          -0.062416           -0.079720
 typeII:deliveryB  typeIII:deliveryB    typeII:deliveryC
        -0.055166          -0.045030            0.006961
typeIII:deliveryC   typeII:deliveryD   typeIII:deliveryD
         0.008646          -0.076974           -0.091368
```

Note the set-to-zero form of parameter estimates.

What else is in `out.lmFULL`?

```
> names(out.lmFULL)
 [1] "coefficients"  "residuals"     "effects"       "rank"
 [5] "fitted.values" "assign"        "qr"            "df.residual"
 [9] "contrasts"     "xlevels"       "call"          "terms"
[13] "model"
```

Full ANOVA: `anova(out.lmFULL)`

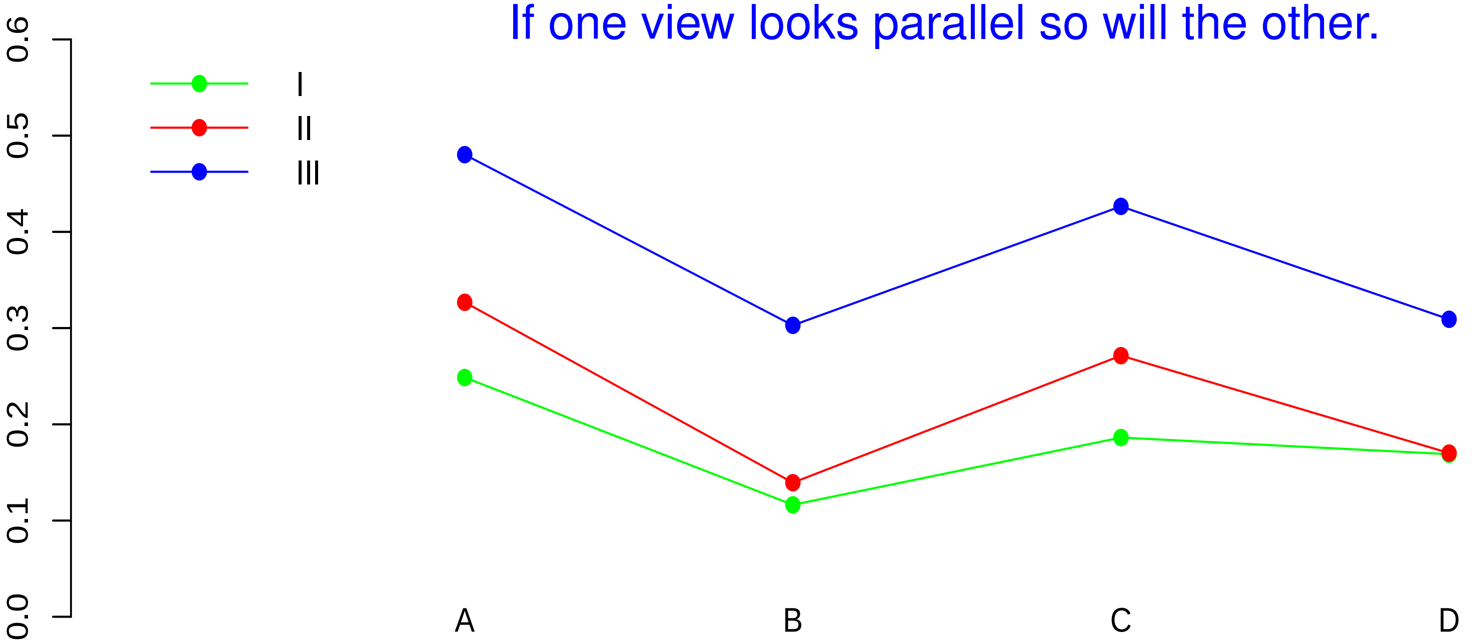
```
Analysis of Variance Table

Response: recip.time
              Df  Sum Sq Mean Sq F value    Pr(>F)
type           2 0.34877 0.17439 72.6347 2.310e-13 ***
delivery       3 0.20414 0.06805 28.3431 1.376e-09 ***
type:delivery  6 0.01571 0.00262  1.0904    0.3867
Residuals     36 0.08643 0.00240
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

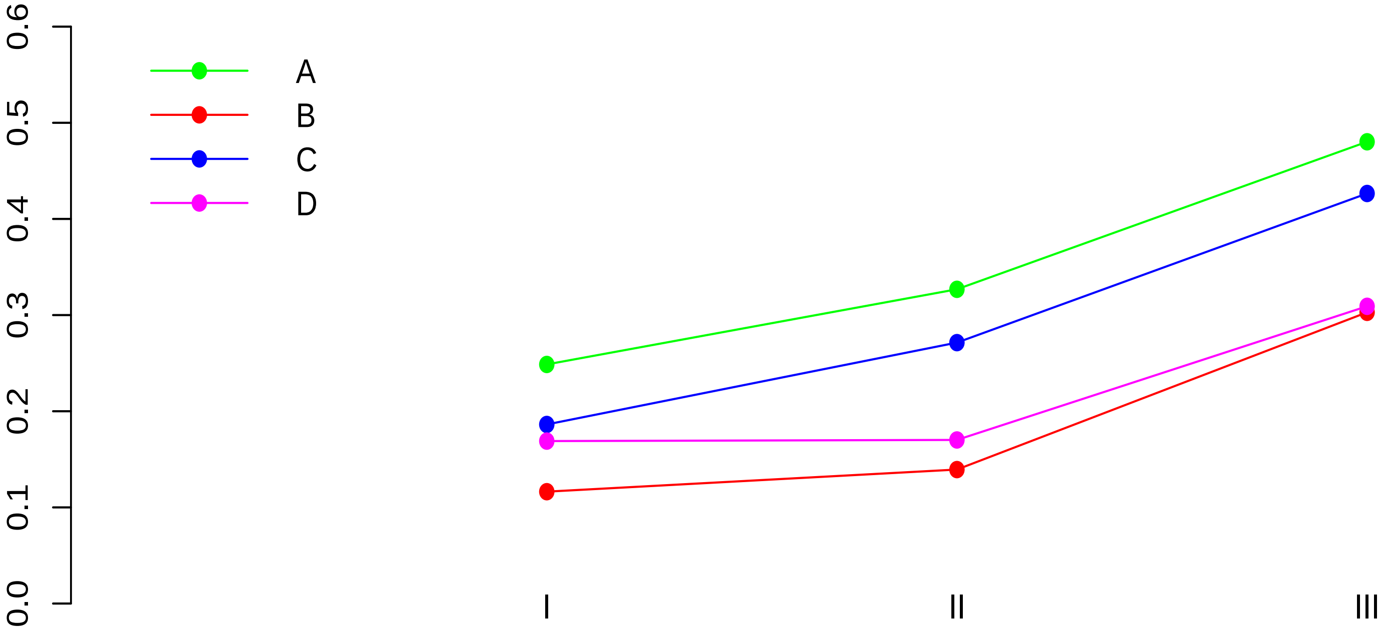
Thus it appears that the additive model is quite acceptable

and that 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.

$$\begin{array}{llll}
\mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} \\
\mu_{11}+d_2 & \mu_{12}+d_2 & \mu_{13}+d_2 & \mu_{14}+d_2 \\
\mu_{11}+d_3 & \mu_{12}+d_3 & \mu_{13}+d_3 & \mu_{14}+d_3
\end{array}$$

Row differences are constant!

$\Longrightarrow$ Column differences are constant as well, i.e.,

$\Delta_2 = \mu_{12} - \mu_{11}$ is the difference between column 2 and column 1

$\Delta_3 = \mu_{13} - \mu_{11}$ is the difference between column 3 and column 1

$\Delta_4 = \mu_{14} - \mu_{11}$ is the difference between column 4 and column 1.

$$\begin{array}{llll}
\mu_{11} & \mu_{11}+\Delta_2 & \mu_{11}+\Delta_3 & \mu_{11}+\Delta_4 \\
\mu_{21} & \mu_{21}+\Delta_2 & \mu_{21}+\Delta_3 & \mu_{21}+\Delta_4 \\
\mu_{31} & \mu_{31}+\Delta_2 & \mu_{31}+\Delta_3 & \mu_{31}+\Delta_4
\end{array}$$

Here we identified $\mu_{11} = \mu_{11}$, $\mu_{21} = \mu_{11} + d_2$, and $\mu_{31} = \mu_{11} + d_3$.

Similarly one argues going the other direction.

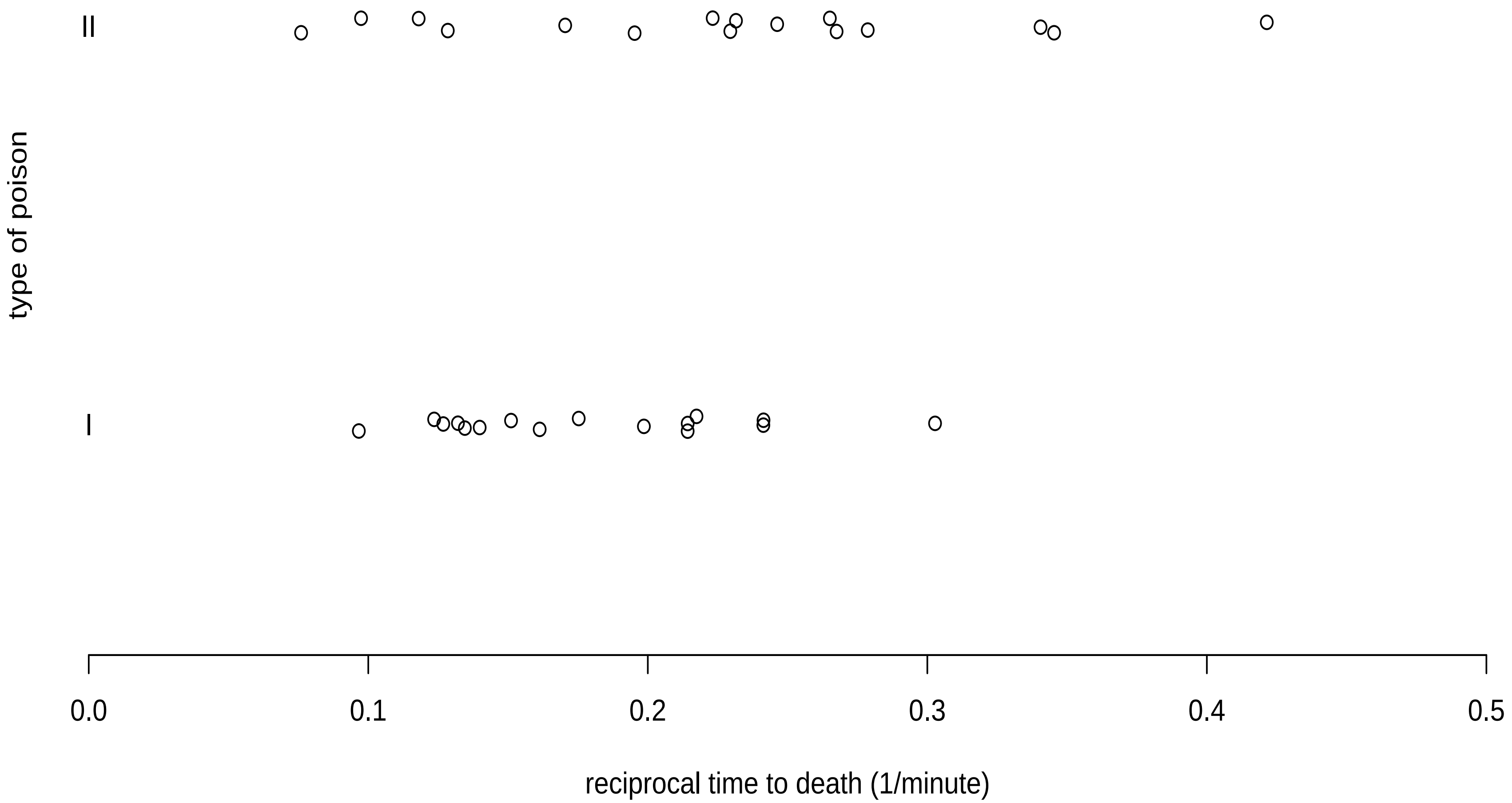
Examining Factor Level Differences

Given that we have strong evidence of factor level differences we may want to examine them individually to see which differences matter.

