2-Factor Balanced Completely Randomized ANOVA

Engineering Statistics II
Section 11.2

Josh Engwer

TTU

2018

PART I:

2-FACTOR BALANCED EXPERIMENTS

Why 2F ANOVA and not two 1F ANOVA's?

Suppose one wishes to analyze a designed experiment involving <u>two</u> factors.

It seems reasonable to conduct two independent 1-Factor ANOVA's
– one on the 1^{st} factor (**factor A**), the other on the 2^{nd} factor (**factor B**).

Unfortunately, this is a poor strategy for the following reasons♣♡:

1. 2F ANOVA tests for an **interaction effect** – two 1F ANOVA's cannot.
   - (Definition and details later in this slide deck.)
2. 2F ANOVA results in more powerful F-tests than two 1F ANOVA's.
   - i.e. 2F ANOVA better explains variability than two 1F ANOVA's.
3. 2F ANOVA is more cost efficient than two 1F ANOVA's.
   - 2F ANOVA requires half as many measurements as two 1F ANOVA's.
4. 3F ANOVA generalizes easily from 2F ANOVA, not from two 1F ANOVA's.

♣ R.G. Lomax, D.L. Hahs-Vaughn, *Statistical Concepts: A 2^{nd} Course*, 4^{th} Ed., 2012.
♡ J.P. Stevens, *Intermediate Statistics: A Modern Approach*, 3^{rd} Ed., 2007.

2-Factor Balanced Experiments

Definition

(2-Factor Balanced Experiment)

A 2-factor experiment with equal group sizes of $K > 1$ is called **balanced**.

A $I \times J$ 2F experiment means Factor A has I levels & Factor B has J levels.

SYNONYMS: Balanced/Orthogonal data/design/model

FACTOR B: $\rightarrow$ **FACTOR A:** $\downarrow$	Level 1 ($x_{\bullet 1}$)	Level 2 ($x_{\bullet 2}$)
Level 1 ($x_{1\bullet}$)	x_{111}, x_{112}	x_{121}, x_{122}
Level 2 ($x_{2\bullet}$)	x_{211}, x_{212}	x_{221}, x_{222}
Level 3 ($x_{3\bullet}$)	x_{311}, x_{312}	x_{321}, x_{322}

Prototype 3×2 balanced experiment with $K = 2$

x_{ijk} $\equiv$ k^{th} measurement at (i, j)-levels of factors (A,B)

$\bar{x}_{ij\bullet}$ $\equiv$ group mean at (i, j)-levels of factors (A,B)

$\bar{x}_{i\bullet\bullet}$ $\equiv$ mean of measurements at i^{th} level of factor A

$\bar{x}_{\bullet j\bullet}$ $\equiv$ mean of measurements at j^{th} level of factor B

2-Factor Balanced Experiments

Definition

(2-Factor Balanced Experiment)

A 2-factor experiment with equal group sizes of $K > 1$ is called **balanced**.

A $I \times J$ experiment means Factor A has I levels & Factor B has J levels.

SYNONYMS: Balanced/Orthogonal data/design/model

FACTOR B: $\rightarrow$ **FACTOR A:** $\downarrow$	Level 1 $(x_{\bullet 1})$	Level 2 $(x_{\bullet 2})$
Level 1 $(x_{1\bullet})$	$x_{111},\ x_{112}$	$x_{121},\ x_{122}$
Level 2 $(x_{2\bullet})$	$x_{211},\ x_{212}$	$x_{221},\ x_{222}$
Level 3 $(x_{3\bullet})$	$x_{311},\ x_{312}$	$x_{321},\ x_{322}$

Prototype 3×2 balanced experiment with $K = 2$

$$
\begin{aligned}
\bar{x}_{11\bullet} &= (x_{111} + x_{112})/2 \\
\bar{x}_{1\bullet\bullet} &= (x_{111} + x_{112} + x_{121} + x_{122})/4 \\
\bar{x}_{\bullet 1\bullet} &= (x_{111} + x_{112} + x_{211} + x_{212} + x_{311} + x_{312})/6
\end{aligned}
$$

2-Factor Balanced Experiments

Definition

(2-Factor Balanced Experiment)

A 2-factor experiment with equal group sizes of $K > 1$ is called **balanced**.

A $I \times J$ 2F experiment means Factor A has I levels & Factor B has J levels.

SYNONYMS: Balanced/Orthogonal data/design/model

FACTOR B: $\rightarrow$ **FACTOR A:** $\downarrow$	Level 1 ($x_{\bullet 1}$)	Level 2 ($x_{\bullet 2}$)
Level 1 ($x_{1\bullet}$)	x_{111}, x_{112}	x_{121}, x_{122}
Level 2 ($x_{2\bullet}$)	x_{211}, x_{212}	x_{221}, x_{222}
Level 3 ($x_{3\bullet}$)	x_{311}, x_{312}	x_{321}, x_{322}

Prototype 3×2 balanced experiment with $K = 2$

$$\bar{x}_{32\bullet} = (x_{321} + x_{322})/2$$
$$\bar{x}_{3\bullet\bullet} = (x_{311} + x_{312} + x_{321} + x_{322})/4$$
$$\bar{x}_{\bullet 2\bullet} = (x_{121} + x_{122} + x_{221} + x_{222} + x_{321} + x_{322})/6$$

Main & Interaction Effects in 2F bcrANOVA

Definition

(Main Effect in 2F bcrANOVA)

Given a 2-Factor balanced completely randomized experiment.

A **main effect** of one factor is present if its effect at a fixed level is the same for all levels of the other factor.

Definition

(Interaction Effect♡ in 2F bcrANOVA)

Given a 2-Factor balanced completely randomized experiment.

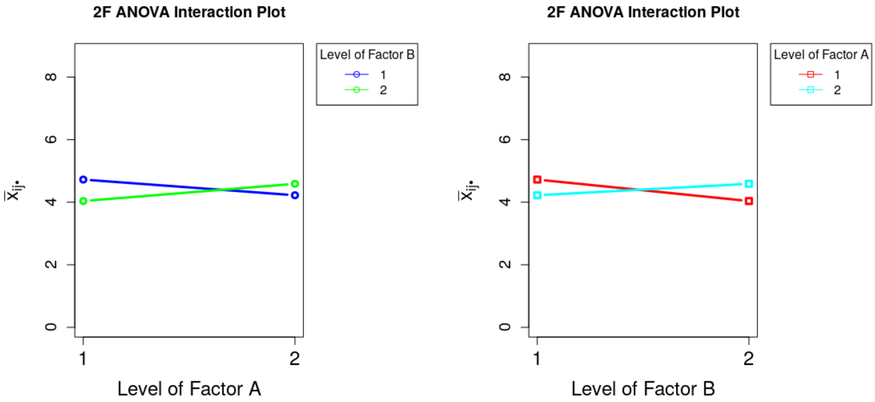
An **interaction (effect)** is present if one factor's effect at a fixed level is not the same for all levels of the other factor.

i.e. An interaction means the combined levels of the two factors results in an effect in addition to any main effects of each factor alone.

i.e. A lack of interaction means the two factors' effects are independent.

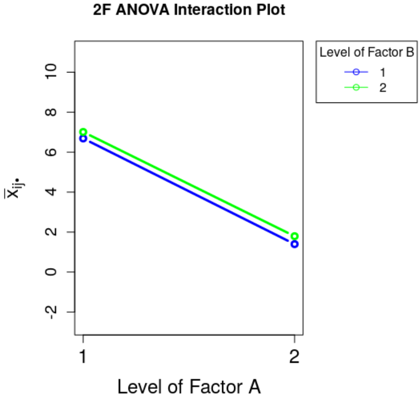
♡ J.P. Stevens, *Intermediate Statistics: A Modern Approach*, 3^{rd} Ed., 2007.

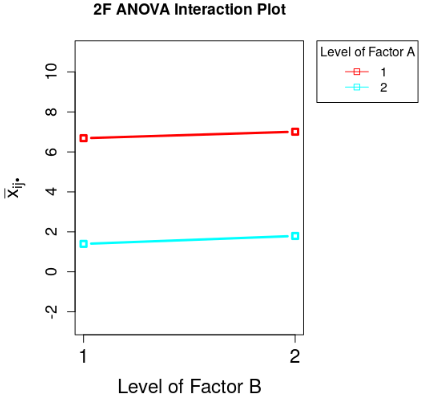
2x2 Interaction Plot (Given: A=no, B=no, AB=no)



(left plot)	A lines nearly parallel	$\Longrightarrow$	AB interaction's absent
(right plot)	B lines nearly parallel	$\Longrightarrow$	AB interaction's absent
(left plot)	A lines nearly horizontal	$\Longrightarrow$	A main effect's absent
(left plot)	A lines nearly coincident	$\Longrightarrow$	B main effect's absent
(right plot)	B lines nearly horizontal	$\Longrightarrow$	B main effect's absent
(right plot)	B lines nearly coincident	$\Longrightarrow$	A main effect's absent

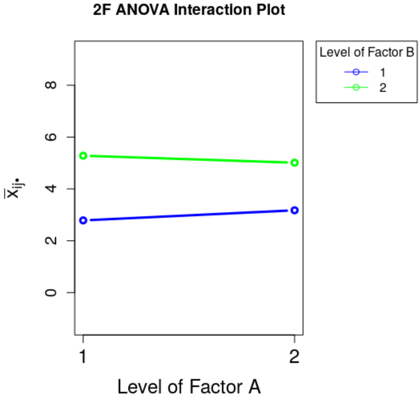
2x2 Interaction Plot (Given: A=yes, B=no, AB=no)

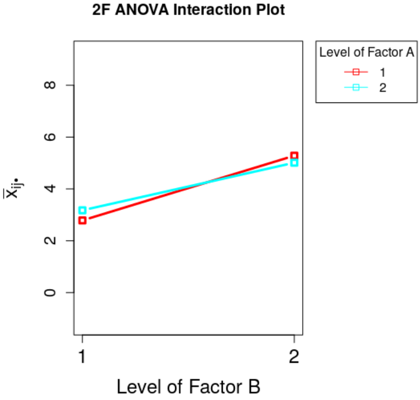




(left plot)	A lines nearly parallel	$\Longrightarrow$	AB interaction's absent
(right plot)	B lines nearly parallel	$\Longrightarrow$	AB interaction's absent
(left plot)	A lines largely slanted	$\Longrightarrow$	A main effect's present
(left plot)	A lines nearly coincident	$\Longrightarrow$	B main effect's absent
(right plot)	B lines nearly horizontal	$\Longrightarrow$	B main effect's absent
(right plot)	B lines largely separate	$\Longrightarrow$	A main effect's present

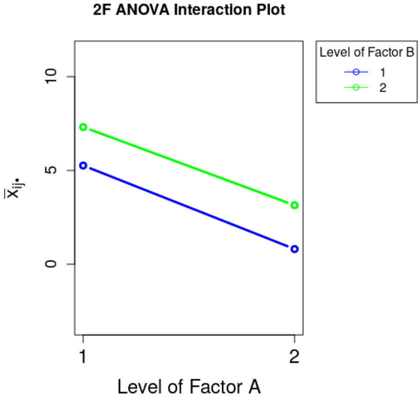
2x2 Interaction Plot (Given: A=no, B=yes, AB=no)

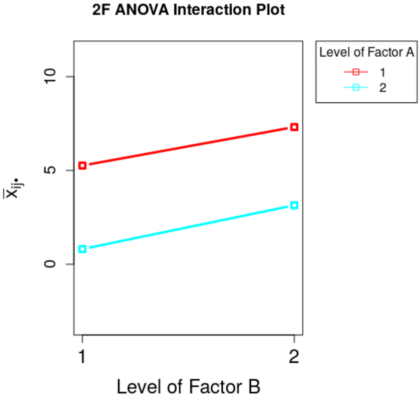




(left plot)	A lines nearly parallel	$\Longrightarrow$	AB interaction's absent
(right plot)	B lines nearly parallel	$\Longrightarrow$	AB interaction's absent
(left plot)	A lines nearly horizontal	$\Longrightarrow$	A main effect's absent
(left plot)	A lines largely separate	$\Longrightarrow$	B main effect's present
(right plot)	B lines largely slanted	$\Longrightarrow$	B main effect's present
(right plot)	B lines nearly coincident	$\Longrightarrow$	A main effect's absent

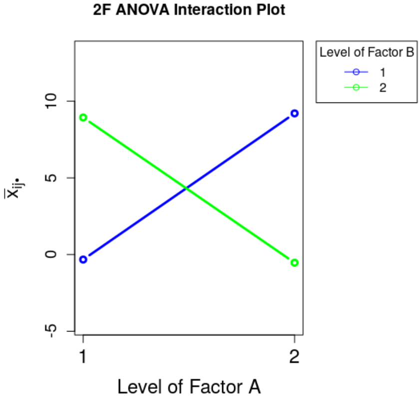
2x2 Interaction Plot (Given: A=yes, B=yes, AB=no)

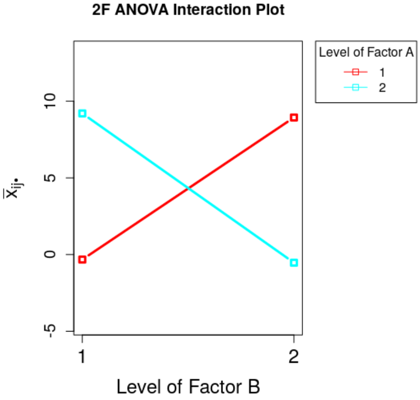




(left plot)	A lines nearly parallel	$\Longrightarrow$	AB interaction's absent
(right plot)	B lines nearly parallel	$\Longrightarrow$	AB interaction's absent
(left plot)	A lines largely slanted	$\Longrightarrow$	A main effect's present
(left plot)	A lines largely separate	$\Longrightarrow$	B main effect's present
(right plot)	B lines largely slanted	$\Longrightarrow$	B main effect's present
(right plot)	B lines largely separate	$\Longrightarrow$	A main effect's present

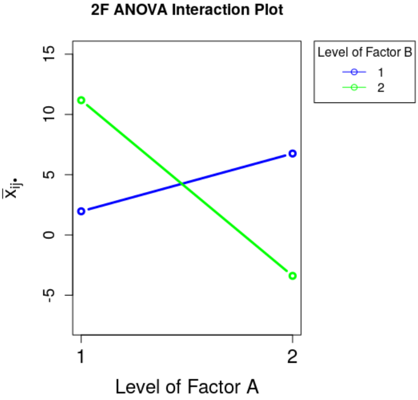
2x2 Interaction Plot (Given: A=no, B=no, AB=yes)

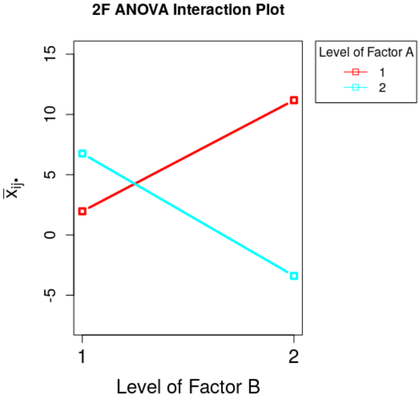




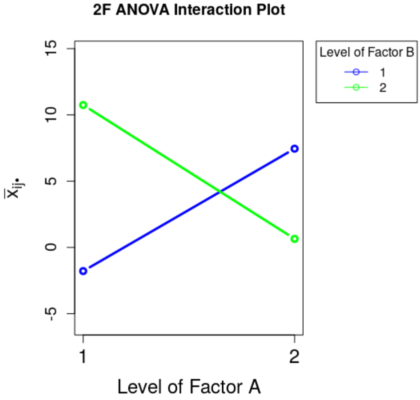
(left plot)	A lines largely non-parallel	$\Longrightarrow$	AB interaction's present
(right plot)	B lines largely non-parallel	$\Longrightarrow$	AB interaction's present
(left plot)	???????		A main effect's absent??
(left plot)	???????		B main effect's absent??
(right plot)	???????		B main effect's absent??
(right plot)	???????		A main effect's absent??

