

Biometrical Best Practices

in light of 2015 working experience with Afghan colleagues

Khaled Al-Sham'aa

1

Annual Planning Meeting for Afghanistan Projects

31st March – 5th April 2016

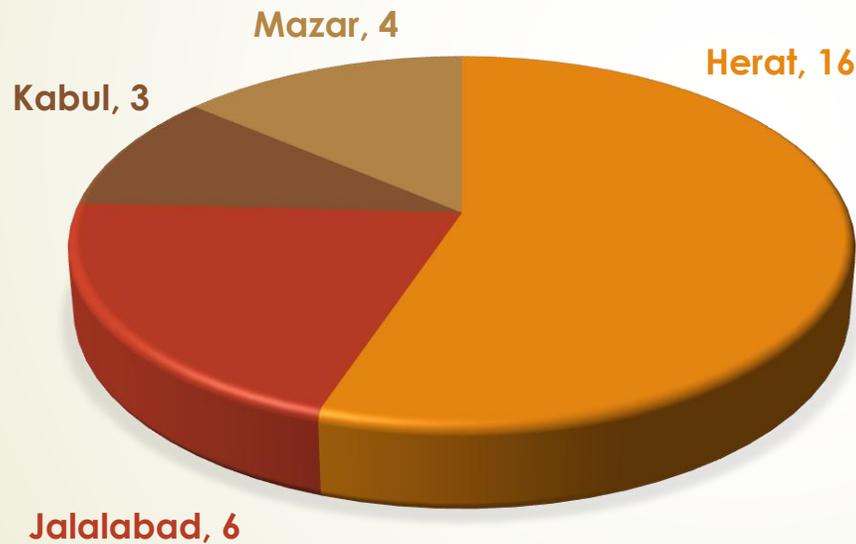
New Delhi, India

Presentation Roadmap

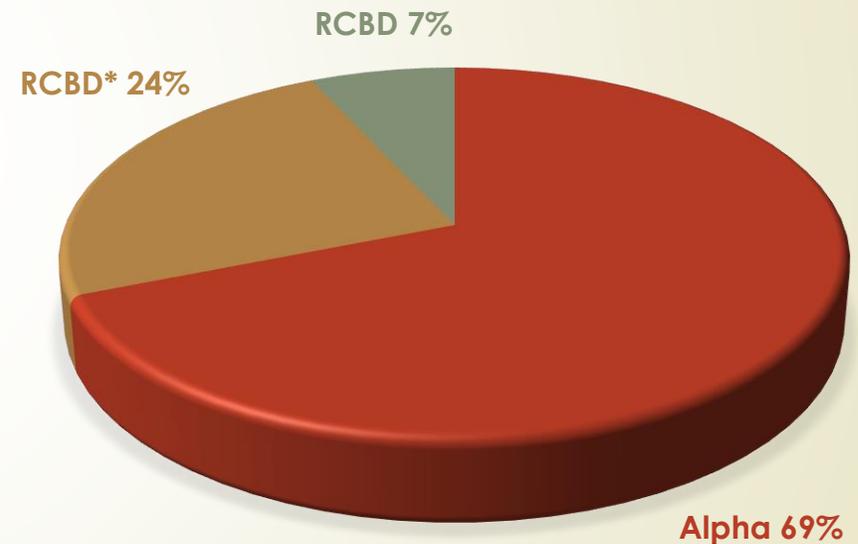
- ▶ 2015 Working Stats
- ▶ Alpha vs. RCBD
- ▶ REML vs. ANOVA
- ▶ Can and Can't in Alpha Design
- ▶ RCBD vs. DiGger
- ▶ What is the Right Block Size?
- ▶ Alpha Implementation
- ▶ Layout Information
- ▶ Meta Data
- ▶ Data Validation and Integrity Check
- ▶ Combine Data for MET Analysis
- ▶ Excel files vs. Database

2015 Working Stats

Trials# per Location



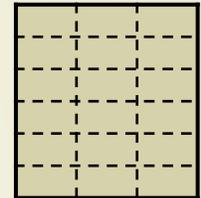
Trials% per Design



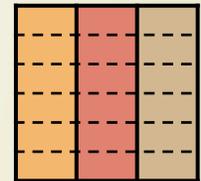
CRD, RCBD, and Alpha Designs

➤ A method which partitions the total variation in the response into the components (sources of variation):

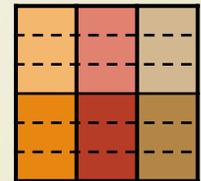
❖ Response = $\mu + \tau + \xi$ (CRD)



❖ Response = $\mu + \tau + \pi + \xi$ (RCBD)