A naive approach: Perform a 2-sample t-test or look at the corresponding confidence intervals, e.g., comparing type I with type II means we get

```
        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 & horizontally

Getting MS_{E} in Types I and II Comparison

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

Response: recip.time[type == "I" | type == "II"]
                                 Df   Sum Sq  Mean Sq F value Pr(>F)
type[type == "I" | type == "II"]  1 0.017570 0.017570  2.6839 0.1118
Residuals                        30 0.196394 0.006546
```

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 $s = \sqrt{MS_{\text{E}}} = \sqrt{0.006546} = .08091$.

This could also have been backed out from previous t-based confidence interval.

What is Wrong?

In the previous 2-sample t-test/interval we treated the observations as i.i.d. from two populations.

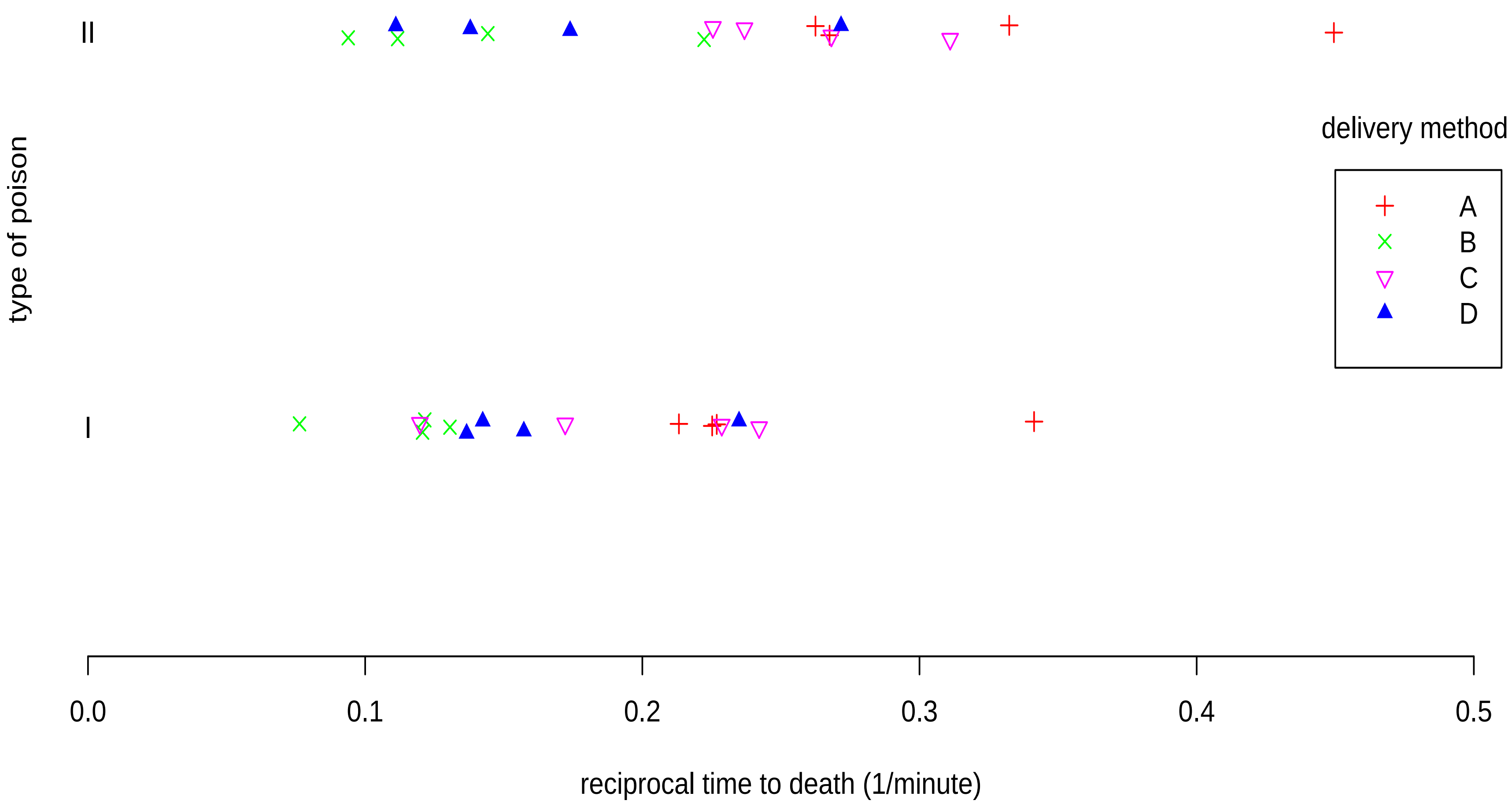
We completely ignored the known variations due to the delivery method.

When we get the pooled sample standard deviation from these 2 “samples” we confound variation between delivery method means with variation (σ) within (delivery,type) combination.

Our “reference distribution” will thus be too dispersed.

Our test will be less discriminating or our confidence interval will be too wide.

Closer Look in Comparing Types I and II



Note the reduced variability within same color clusters.

Correct Approach

According to our (accepted) additive model we have

$Y_{1jk} = \mu + a_1 + b_j + \varepsilon_{1jk}$ and $Y_{2jk} = \mu + a_2 + b_j + \varepsilon_{2jk}$ with $\varepsilon_{ijk} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$

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:

$$\bar{\mu}_{1\bullet} - \bar{\mu}_{2\bullet} = \frac{\sum_j \mu_{1j}}{4} - \frac{\sum_j \mu_{2j}}{4} = \frac{\sum_j (\mu + a_1 + b_j) - \sum_j (\mu + a_2 + b_j)}{4} = a_1 - a_2$$

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 out other effects.

Contrast in Full Model

Even in the full model with interactions the previous contrast in means stays the same since

$$
\begin{aligned}
\bar{\mu}_{1\bullet} - \bar{\mu}_{2\bullet} &= \frac{\sum_j \mu_{1j}}{4} - \frac{\sum_j \mu_{2j}}{4} \\
&= \frac{\sum_j(\mu + a_1 + b_j + c_{1j}) - \sum_j(\mu + a_2 + b_j + c_{2j})}{4} \\
&= \mu + a_1 + \frac{\sum_j(b_j + c_{1j})}{4} - \mu - a_2 - \frac{\sum_j(b_j + c_{2j})}{4} \\
&= a_1 - a_2
\end{aligned}
$$

μ, b_j, c_{1j} and c_{2j} are canceled out.

Estimated Contrast

The natural estimate of $\bar{\mu}_{1\bullet} - \bar{\mu}_{2\bullet} = a_1 - a_2$ is

$$\hat{\mu}_{1\bullet} - \hat{\mu}_{2\bullet} = \hat{a}_1 - \hat{a}_2 = (\bar{Y}_{1\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}) - (\bar{Y}_{2\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet}) = \bar{Y}_{1\bullet\bullet} - \bar{Y}_{2\bullet\bullet}$$

the same as used in the previous (naive) 2-sample t-test/interval.

This estimate can also be viewed as the contrast of estimated cell averages

$$\hat{a}_1 - \hat{a}_2 = \bar{Y}_{1\bullet\bullet} - \bar{Y}_{2\bullet\bullet} \quad = \quad \frac{\bar{Y}_{11\bullet} + \bar{Y}_{12\bullet} + \bar{Y}_{13\bullet} + \bar{Y}_{14\bullet}}{4} - \frac{\bar{Y}_{21\bullet} + \bar{Y}_{22\bullet} + \bar{Y}_{23\bullet} + \bar{Y}_{24\bullet}}{4}$$

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

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

because $\hat{c}_{i1} + \hat{c}_{i2} + \hat{c}_{i3} + \hat{c}_{i4} = 0$ in $\sum_j \hat{\hat{\mu}}_{ij} = \sum_j (\hat{\mu}_{ij} + \hat{c}_{ij})$.

Correct Approach (continued)

However, $$\operatorname{var}(\bar{Y}_{1\bullet\bullet} - \bar{Y}_{2\bullet\bullet}) = \operatorname{var}(\bar{Y}_{1\bullet\bullet}) + \operatorname{var}(\bar{Y}_{2\bullet\bullet}) = \frac{\sigma^2}{nt_2} + \frac{\sigma^2}{nt_2} = \frac{2\sigma^2}{nt_2}$$

$$\frac{(\bar{Y}_{1\bullet\bullet} - \bar{Y}_{2\bullet\bullet} - (a_1 - a_2)) / \left(\sigma\sqrt{\frac{2}{nt_2}}\right)}{s/\sigma} = \frac{(\bar{Y}_{1\bullet\bullet} - \bar{Y}_{2\bullet\bullet} - (a_1 - a_2))}{s\sqrt{2/(nt_2)}} = \frac{\hat{a}_1 - \hat{a}_2 - (a_1 - a_2)}{s\sqrt{2/(nt_2)}} \sim t_f \,.$$

Here we have two options in choosing s and the corresponding f:

$$s^2 = s_\times^2 = MS_{\text{E}}^{\text{full model}} = \frac{\sum_{ijk}(Y_{ijk} - \bar{Y}_{ij\bullet})^2}{t_1 t_2 (n-1)}$$ and thus $f = t_1 t_2 (n-1)$.