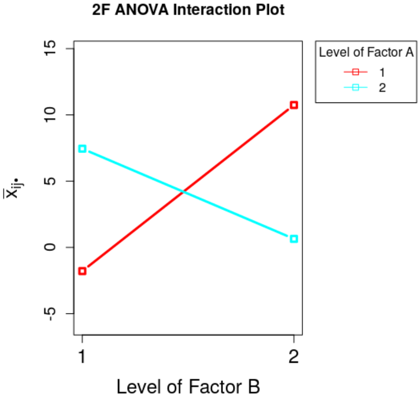
2x2 Interaction Plot (Given: A=yes, B=no, AB=yes)





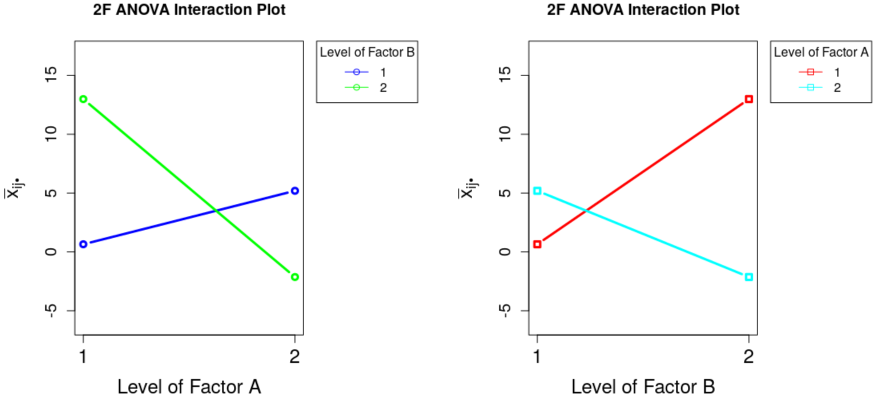
(left plot)	A lines largely non-parallel	$\Longrightarrow$	AB interaction's present
(right plot)	B lines largely non-parallel	$\Longrightarrow$	AB interaction's present
(left plot)	???????		A main effect's present??
(left plot)	???????		B main effect's absent??
(right plot)	???????		B main effect's absent??
(right plot)	???????		A main effect's present??





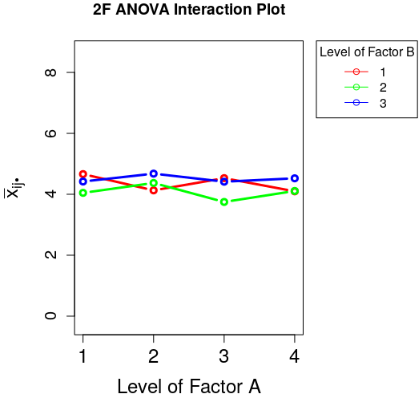
(*left plot*)	A lines largely non-parallel	$\Longrightarrow$	AB interaction's present
(*right plot*)	B lines largely non-parallel	$\Longrightarrow$	AB interaction's present
(*left plot*)	???????		A main effect's absent??
(*left plot*)	???????		B main effect's present??
(*right plot*)	???????		B main effect's present??
(*right plot*)	???????		A main effect's absent??

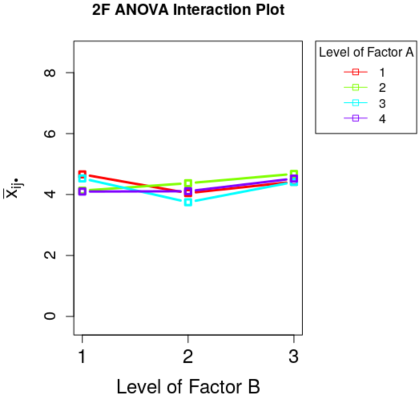
2x2 Interaction Plot (Given: A=yes, B=yes, AB=yes)



(*left plot*)	A lines largely non-parallel	$\Longrightarrow$	AB interaction's present
(*right plot*)	B lines largely non-parallel	$\Longrightarrow$	AB interaction's present
(*left plot*)	???????		A main effect's present??
(*left plot*)	???????		B main effect's present??
(*right plot*)	???????		B main effect's present??
(*right plot*)	???????		A main effect's present??

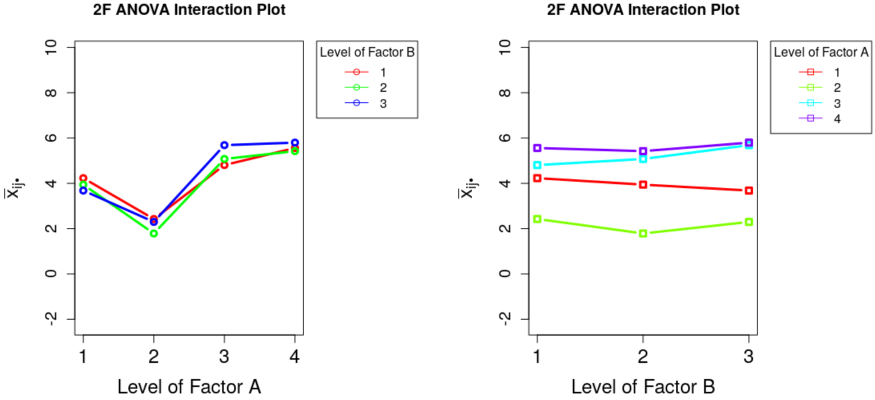
4x3 Interaction Plot (Given: A=no, B=no, AB=no)





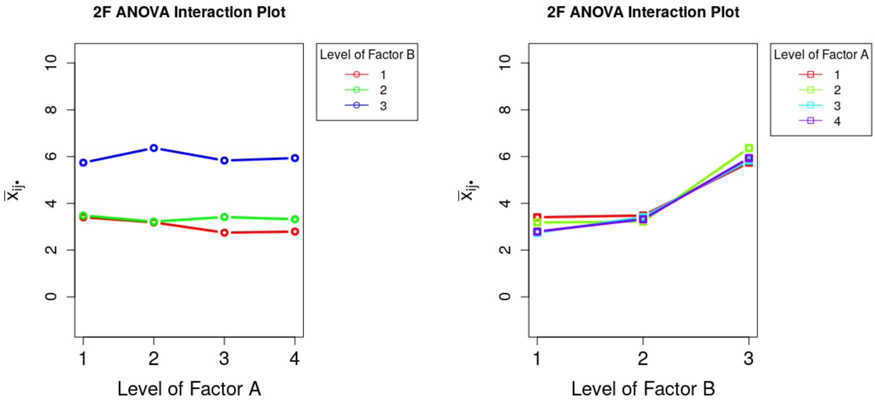
(left plot)	A lines nearly parallel	$\Longrightarrow$	AB interaction's absent
(right plot)	B lines nearly parallel	$\Longrightarrow$	AB interaction's absent
(left plot)	A lines nearly horizontal	$\Longrightarrow$	A main effect's absent
(left plot)	A lines nearly coincident	$\Longrightarrow$	B main effect's absent
(right plot)	B lines nearly horizontal	$\Longrightarrow$	B main effect's absent
(right plot)	B lines nearly coincident	$\Longrightarrow$	A main effect's absent

4x3 Interaction Plot (Given: A=yes, B=no, AB=no)



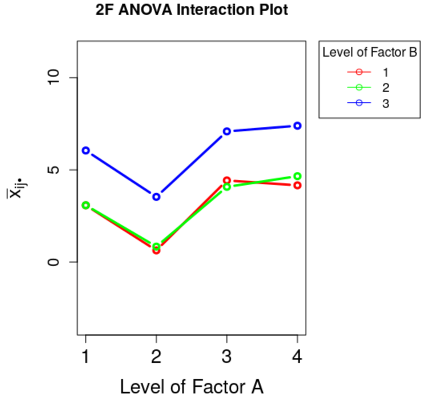
(left plot)	A lines nearly parallel	$\Longrightarrow$	AB interaction's absent
(right plot)	B lines nearly parallel	$\Longrightarrow$	AB interaction's absent
(left plot)	A lines largely slanted	$\Longrightarrow$	A main effect's present
(left plot)	A lines nearly coincident	$\Longrightarrow$	B main effect's absent
(right plot)	B lines nearly horizontal	$\Longrightarrow$	B main effect's absent
(right plot)	B lines largely separate	$\Longrightarrow$	A main effect's present

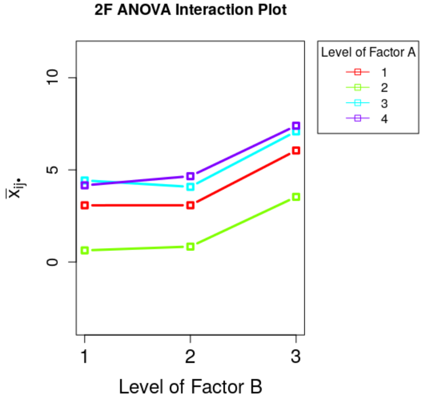
4x3 Interaction Plot (Given: A=no, B=yes, AB=no)



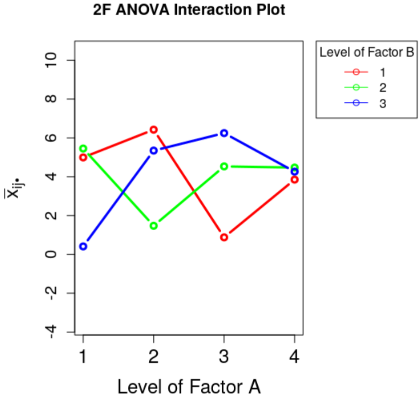
(left plot)	A lines nearly parallel	$\Longrightarrow$	AB interaction's absent
(right plot)	B lines nearly parallel	$\Longrightarrow$	AB interaction's absent
(left plot)	A lines nearly horizontal	$\Longrightarrow$	A main effect's absent
(left plot)	A lines largely separate	$\Longrightarrow$	B main effect's present
(right plot)	B lines largely slanted	$\Longrightarrow$	B main effect's present
(right plot)	B lines nearly coincident	$\Longrightarrow$	A main effect's absent

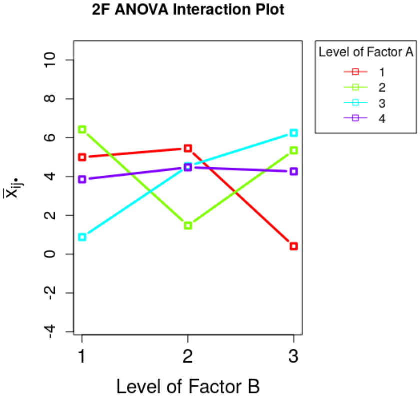
4x3 Interaction Plot (Given: A=yes, B=yes, AB=no)





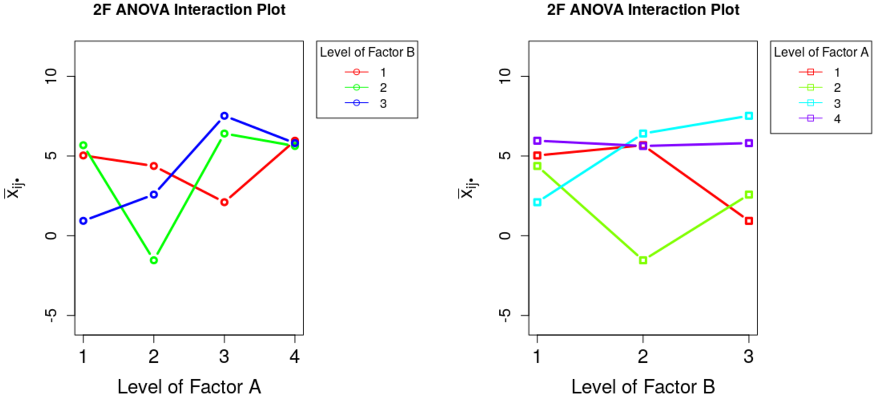
(left plot)	A lines nearly parallel	$\Longrightarrow$	AB interaction's absent
(right plot)	B lines nearly parallel	$\Longrightarrow$	AB interaction's absent
(left plot)	A lines largely slanted	$\Longrightarrow$	A main effect's present
(left plot)	A lines largely separate	$\Longrightarrow$	B main effect's present
(right plot)	B lines largely slanted	$\Longrightarrow$	B main effect's present
(right plot)	B lines largely separate	$\Longrightarrow$	A main effect's present





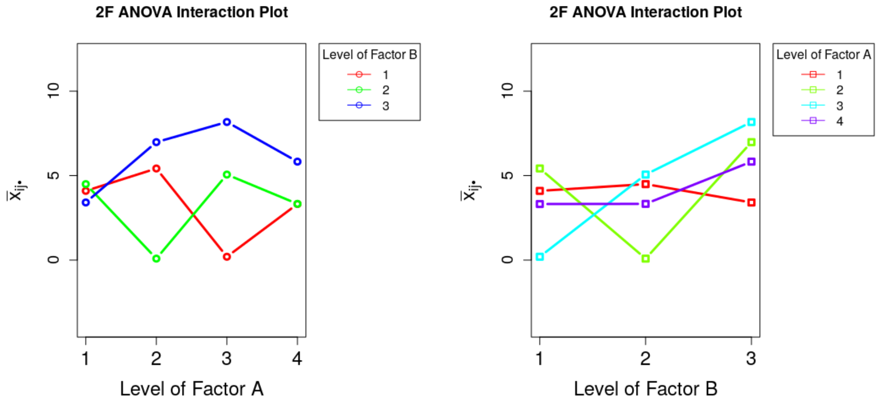
(*left plot*)	A lines largely non-parallel	$\Longrightarrow$	AB interaction's present
(*right plot*)	B lines largely non-parallel	$\Longrightarrow$	AB interaction's present
(*left plot*)	???????		A main effect's absent??
(*left plot*)	???????		B main effect's absent??
(*right plot*)	???????		B main effect's absent??
(*right plot*)	???????		A main effect's absent??