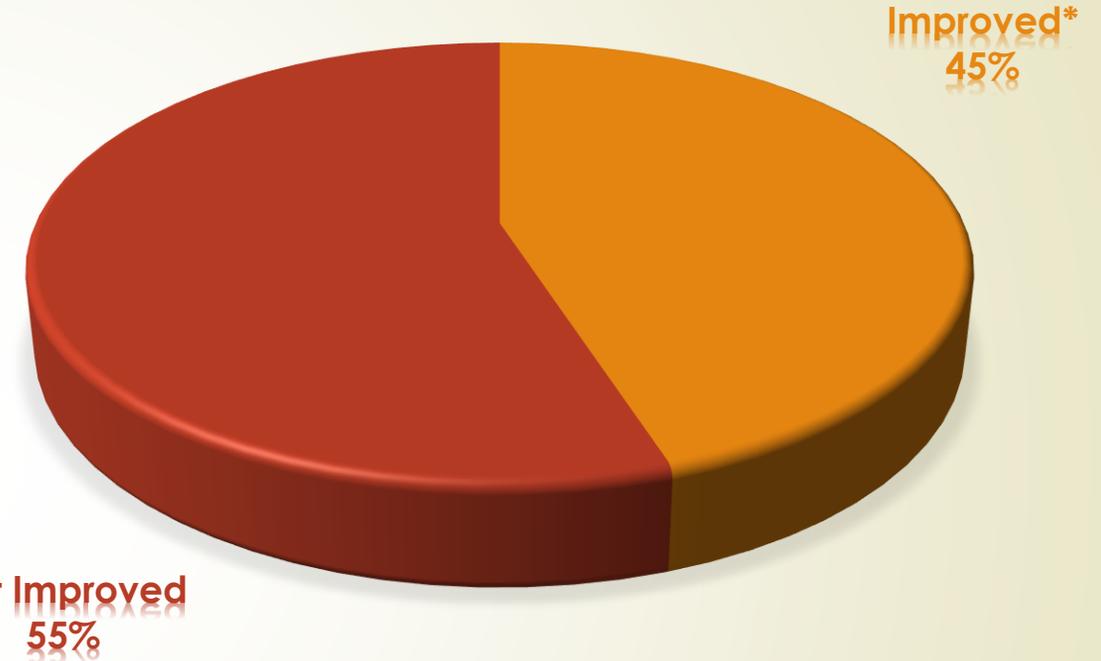
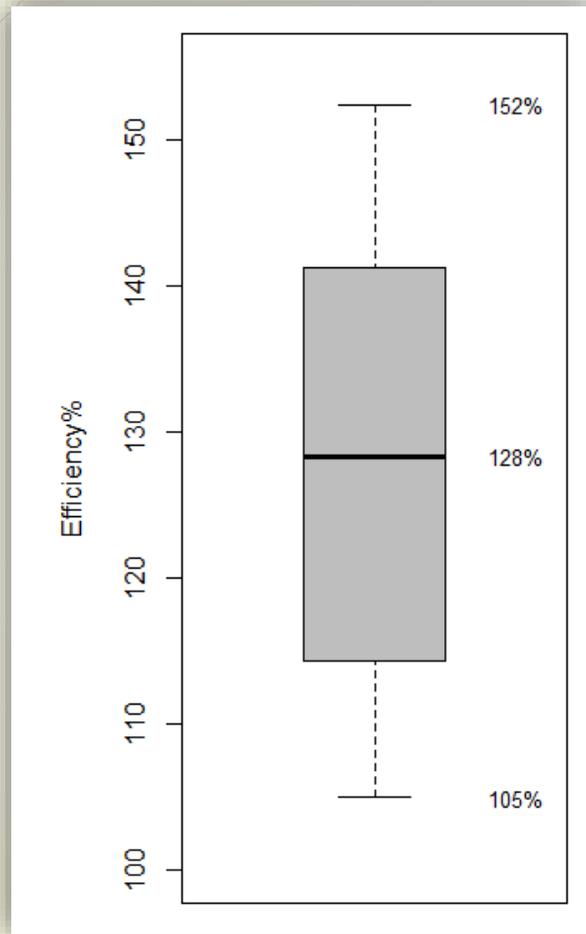


❖ Response = $\mu + \tau + \pi + \beta + \xi$ (Alpha)



μ) grand mean τ) effect of treatments π) effect of replicates β) effect of blocks ξ) experimental error

REML vs. ANOVA



* 4 out of 9 trials become significant

$$Efficiency \% = \frac{100 \times (SED \text{ under } RCB)^2}{(SED \text{ under } Alpha)^2}$$

Can and Can't in Alpha Design



▶ You Can

- ▶ Get more precision without increase any resources in the field (e.g. land, seeds, labors)
- ▶ Analysis your data using simple ANOVA and ignore block information!
- ▶ Generate it for free (e.g. ICARDA BioComuting online service, Agricola R package from CIP, CIMMYT ALPHAGEN software, etc..)



▶ You Can't

- ▶ Convert RCBD randomization into Alpha design simply by add an extra column to split your replications into hypothesis blocks!
- ▶ Lose precision comparing to RCBD (i.e. worst scenario when blocks within replicates are homogeneous)

RCBD vs. DiGGer

3	1	2
7	14	6
17	8	17
5	24	14
21	5	12
6	12	15
20	20	5
23	11	7
9	21	10
12	7	21
19	15	23
2	2	3
24	19	18
15	17	8
14	6	24
4	10	4
10	4	1
16	9	13
1	3	9
13	18	11
11	23	20
18	22	16
22	13	19
8	16	22

21	5	10
19	14	15
24	7	9
13	6	2
22	8	4
18	16	12
23	17	11
1	3	20
20	23	14
6	9	21
4	13	1
15	11	24
10	19	17
7	2	18
16	22	5
3	12	8
2	10	3
17	4	7
8	18	19
14	24	6
9	20	22
12	21	23
5	15	13
11	1	16

What is the Right Block Size?