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

and thus $f = t_1 t_2 (n-1) + (t_1 - 1)(t_2 - 1) = t_1 t_2 n - t_1 - t_2 + 1$.

t-Test for $H_0 : a_1 = a_2$

Reject $H_0 : a_1 = a_2$ when

$$\left| \frac{\hat{a}_1 - \hat{a}_2}{\sqrt{MS_{\mathrm{E}} \times 2/(nt_2)}} \right| > t_{1-\alpha/2,f}$$

or

$$|\hat{a}_1 - \hat{a}_2| > \sqrt{MS_{\mathrm{E}} \times 2/(nt_2)} \times t_{1-\alpha/2,f}$$

or

$$|\hat{a}_1 - \hat{a}_2| > SE(\hat{a}_1 - \hat{a}_2) \times t_{1-\alpha/2,f} = LSD_{\mathrm{A}} \, ,$$

where LSD_{A} is the least significant difference against which any estimated differences $\hat{a}_i - \hat{a}_{i'}$ $(i \neq i')$ in levels of factor A can be compared.

This is in 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

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

or

$$|\hat{b}_1 - \hat{b}_2| > \sqrt{MS_{\mathrm{E}} \times 2/(nt_1)} \times t_{1-\alpha/2,f}$$

or

$$|\hat{b}_1 - \hat{b}_2| > SE(\hat{b}_1 - \hat{b}_2) \times t_{1-\alpha/2,f} = LSD_{\mathrm{B}} \ ,$$

where LSD_{B} is the least significant difference against which any estimated differences $\hat{b}_j - \hat{b}_{j'}$ ($j \neq j'$) in levels of factor B can be compared.

Note the change $(t_2 \longleftrightarrow t_1)$ between LSD_{A} and LSD_{B}.