4x3 Interaction Plot (Given: A=yes, B=no, AB=yes)

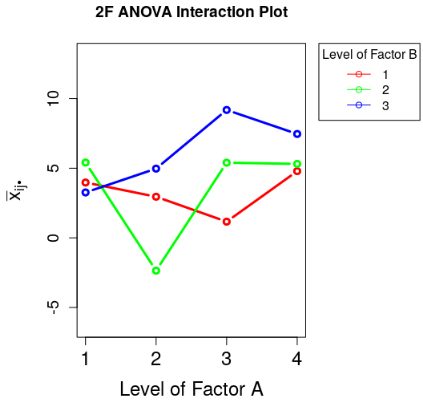


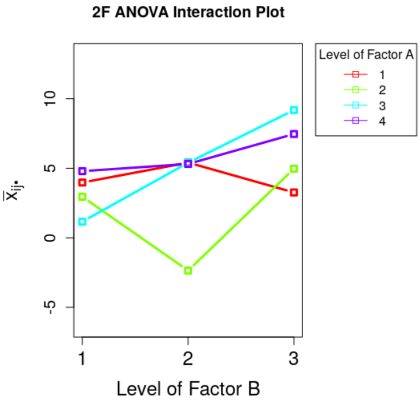
(*left plot*)	A lines largely non-parallel	$\Longrightarrow$	AB interaction's present
(*right plot*)	B lines largely non-parallel	$\Longrightarrow$	AB interaction's present
(*left plot*)	???????		A main effect's present??
(*left plot*)	???????		B main effect's absent??
(*right plot*)	???????		B main effect's absent??
(*right plot*)	???????		A main effect's present??

4x3 Interaction Plot (Given: A=no, B=yes, AB=yes)



(left plot)	A lines largely non-parallel	$\Longrightarrow$	AB interaction's present
(right plot)	B lines largely non-parallel	$\Longrightarrow$	AB interaction's present
(left plot)	???????		A main effect's absent??
(left plot)	???????		B main effect's present??
(right plot)	???????		B main effect's present??
(right plot)	???????		A main effect's absent??





(left plot)	A lines largely non-parallel	$\Longrightarrow$	AB interaction's present
(right plot)	B lines largely non-parallel	$\Longrightarrow$	AB interaction's present
(left plot)	???????		A main effect's present??
(left plot)	???????		B main effect's present??
(right plot)	???????		B main effect's present??
(right plot)	???????		A main effect's present??

Moral of the Story regarding Interaction Plots

1. Use interaction plots to infer the presence of a significant interaction.
   - Widen plot's vertical axis limits by four times the estimated std deviation.
   - Otherwise, an interaction may appear when the vertical axis scale is small.
2. If there's no significant interaction present:
   - The presence of main effects can be inferred.
3. If there is a significant interaction present:
   - It's too hard to infer presence of main effects visually.
   - However, the actual 2F ANOVA may infer presence of main effects...
     - ...but proper interpretation of any main effects given an interaction is hard.
   - Moreover, 2F ANOVA can infer the presence of an interaction.

All this said, interaction plots are mainly used to determine the presence of a significant interaction <u>before</u> performing an ANOVA when the corresponding assumptions call for the presence or lack of said interaction.

PART II:

2-Factor Linear (Statistical) Models:

Definitions, Examples

Least Squares Estimators (LSE's)

Best Linear Unbiased Estimators (BLUE's)

Gauss-Markov Theorem

2-Factor Linear (Statistical) Models (Definition)

With many-sample inference, it's convenient to use a **linear model**:

Definition

(2-Factor Linear Model)

Given a 2-factor balanced experiment with IJ groups, each of size $K > 1$.
In particular, factor A has I levels and factor B has J levels.

Let $X_{ijk} \equiv$ rv for k^{th} measurement at (i, j)-level of factors A & B.

Then, the **linear (statistical) model** for the experiment is defined as:

$$X_{ijk} = \mu + \alpha_i^A + \alpha_j^B + \gamma_{ij}^{AB} + E_{ijk} \quad \text{where} \quad E_{ijk} \overset{iid}{\sim} \text{Normal}(0, \sigma^2)$$

where:

$$\begin{aligned} \mu \quad &\equiv \quad \text{Population grand mean of all } IJ \text{ population means} \\ (\alpha_i^A, \alpha_j^B) \quad &\equiv \quad \text{Effect of } (i^{th}\text{-level factor A, } j^{th}\text{-level factor B)} \\ \gamma_{ij}^{AB} \quad &\equiv \quad \text{Interaction between } (i, j)\text{-level factors A \& B} \\ E_{ijk} \quad &\equiv \quad \text{Deviation of } X_{ijk} \text{ from } \mu \text{ due to random error} \end{aligned}$$

2-Factor Linear Models (Least-Squares Estimators)

Proposition

Given a 2-factor linear model:

$$X_{ijk} = \mu + \alpha_i^A + \alpha_j^B + \gamma_{ij}^{AB} + E_{ijk} \quad \textit{where} \quad E_{ijk} \overset{iid}{\sim} \textit{Normal}(0, \sigma^2)$$

*(a) The **least-squares♠♣ estimators (LSE's)** for the model parameters are:*

$$\begin{array}{rcl} \hat{\mu} & = & \bar{x}_{\bullet\bullet\bullet} \\ \hat{\alpha}_i^A & = & \bar{x}_{i\bullet\bullet} - \bar{x}_{\bullet\bullet\bullet} \\ \hat{\alpha}_j^B & = & \bar{x}_{\bullet j\bullet} - \bar{x}_{\bullet\bullet\bullet} \\ \hat{\gamma}_{ij}^{AB} & = & \bar{x}_{ij\bullet} - \bar{x}_{i\bullet\bullet} - \bar{x}_{\bullet j\bullet} + \bar{x}_{\bullet\bullet\bullet} \end{array} , \quad \begin{array}{rcl} \bar{x}_{\bullet\bullet\bullet} & \equiv & \textit{Grand sample mean} \\ \bar{x}_{i\bullet\bullet} & \equiv & \textit{Mean of groups at } i^{th}\textit{-lvl A} \\ \bar{x}_{\bullet j\bullet} & \equiv & \textit{Mean of groups at } j^{th}\textit{-lvl B} \\ \bar{x}_{ij\bullet} & \equiv & \textit{Mean of } (i,j)\textit{-lvl group} \end{array}$$

(b) For these LSE's, it's required that $\sum_i \alpha_i^A = \sum_j \alpha_j^B = \sum_i \gamma_{ij}^{AB} = \sum_j \gamma_{ij}^{AB} = 0$.

(c) These least-squares estimators are all unbiased.

PROOF: The general case is left as an ungraded exercise for the reader.

♠ A.M. Legendre, *Nouvelles Méthodes pour la Détermination des Orbites des Comètes*, 1806.

♣ Gauss, *Theoria Motus Corporum Coelestrium in Sectionibus Conicis Solem Ambientium*, 1809.

Definition

(Predicted Responses)

Given a 2-factor linear model:

$$X_{ijk} = \mu + \alpha_i^A + \alpha_j^B + \gamma_{ij}^{AB} + E_{ijk} \quad \text{where} \quad E_{ijk} \overset{iid}{\sim} \text{Normal}(0, \sigma^2)$$

Then the corresponding **predicted responses**, denoted $\hat{x}_{ijk}$, are:

$$\hat{x}_{ijk} := \hat{\mu} + \hat{\alpha}_i^A + \hat{\alpha}_j^B + \hat{\gamma}_{ij}^{AB} = \bar{x}_{ij\bullet}$$

SYNONYMS: Predicted values, fitted values

2-Factor Linear Models (Residuals)

Definition

(Residuals)

Given a 2-factor linear model:

$$X_{ijk} = \mu + \alpha_i^A + \alpha_j^B + \gamma_{ij}^{AB} + E_{ijk} \quad \text{where} \quad E_{ijk} \overset{iid}{\sim} \text{Normal}(0, \sigma^2)$$

Then the corresponding predicted responses, denoted $\hat{x}_{ijk}$, are:

$$\hat{x}_{ijk} := \hat{\mu} + \hat{\alpha}_i^A + \hat{\alpha}_j^B + \hat{\gamma}_{ij}^{AB} = \overline{x}_{ij\bullet}$$

Moreover, the corresponding **residuals**, denoted x_{ijk}^{res}, are:

$$x_{ijk}^{res} := x_{ijk} - \hat{x}_{ijk} = x_{ijk} - \overline{x}_{ij\bullet}$$

Linear Models (Best Linear Unbiased Estimators)

Point estimators for a linear model should be ideal ones:

Definition

(Best Linear Unbiased Estimators – BLUE's)

A point estimator $\hat{\theta}$ is called a **best linear unbiased estimator (BLUE)** if:

- It estimates a parameter θ of a linear model.
- It is a linear combination of the data points: $\hat{\theta} := \sum_{k=1}^{n} c_k x_k$
- It is an unbiased estimator: $\mathbb{E}[\hat{\theta}] = \theta$
- It has minimum variance of all such unbiased estimators.

<ins>REMARK:</ins> BLUE's are generally easier to construct & prove than UMVUE's.

For a 2-factor linear model: $X_{ijk} = \mu + \alpha_i^A + \alpha_j^B + \gamma_{ij}^{AB} + E_{ijk}$

$\hat{\mu}, \hat{\alpha}_i^A, \hat{\alpha}_j^B, \hat{\gamma}_{ij}^{AB}$ are each linear combinations of data points in the linear model.

Theorem

(Gauss[†]-Markov[‡] Theorem)

Given a 2-factor linear model: $X_{ijk} = \mu + \alpha_i^A + \alpha_j^B + \gamma_{ij}^{AB} + E_{ijk}$
Moreover, suppose the following conditions are all satisfied:

$$\begin{aligned} \mathbb{E}[E_{ijk}] &= 0 && (\textit{errors are all centered at zero}) \\ \mathbb{V}[E_{ijk}] &= \sigma^2 && (\textit{errors all have the same finite variance}) \\ \mathbb{C}[E_{ijk}, E_{i'j'k'}] &= 0 && (\textit{errors are uncorrelated when } i \neq i' \textit{ or } j \neq j' \textit{ or } k \neq k') \end{aligned}$$

Then, the least-squares estimators (LSE's) $\hat{\mu}, \hat{\alpha}_i^A, \hat{\alpha}_j^B, \hat{\gamma}_{ij}^{AB}$ *are all BLUE's.*

PROOF: Omitted due to time.

[†] C.F. Gauss, “Theoria Combinationis Observationum Erroribus Minimis Obnoxiae”, (1823), 1-58.

[‡] A.A. Markov, *Calculus of Probabilities*, 1^{st} Edition, 1900.
Андрей Андреевич Марков, Исчисление Вероятностей, Первое издание, 1900.

PART III:

2-FACTOR BALANCED COMPLETELY RANDOMIZED ANOVA
(2F bcrANOVA)

Motivation

Visual Dotplots

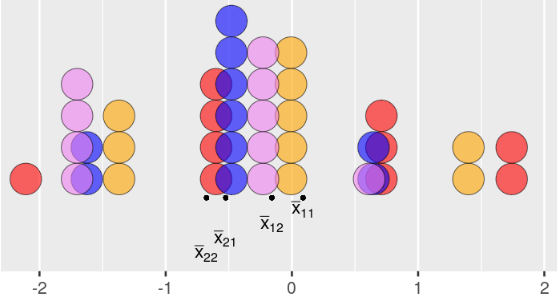
$$\bar{x}_{A_1} := \frac{\bar{x}_{11}+\bar{x}_{12}}{2}, \quad \bar{x}_{A_2} := \frac{\bar{x}_{21}+\bar{x}_{22}}{2}, \quad \bar{x}_{B_1} := \frac{\bar{x}_{11}+\bar{x}_{21}}{2}, \quad \bar{x}_{B_2} := \frac{\bar{x}_{12}+\bar{x}_{22}}{2}$$

$$\bar{x}_{AB_1} := \frac{\bar{x}_{11}+\bar{x}_{22}}{2}, \quad \bar{x}_{AB_2} := \frac{\bar{x}_{12}+\bar{x}_{21}}{2}$$

$s_A^2 :=$ Variance of the sample consisting of values $\bar{x}_{A_1}$ & $\bar{x}_{A_2}$

$s_B^2 :=$ Variance of the sample consisting of values $\bar{x}_{B_1}$ & $\bar{x}_{B_2}$

$s_{AB}^2 :=$ Variance of the sample consisting of values $\bar{x}_{AB_1}$ & $\bar{x}_{AB_2}$

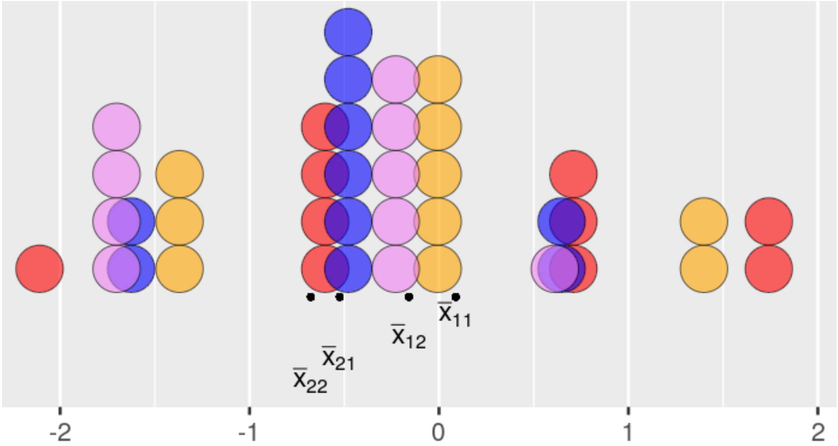


$s_A^2 / s_{within}^2 \ll 1 \implies$ Factor A clearly has no significant main effect!