➤ Do **NOT** ask ME that question! It is constrained by the physical arrangement of plots in the field, for example:

- *Total number of genotypes*
- *Field homogeneity*
- *Plot size*
- *Field layout (i.e. rows & columns)*

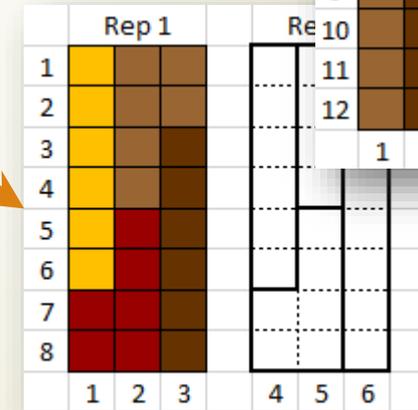
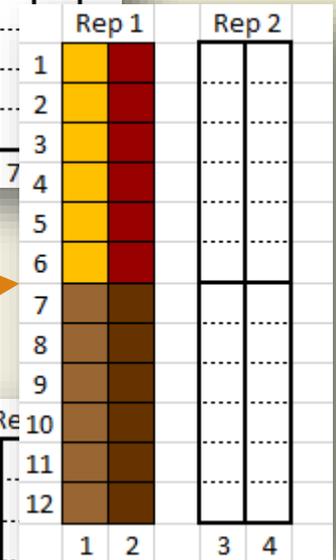
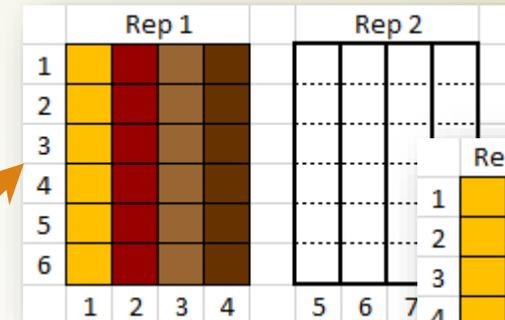
	Rep 1				Rep 2			
1	Yellow	Red	Light Brown	Dark Brown				
2	Yellow	Red	Light Brown	Dark Brown				
3	Yellow	Red	Light Brown	Dark Brown				
4	Yellow	Red	Light Brown	Dark Brown				
5	Yellow	Red	Light Brown	Dark Brown				
6	Yellow	Red	Light Brown	Dark Brown				
	1	2	3	4	5	6	7	8

	Rep 1				Rep 2			
1	Yellow	Yellow	Dark Brown	Dark Brown				
2	Yellow	Yellow	Dark Brown	Dark Brown				
3	Yellow	Yellow	Dark Brown	Dark Brown				
4	Yellow	Yellow	Dark Brown	Dark Brown				
5	Yellow	Yellow	Dark Brown	Dark Brown				
6	Yellow	Yellow	Dark Brown	Dark Brown				
	1	2	3	4	5	6	7	8