Factor Level Means

```
> 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

$$LSD_{\mathrm{A}} = t_{.975,f} \times SE(\hat{a}_1 - \hat{a}_2) = 2.028094 \times \sqrt{.00240 \times \frac{2}{4\times 4}} = 0.03513$$

for the full model with $f = 36$.

$$LSD_{\mathrm{A}} = t_{.975,f} \times SE(\hat{a}_1 - \hat{a}_2) = 2.018082 \times \sqrt{.00243 \times \frac{2}{4\times 4}} = 0.03517$$

for the additive model with $f = 42$.

For the Delivery Factor

$$LSD_{\mathrm{B}} = t_{.975,f} \times SE(\hat{b}_1 - \hat{b}_2) = 2.028094 \times \sqrt{.00240 \times \frac{2}{3\times 4}} = 0.04056$$

for the full model with $f = 36$.

$$LSD_{\mathrm{B}} = t_{.975,f} \times SE(\hat{b}_1 - \hat{b}_2) = 2.018082 \times \sqrt{.00243 \times \frac{2}{3\times 4}} = 0.04061$$

for the additive model with $f = 42$.

LSD Groupings

Type

$LSD_{\mathrm{A}} = .0352$

Poison Type	Mean $\hat{\mu}_{i\bullet}$	*LSD* Grouping
I	0.180	1
II	0.227	2
III	0.380	3

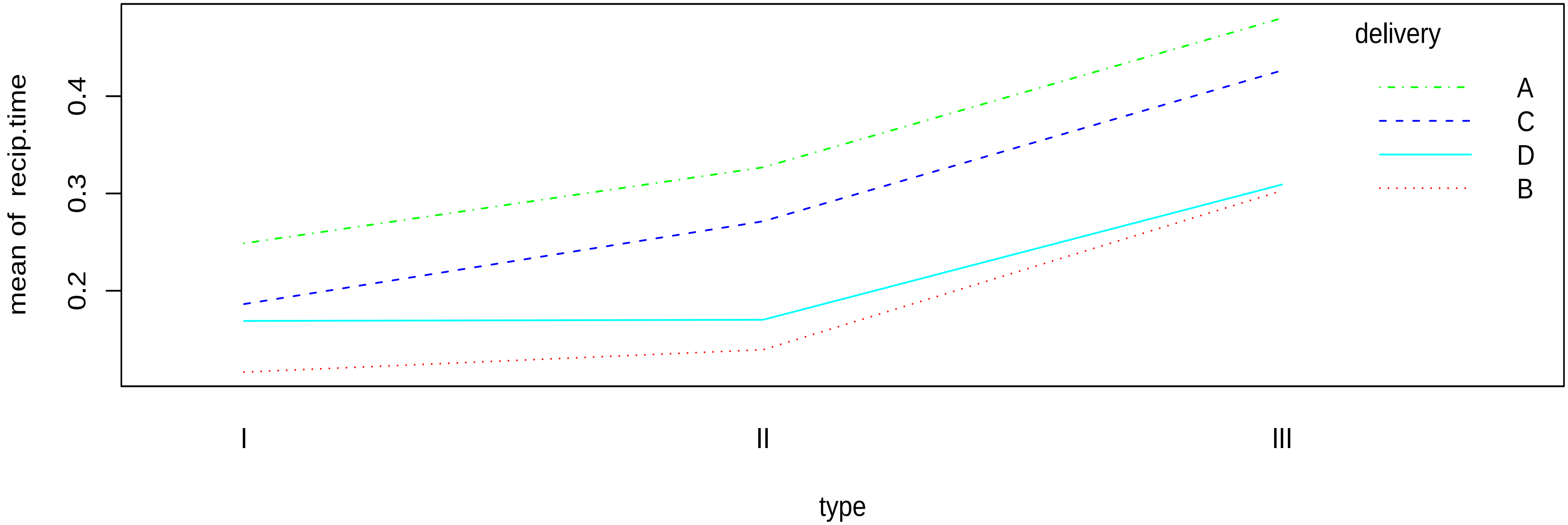
Note difference to naive 2-sample t-test.

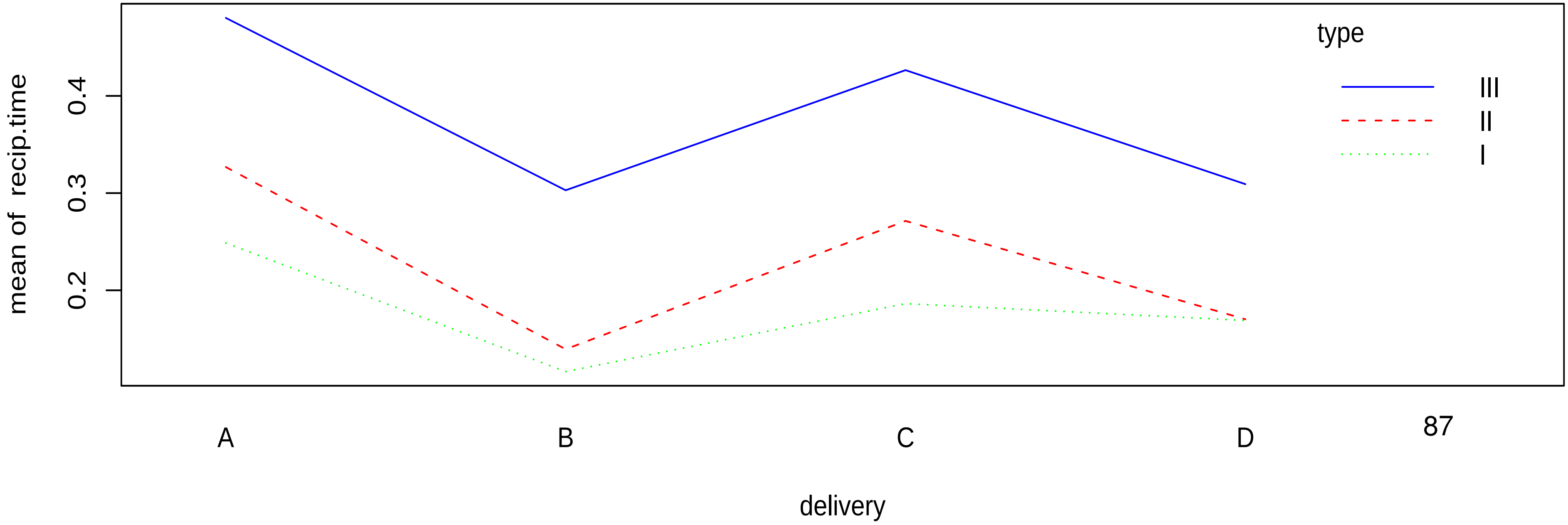
Delivery

$LSD_{\mathrm{B}} = .0406$

Poison Delivery	Mean $\hat{\mu}_{\bullet j}$	*LSD* Grouping
B	0.186	1
D	0.216	1
C	0.295	2
A	0.352	3

Looking Back





`interaction.plot`

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 margins around the plots, and `mfrow=c(2,1)` sets up plotting for two plots per page, one above the other.

The Full Model Revisited

The full model is $Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$ and we can reparametrize it as

$$Y_{ijk} = \mu + a_i + b_j + c_{ij} + \varepsilon_{ijk} \quad \text{with replications } k = 1, \ldots, n \quad \text{where}$$

- $\mu = \sum_{ij} \mu_{ij}/(t_1 t_2) = \bar{\mu}_{\bullet\bullet}$ is the grand mean.

- $a_i = \sum_j (\mu_{ij} - \bar{\mu}_{\bullet\bullet})/t_2 = \bar{\mu}_{i\bullet} - \bar{\mu}_{\bullet\bullet}$ with $\sum_i a_i = 0.$

- $b_j = \sum_i (\mu_{ij} - \bar{\mu}_{\bullet\bullet})/t_2 = \bar{\mu}_{\bullet j} - \bar{\mu}_{\bullet\bullet}$ with $\sum_j b_j = 0.$

  a_i and b_j are also referred to as main effects.

- $c_{ij} = \mu_{ij} - \bar{\mu}_{i\bullet} - \bar{\mu}_{\bullet j} + \bar{\mu}_{\bullet\bullet} = \mu_{ij} - (\mu + a_i + b_j)$ with $\sum_i c_{ij} = \sum_j c_{ij} = 0.$

In the additive model nothing in the above relations of μ, a_i, b_j relative to the μ_{ij} changes except the last •, which becomes: • $c_{ij} = 0 \ \forall \ i, j.$

The Interpretation of $a_i - a_{i'}$

In the additive or main effects model we have

$$a_i - a_{i'} = \mu_{ij} - \mu_{i'j} = \mu + a_i + b_j - (\mu + a_{i'} + b_j) \;\; \forall j = 1, \ldots, t_2 \, ,$$

i.e., $a_i - a_{i'}$ represents the 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 we have (since $b_\bullet = \bar{c}_{i\bullet} = \bar{c}_{i'\bullet} = 0$)

$$a_i - a_{i'} = \mu_{i\bullet} - \mu_{i'\bullet} = \mu + a_i + b_\bullet + \bar{c}_{i\bullet} - (\mu + a_{i'} + b_\bullet + \bar{c}_{i'\bullet}) \;\; \forall j = 1, \ldots, t_2 \, ,$$

i.e., $a_i - a_{i'}$ represents the 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 $b_j - b_{j'}$.

Understanding Interactions

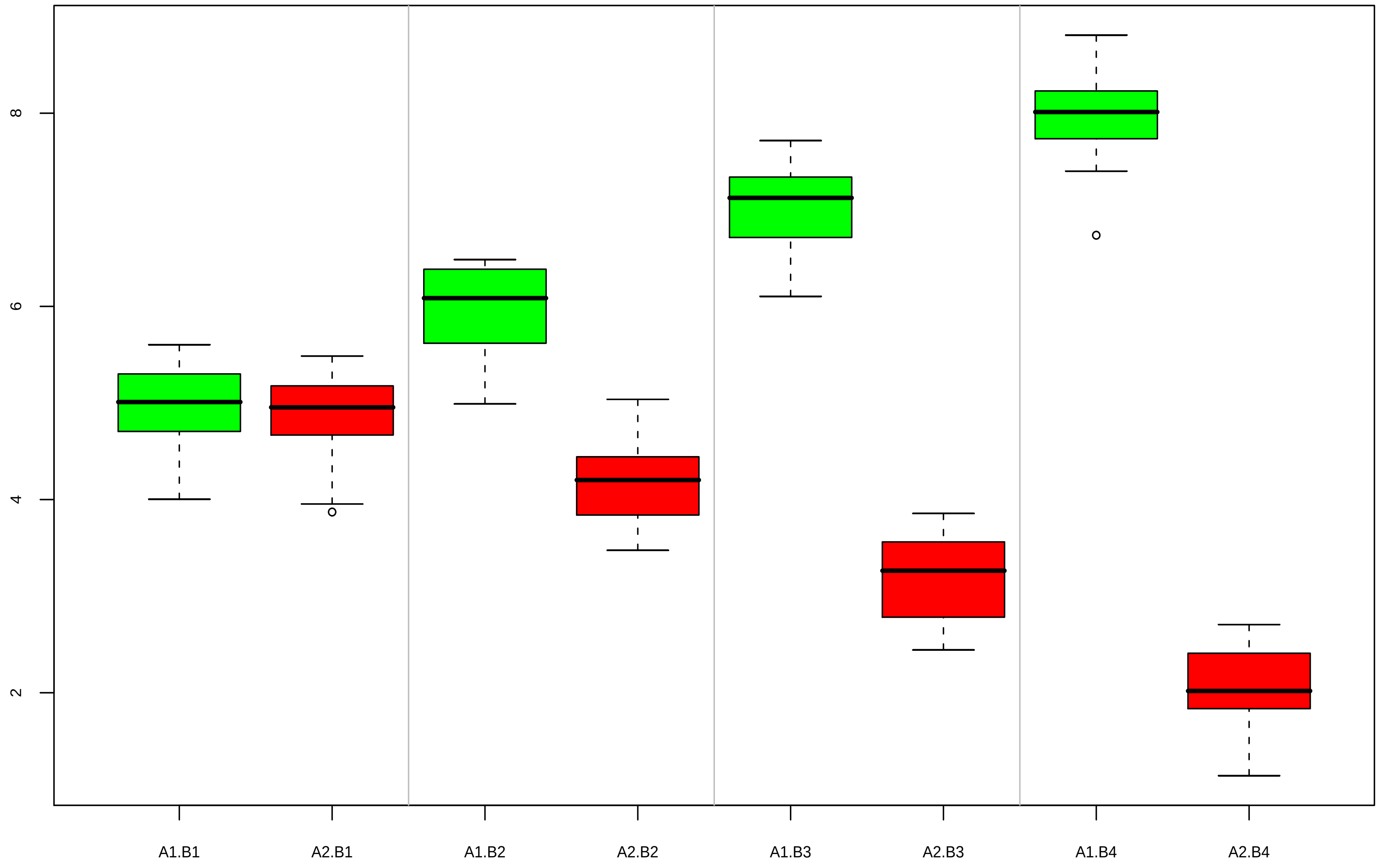
Main effects a_i and b_j are called that way because their additive effects on the mean μ_{ij} are easily understood.

Interactions (as distinguished from main effects) 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



Observations about Pattern 1

In pattern 1 the differences in means $\mu_{1j} - \mu_{2j}$ appear to increase linearly in j.

Without interactions this difference would be constant.

One could look at orthogonal polynomial contrasts in such differences.

For a contrast vector $(d_1, \ldots, d_4)$, i.e., $d_1 + \ldots + d_4 = 0$, consider the contrast

$$D = d_1(\mu_{11} - \mu_{21}) + d_2(\mu_{12} - \mu_{22}) + d_3(\mu_{13} - \mu_{23}) + d_4(\mu_{14} - \mu_{24})$$

Since

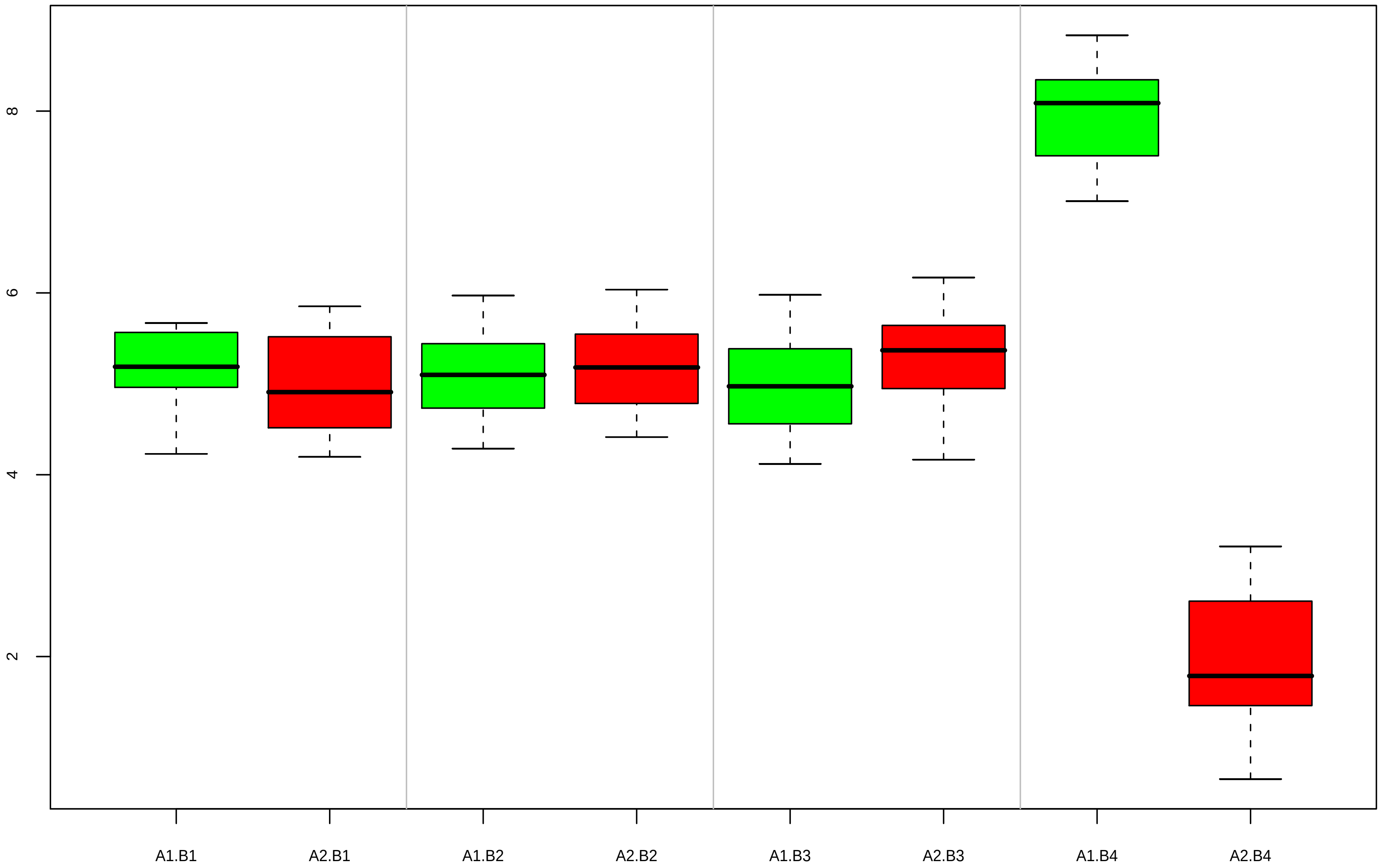
$$c_{1j} - c_{2j} = \mu_{1j} - \bar{\mu}_{1\bullet} - \bar{\mu}_{\bullet j} + \bar{\mu}_{\bullet\bullet} - (\mu_{2j} - \bar{\mu}_{2\bullet} - \bar{\mu}_{\bullet j} + \bar{\mu}_{\bullet\bullet}) = \mu_{1j} - \mu_{2j} - \bar{\mu}_{1\bullet} + \bar{\mu}_{2\bullet}$$

$$\Longrightarrow \quad D = d_1(c_{11} - c_{21}) + d_2(c_{12} - c_{22}) + d_3(c_{13} - c_{23}) + d_4(c_{14} - c_{24})$$