$s_B^2 / s_{within}^2 \ll 1 \implies$ Factor B clearly has no significant main effect!

$s_{AB}^2 / s_{within}^2 \ll 1 \implies$ Factors A & B clearly have no interactive effect!

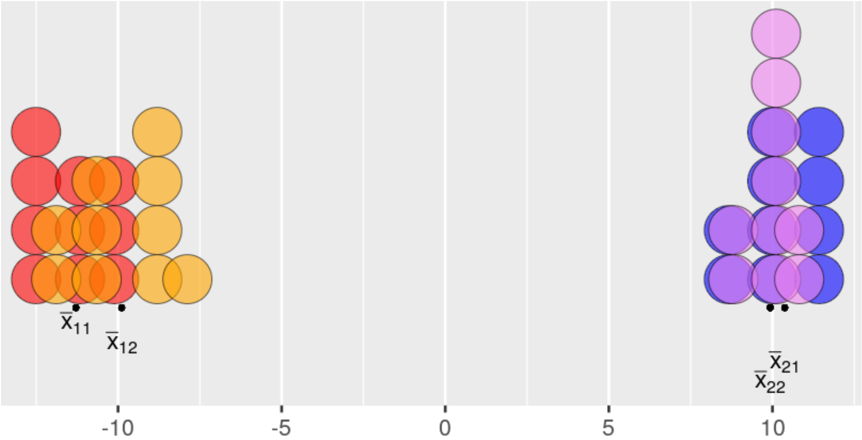
2F bcrANOVA (Motivation)



$s_A^2/s_{within}^2 \ll 1 \implies$ Factor A clearly has <u>no</u> significant main effect!

$s_B^2/s_{within}^2 \ll 1 \implies$ Factor B clearly has <u>no</u> significant main effect!

$s_{AB}^2/s_{within}^2 \ll 1 \implies$ Factors A & B clearly have <u>no</u> interactive effect!

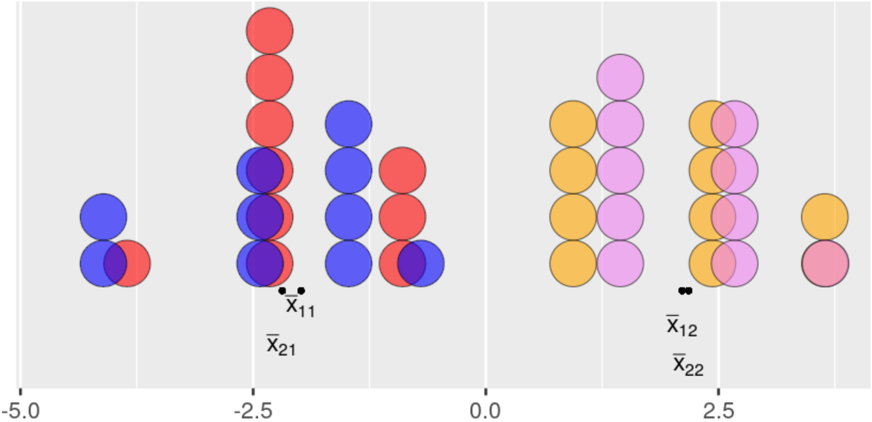


$s_A^2/s_{within}^2 \gg 1 \implies$ Factor A clearly has a significant main effect!

$s_B^2/s_{within}^2 \ll 1 \implies$ Factor B clearly has no significant main effect!

$s_{AB}^2/s_{within}^2 \ll 1 \implies$ Factors A & B clearly have no interactive effect!

2F bcrANOVA (Motivation)

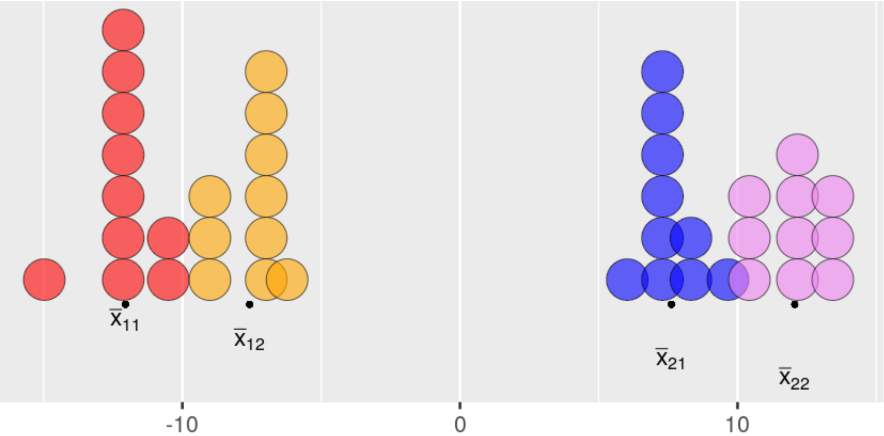


$s_A^2/s_{within}^2 \ll 1 \implies$ Factor A clearly has <u>no</u> significant main effect!

$s_B^2/s_{within}^2 \gg 1 \implies$ Factor B clearly <u>has</u> a significant main effect!

$s_{AB}^2/s_{within}^2 \ll 1 \implies$ Factors A & B clearly have <u>no</u> interactive effect!

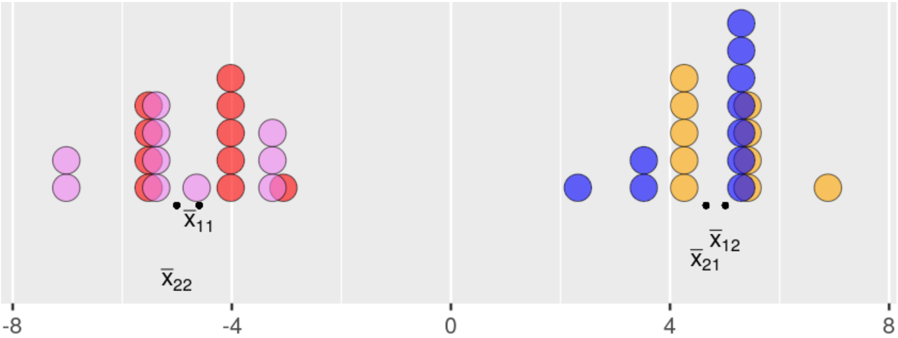
2F bcrANOVA (Motivation)



$s_A^2 / s_{within}^2 \gg 1 \implies$ Factor A clearly has a significant main effect!

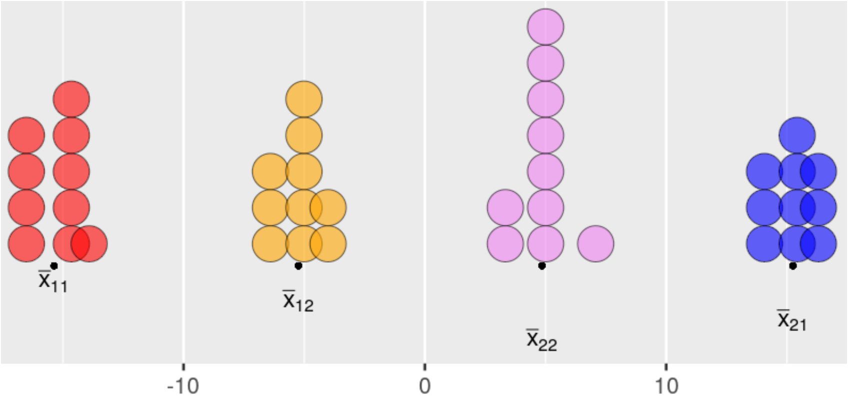
$s_B^2 / s_{within}^2 \gg 1 \implies$ Factor B clearly has a significant main effect!

$s_{AB}^2 / s_{within}^2 \ll 1 \implies$ Factors A & B clearly have no interactive effect!



s^2_A / s^2_{within}	$\ll$	1	$\Longrightarrow$	Factor A clearly has no significant main effect!
s^2_B / s^2_{within}	$\ll$	1	$\Longrightarrow$	Factor B clearly has no significant main effect!
s^2_{AB} / s^2_{within}	$\gg$	1	$\Longrightarrow$	Factors A & B clearly have an interactive effect!

2F bcrANOVA (Motivation)

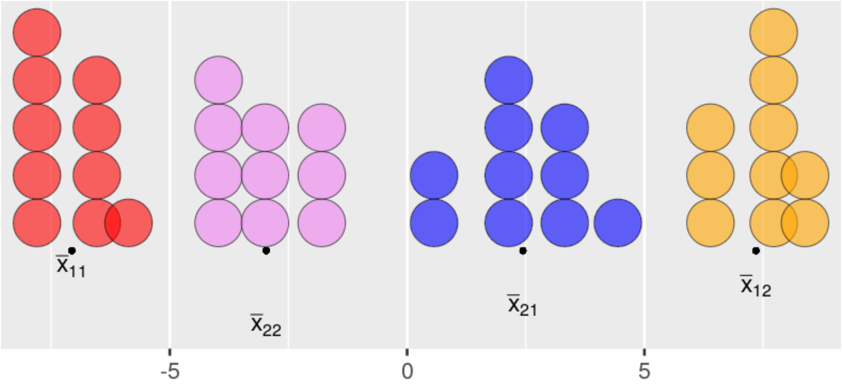


$$s_A^2/s_{within}^2 \gg 1 \implies$$ Factor A clearly <u>has</u> a significant main effect!

$$s_B^2/s_{within}^2 \ll 1 \implies$$ Factor B clearly has <u>no</u> significant main effect!

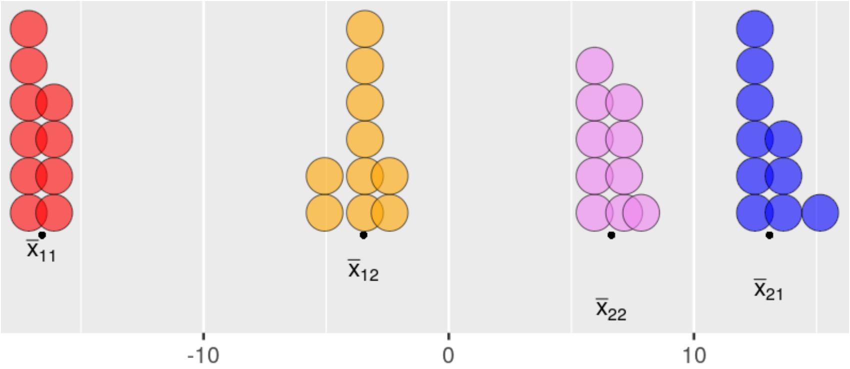
$$s_{AB}^2/s_{within}^2 \gg 1 \implies$$ Factors A & B clearly <u>have</u> an interactive effect!

2F bcrANOVA (Motivation)



s_A^2 / s_{within}^2	$\ll$	1	$\Longrightarrow$	Factor A clearly has no significant main effect!
s_B^2 / s_{within}^2	$\gg$	1	$\Longrightarrow$	Factor B clearly has a significant main effect!
s_{AB}^2 / s_{within}^2	$\gg$	1	$\Longrightarrow$	Factors A & B clearly have an interactive effect!

2F bcrANOVA (Motivation)



$s_A^2/s_{within}^2 \gg 1 \implies$ Factor A clearly <u>has</u> a significant main effect!

$s_B^2/s_{within}^2 \gg 1 \implies$ Factor B clearly <u>has</u> a significant main effect!

$s_{AB}^2/s_{within}^2 \gg 1 \implies$ Factors A & B clearly <u>have</u> an interactive effect!

PART IV:

2-FACTOR BALANCED COMPLETELY RANDOMIZED ANOVA (2F bcrANOVA)

2-Factor Completely Randomized Design

Fixed Effects Model Assumptions

Fixed Effects Linear Model

Sums of Squares Partitioning

F-Test Procedure

Expected Mean Squares

Point Estimators of σ^2

Effect Size Measures

Post-Hoc Comparisons

2-Factor Completely Randomized Design

An example **completely randomized design** entails the following:

- Collect 12 relevant experimental units (EU's): $\text{EU}_1, \text{EU}_2, \cdots, \text{EU}_{12}$
- Produce a random shuffle sequence using software:
  $(6, 10;\ 3, 1;\ 5, 8;\ 11, 9;\ 7, 2;\ 12, 4)$
- Use random shuffle sequence to assign the EU's into the IJ groups:

FACTOR B: $\rightarrow$ / **FACTOR A:** $\downarrow$	Level 1	Level 2	Level 3
Level 1	$\text{EU}_6,\ \text{EU}_{10}$	$\text{EU}_3,\ \text{EU}_1$	$\text{EU}_5,\ \text{EU}_8$
Level 2	$\text{EU}_{11},\ \text{EU}_9$	$\text{EU}_7,\ \text{EU}_2$	$\text{EU}_{12}, \text{EU}_4$

- Measure each EU appropriately (note the change in notation):

FACTOR B: $\rightarrow$ / **FACTOR A:** $\downarrow$	Level 1 $(x_{\bullet 1})$	Level 2 $(x_{\bullet 2})$	Level 3 $(x_{\bullet 3})$
Level 1 $(x_{1\bullet})$	$x_{111},\ x_{112}$	$x_{121},\ x_{122}$	$x_{131},\ x_{132}$
Level 2 $(x_{2\bullet})$	$x_{211},\ x_{212}$	$x_{221},\ x_{222}$	$x_{231},\ x_{232}$

$$\text{EU}_k \equiv \left(k^{th} \text{ experimental unit collected}\right)$$
$$x_{ijk} \equiv \left(\text{Measurement of } k^{th} \text{ EU in } (i,j)\text{-levels of factors A \& B}\right)$$

2F bcrANOVA Fixed Effects Model Assumptions

Proposition

(2F bcrANOVA Fixed Effects Model Assumptions)

- *(**2 Desired Factors**) Factor A has I levels & Factor B has J levels.*
- *(**All Factor Levels are Considered**) AKA Fixed Effects.*
- *(**Factors are Crossed**) IJ groups – one per (i,j)-level factor combination.*
- *(**Balanced Replication in Groups**) Each group has $K > 1$ units.*
- *(**Distinct Exp. Units**) All IJK units are distinct from each other.*