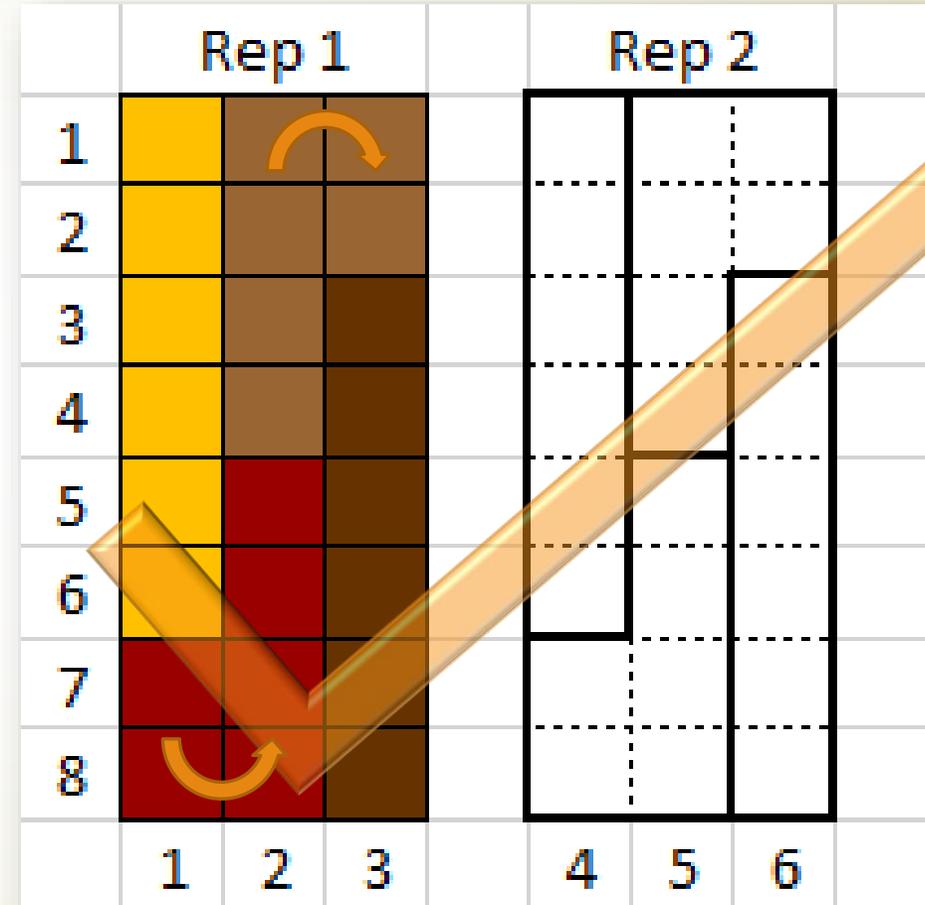
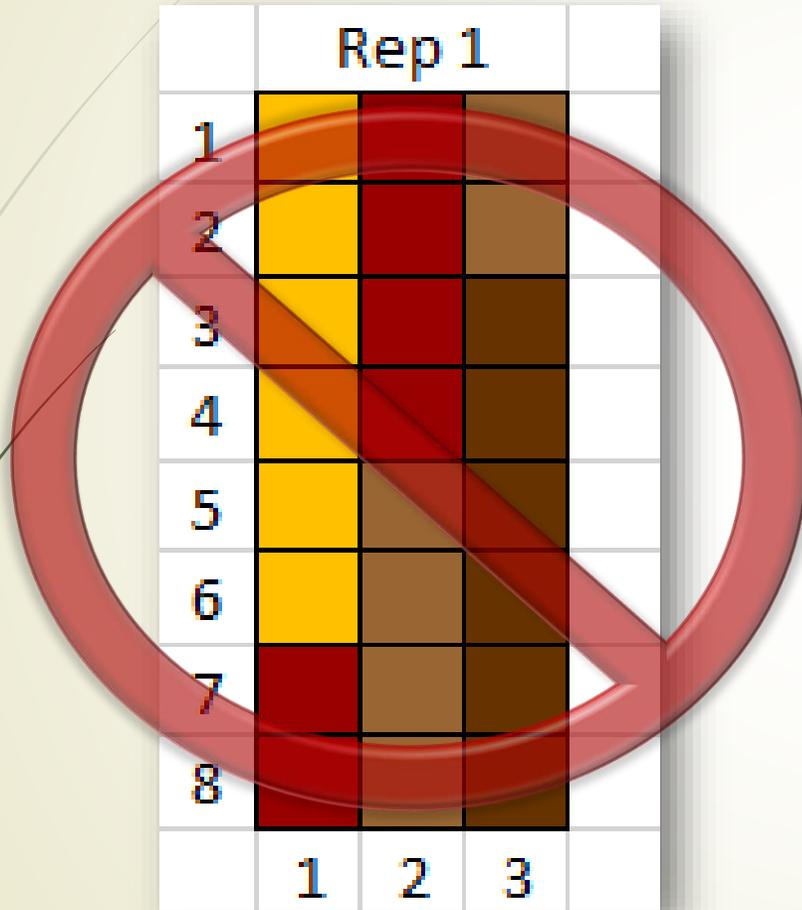
	Rep 1			
1	Yellow	Red	Light Brown	Dark Brown
2	Yellow	Red	Light Brown	Dark Brown
3	Yellow	Red	Light Brown	Dark Brown
4	Yellow	Red	Light Brown	Dark Brown
5	Yellow	Red	Light Brown	Dark Brown
6	Yellow	Red	Light Brown	Dark Brown
	1	2	3	4

Alpha Design Implementation (1)

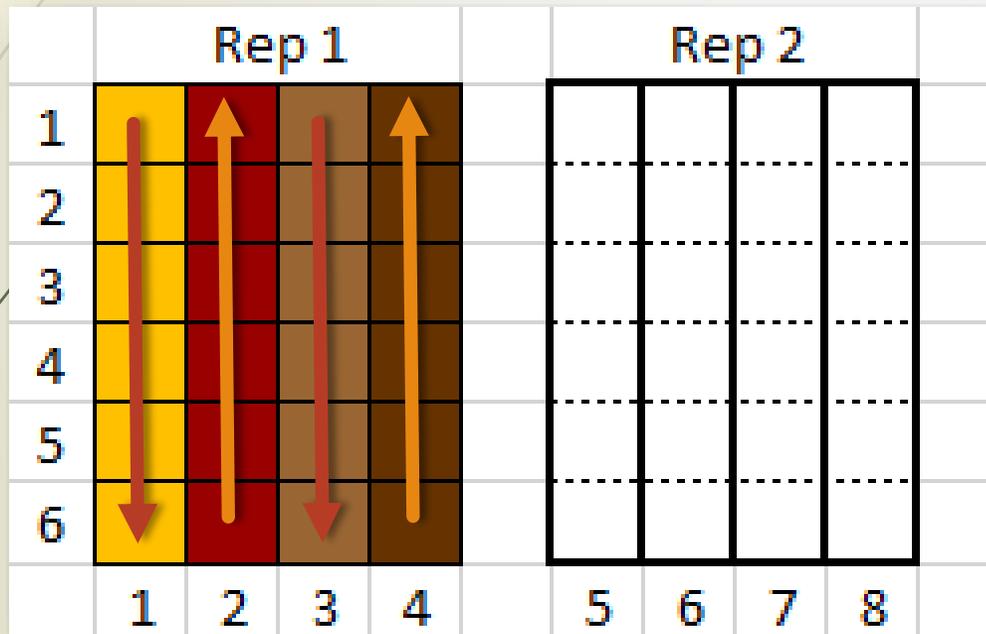
<i>Rep</i>	<i>Block</i>	<i>Plot</i>
1	1	1
1	1	:
1	1	6
1	2	7
1	2	:
1	2	12
1	3	13
1	3	:
1	3	18
1	4	19
1	4	:
1	4	24