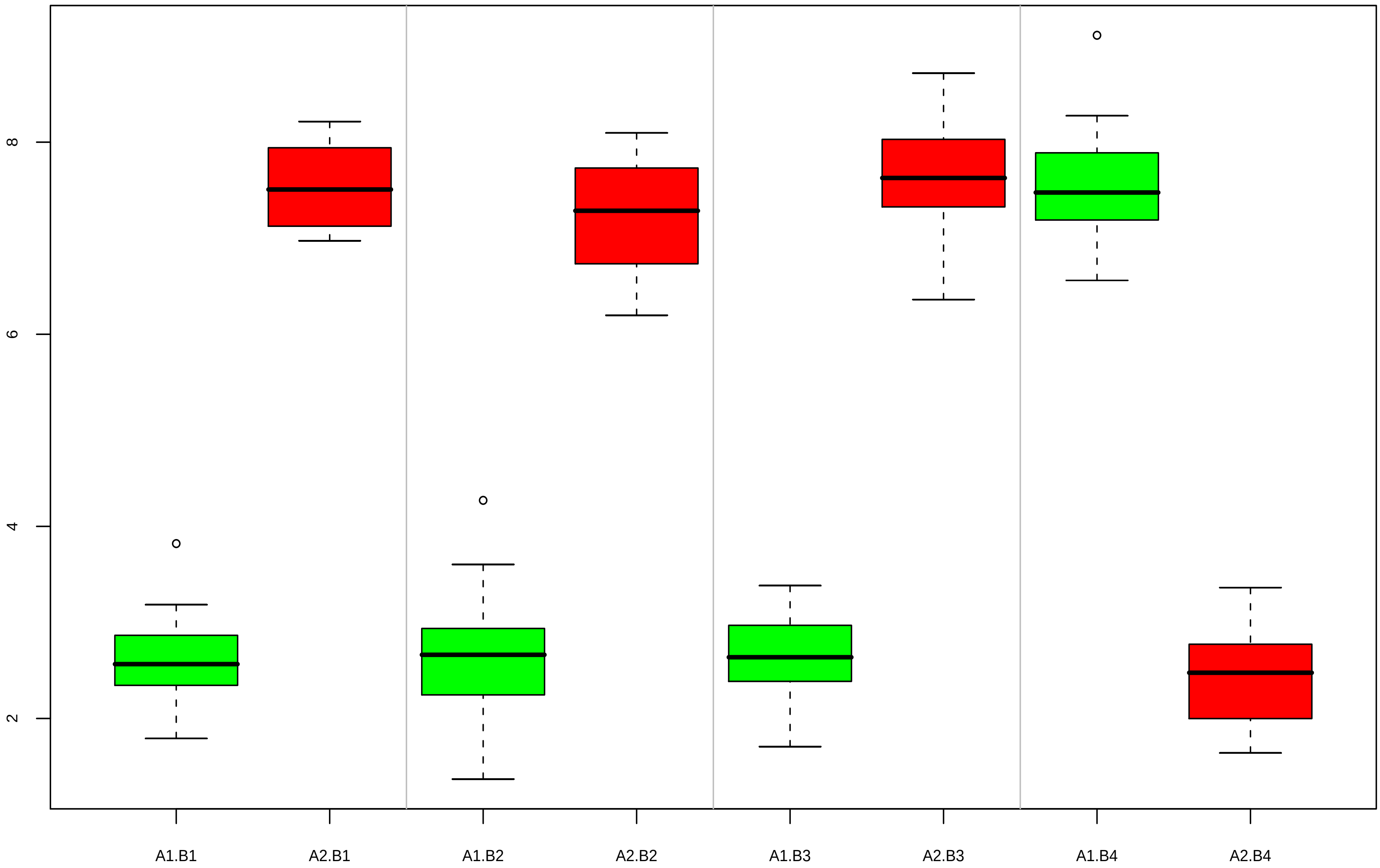
is the same contrast in the corresponding interaction differences.

$(d_1, d_2, d_3, d_4) = (-3, -1, 1, 3)$ would resonate with a linear trend.

Interaction Pattern 2



Interaction Pattern 3



Comments on Interaction Patterns

Pattern 1 seems to show a linear interaction trend along the levels of factor B. Furthermore, 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 when we have level B_4 for factor B is there a clear change in factor A. Hence B_4 acts as an interaction switch.

Pattern 3 again 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.

Randomized Complete Block Designs

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

While we cannot influence the size of treatment effects, $|\mu_i - \bar{\mu}|$, we can influence the sample sizes n_i and possibly σ. Note $\sigma \longrightarrow \sigma/2 \quad \Longleftrightarrow \quad n_i \longrightarrow 4 \times n_i$!!

How can we influence σ? We need to understand what may affect σ.

Often σ is caused by the 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):

Source	SS	df	MS	F
A	SS_{A}	$t_1 - 1$	$SS_{\mathrm{A}}/(t_1 - 1)$	$MS_{\mathrm{A}}/MS_{\mathrm{E}}$
B	SS_{B}	$t_2 - 1$	$SS_{\mathrm{B}}/(t_2 - 1)$	$MS_{\mathrm{B}}/MS_{\mathrm{E}}$
Error	SS_{E}	df_{E}	$SS_{\mathrm{E}}/\mathrm{df}_{\mathrm{E}}$	
Total	SS_{T}	$t_1 t_2 n - 1$		

If we ignored factor B, we would treat $SS'_{\mathrm{E}} = SS_B + SS_{\mathrm{E}}$ as our error sum of squares and $\mathrm{df}'_{\mathrm{E}} = \mathrm{df}_{\mathrm{E}} + t_2 - 1$ as the corresponding degrees of freedom. $SS'_{\mathrm{E}}/\mathrm{df}'_{\mathrm{E}} = MS'_{\mathrm{E}}$ would be a legitimate estimate of σ^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

Blocking consists of stratifying experimental units into groups that will have more homogeneous responses within groups, and 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.

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.

Typical Blocking Criteria

- **Location:** If the experiment is conducted over varying locations and if location is judged to have an effect on the response.
- **Time:** If time of day, month, or year are likely to affect response and if the 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:** If variations in the process for creating the 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.

Sir Ronald Aylmer Fisher

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`

for an account of 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:

`{\tt http://www.tufts.edu/˜gdallal/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 (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.

It is assumed that this moisture gradient is mainly across the width of the field.

This suggest rows along the field length as blocks.

Field Moisture Gradient

The Data

The field was subdivided into 4 rows (or blocks) with 6 plots each.

The 6 treatments were randomly assigned to each row.

Any remaining moisture variability within rows is absorbed into MS_{E}.

row	treatment response					
1	2 40.89	5 37.99	4 37.18	1 34.98	6 34.89	3 42.07
2	1 41.22	3 49.42	4 45.85	6 50.15	5 41.99	2 46.69
3	6 44.57	3 52.68	5 37.61	1 36.94	2 46.65	4 40.23
4	2 41.90	4 39.20	6 43.29	5 40.45	3 42.91	1 39.97

Also available as `fertilizerdata.csv` on web site.

Randomized Complete Block (RCB) Design

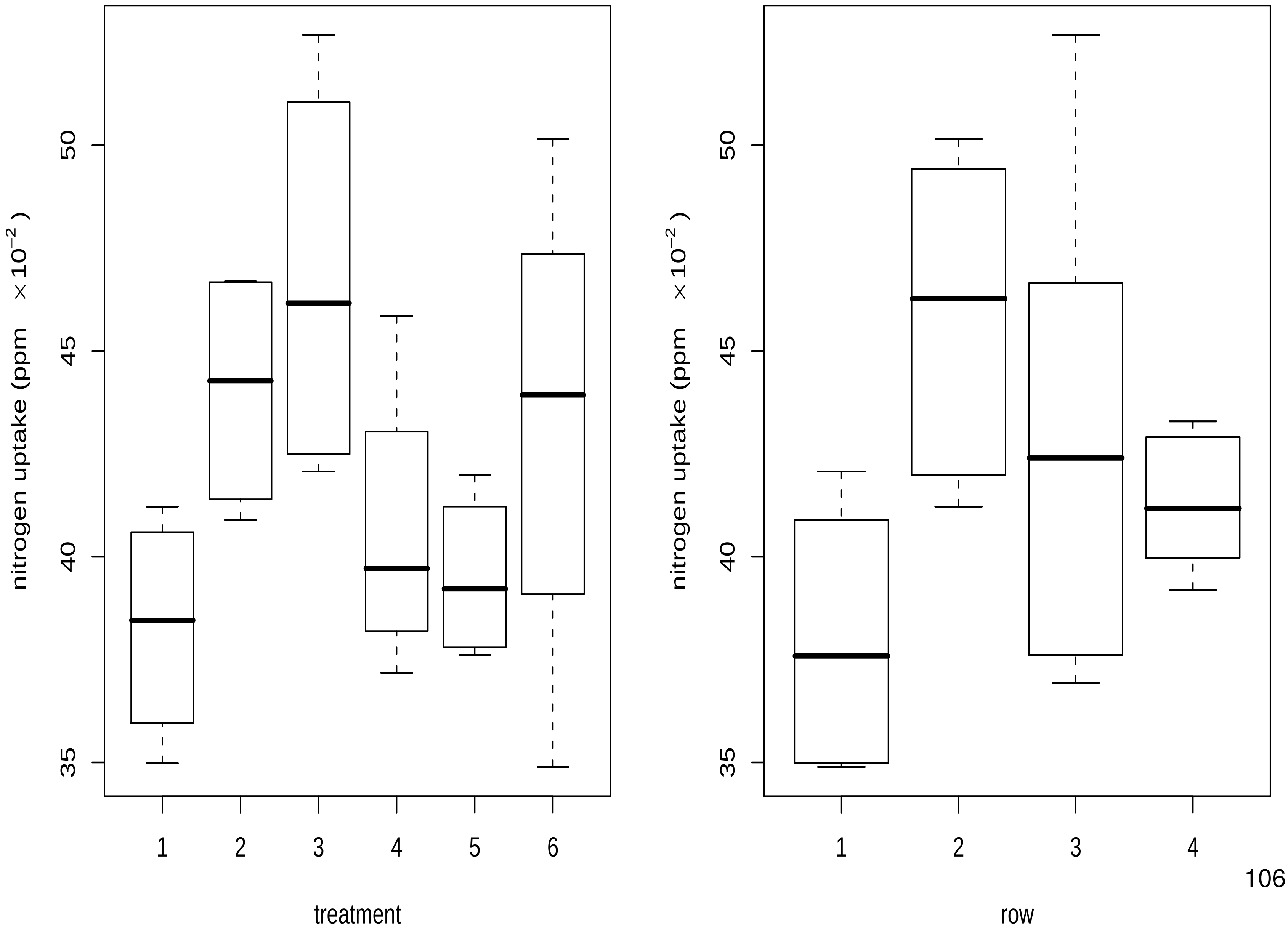
1. Experimental units are blocked into presumably more homogeneous groups.

2. The blocks are complete, i.e., each treatment appears in each block.

3. 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 level,treatment level) combination.

Nitrogen Fertilizer Box Plots



ANOVA Table

Source	SS	df	MS	F
Block	SS_{Block}	$t_1 - 1$	$SS_{\text{Block}}/(t_1 - 1)$	$MS_{\text{Block}}/MS_{\text{E}}$
Treatment	SS_{Treat}	$t_2 - 1$	$SS_{\text{Treat}}/(t_2 - 1)$	$MS_{\text{Treat}}/MS_{\text{E}}$
Error	SS_{E}	df_{E}	$SS_{\text{E}}/\text{df}_{\text{E}}$	
Total	SS_{T}	$t_1 t_2 n - 1$		

Note that here $n = 1$, thus $t_1 t_2 n - 1 = t_1 t_2 - 1,$ and

$$\text{df}_{\text{E}} = (t_1 - 1)(t_2 - 1) + t_1 t_2 (n - 1) = (t_1 - 1)(t_2 - 1).$$

The table is the same as in the additive 2-factor ANOVA situation,

just with relabeling factors A and B to Block and Treatment.

Nitrogen Fertilizer Results