- *(**Random Assignment across Groups**)*

- *(**Independence**) All measurements on units are independent.*
- *(**Normality**) All groups are approximately normally distributed.*
- *(**Equal Variances**) All groups have approximately same variance.*

Mnemonic: ***2DF AFLaC FaC BRiG DEU** | **RAaG** | **I.N.EV***

2F bcrANOVA Linear Model (Fixed Effects)

2F bcrANOVA Fixed Effects Linear Model

(I, J)	$\equiv$	(# levels of factor A, # levels of factor B)
K	$\equiv$	# observations (replications) at each (i, j)-level of factors A & B
X_{ijk}	$\equiv$	rv for k^{th} observation at (i, j)-level of factors A & B
μ	$\equiv$	Mean average response over all levels of factors A & B
(α_i^A, α_j^B)	$\equiv$	(Effect of i^{th}-level factor A, Effect of j^{th}-level factor B)
γ_{ij}^{AB}	$\equiv$	Interaction between (i, j)-level factors A & B
E_{ijk}	$\equiv$	Deviation from μ due to random error

ASSUMPTIONS: $E_{ijk} \overset{iid}{\sim}$ Normal $\left(0, \sigma^2\right)$

$$X_{ijk} = \mu + \alpha_i^A + \alpha_j^B + \gamma_{ij}^{AB} + E_{ijk} \quad \text{where} \quad \begin{cases} \sum_i \alpha_i^A = \sum_j \alpha_j^B = 0 \\ \sum_i \gamma_{ij}^{AB} = \sum_j \gamma_{ij}^{AB} = 0 \end{cases}$$

H_0^A :	All	$\alpha_i^A = 0$	H_0^B :	All	$\alpha_j^B = 0$	H_0^{AB} :	All	$\gamma_{ij}^{AB} = 0$
H_A^A :	Some	$\alpha_i^A \neq 0$	H_A^B :	Some	$\alpha_j^B \neq 0$	H_A^{AB} :	Some	$\gamma_{ij}^{AB} \neq 0$

$X_{ijk} \overset{IND}{\sim} \ldots \equiv$ rv's X_{ijk} are independently distributed as ...

$E_{ijk} \overset{iid}{\sim} \ldots \equiv$ rv's E_{ijk} are independently and identically distributed as ...

Sums of Squares as a "Partitioning" of Variation Explanation for 2F bcrANOVA

$$\underbrace{\text{SS}_{total}}_{\substack{Total\ Variation\\ in\ Experiment}} = \underbrace{\text{SS}_{A}}_{\substack{Variation\ due\\ to\ Factor\ A}} + \underbrace{\text{SS}_{B}}_{\substack{Variation\ due\\ to\ Factor\ B}} + \underbrace{\text{SS}_{AB}}_{\substack{Variation\ due\\ to\ Interaction}} + \underbrace{\text{SS}_{res}}_{\substack{Unexplained\\ Variation}}$$

$$\sum_{ijk}(x_{ijk} - \hat{\mu})^2 = \sum_{ijk}(\hat{\alpha}_i^A)^2 + \sum_{ijk}(\hat{\alpha}_j^B)^2 + \sum_{ijk}(\hat{\gamma}_{ij}^{AB})^2 + \sum_{ijk}(x_{ijk}^{res})^2$$

$$\underbrace{\nu}_{\substack{Total\ dof's\ in\\ Experiment}} = \underbrace{\nu_A}_{\substack{Factor\ A\\ dof's}} + \underbrace{\nu_B}_{\substack{Factor\ B\\ dof's}} + \underbrace{\nu_{AB}}_{\substack{Interaction\\ dof's}} + \underbrace{\nu_{res}}_{\substack{'Within\ Groups'\\ dof's}}$$

$$\nu = IJK - 1, \quad \nu_A = I - 1, \quad \nu_B = J - 1, \quad \nu_{AB} = (I-1)(J-1), \quad \nu_{res} = IJ(K-1)$$

2F bcrANOVA F-Test

1. $\nu_A = I - 1,\ \nu_B = J - 1,\ \nu_{AB} = (I-1)(J-1),\ \nu_{res} = IJ(K-1)$

2. $\bar{x}_{i\bullet\bullet} := \frac{1}{JK}\sum_j\sum_k x_{ijk}, \quad \bar{x}_{\bullet j\bullet} := \frac{1}{IK}\sum_i\sum_k x_{ijk}, \quad \bar{x}_{ij\bullet} := \frac{1}{K}\sum_k x_{ijk}$
3. $\bar{x}_{\bullet\bullet\bullet} := \frac{1}{IJK}\sum_i\sum_j\sum_k x_{ijk}$

4. $$\begin{cases} \text{SS}_{res} & := & \sum_{ijk}(x_{ijk}^{res})^2 & = & \sum_i\sum_j\sum_k(x_{ijk} - \bar{x}_{ij\bullet})^2 \\ \text{SS}_{A} & := & \sum_{ijk}(\hat{\alpha}_i^A)^2 & = & \sum_i\sum_j\sum_k(\bar{x}_{i\bullet\bullet} - \bar{x}_{\bullet\bullet\bullet})^2 \\ \text{SS}_{B} & := & \sum_{ijk}(\hat{\alpha}_j^B)^2 & = & \sum_i\sum_j\sum_k(\bar{x}_{\bullet j\bullet} - \bar{x}_{\bullet\bullet\bullet})^2 \\ \text{SS}_{AB} & := & \sum_{ijk}(\hat{\gamma}_{ij}^{AB})^2 & = & \sum_i\sum_j\sum_k(\bar{x}_{ij\bullet} - \bar{x}_{i\bullet\bullet} - \bar{x}_{\bullet j\bullet} + \bar{x}_{\bullet\bullet\bullet})^2 \end{cases}$$

(Optional) $\text{SS}_{total} := \sum_{ijk}(x_{ijk} - \hat{\mu})^2 = \sum_i\sum_j\sum_k(x_{ijk} - \bar{x}_{\bullet\bullet\bullet})^2$

2F bcrANOVA F-Test

5. $\text{MS}_A = \dfrac{\text{SS}_A}{\nu_A}, \quad \text{MS}_B = \dfrac{\text{SS}_B}{\nu_B}, \quad \text{MS}_{AB} = \dfrac{\text{SS}_{AB}}{\nu_{AB}}, \quad \text{MS}_{res} = \dfrac{\text{SS}_{res}}{\nu_{res}}$

6. $f_A = \dfrac{\text{MS}_A}{\text{MS}_{res}}, \quad f_B = \dfrac{\text{MS}_B}{\text{MS}_{res}}, \quad f_{AB} = \dfrac{\text{MS}_{AB}}{\text{MS}_{res}}$

7. (if using software): $\begin{cases} p_A & := & \mathbb{P}(F > f_A) & \approx & 1 - \Phi_F(f_A; \nu_A, \nu_{res}) \\ p_B & := & \mathbb{P}(F > f_B) & \approx & 1 - \Phi_F(f_B; \nu_B, \nu_{res}) \\ p_{AB} & := & \mathbb{P}(F > f_{AB}) & \approx & 1 - \Phi_F(f_{AB}; \nu_{AB}, \nu_{res}) \end{cases}$

8.

$$\begin{cases} \text{If} & p_A \leq \alpha & \text{or} & f_A > f^*_{\nu_A, \nu_{res}; \alpha} & \text{then reject} & H_0^A & \text{else accept} & H_0^A \\ \text{If} & p_B \leq \alpha & \text{or} & f_B > f^*_{\nu_B, \nu_{res}; \alpha} & \text{then reject} & H_0^B & \text{else accept} & H_0^B \\ \text{If} & p_{AB} \leq \alpha & \text{or} & f_{AB} > f^*_{\nu_{AB}, \nu_{res}; \alpha} & \text{then reject} & H_0^{AB} & \text{else accept} & H_0^{AB} \end{cases}$$

2F bcrANOVA Table

2F bcrANOVA Table $(\text{Significance Level } \alpha)$						
Variation Source	**df**	**Sum of Squares**	**Mean Square**	**F Stat Value**	**P-value**	**Decision**
A	ν_A	SS_A	MS_A	f_A	p_A	Acc/Rej H_0^A
B	ν_B	SS_B	MS_B	f_B	p_B	Acc/Rej H_0^B
AB	ν_{AB}	SS_{AB}	MS_{AB}	f_{AB}	p_{AB}	Acc/Rej H_0^{AB}
Unknown	ν_{res}	SS_{res}	MS_{res}			
Total	ν	SS_{total}				

Proposition

Given 2-factor experiment satisfying the 2F bcrANOVA assumptions. Then:

$$
\begin{aligned}
(i) \quad \mathbb{E}[MS_{res}] &= \sigma^2 \\
(ii) \quad \mathbb{E}[MS_A] &= \sigma^2 + \frac{JK}{I-1} \sum_i (\alpha_i^A)^2 \\
(iii) \quad \mathbb{E}[MS_B] &= \sigma^2 + \frac{IK}{J-1} \sum_j (\alpha_j^B)^2 \\
(iv) \quad \mathbb{E}[MS_{AB}] &= \sigma^2 + \frac{K}{(I-1)(J-1)} \sum_i \sum_j (\gamma_{ij}^{AB})^2
\end{aligned}
$$

2F bcrANOVA Expected Mean Squares: Proof of (i)

$$
\begin{aligned}
\mathbb{E}[\mathsf{SS}_{res}] &:= \mathbb{E}\left[\sum_{ijk}(X^{res}_{ijk})^2\right] \\
&= \mathbb{E}\left[\sum_{ijk}(X_{ijk} - \hat{X}_{ijk})^2\right] \\
&= \mathbb{E}\left[\sum_{ijk}\left(X_{ijk} - (\hat{\mu} + \hat{\alpha}^A_i + \hat{\alpha}^B_j + \hat{\gamma}^{AB}_{ij})\right)^2\right] \\
&\overset{BLUE}{=} \mathbb{E}\left[\sum_{ijk}(X_{ijk} - \overline{X}_{ij\bullet})^2\right] \\
&\overset{CIO}{=} \tfrac{K-1}{K-1} \cdot \mathbb{E}\left[\sum_i \sum_j \sum_k (X_{ijk} - \overline{X}_{ij\bullet})^2\right] \\
&= (K-1) \cdot \sum_i \sum_j \mathbb{E}\left[\tfrac{1}{K-1} \sum_k (X_{ijk} - \overline{X}_{ij\bullet})^2\right] \\
&= (K-1) \cdot \sum_i \sum_j \mathbb{E}\left[S^2_{ij}\right] \\
&= (K-1) \cdot \sum_i \sum_j \sigma^2 \\
&= IJ(K-1)\sigma^2
\end{aligned}
$$

$$
\Longrightarrow \mathbb{E}\left[\mathsf{MS}_{res}\right] := \mathbb{E}\left[\frac{\mathsf{SS}_{res}}{\nu_{res}}\right] = \frac{\mathbb{E}[\mathsf{SS}_{res}]}{IJ(K-1)} = \frac{IJ(K-1)\sigma^2}{IJ(K-1)} = \sigma^2 \qquad \square
$$

CIO $\equiv$ "**C**lever **I**nsertion of **O**ne"

$S^2_{ij} \equiv$ Variance of (i,j)-level group

2F bcrANOVA Expected Mean Squares: Proof of (ii)

$$\begin{aligned}
\text{Given} \quad X_{ijk} &= \mu + \alpha_i^A + \alpha_j^B + \gamma_{ij}^{AB} + E_{ijk} \quad \text{s.t.} \quad E_{ijk} \overset{IND}{\sim} \text{Normal}(0, \sigma^2) \\
\implies \quad \overline{X}_{i\bullet\bullet} &= \mu + \alpha_i^A + \overline{E}_{i\bullet\bullet} \quad \overset{CLT}{\implies} \quad \overline{E}_{i\bullet\bullet} \overset{IND}{\sim} \text{Normal}(0, \tfrac{\sigma^2}{JK}) \\
\implies \quad \overline{X}_{\bullet\bullet\bullet} &= \mu + \overline{E}_{\bullet\bullet\bullet} \quad \overset{CLT}{\implies} \quad \overline{E}_{\bullet\bullet\bullet} \sim \text{Normal}(0, \tfrac{\sigma^2}{IJK})
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[\text{SS}_A] &:= \mathbb{E}\left[\textstyle\sum_{ijk}(\hat{\alpha}_i^A)^2\right] \overset{BLUE}{=} \textstyle\sum_{ijk} \mathbb{E}\left[(\overline{X}_{i\bullet\bullet} - \overline{X}_{\bullet\bullet\bullet})^2\right] \\
&= \textstyle\sum_{ijk} \mathbb{E}\left[(\alpha_i^A + \overline{E}_{i\bullet\bullet} - \overline{E}_{\bullet\bullet\bullet})^2\right] \\
&\overset{(1)}{=} JK \cdot \textstyle\sum_i \mathbb{E}\left[(\alpha_i^A)^2\right] + JK \cdot \textstyle\sum_i \mathbb{E}\left[(\overline{E}_{i\bullet\bullet})^2 - 2(\overline{E}_{i\bullet\bullet}\overline{E}_{\bullet\bullet\bullet}) + (\overline{E}_{\bullet\bullet\bullet})^2\right] \\
&\overset{(2)}{=} JK \cdot \textstyle\sum_i (\alpha_i^A)^2 + JK \cdot \textstyle\sum_i \mathbb{E}\left[(\overline{E}_{i\bullet\bullet})^2\right] + \mathbb{E}\left[-IJK(\overline{E}_{\bullet\bullet\bullet})^2\right] \\
&\overset{(3)}{=} JK \cdot \textstyle\sum_i (\alpha_i^A)^2 + JK \cdot \textstyle\sum_i \left[(0)^2 + \frac{\sigma^2}{JK}\right] - IJK \cdot \mathbb{E}\left[(\overline{E}_{\bullet\bullet\bullet})^2\right] \\
&\overset{(3)}{=} JK \cdot \textstyle\sum_i (\alpha_i^A)^2 + I\sigma^2 - IJK \cdot \left((0)^2 + \frac{\sigma^2}{IJK}\right) \\
&= JK \cdot \textstyle\sum_i (\alpha_i^A)^2 + (I-1)\sigma^2
\end{aligned}$$

$$\therefore \; \mathbb{E}[\text{MS}_A] := \mathbb{E}\left[\frac{\text{SS}_A}{\nu_A}\right] = \frac{\mathbb{E}[\text{SS}_A]}{I-1} = \sigma^2 + \frac{JK}{I-1} \cdot \textstyle\sum_i (\alpha_i^A)^2 \qquad \square$$