Alpha Design Implementation (2)



Layout Information (1)



<i>Rep</i>	<i>Block</i>	<i>Col</i>	<i>Row</i>	<i>Plot</i>
1	1	1	1	1
1	1	1	:	:
1	1	1	6	6
1	2	2	6	7
1	2	2	:	:
1	2	2	1	12
1	3	3	1	13
1	3	3	:	:
1	3	3	6	18
1	4	4	6	19
1	4	4	:	:
1	4	4	1	24

Layout Information (2)

	Rep 1		Rep 2	
1	↓	↑		
2	↓	↑		
3	↓	↑		
4	↓	↑		
5	↓	↑		
6	↓	↑		
7	↓	↑		
8	↓	↑		
9	↓	↑		
10	↓	↑		
11	↓	↑		
12	↓	↑		
	1	2	3	4

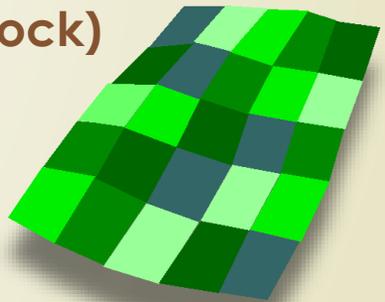
Rep	Block	Col	Row	Plot
1	1	1	1	1
1	1	1	:	:
1	1	1	6	6
1	2	1	7	7
1	2	1	:	:
1	2	1	12	12
1	3	2	12	13
1	3	2	:	:
1	3	2	7	18
1	4	2	6	19
1	4	2	:	:
1	4	2	1	24

Meta Data

- **Experiment name**
- **Description**
- **Coordinator name**
- **Coordinator institute**
- **Coordinator contact information**
- **Location**
- **Province (state)**
- **Country**
- **Latitude**
- **Longitude**