```
> fertilizer.analysis() # see class web page
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
                      Df  Sum Sq Mean Sq F value  Pr(>F)
as.factor(treatment)   5 201.316  40.263  2.3761 0.08024 .
Residuals             18 305.012  16.945
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Not significant at $.05$, because we are affected by the extra variation induced by the wetness gradient. Less discrimination power to see the fertilizer timing effect.

F-Statistics in the Last Two ANOVAs

It is worthwhile to compare the corresponding F-statistics for the last two ANOVAs.

$$F_{\text{Treat}} = \frac{\sum_{ij}(\bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet})^2/5}{\sum_{ij}(Y_{ij} - \bar{Y}_{\bullet j} - [\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}])^2/15} \qquad \text{and} \qquad F_{\text{Treat}} = \frac{\sum_{ij}(\bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet})^2/5}{\sum_{ij}(Y_{ij} - \bar{Y}_{\bullet j})^2/18}$$

While in the former F-statistic the variation of the Y_{ij} around $\bar{Y}_{\bullet j}$ is corrected for the variation due to $\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}$, i.e., the row variation, such a correction is omitted in the latter F-statistic, i.e., the row variation is ignored and absorbed as part of the MS_{E}:

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

The Last Identity

The previous decomposition could use some elaboration:

$$\begin{aligned}\sum_{ij}(Y_{ij}-\bar{Y}_{\bullet j}-[\bar{Y}_{i\bullet}-\bar{Y}_{\bullet\bullet}])^2 &= \sum_{ij}(Y_{ij}-\bar{Y}_{\bullet j})^2+\sum_{ij}(\bar{Y}_{i\bullet}-\bar{Y}_{\bullet\bullet})^2-2\sum_{ij}(Y_{ij}-\bar{Y}_{\bullet j})(\bar{Y}_{i\bullet}-\bar{Y}_{\bullet\bullet}) \\ &= \sum_{ij}(Y_{ij}-\bar{Y}_{\bullet j})^2+\sum_{ij}(\bar{Y}_{i\bullet}-\bar{Y}_{\bullet\bullet})^2-2\sum_{ij}(\bar{Y}_{i\bullet}-\bar{Y}_{\bullet\bullet})^2 \\ &= \sum_{ij}(Y_{ij}-\bar{Y}_{\bullet j})^2-\sum_{ij}(\bar{Y}_{i\bullet}-\bar{Y}_{\bullet\bullet})^2\end{aligned}$$

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

A Look at the Block Adjustment

Model: $Y_{ij} = \mu + b_i + \tau_j + \varepsilon_{ij}$ with estimated block effect $\hat{b}_i = \bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}$

Consider the block adjusted observations $Z_{ij} = Y_{ij} - \hat{b}_i = Y_{ij} - (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})$

These Z_{ij}, when plotted on separate levels for each block identifier i, will look more aligned, more like “replicates” (see next 2 slides). Thus it makes sense to look at the average $\bar{Z}_{\bullet j} - \bar{Z}_{\bullet\bullet}$ over those blocks as estimate for the j^{th} treatment effect.

Note that $\bar{Z}_{\bullet j} - \bar{Z}_{\bullet\bullet} = \bar{Y}_{\bullet j} - 0 - (\bar{Y}_{\bullet\bullet} - 0) = \bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet}$ and

$Z_{ij} - \bar{Z}_{\bullet j} = Y_{ij} - (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}) - (\bar{Y}_{\bullet j} - 0) = Y_{ij} - \bar{Y}_{i\bullet} - \bar{Y}_{\bullet j} + \bar{Y}_{\bullet\bullet}$

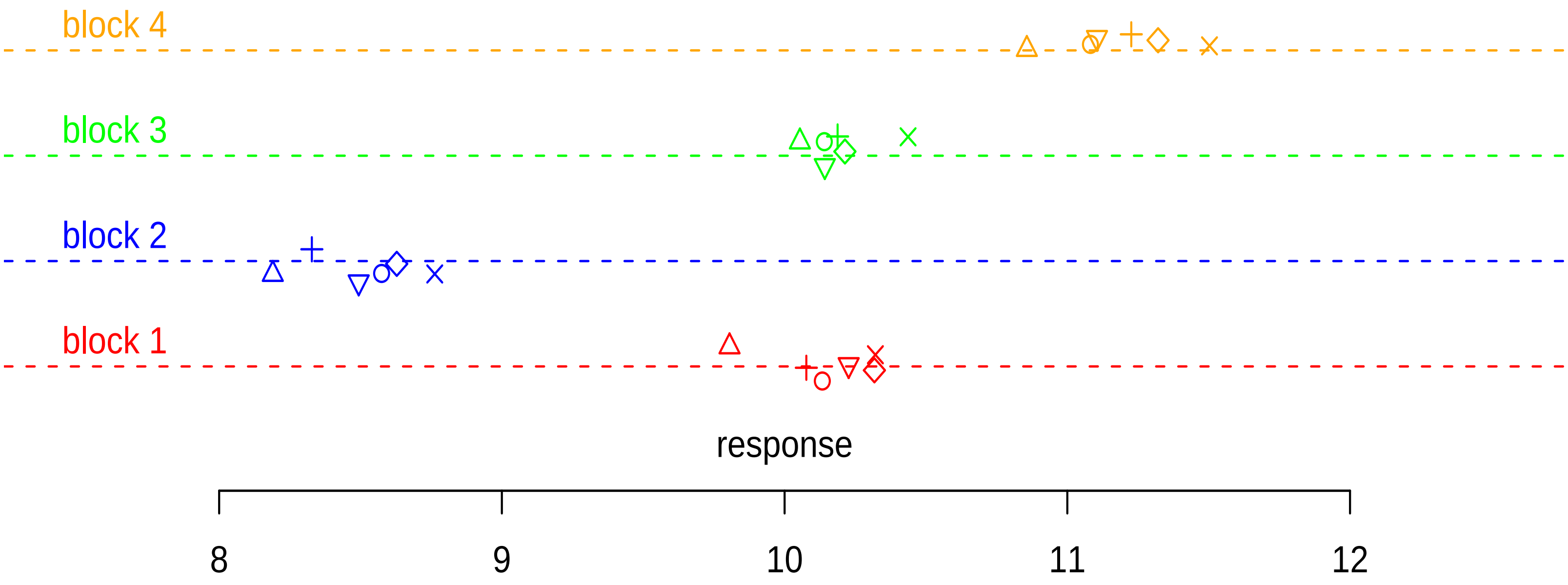
Compare the treatment effect dispersion $\sum_{ij}(\bar{Z}_{\bullet j} - \bar{Z}_{\bullet\bullet})^2 = \sum_{ij}(\bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet})^2$

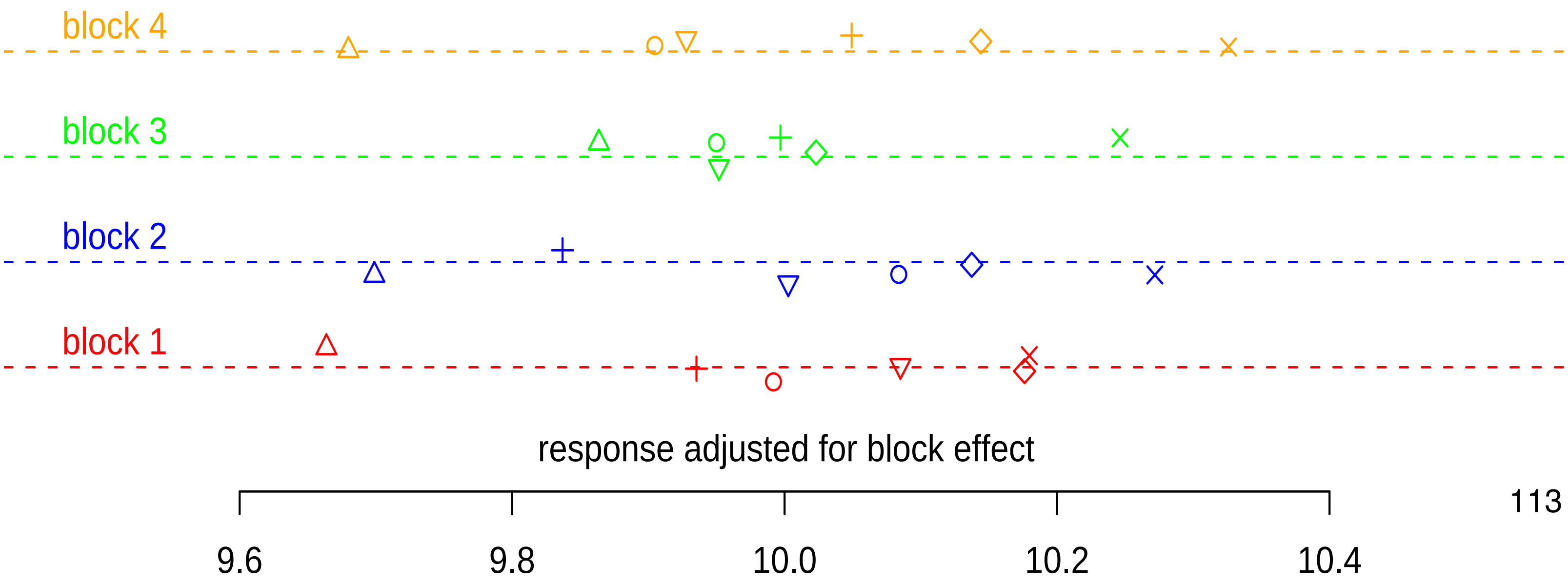
against the within “replication” dispersion:

$\sum_{ij}(Z_{ij} - \bar{Z}_{\bullet j})^2 = \sum_{ij}(Y_{ij} - \bar{Y}_{i\bullet} - \bar{Y}_{\bullet j} + \bar{Y}_{\bullet\bullet})^2$ which leads back to our