(1) $\sum_i(\overline{E}_{i\bullet\bullet} - \overline{E}_{\bullet\bullet\bullet}) = 0$, (2) $\sum_i \overline{E}_{i\bullet\bullet} = I \cdot \overline{E}_{\bullet\bullet\bullet}$, (3) $\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$

2F bcrANOVA Expected Mean Squares: Proof of (iii)

$$\begin{aligned}
\text{Given} \quad X_{ijk} &= \mu + \alpha_i^A + \alpha_j^B + \gamma_{ij}^{AB} + E_{ijk} \quad \text{s.t.} \quad E_{ijk} \overset{IND}{\sim} \text{Normal}(0, \sigma^2) \\
\Longrightarrow \quad \overline{X}_{\bullet j \bullet} &= \mu + \alpha_j^B + \overline{E}_{\bullet j \bullet} \quad \overset{CLT}{\Longrightarrow} \quad \overline{E}_{\bullet j \bullet} \overset{IND}{\sim} \text{Normal}(0, \tfrac{\sigma^2}{IK}) \\
\Longrightarrow \quad \overline{X}_{\bullet\bullet\bullet} &= \mu + \overline{E}_{\bullet\bullet\bullet} \quad \overset{CLT}{\Longrightarrow} \quad \overline{E}_{\bullet\bullet\bullet} \sim \text{Normal}(0, \tfrac{\sigma^2}{IJK})
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[\text{SS}_B] &:= \mathbb{E}\left[\textstyle\sum_{ijk} (\hat{\alpha}_j^B)^2\right] \overset{BLUE}{=} \textstyle\sum_{ijk} \mathbb{E}\left[(\overline{X}_{\bullet j \bullet} - \overline{X}_{\bullet\bullet\bullet})^2\right] \\
&= \textstyle\sum_{ijk} \mathbb{E}\left[(\alpha_j^B + \overline{E}_{\bullet j \bullet} - \overline{E}_{\bullet\bullet\bullet})^2\right] \\
&\overset{(1)}{=} IK \cdot \textstyle\sum_j \mathbb{E}\left[(\alpha_j^B)^2\right] + IK \cdot \textstyle\sum_j \mathbb{E}\left[(\overline{E}_{\bullet j \bullet})^2 - 2(\overline{E}_{\bullet j \bullet}\overline{E}_{\bullet\bullet\bullet}) + (\overline{E}_{\bullet\bullet\bullet})^2\right] \\
&\overset{(2)}{=} IK \cdot \textstyle\sum_j (\alpha_j^B)^2 + IK \cdot \textstyle\sum_j \mathbb{E}\left[(\overline{E}_{\bullet j \bullet})^2\right] + \mathbb{E}\left[-IJK(\overline{E}_{\bullet\bullet\bullet})^2\right] \\
&\overset{(3)}{=} IK \cdot \textstyle\sum_j (\alpha_j^B)^2 + IK \cdot \textstyle\sum_j \left[(0)^2 + \frac{\sigma^2}{IK}\right] - IJK \cdot \mathbb{E}\left[(\overline{E}_{\bullet\bullet\bullet})^2\right] \\
&\overset{(3)}{=} IK \cdot \textstyle\sum_j (\alpha_j^B)^2 + J\sigma^2 - IJK \cdot \left((0)^2 + \frac{\sigma^2}{IJK}\right) \\
&= IK \cdot \textstyle\sum_j (\alpha_j^B)^2 + (J-1)\sigma^2
\end{aligned}$$

$$\therefore \ \mathbb{E}[\text{MS}_B] := \mathbb{E}\left[\frac{\text{SS}_B}{\nu_B}\right] = \frac{\mathbb{E}[\text{SS}_B]}{J-1} = \sigma^2 + \frac{IK}{J-1} \cdot \textstyle\sum_j (\alpha_j^B)^2 \qquad \square$$

$(1)\ \sum_j (\overline{E}_{\bullet j \bullet} - \overline{E}_{\bullet\bullet\bullet}) = 0$, $\quad (2)\ \sum_j \overline{E}_{\bullet j \bullet} = J \cdot \overline{E}_{\bullet\bullet\bullet}$, $\quad (3)\ \mathbb{V}[X] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2$

2F bcrANOVA Expected Mean Squares: Proof of (iv)

$$\text{Given} \quad X_{ijk} = \mu + \alpha_i^A + \alpha_j^B + \gamma_{ij}^{AB} + E_{ijk} \quad \text{s.t.} \quad E_{ijk} \overset{IND}{\sim} \text{Normal}(0, \sigma^2)$$

$$\Longrightarrow \quad \overline{X}_{ij\bullet} = \mu + \alpha_i^A + \alpha_j^B + \gamma_{ij}^{AB} + \overline{E}_{ij\bullet} \quad \overset{CLT}{\Longrightarrow} \quad \overline{E}_{ij\bullet} \overset{IND}{\sim} \text{Normal}(0, \tfrac{\sigma^2}{K})$$

$$\Longrightarrow \quad \overline{X}_{\bullet\bullet\bullet} = \mu + \overline{E}_{\bullet\bullet\bullet} \quad \overset{CLT}{\Longrightarrow} \quad \overline{E}_{\bullet\bullet\bullet} \sim \text{Normal}(0, \tfrac{\sigma^2}{IJK})$$

$$\mathbb{E}[\text{SS}_{AB}] := \mathbb{E}\left[\sum_{ijk} (\hat{\gamma}_{ij}^{AB})^2\right] \overset{BLUE}{=} \sum_{ijk} \mathbb{E}\left[(\overline{X}_{ij\bullet} - \overline{X}_{i\bullet\bullet} - \overline{X}_{\bullet j\bullet} + \overline{X}_{\bullet\bullet\bullet})^2\right]$$

$$= \sum_{ijk} \mathbb{E}\left[(\gamma_{ij}^{AB} + \overline{E}_{ij\bullet} - \overline{E}_{i\bullet\bullet} - \overline{E}_{\bullet j\bullet} + \overline{E}_{\bullet\bullet\bullet})^2\right]$$

$$\overset{(*)}{=} K \cdot \sum_{ij} (\gamma_{ij}^{AB})^2 + K \cdot \sum_{ij} \mathbb{E}\left[(\overline{E}_{ij\bullet} - \overline{E}_{i\bullet\bullet} - \overline{E}_{\bullet j\bullet} + \overline{E}_{\bullet\bullet\bullet})^2\right]$$

$$\overset{(\clubsuit)}{=} K \cdot \sum_{ij} (\gamma_{ij}^{AB})^2 + K \cdot \sum_{ij} \mathbb{E}\left[(\overline{E}_{ij\bullet})^2 + (\overline{E}_{i\bullet\bullet})^2 + (\overline{E}_{\bullet j\bullet})^2 + (\overline{E}_{\bullet\bullet\bullet})^2\right]$$

$$+ K \cdot \sum_{ij} \mathbb{E}\left[-2(\overline{E}_{ij\bullet})(\overline{E}_{i\bullet\bullet}) - 2(\overline{E}_{ij\bullet})(\overline{E}_{\bullet j\bullet}) + 2(\overline{E}_{ij\bullet})(\overline{E}_{\bullet\bullet\bullet})\right]$$

$$+ K \cdot \sum_{ij} \mathbb{E}\left[2(\overline{E}_{i\bullet\bullet})(\overline{E}_{\bullet j\bullet}) - 2(\overline{E}_{i\bullet\bullet})(\overline{E}_{\bullet\bullet\bullet}) - 2(\overline{E}_{\bullet j\bullet})(\overline{E}_{\bullet\bullet\bullet})\right]$$

$(*)$ $\sum_{ij} (\overline{E}_{ij\bullet} - \overline{E}_{i\bullet\bullet} - \overline{E}_{\bullet j\bullet} + \overline{E}_{\bullet\bullet\bullet}) = IJ \cdot \overline{E}_{\bullet\bullet\bullet} - IJ \cdot \overline{E}_{\bullet\bullet\bullet} - IJ \cdot \overline{E}_{\bullet\bullet\bullet} + IJ \cdot \overline{E}_{\bullet\bullet\bullet} = 0$

$(\clubsuit)$ $(w - x - y + z)^2 = w^2 + x^2 + y^2 + z^2 - 2wx - 2wy + 2wz + 2xy - 2xz - 2yz$

2F bcrANOVA Expected Mean Squares: Proof of (iv)

$$
\begin{array}{lllll}
\text{Given} & X_{ijk} = \mu + \alpha_i^A + \alpha_j^B + \gamma_{ij}^{AB} + E_{ijk} & \text{s.t.} & E_{ijk} \overset{IND}{\sim} \text{Normal}(0, \sigma^2) \\
\Longrightarrow & \overline{X}_{ij\bullet} = \mu + \alpha_i^A + \alpha_j^B + \gamma_{ij}^{AB} + \overline{E}_{ij\bullet} & \overset{CLT}{\Longrightarrow} & \overline{E}_{ij\bullet} \overset{IND}{\sim} \text{Normal}(0, \frac{\sigma^2}{K}) \\
\Longrightarrow & \overline{X}_{i\bullet\bullet} = \mu + \alpha_i^A + \overline{E}_{i\bullet\bullet} & \overset{CLT}{\Longrightarrow} & \overline{E}_{i\bullet\bullet} \overset{IND}{\sim} \text{Normal}(0, \frac{\sigma^2}{JK}) \\
\Longrightarrow & \overline{X}_{\bullet j\bullet} = \mu + \alpha_j^B + \overline{E}_{\bullet j\bullet} & \overset{CLT}{\Longrightarrow} & \overline{E}_{\bullet j\bullet} \overset{IND}{\sim} \text{Normal}(0, \frac{\sigma^2}{IK}) \\
\Longrightarrow & \overline{X}_{\bullet\bullet\bullet} = \mu + \overline{E}_{\bullet\bullet\bullet} & \overset{CLT}{\Longrightarrow} & \overline{E}_{\bullet\bullet\bullet} \sim \text{Normal}(0, \frac{\sigma^2}{IJK})
\end{array}
$$

$$
\begin{aligned}
\mathbb{E}[\text{SS}_{AB}] &= K \cdot \textstyle\sum_{ij} (\gamma_{ij}^{AB})^2 + K \cdot \sum_{ij} \mathbb{E}\left[(\overline{E}_{ij\bullet})^2 + (\overline{E}_{i\bullet\bullet})^2 + (\overline{E}_{\bullet j\bullet})^2 + (\overline{E}_{\bullet\bullet\bullet})^2\right] \\
&\quad + K \cdot \textstyle\sum_{ij} \mathbb{E}\left[-2(\overline{E}_{ij\bullet})(\overline{E}_{i\bullet\bullet}) - 2(\overline{E}_{ij\bullet})(\overline{E}_{\bullet j\bullet}) + 2(\overline{E}_{ij\bullet})(\overline{E}_{\bullet\bullet\bullet})\right] \\
&\quad + K \cdot \textstyle\sum_{ij} \mathbb{E}\left[2(\overline{E}_{i\bullet\bullet})(\overline{E}_{\bullet j\bullet}) - 2(\overline{E}_{i\bullet\bullet})(\overline{E}_{\bullet\bullet\bullet}) - 2(\overline{E}_{\bullet j\bullet})(\overline{E}_{\bullet\bullet\bullet})\right] \\
&\overset{(\spadesuit)}{=} K \cdot \textstyle\sum_{ij} (\gamma_{ij}^{AB})^2 + K \cdot \sum_{ij} \left(\frac{\sigma^2}{K} + \frac{\sigma^2}{JK} + \frac{\sigma^2}{IK} + \frac{\sigma^2}{IJK}\right) \\
&\quad + K \cdot \left(-2 \cdot \frac{I\sigma^2}{K} - 2 \cdot \frac{J\sigma^2}{K} + 2 \cdot \frac{\sigma^2}{K}\right) + K \cdot \left(2 \cdot \frac{\sigma^2}{K} - 2 \cdot \frac{\sigma^2}{K} - 2 \cdot \frac{\sigma^2}{K}\right)
\end{aligned}
$$

$$(\spadesuit)\ \textstyle\sum_{ij} \mathbb{E}\left[(\overline{E}_{i\bullet\bullet})(\overline{E}_{\bullet j\bullet})\right] = \mathbb{E}\left[\sum_i (\overline{E}_{i\bullet\bullet}) \cdot \sum_j (\overline{E}_{\bullet j\bullet})\right] = IJ \cdot \mathbb{E}\left[(\overline{E}_{\bullet\bullet\bullet})^2\right] = \frac{\sigma^2}{K}$$

$$\textstyle\sum_{ij} \mathbb{E}\left[(\overline{E}_{ij\bullet})(\overline{E}_{i\bullet\bullet})\right] = \mathbb{E}\left[\sum_i ((\overline{E}_{i\bullet\bullet}) \cdot \sum_j (\overline{E}_{ij\bullet}))\right] = J \cdot \sum_i \mathbb{E}\left[(\overline{E}_{i\bullet\bullet})^2\right] = IJ \cdot \frac{\sigma^2}{JK} = \frac{I\sigma^2}{K}$$

$$\textstyle\sum_{ij} \mathbb{E}\left[(\overline{E}_{ij\bullet})(\overline{E}_{\bullet j\bullet})\right] = \mathbb{E}\left[\sum_j ((\overline{E}_{\bullet j\bullet}) \cdot \sum_i (\overline{E}_{ij\bullet}))\right] = I \cdot \sum_j \mathbb{E}\left[(\overline{E}_{\bullet j\bullet})^2\right] = IJ \cdot \frac{\sigma^2}{IK} = \frac{J\sigma^2}{K}$$