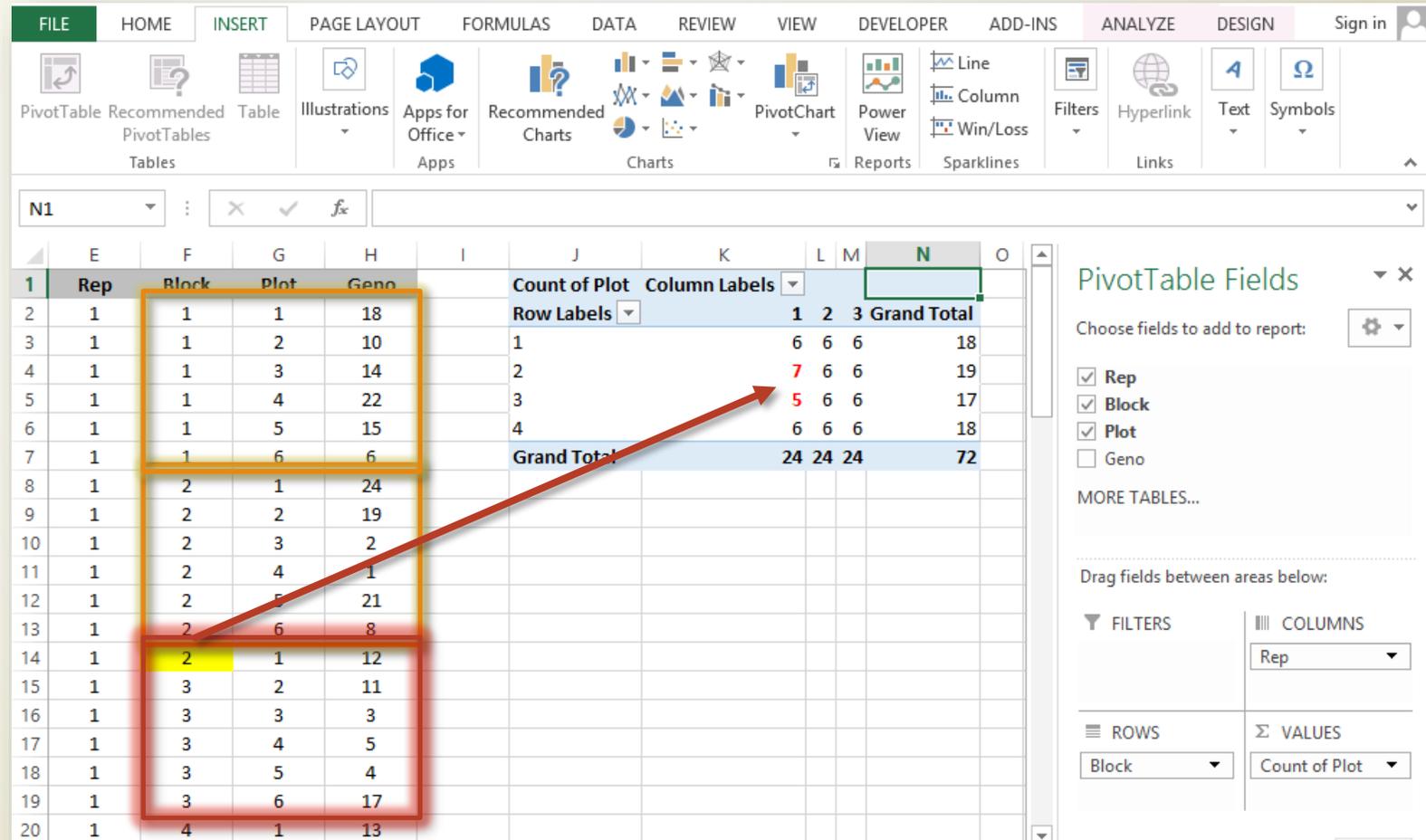


- **Crop**
- **List of entries / genotypes**
- **Season and cycle**
- **Experiment design**
- **Total number of entries**
- **Total number of plots**
- **Number of replications**
- **Block size (plots per block)**
- **Number of rows**
- **Number of columns**



Data Validation and Integrity Check (1)

Use Pivot Table Functionality



The screenshot displays an Excel spreadsheet with a PivotTable and the PivotTable Fields task pane. The PivotTable summarizes the 'Count of Plot' by 'Rep' and 'Block'. The task pane shows 'Rep', 'Block', and 'Plot' checked as filters and 'Count of Plot' as a value field.

Rep	Block	Plot	Geno	Count of Plot
1	1	1	18	6
1	1	2	10	6
1	1	3	14	6
1	1	4	22	6
1	1	5	15	6
1	1	6	6	6
1	2	1	24	7
1	2	2	19	6
1	2	3	2	6
1	2	4	1	6
1	2	5	21	6
1	2	6	8	6
1	3	1	12	5
1	3	2	11	4
1	3	3	3	5
1	3	4	5	4
1	3	5	4	6
1	3	6	17	6
1	4	1	13	6
Grand Total				72

Data Validation and Integrity Check (2)

Use Pivot Table Functionality

Rep	Block	Plot	Geno	Count of Plot	Column Labels			
1	1	1	18		1	2	3	Grand Total
1	1	2	10	1	2	1	1	4
1	1	3	14	2	1	1	1	3
1	1	4	22	3	1	1	1	3
1	1	5	15	4	1	1	1	3
1	1	6	6	5	1	1	1	3
1	2	1	24	6	1	1	1	3
1	2	2	19	7	1	1	1	3
1	2	3	2	8	1	1	1	3
1	2	4	1	9	1	1	1	3
1	2	5	21	10	1	1	1	3
1	2	6	8	11	1	1	1	2
1	3	1	12	12	1	1	1	3
1	3	2	1	13	1	1	1	3
1	3	3	3	14	1	1	1	3
1	3	4	5	15	1	1	1	3
1	3	5	4	16	1	1	1	3
1	3	6	17	17	1	1	1	3
1	4	1	13	18	1	1	1	3
1	4	2	16	19	1	1	1	3
1	4	3	20	20	1	1	1	3
1	4	4	7	21	1	1	1	3
1	4	5	9	22	1	1	1	3
1	4	6	23	23	1	1	1	3
2	1	1	13	24	1	1	1	3
2	1	2	23	Grand Total	24	24	24	72

PivotTable Fields

Choose fields to add to report: ⚙️

- Rep
- Block
- Plot
- Geno

MORE TABLES...

Drag fields between areas below:

FILTERS

COLUMNS

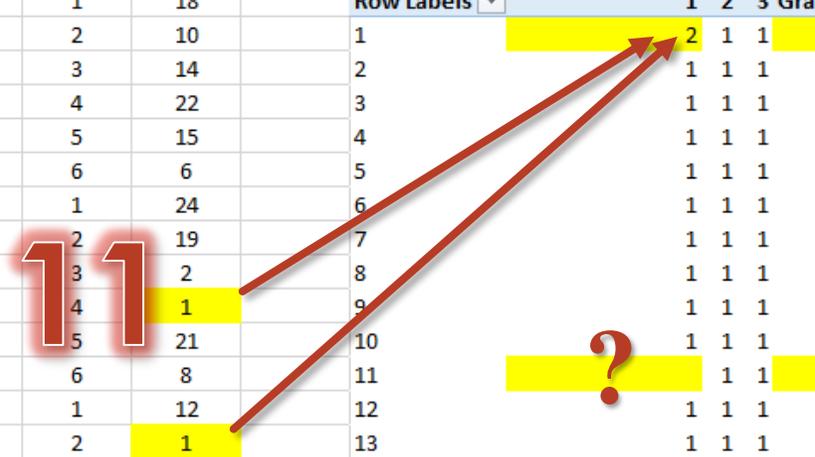
Rep

ROWS

Geno

VALUES

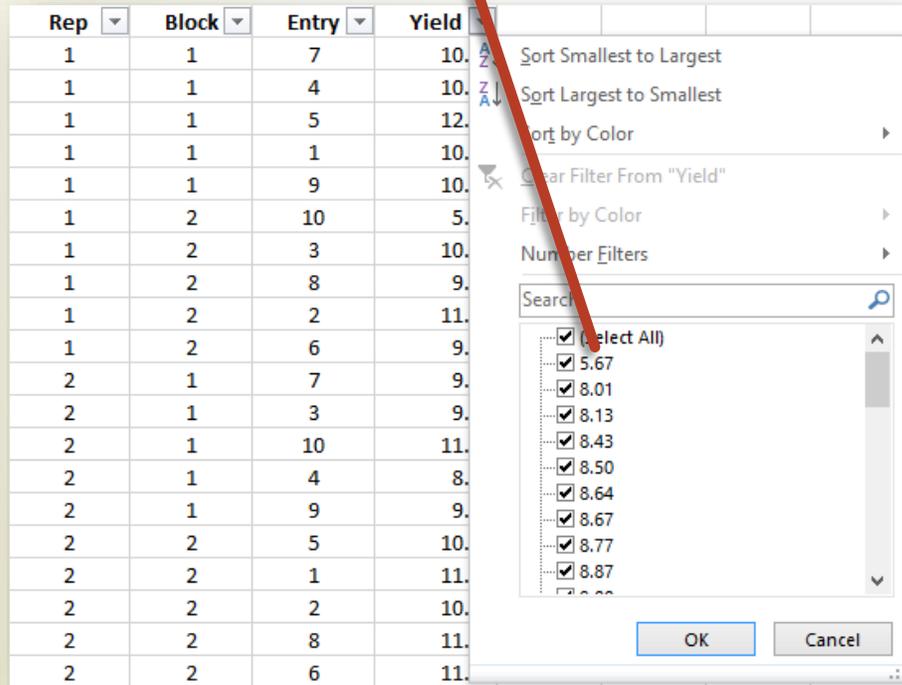
Count of Plot



Data Validation and Integrity Check (3)

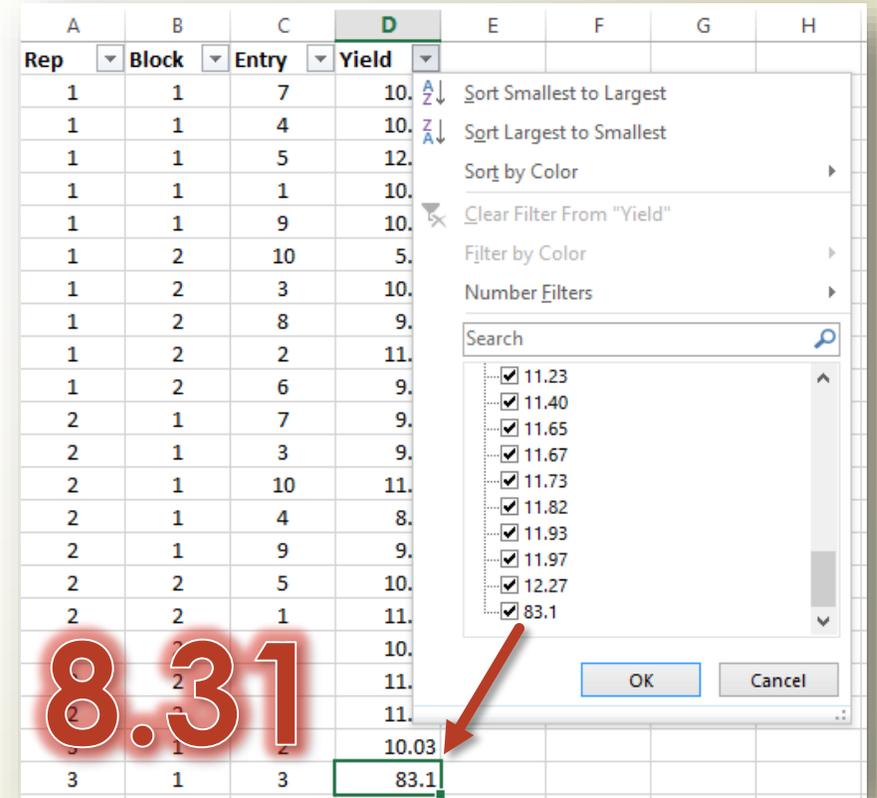
Use Filter Functionality

	A	B	C	D
1	Rep	Block	Entry	Yield
7	1	2	10	5.67
14	2	1	10	11.23
31	3	2	10	9.80
39	4	2	10	11.93
42				



Sort dialog box showing the 'Yield' column selected for sorting. The list of values to be sorted includes 5.67, 8.01, 8.13, 8.43, 8.50, 8.64, 8.67, 8.77, and 8.87. A red asterisk is placed to the left of the dialog box.