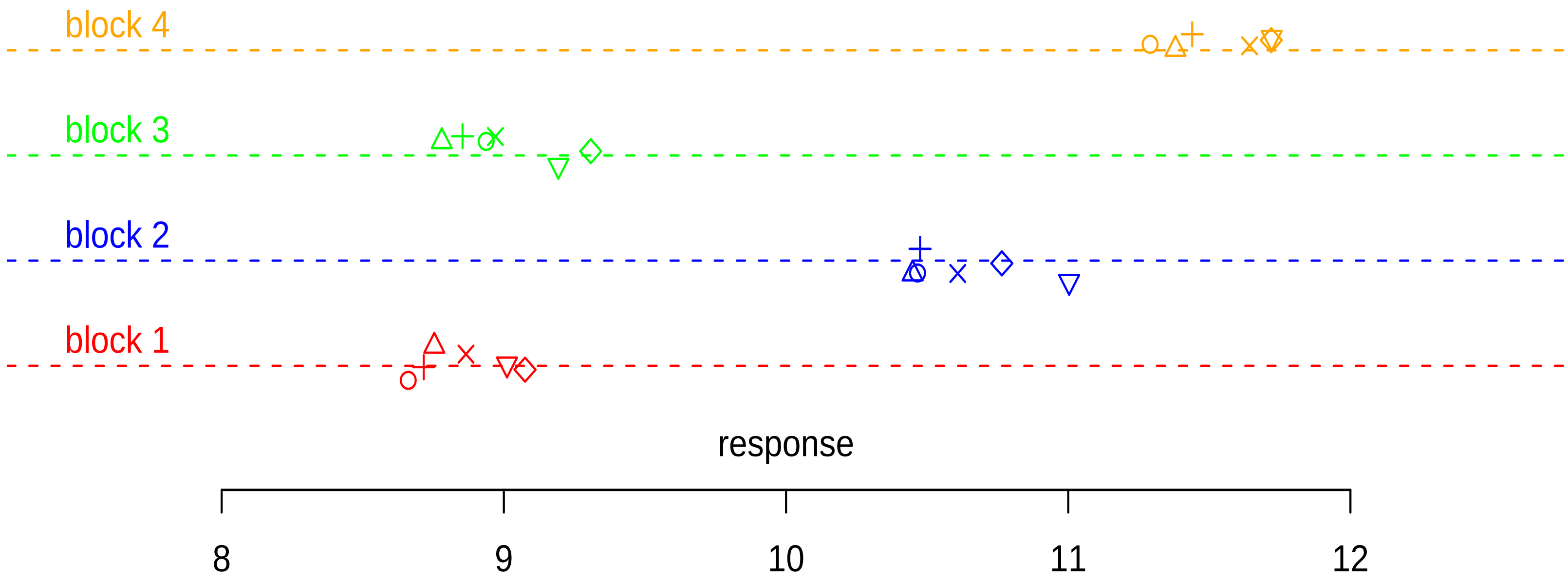
$$F_{\text{Treat}} = \left[\sum_{ij}(\bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet})^2/(t_2 - 1)\right] \Big/ \left[\sum_{ij}(Y_{ij} - \bar{Y}_{i\bullet} - \bar{Y}_{\bullet j} + \bar{Y}_{\bullet\bullet})^2/((t_1 - 1)(t_2 - 1))\right]$$

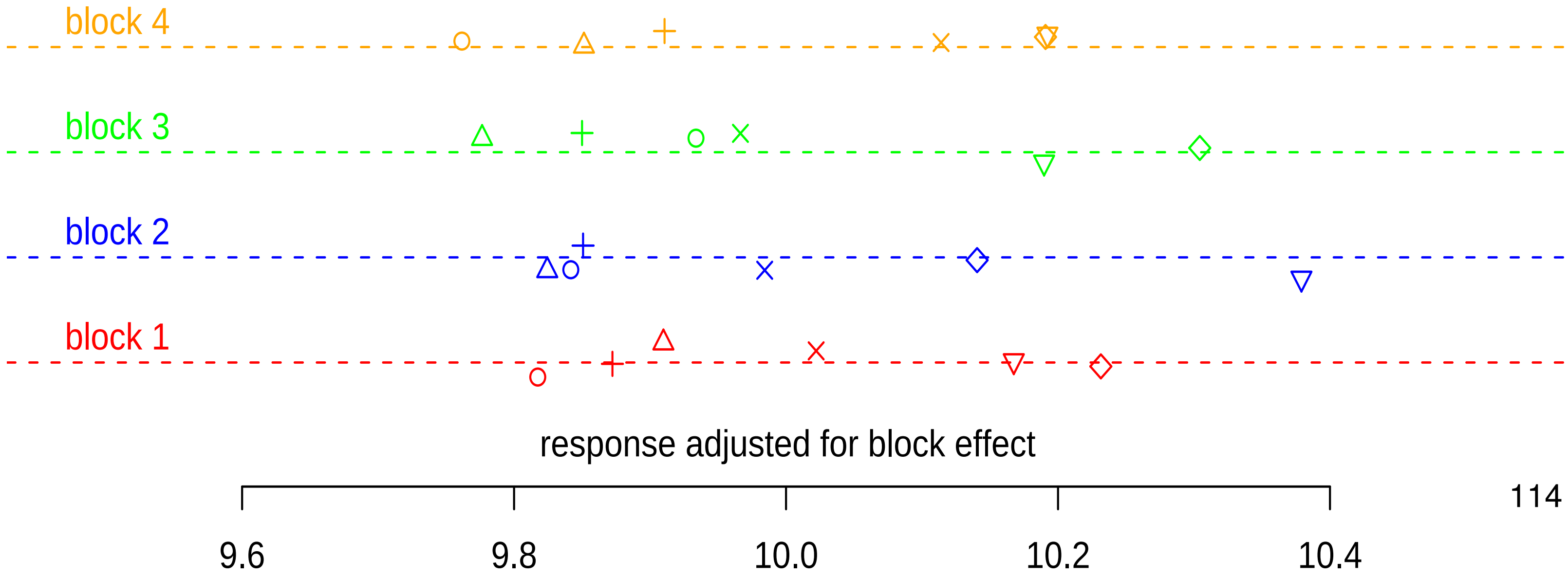
Visual Block Adjustment 1





Visual Block Adjustment 2





Comments on the Visual Block Adjustments

Note the clear block effect: variability between blocks $\longleftrightarrow$ variability within block.

After the blocks are shifted to remove the individual block effects

the data sets appear aligned on top of each other.

The relative pattern relationship within each block is undisturbed.

We are just using a magnified scale.

It is easy to discern the differences between the various symbol groups (treatments)

relative to the variation within symbol groups.

It might have been more difficult to see without prior alignment.

No Replication in Fertilizer Data!

```
> anova(lm(response~as.factor(treatment):as.factor(row),
+ data=fertilizerdata))
Analysis of Variance Table

Response: response
                                    Df Sum Sq Mean Sq F value Pr(>F)
as.factor(treatment):as.factor(row) 23 506.33   22.01
Residuals                            0   0.00
```

Here we are trying to test for the significance of all (treatment, row) combinations.

Without success, since $SS_{\mathrm{E}} = 0$ because of no replication, $\mathrm{df}_{\mathrm{E}} = 0$.

No Replication!

```
> anova(lm(response~as.factor(treatment)*as.factor(row),
+ data=fertilizerdata))
Analysis of Variance Table

Response: response
                                     Df  Sum Sq Mean Sq F value Pr(>F)
as.factor(treatment)                  5 201.316  40.263
as.factor(row)                        3 197.004  65.668
as.factor(treatment):as.factor(row)  15 108.008   7.201
Residuals                             0   0.000
```

Here we are trying to test for the significance of the full interaction model.

Without success, since $SS_{\mathrm{E}} = 0$ because of no replication, $\mathrm{df}_{\mathrm{E}} = 0$.

The difference to the previous analysis is that we express the 24 means μ_{ij}

via structure (grand mean, main effects and interactions).

SS for the No Replication Cases

In the first case, testing all (treatment level,block level) combinations, we have

$$SS_{\mathrm{T}} = \sum_{ijk}(Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet})^2 \quad \text{and} \quad SS_{\mathrm{E}} = \sum_{ijk}(Y_{ijk} - \bar{Y}_{ij\bullet})^2 = 0$$

since $k \equiv 1$ and thus $Y_{ijk} = \bar{Y}_{ij\bullet}$, and in the second case (full model)

$$\begin{aligned} SS_T &= \sum_{ij}(Y_{ij} - \bar{Y}_{\bullet\bullet})^2 = \sum_{ij}(\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 + \sum_{ij}(\bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet})^2 + \sum_{ij}(Y_{ij} - \bar{Y}_{i\bullet} - \bar{Y}_{\bullet j} + \bar{Y}_{\bullet\bullet})^2 \\ &= \sum_{ijk}(Y_{ijk} - \bar{Y}_{\bullet\bullet\bullet})^2 = \sum_{ijk}(\bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 + \sum_{ijk}(\bar{Y}_{\bullet j\bullet} - \bar{Y}_{\bullet\bullet\bullet})^2 + \sum_{ijk}(\bar{Y}_{ij\bullet} - \bar{Y}_{i\bullet\bullet} - \bar{Y}_{\bullet j\bullet} + \bar{Y}_{\bullet\bullet\bullet})^2 \\ &\qquad + \sum_{ijk}(Y_{ijk} - \bar{Y}_{ij\bullet})^2 \end{aligned}$$

with

$$SS_{\mathrm{E}} = \sum_{ijk}(Y_{ijk} - \bar{Y}_{ij\bullet})^2 = 0$$

since $k \equiv 1$ collapses the summation over k, i.e, $Y_{ijk} = Y_{ij1} = \bar{Y}_{ij\bullet} = Y_{ij} \ \forall i, j$.

With Replication?

With replication we could test for interaction. Would we be interested in interaction effects between treatment and the blocking factor?

The blocking factor is only used experimentally in order to make treatment effects stand out more within a block due to anticipated reduced variability within each block.

In real life use of the treatment we will not manipulate the blocking factor.
It is just a nuisance variation factor that we have to live with.

However, if variation as caused by the uncontrolled blocking factor is substantial, any benefit of the treatment may become swamped in real life application. This may be an instance of the old wisdom: Don’t sweat the small stuff. On the other hand, from a patient’s point of view a small pain reduction may be big. The patient does not care about the wide variability of pain in others.