2F bcrANOVA Expected Mean Squares: Proof of (iv)

$$
\begin{aligned}
\mathbb{E}[\mathrm{SS}_{AB}] &= K \cdot \textstyle\sum_{ij}(\gamma_{ij}^{AB})^2 + K \cdot \sum_{ij}\left(\frac{\sigma^2}{K} + \frac{\sigma^2}{JK} + \frac{\sigma^2}{IK} + \frac{\sigma^2}{IJK}\right) \\
&\quad + K \cdot \left(-2 \cdot \frac{I\sigma^2}{K} - 2 \cdot \frac{J\sigma^2}{K} + 2 \cdot \frac{\sigma^2}{K}\right) + \ K \cdot \left(2 \cdot \frac{\sigma^2}{K} - 2 \cdot \frac{\sigma^2}{K} - 2 \cdot \frac{\sigma^2}{K}\right) \\
&= K \cdot \textstyle\sum_{ij}(\gamma_{ij}^{AB})^2 + IJ\sigma^2 + I\sigma^2 + J\sigma^2 + \sigma^2 \\
&\quad - 2(I\sigma^2) - 2(J\sigma^2) + 2\sigma^2 + 2\sigma^2 - 2\sigma^2 - 2\sigma^2 \\
&= IJ\sigma^2 - I\sigma^2 - J\sigma^2 + \sigma^2 + K \cdot \textstyle\sum_{ij}(\gamma_{ij}^{AB})^2 \\
&= I(J-1)\sigma^2 - (J-1)\sigma^2 + K \cdot \textstyle\sum_{ij}(\gamma_{ij}^{AB})^2 \\
&= (I-1)(J-1)\sigma^2 + K \cdot \textstyle\sum_{ij}(\gamma_{ij}^{AB})^2
\end{aligned}
$$

$$
\therefore\ \mathbb{E}\left[\mathrm{MS}_{AB}\right] := \mathbb{E}\left[\frac{\mathrm{SS}_{AB}}{\nu_{AB}}\right] = \frac{\mathbb{E}[\mathrm{SS}_{AB}]}{(I-1)(J-1)} = \sigma^2 + \frac{K}{(I-1)(J-1)} \cdot \textstyle\sum_{ij}(\gamma_{ij}^{AB})^2 \qquad \square
$$

Mean Squares as Point Estimators of σ^2

Proposition

(Point Estimation of Mean Squares)

Given a 2F balanced exp. satisfying the 2F bcrANOVA assumptions. Then:

$$\begin{array}{llllll} (i) & \text{Regardless of the truthness of } H_0^A, H_0^B, H_0^{AB} & & & \Longrightarrow & \mathbb{E}[MS_{res}] = \sigma^2 \\ (ii) & H_0^A \text{ is true} & \Longrightarrow & \mathbb{E}[MS_A] = \sigma^2, \quad H_0^A \text{ is false} & \Longrightarrow & \mathbb{E}[MS_A] > \sigma^2 \\ (iii) & H_0^B \text{ is true} & \Longrightarrow & \mathbb{E}[MS_B] = \sigma^2, \quad H_0^B \text{ is false} & \Longrightarrow & \mathbb{E}[MS_B] > \sigma^2 \\ (iv) & H_0^{AB} \text{ is true} & \Longrightarrow & \mathbb{E}[MS_{AB}] = \sigma^2, \quad H_0^{AB} \text{ is false} & \Longrightarrow & \mathbb{E}[MS_{AB}] > \sigma^2 \end{array}$$

PROOF:

(i) Follows immediately from the Expected Mean Squares proposition.

MS_A as Point Estimator of σ^2: Proof of (ii)

Recall from the Expected Mean Squares proposition that

$$\mathbb{E}[\text{MS}_A] = \sigma^2 + \frac{JK}{I-1} \cdot \sum_i (\alpha_i^A)^2$$

Then:

$$\begin{aligned} H_0^A \text{ is true} \quad &\Longrightarrow \quad \alpha_1^A = \alpha_2^A = \cdots = \alpha_I^A = 0 \\ &\Longrightarrow \quad \textstyle\sum_i (\alpha_i^A)^2 = 0 \\ &\Longrightarrow \quad \mathbb{E}[\text{MS}_A] = \sigma^2 \end{aligned}$$

$$\begin{aligned} H_0^A \text{ is false} \quad &\Longrightarrow \quad \text{At least two of the } \alpha^A\text{'s } \neq 0 \\ &\Longrightarrow \quad \textstyle\sum_i (\alpha_i^A)^2 > 0 \\ &\Longrightarrow \quad \mathbb{E}[\text{MS}_A] > \sigma^2 \qquad \square \end{aligned}$$

MS_B as Point Estimator of σ^2: Proof of (iii)

Recall from the Expected Mean Squares proposition that

$$\mathbb{E}[\text{MS}_B] = \sigma^2 + \frac{IK}{J-1} \cdot \sum_j (\alpha_j^B)^2$$

Then:

$$\begin{aligned} H_0^B \text{ is true} \quad &\implies \quad \alpha_1^B = \alpha_2^B = \cdots = \alpha_J^B = 0 \\ &\implies \quad \sum_j (\alpha_j^B)^2 = 0 \\ &\implies \quad \mathbb{E}[\text{MS}_B] = \sigma^2 \end{aligned}$$

$$\begin{aligned} H_0^B \text{ is false} \quad &\implies \quad \text{At least two of the } \alpha^B\text{'s } \neq 0 \\ &\implies \quad \sum_j (\alpha_j^B)^2 > 0 \\ &\implies \quad \mathbb{E}[\text{MS}_B] > \sigma^2 \qquad \square \end{aligned}$$

MS_{AB} as Point Estimator of σ^2: Proof of (iv)

Recall from the Expected Mean Squares proposition that

$$\mathbb{E}[\mathsf{MS}_{AB}] = \sigma^2 + \frac{K}{(I-1)(J-1)} \cdot \textstyle\sum_{ij}(\gamma_{ij}^{AB})^2$$

Then:

$$\begin{aligned} H_0^{AB} \text{ is true} \quad &\Longrightarrow \quad \gamma_{11}^{AB} = \gamma_{12}^{AB} = \cdots = \gamma_{1J}^{AB} = \gamma_{21}^{AB} = \cdots = \gamma_{IJ}^{AB} = 0 \\ &\Longrightarrow \quad \textstyle\sum_{ij}(\gamma_{ij}^{AB})^2 = 0 \\ &\Longrightarrow \quad \mathbb{E}[\mathsf{MS}_{AB}] = \sigma^2 \end{aligned}$$

$$\begin{aligned} H_0^{AB} \text{ is false} \quad &\Longrightarrow \quad \text{At least two of the } \gamma^{AB}\text{'s} \neq 0 \\ &\Longrightarrow \quad \textstyle\sum_{ij}(\gamma_{ij}^{AB})^2 > 0 \\ &\Longrightarrow \quad \mathbb{E}[\mathsf{MS}_{AB}] > \sigma^2 \qquad \square \end{aligned}$$

2F bcrANOVA (Effect Size Measures)

YEAR	NAME	MEASURE
1925[†]	Fisher (Eta-Squared)	$\hat{\eta}^2_A := \frac{\text{SS}_A}{\text{SS}_{total}} = \frac{\nu_A f_A}{\nu_A f_A + \nu_B f_B + \nu_{AB} f_{AB} + \nu_{res}}$ $\hat{\eta}^2_B := \frac{\text{SS}_B}{\text{SS}_{total}} = \frac{\nu_B f_B}{\nu_A f_A + \nu_B f_B + \nu_{AB} f_{AB} + \nu_{res}}$ $\hat{\eta}^2_{AB} := \frac{\text{SS}_{AB}}{\text{SS}_{total}} = \frac{\nu_{AB} f_{AB}}{\nu_A f_A + \nu_B f_B + \nu_{AB} f_{AB} + \nu_{res}}$ $\hat{\eta}^2_{res} := \frac{\text{SS}_{res}}{\text{SS}_{total}}$
1965[‡]	Cohen[♠♣] (Partial η^2)	$\hat{\eta}^2_{(A)} := \frac{\text{SS}_A}{\text{SS}_A + \text{SS}_{res}} = \frac{\nu_A f_A}{\nu_A f_A + \nu_{res}}$ $\hat{\eta}^2_{(B)} := \frac{\text{SS}_B}{\text{SS}_B + \text{SS}_{res}} = \frac{\nu_B f_B}{\nu_B f_B + \nu_{res}}$ $\hat{\eta}^2_{(AB)} := \frac{\text{SS}_{AB}}{\text{SS}_{AB} + \text{SS}_{res}} = \frac{\nu_{AB} f_{AB}}{\nu_{AB} f_{AB} + \nu_{res}}$

$$\hat{\eta}^2_A + \hat{\eta}^2_B + \hat{\eta}^2_{AB} + \hat{\eta}^2_{res} = 1 \quad \text{but} \quad \hat{\eta}^2_{(A)} + \hat{\eta}^2_{(B)} + \hat{\eta}^2_{(AB)} > 1$$

[†] R.A. Fisher, *Statistical Methods for Reasearch Workers*, 1925.
[‡] B.B. Wolman (Ed.), *Handbook of Clinical Psychology*, 1965. (§5 by J. Cohen)
[♠] F.J. Gravetter, L.B. Wallnau, *Statistics for the Behavioral Sciences*, 7^{th} Ed., 2007.
[♣] R.G. Lomax, D.L. Hahs-Vaughn, *Statistical Concepts: A* 2^{nd} *Course*, 4^{th} Ed., 2012.

EFFECT SIZE VALUE:	INTERPRETATION:
$\hat{\eta}_A^2 := \frac{SS_A}{SS_A+SS_B+SS_{AB}+SS_{res}} = 0.38$	38% of the variation in the reponse is due to Factor A
$\hat{\eta}_B^2 := \frac{SS_B}{SS_A+SS_B+SS_{AB}+SS_{res}} = 0.02$	2% of the variation in the reponse is due to Factor B
$\hat{\eta}_{AB}^2 := \frac{SS_{AB}}{SS_A+SS_B+SS_{AB}+SS_{res}} = 0.27$	27% of the variation in the reponse is due to Interaction AB
$\hat{\eta}_{res}^2 := \frac{SS_{res}}{SS_A+SS_B+SS_{AB}+SS_{res}} = 0.33$	33% of the variation in the reponse is unexplained with experiment
$\hat{\eta}_{(A)}^2 := \frac{SS_A}{SS_A+SS_{res}} = 0.43$	43% of the variation possibly due to A is actually due to A
$\hat{\eta}_{(B)}^2 := \frac{SS_B}{SS_B+SS_{res}} = 0.65$	65% of the variation possibly due to B is actually due to B
$\hat{\eta}_{(AB)}^2 := \frac{SS_{AB}}{SS_{AB}+SS_{res}} = 0.31$	31% of the variation possibly due to AB is actually due to AB

2F bcrANOVA (More Effect Size Measures)

YEAR	NAME	MEASURE
1963[†]	Hays (Omega-Squared)	$\hat{\omega}^2_A := \frac{\text{SS}_A - \nu_A \text{MS}_{res}}{\text{SS}_{total} + \text{MS}_{res}} = \frac{\nu_A f_A - \nu_A}{\nu_A f_A + \nu_B f_B + \nu_{AB} f_{AB} + n}$ $\hat{\omega}^2_B := \frac{\text{SS}_B - \nu_B \text{MS}_{res}}{\text{SS}_{total} + \text{MS}_{res}} = \frac{\nu_B f_B - \nu_B}{\nu_A f_A + \nu_B f_B + \nu_{AB} f_{AB} + n}$ $\hat{\omega}^2_{AB} := \frac{\text{SS}_{AB} - \nu_{AB} \text{MS}_{res}}{\text{SS}_{total} + \text{MS}_{res}} = \frac{\nu_{AB} f_{AB} - \nu_{AB}}{\nu_A f_A + \nu_B f_B + \nu_{AB} f_{AB} + n}$
1979[‡]	Keren-Lewis (Partial ω^2)	$\hat{\omega}^2_{(A)} := \frac{\text{SS}_A - \nu_A \text{MS}_{res}}{\text{SS}_A + (n - \nu_A) \text{MS}_{res}} = \frac{\nu_A (f_A - 1)}{\nu_A (f_A - 1) + n}$ $\hat{\omega}^2_{(B)} := \frac{\text{SS}_B - \nu_B \text{MS}_{res}}{\text{SS}_B + (n - \nu_B) \text{MS}_{res}} = \frac{\nu_B (f_B - 1)}{\nu_B (f_B - 1) + n}$ $\hat{\omega}^2_{(AB)} := \frac{\text{SS}_{AB} - \nu_{AB} \text{MS}_{res}}{\text{SS}_{AB} + (n - \nu_{AB}) \text{MS}_{res}} = \frac{\nu_{AB} (f_{AB} - 1)}{\nu_{AB} (f_{AB} - 1) + n}$

$$n := IJK = (1 + \nu_A)(1 + \nu_B)K$$

[†] W.L. Hays, *Statistics for Psychologists*, 1963.

[‡] G. Keren, C. Lewis, "Partial Omega Squared for ANOVA Designs", *Educational & Psychological Measurement*, **39** (1979), 119-128.

Eta-Squared or Partial Eta-Squared??

There has been discussion regarding which effect size measure (eta-squared & partial eta-squared) is better for multi-factor ANOVA – the short answer being it depends on the particular multi-factor design(s) and whether meta-analyses will be performed[†‡◇♣].

To play it safe, we shall always report both η^2 & $\eta^2_{(\cdot)}$. Ditto for ω^2 & $\omega^2_{(\cdot)}$.

[†] J. Cohen, "Eta-Squared and Partial Eta-Squared in Fixed Factor ANOVA Designs", *Educational & Psychological Measurement*, **33** (1973), 107-112.

[‡] T.R. Levine, C.R. Hullett, "Eta Squared, Partial Eta Squared, and Misreporting of Effect Size in Communication Research", *Human Communication Research*, **28** (2002), 612-625.

[◇] S. Olejnik, J. Algina, "Generalized Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs", *Psychological Methods*, **8** (2003), 434-447.

[♣] C.A. Pierce, R.A. Block, H. Aguinis, "Cautionary Note on Reporting Eta-Squared Values from Multifactor ANOVA Designs", *Educational and Psychological Measurement*, **64** (2004), 916-924.

2F bcrANOVA (Post-Hoc Comparisons)

Proposition

Given a 2-factor experiment with I levels of factor A, J levels of factor B, and each group has $K > 1$ measurements. $[\nu_{res} := IJ(K-1)]$

Moreover, 2F bcrANOVA accepts H_0^{AB} and rejects H_0^{A} at significance level α.