Rep	Block	Entry	Yield
1	1	7	10.03
1	1	4	10.03
1	1	5	12.27
1	1	1	10.03
1	1	9	10.03
1	2	10	5.67
1	2	3	10.03
1	2	8	9.80
1	2	2	11.93
1	2	6	9.80
2	1	7	9.80
2	1	3	9.80
2	1	10	11.23
2	1	4	8.01
2	1	9	9.80
2	2	5	10.03
2	2	1	11.23
2	2	2	10.03
2	2	8	11.23
2	2	6	11.23



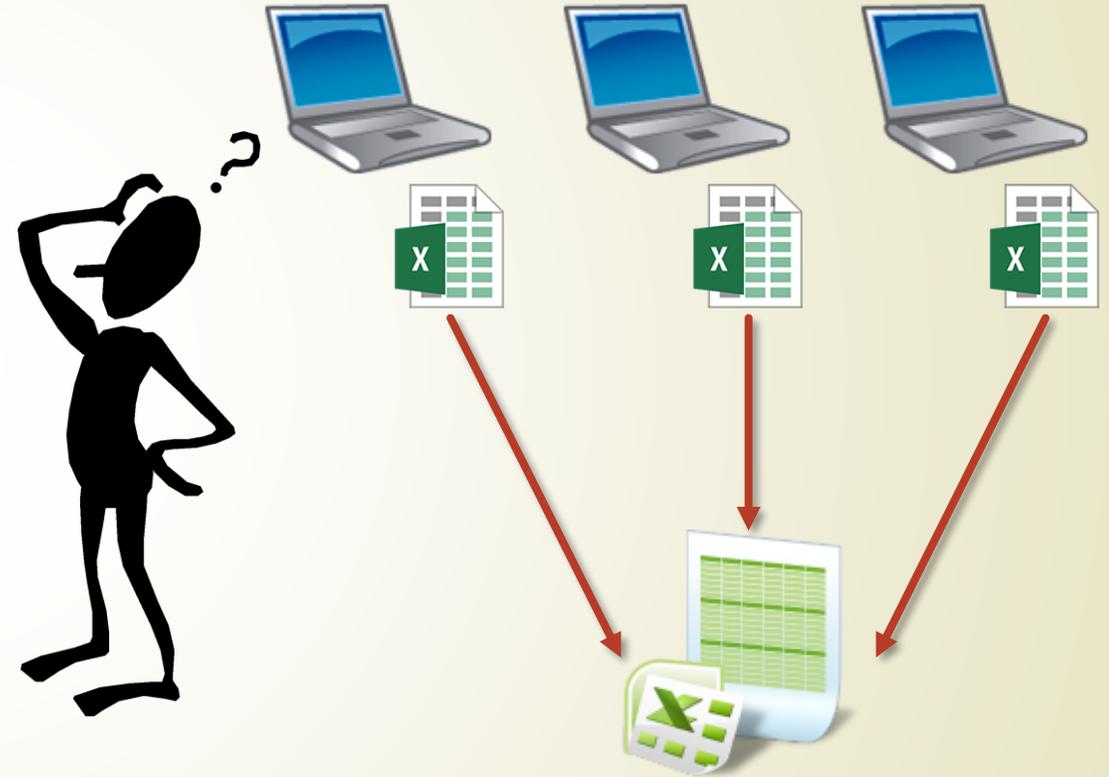
Filter dialog box showing the 'Yield' column selected for filtering. The list of values to be filtered includes 11.23, 11.40, 11.65, 11.67, 11.73, 11.82, 11.93, 11.97, 12.27, and 83.1. A red asterisk is placed to the left of the dialog box.

Rep	Block	Entry	Yield
1	1	7	10.03
1	1	4	10.03
1	1	5	12.27
1	1	1	10.03
1	1	9	10.03
1	2	10	5.67
1	2	3	10.03
1	2	8	9.80
1	2	2	11.93
1	2	6	9.80
2	1	7	9.80
2	1	3	9.80
2	1	10	11.23
2	1	4	8.01
2	1	9	9.80
2	2	5	10.03
2	2	1	11.23
2	2	2	10.03
2	2	8	11.23
2	2	6	11.23
3	1	3	10.03
3	1	3	83.1

8.31

Combine Data for MET Analysis

- ▶ Genotype name as a join key:
 - ▶ FLIP 96-15L (Ibla 1)
 - ▶ Ibla 1
 - ▶ FLIP 96-15L
 - ▶ FLIP96-15L
 - ▶ FLIP 1996-15L
 - ▶ Leading and trailing space
 - ▶ Letter case (i.e. FLIP, Flip, or flip)
 - ▶ Typo (e.g. FLOP ~ FLIP)
- ▶ Naming style (e.g. Yield ~ GY or BY)
- ▶ Units (e.g. T/ha, Kg/ha, or gr/m²)
- ▶ Abbreviation & acronym (e.g. DFLR and DF)



Excel Files vs. Database

DB or not DB, is it really your question!?



➤ DB Pros

- Data integrity (*independent copies causing duplicate and often outdated versions*)
- Query and retrieve data offers a greater range of complexity (SQL)
 - * *What was planted in this field during last **n** years?*
 - * *Who planted **x** genotype in this area and what was the performance?*
- Better access control (*multi-user, security privileges, web access/open access*)
- Better for long-term storage vs. Excel work-in-progress module



➤ DB Cons

- Excel is really widely used and more user friendly



Thank you