Randomization Tests Revisited

Having assigned the 6 treatments randomly within each row, there are

$$(6!)^4 = 720^4 = 268,738,560,000 \approx 2.7 \cdot 10^{11}$$

such assignments. This is roughly $1/10,000$ of the number of treatment patterns had we assigned the 6 treatments in groups of 4 without the row blocking restriction

$$\binom{24}{4} \times \binom{20}{4} \times \binom{16}{4} \times \binom{12}{4} \times \binom{8}{4} \times \binom{4}{4} = 3.246671 \cdot 10^{15}$$

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 $H_0 : \tau_1 = \ldots = \tau_6 (= 0)$ (no treatment effect) using as our randomization test statistics

$$F_{\text{Treat}} = \frac{\sum_{ij}(\bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet})^2/(6-1)}{\sum_{ij}(Y_{ij} - \bar{Y}_{i\bullet} - \bar{Y}_{\bullet j} + \bar{Y}_{\bullet\bullet})^2/((6-1)\times(4-1))} \Rightarrow$$ reference distribution

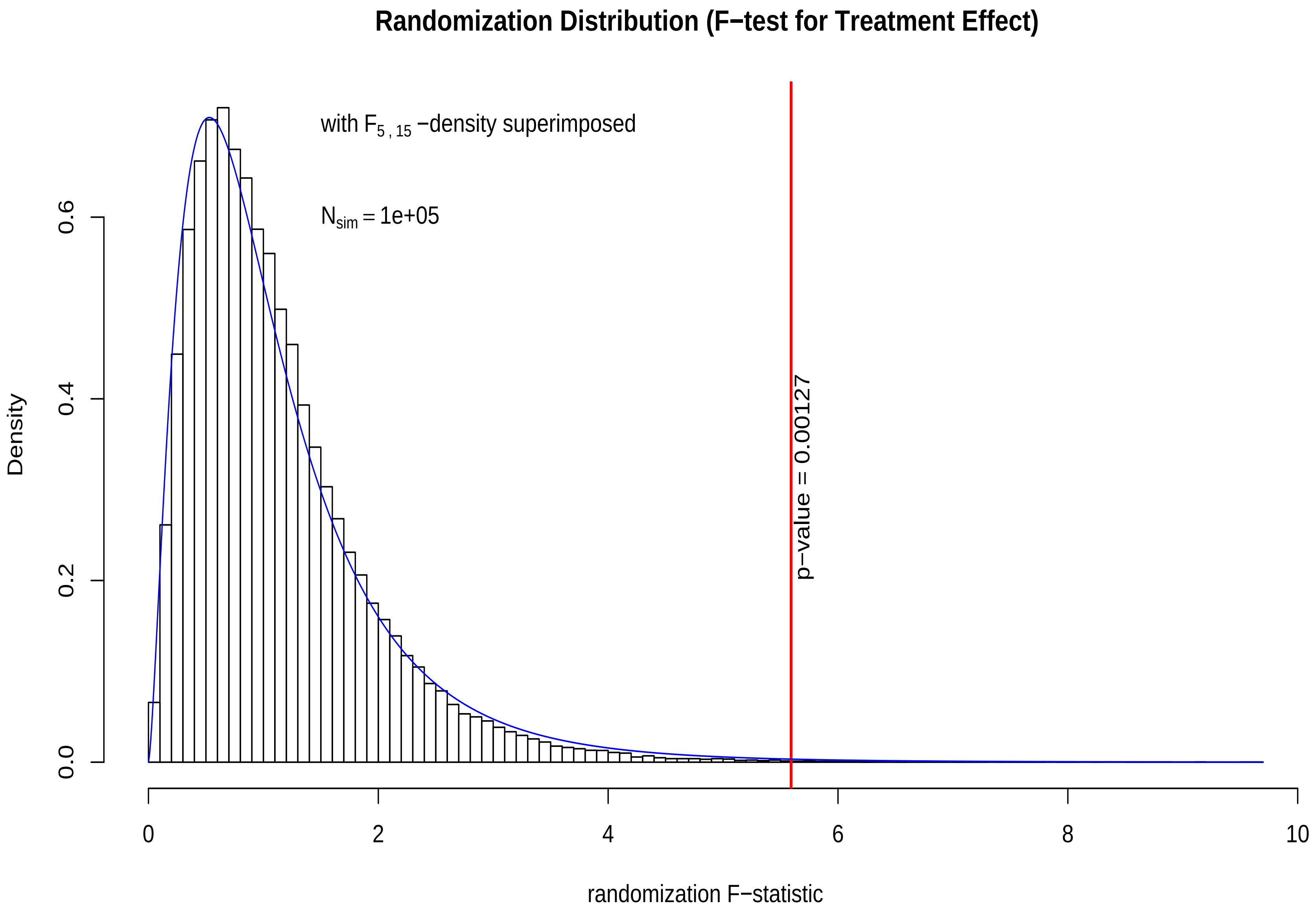
A Large Value of F_{Treat}

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

The value of F_{Treat} observed, namely $F_{\text{Treat}}^{\text{obs}}$, would then be just one of these equally likely values.

If the value $F_{\text{Treat}}^{\text{obs}}$ is in the far upper tail of the reference distribution, i.e., the p-value $= P(F_{\text{Treat}} \geq F_{\text{Treat}}^{\text{obs}})$ is (very) small ($\leq .05$ or $\leq .01$), then we can either take the position that we saw something rare by accident in our random assignment of treatments, or the assumption that treatment does not matter is not true. In the latter case this would be manifested quite naturally in our large value of $F_{\text{Treat}}^{\text{obs}}$.

Randomization Test for Treatment Effect



Comparing Results

The randomization reference distribution yielded a p-value of $.00127 = 127/100000$ while the normal theory test for the same hypothesis and using the same test statistic came up with a p-value of $.004191$ 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 `qgamma(.95,127+1)/100000=0.001471603` or via `qbeta(.95,127+1,100000-127)=0.001471455`.

That still leaves a wide gap to $.004191$, which is probably due to the possibility that the $F_{5,15}$ approximation is not quite a good enough fit out in the far tail.
The superimposed F-density looks mostly like a good fit.

Randomization Test for Block Effect?

Can we carry out a randomization test for the block effect?

We did not randomize our experimental units $(4 \times 6 = 24$ plots) over the rows.

After assigning the 6 treatments per row in a random fashion, it is physically not feasible to take the 4 plots with treatment 5, one in each row, and randomly assign these 4 plots to one of the $4! = 24$ different row permutations.

We cannot transplant a plot from one row to another. This would run counter to our intent of exploiting the block (row) variation to better see treatment effects.

What Can We Do?

If in fact the block or row effect does not exist, then it would have been conceptually OK to transplant any plot to a different location (row).

In that case (of no row effect), transplanting or not transplanting would not affect our test for a treatment effect.

However, we did not transplant plots in the hope of between block (row) variation.

Thus we can only pretend that our particular arrangement of 24 (treatment, row) combinations was completely random and formally carry out a randomization test for the block effect.

How Do We Get the Randomization Distribution?

Take the data as given, with 6 treatments as assigned to the 6 plots in each row.

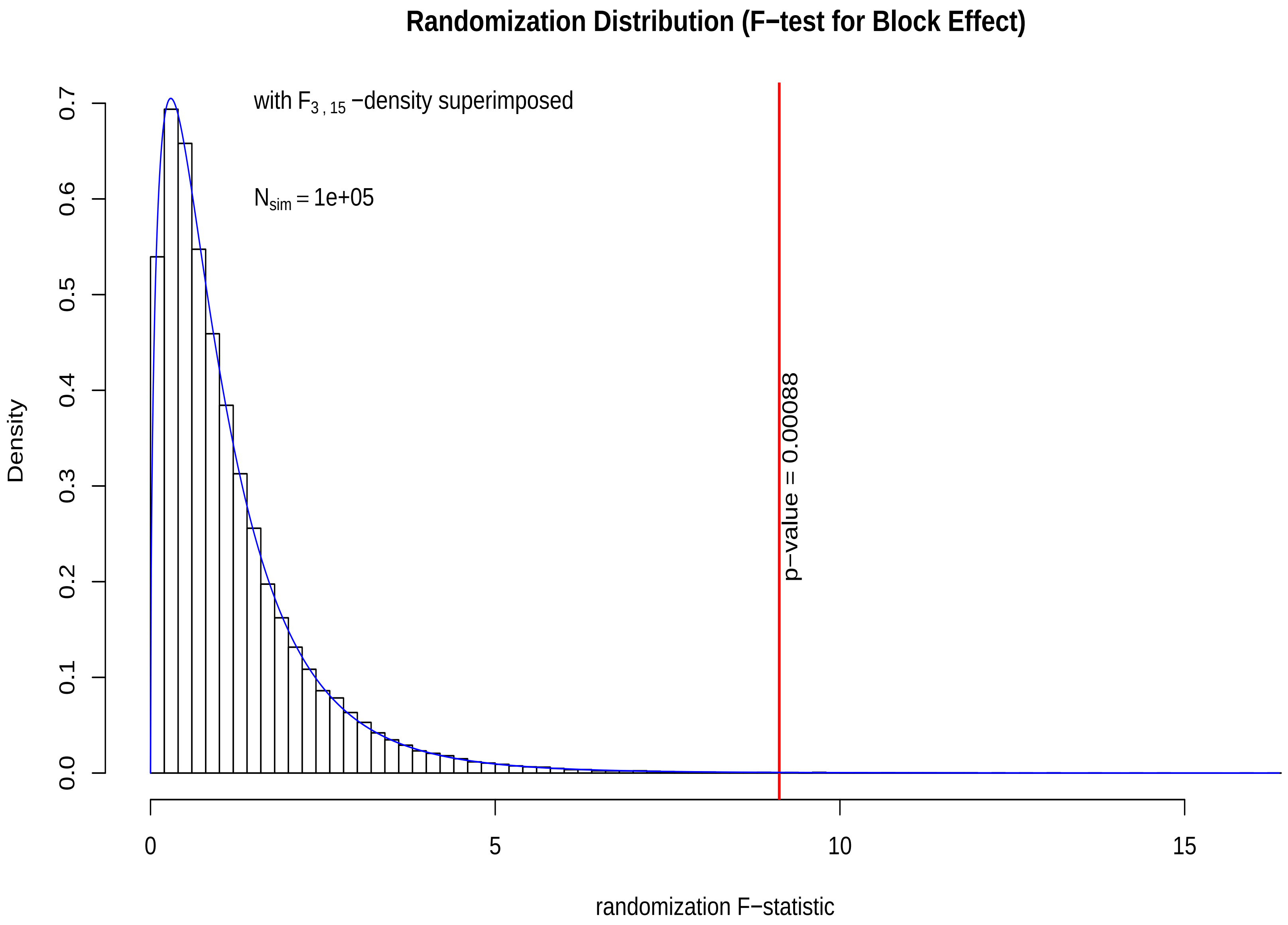
Since we assume no row effect we can take the row labels assigned to the 4 cases with treatment j and permute them in all 24 possible ways. Do this independently for each $j = 1, \ldots, 6$.

This would give us $24^6 = 191102976$ possible assignments and lead to the full F_{Block} reference distribution. By theory it is $\approx F_{3,15}$.

Doing this is not very practical and thus we simulate this distribution by computing F_{Block} for $N_{\text{sim}} = 100,000$ randomly chosen row label assignments.

Calculate or estimate the p-value of $F_{\text{Block}}^{\text{obs}}$ as the proportion of simulated F_{Block} values $\geq F_{\text{Block}}^{\text{obs}}$. The results are shown on the next slide.

Randomization Test for Block Effect



Comparing Results

The randomization reference distribution yielded a p-value of $.00088 = 88/100000$ while the normal theory test for the same hypothesis (no row factor effect) and using the same test statistic came up with a p-value of $.001116$ from $F_{3,15}$.

Question: Is the difference explainable by statistical variation alone? Such statistical variation results from approximating the true randomization reference distribution by simulating $N_{\text{sim}} = 100,000$ row label permutations.

A 95% upper confidence bound for the p-value resulting from the full true reference distribution can be obtained from R via `qgamma(.95,88+1)/100000= 0.001050649`. This is not too far off from .001116.

The superimposed F-density looks mostly like a good fit.