Then, to determine which levels of factor A significantly differ:

1. *Compute the factor A significant difference width:*

$$w_A = q^*_{I,\nu_{res};\alpha} \cdot \sqrt{MS_{res}/(JK)}$$

2. *Sort the I factor A level means in ascending order:*

$$\overline{x}_{(1)\bullet\bullet} \leq \overline{x}_{(2)\bullet\bullet} \leq \cdots \leq \overline{x}_{(I)\bullet\bullet}$$

3. *For each sorted factor A level mean $\overline{x}_{(i)\bullet\bullet}$:*
   - *If $\overline{x}_{(i+1)\bullet\bullet} \notin \left[\overline{x}_{(i)\bullet\bullet},\ \overline{x}_{(i)\bullet\bullet} + w_A\right]$, repeat STEP 3 with next sorted mean.*
   - *Else, underline $\overline{x}_{(i)\bullet\bullet}$ and all larger means within a distance of w_A with new line.*

Proposition

Given a 2-factor experiment with I levels of factor A, J levels of factor B, and each group has $K > 1$ measurements. $[\nu_{res} := IJ(K-1)]$

Moreover, 2F bcrANOVA accepts H_0^{AB} and rejects H_0^{B} at significance level α.

Then, to determine which levels of factor B significantly differ:

1. *Compute the factor B significant difference width:*
$$w_B = q^*_{J,\nu_{res};\alpha} \cdot \sqrt{MS_{res}/(IK)}$$
2. *Sort the J factor B level means in ascending order:*
$$\overline{x}_{\bullet(1)\bullet} \leq \overline{x}_{\bullet(2)\bullet} \leq \cdots \leq \overline{x}_{\bullet(J)\bullet}$$
3. *For each sorted factor B level mean $\overline{x}_{\bullet(j)\bullet}$:*
   - *If $\overline{x}_{\bullet(j+1)\bullet} \notin \left[\overline{x}_{\bullet(j)\bullet},\ \overline{x}_{\bullet(j)\bullet} + w_B\right]$, repeat STEP 3 with next sorted mean.*
   - *Else, underline $\overline{x}_{\bullet(j)\bullet}$ and all larger means within a distance of w_B with new line.*

Post-Hoc Comparisons with a Significant Interaction

Post-hoc comparisons when there is a statistically significant interaction (i.e. 2F bcrANOVA rejects H_0^{AB}) are far trickier and, hence, beyond the scope of this course.

Interested readers may consult any of the following:

L.E. Toothaker, *Multiple Comparison Procedures*, SAGE, 1992. (Ch 5)

P.H. Westfall *et al*, *Multiple Comparisons & Multiple Tests using SAS*, SAS Inst., 1999. (§9.2.4)

Y. Hochberg *et al*, *Multiple Comparison Procedures*, Wiley, 1987. (§10.5)

G. Keppel, *Design and Analysis: A Researcher's Handbook*, Pearson, 1991.

R.J. Boik, "The Analysis of Two-Factor Interactions in Fixed Effects Linear Models", *Journal of Educational Statistics*, **18** (1993), 1-40.

PART V:

2-Factor ANOVA Model (Adequacy) Checking

Standardized Residuals

Checking for Outliers

Checking Normality Assumption

Checking Independence Assumption

Checking Equal Variances Assumption

ANOVA Model Checking: Standardized Residuals

Definition

(Standardized Residuals)

Given a balanced 2-factor experiment:

$$X_{ijk} = \mu + \alpha_i^A + \alpha_j^B + \gamma_{ij}^{AB} + E_{ijk}$$

Moreover, suppose 2F bcrANOVA was performed accordingly.

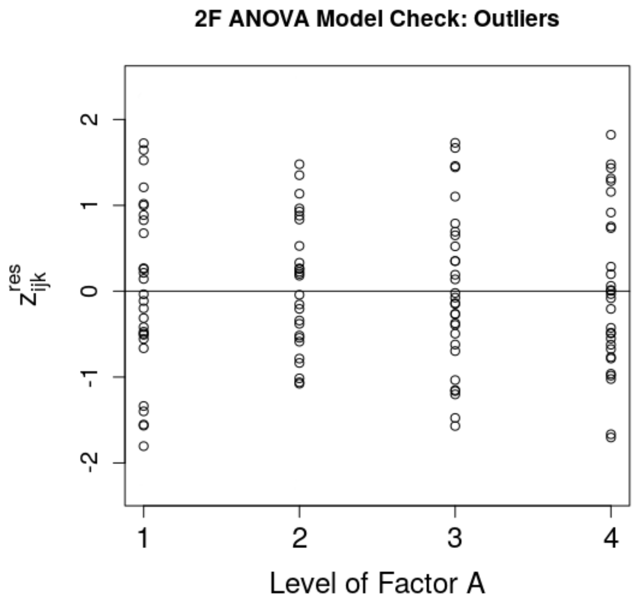
Then, the **standardized residuals**[†] are defined to be:

$$z_{ijk}^{res} := \frac{x_{ijk}^{res}}{\sqrt{\text{SS}_{res}/(n-1)}}$$

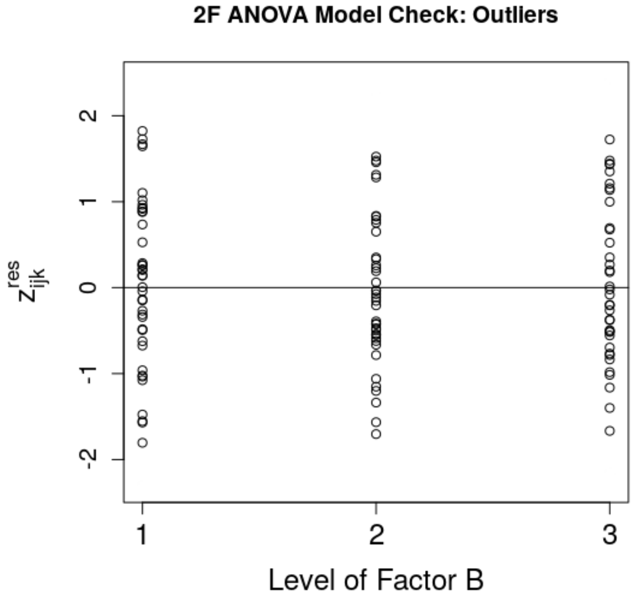
An alternative definition[‡] that's reasonable but not used here is: $\dfrac{x_{ijk}^{res}}{\sqrt{\text{MS}_{res}}}$

[†] Dean, Voss, Draguljić, *Design & Analysis of Experiments*, 2^{nd} Ed, 2017. (§6.2.3)
[‡] D.C. Montgomery, *Design & Analysis of Experiments*, 7^{th} Ed, Wiley, 2009. (§5.3.3)

ANOVA Model Checking: No Outliers



ANOVA Model Checking: No Outliers



ANOVA Model Checking: Outlier Mitigation

Q: How to handle outliers when performing 2F ANOVA?

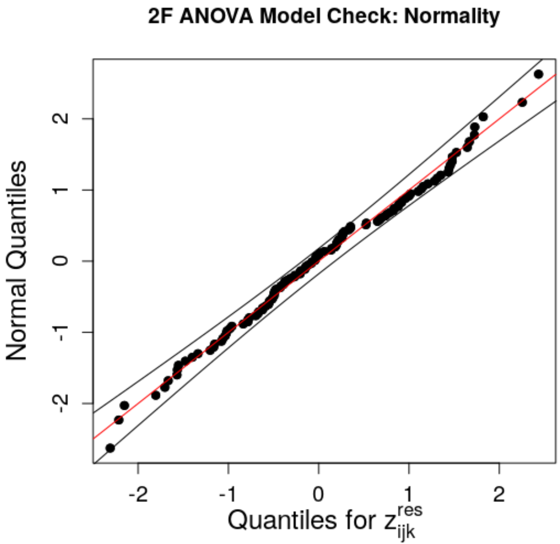
A: For each outlier:

- If outlier was due to measurement/calculation error, correct it[†‡].
- Else, outlier may be due to violation(s) of the ANOVA assumptions[†].
- Else, the 2-factor linear model may be insufficient[†]:
  - Consider building a 3-Factor ANOVA model... (beyond scope of course)
  - ...or an Analysis of Covariance (ANCOVA) model (beyond scope of course)

"We should be careful not to reject or discard an outlying observation unless we have reasonably non-statistical grounds for doing so. At worst, you may end up with two analyses; one with the outlier and one without."[‡]

[†] Dean, Voss, Draguljić, *Design & Analysis of Experiments*, 2^{nd} Ed, 2017. (§5.4)
[‡] D.C. Montgomery, *Design & Analysis of Experiments*, 7^{th} Ed, Wiley, 2009. (§3.4.1)

ANOVA Model Checking: Normality Satisfied



ANOVA Model Checking: Normality Mitigation

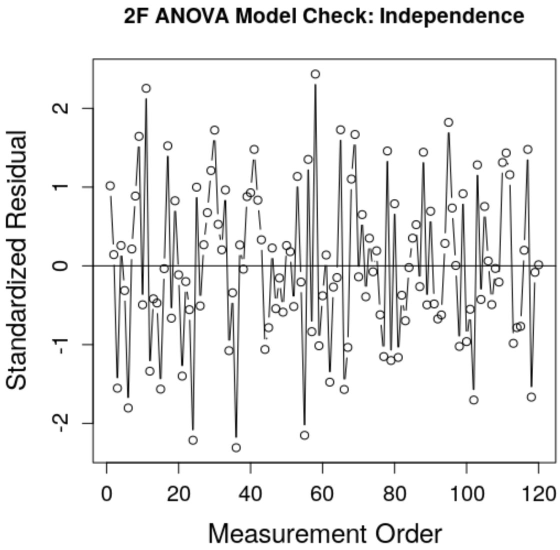
Q: How to perform a 2F ANOVA when the Normality Assumption is violated?

A: Alas, there's no 2F non-parametric ANOVA due to presence of interaction.

- Instead, consider using a regression model with dummy variables♠.

♠ Mendenhall, Sincich, *A 2nd Course in Statistics: Regression Analysis*, 7^{th} Ed, Pearson, 2012. (§5.8, §11.5, §12.5)

ANOVA Model Checking: Independence Satisfied



ANOVA Model Checking: Independence Mitigation

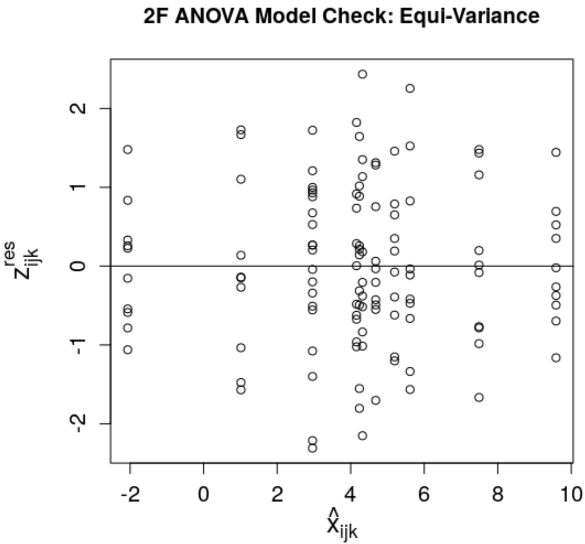
Q: How to perform a 2F ANOVA when Independence is violated?

A: This is where things become frustrating:

- If randomization was <u>not</u> used, redo the experiment using randomization‡.
- If randomization <u>was</u> used, then use a more complicated model†:
  - 3-Factor ANOVA – beyond scope of course (but covered in §11.3 of Devore)
  - Analysis of Covariance (ANCOVA) – beyond scope of this course

† Dean, Voss, Draguljić, *Design & Analysis of Experiments*, 2^{nd} Ed, 2017. (§5.5)
‡ D.C. Montgomery, *Design & Analysis of Experiments*, 7^{th} Ed, Wiley, 2009. (§3.4.2)

ANOVA Model Checking: Equi-Variance Satisfied



ANOVA Model Checking: Equi-Variance Mitigation

Q: How to perform 2F ANOVA when Equi-Variance Assumption is violated?

A: Perform an appropriate **variance-stabilizing data transformation**†‡♣ first:

$$\log X,\ \log(1+X),\ \log(1+\min x_{ij}+X),$$
$$\sqrt{X},\ \sqrt{0.5+X},\ \sqrt{X}+\sqrt{1+X},$$
$$1/X,\ 1/\sqrt{X},\ \arcsin(\sqrt{X}),\ 2\arcsin(\sqrt{X\pm 1/2m})$$

If data are counts or Poisson-like, use a square-root transformation†‡♣.
If data are proportions or Binomial-like, use an arcsine transformation†♣.
When in doubt, plot $\log s_i$ vs. $\log(\bar{x}_{i\bullet})$ to help determine data transformation†‡.
If data transformations don't help much, a more robust method is necessary♡.

<u>NOTE:</u> Data transformations are beyond the scope of this course.

† Dean, Voss, Draguljić, *Design & Analysis of Experiments*, 2^{nd} Ed, 2017. (§5.6.2)
‡ D.C. Montgomery, *Design & Analysis of Experiments*, 7^{th} Ed, 2009. (§3.4.3)
♣ D.C. Howell, *Statistical Methods for Psychology*, 7^{th} Ed, 2010. (§11.9)
♡ Grissom, "Heterogeneity of Variance in Clinical Data", *J. Cons. & Clin. Psy.*, **68** (2000), 155-165.

Textbook Logistics for Section 11.2

CONCEPT	TEXTBOOK NOTATION	SLIDES/OUTLINE NOTATION
Sum of Squares of Factor A	SSTr	SS_A
Mean Square of Factor A	MSTr	MS_A
Sum of Squares of Residuals	SSE	SS_{res}
Mean Square of Residuals	MSE	MS_{res}
Effect of i^{th} Factor A	α_i	α_i^A
Interaction of Factors A & B	γ_{ij}	γ_{ij}^{AB}
Null Hypothesis for Factor A	H_{0A}	H_0^A
Alt. Hypothesis for Factor A	H_{aA}	H_A^A
Null Hypothesis for Interaction AB	H_{0AB}	H_0^{AB}
Alt. Hypothesis for Interaction AB	H_{aAB}	H_A^{AB}
Expected Value	$E(X)$	$\mathbb{E}[X]$
Variance	$V(X)$	$\mathbb{V}[X]$

- Ignore "Models with Mixed and Random Effects" section.
  - The ANOVA procedure is identical as for fixed effects linear models.
  - However, model assumption checking is subtler and trickier.
  - Also, expected mean squares differ in expression.

Fin.