CV values from historical trials as a quality measure: a tool

CV-Werte von historischen Versuchen als Qualitätsmesslatte für Bewertungen: ein Werkzeug

Sommertagung AG Landwirtschaftliches Versuchswesen 27-28 June 2019, IPK Gatersleben

Peter Lancashire and Yinfei Li
Bayer AG Crop Science
Field Solutions
Monheim, Germany

Agenda

- Bayer field trials
- EPPO standards
- Coefficient of variation
- Scout field trials system
- Examples of reference distributions
- CV inspector tool
- Discussion and questions

Bayer field trials

Bayer conducts around 20000 crop protection field trials globally each year.

They are used to investigate, develop, register and support new product solutions for farmers and growers.

Tea plots in Japan

Sugar beet damaged by pests

Potato infected with late blight

European and Mediterranean Plant Protection Organization (EPPO) Standards

The EPPO Standards for the efficacy evaluation of plant protection products (PP1) describe the **conduct of trials** carried out to assess the **efficacy** of plant protection products against specific pests.

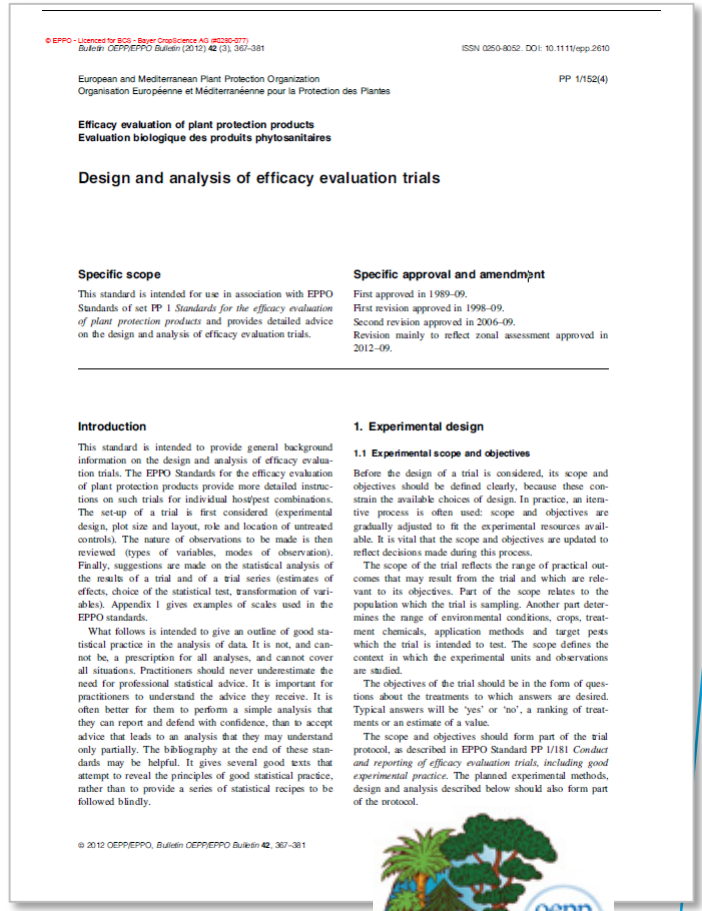
They are addressed to all institutions, official registration authorities, public institutes or **private firms** carrying out such trials.

In Europe, they now constitute a reference in Commission Regulation (EU) No 545/2011 of 10 June 2011, implementing Regulation (EC) No 1107/2009 of the European Parliament and of the Council regarding the **data requirements** for plant protection products.

PP 1/152(4) Design and analysis of efficacy evaluation trials

// Relevant standard for statistics

// Implemented by Bayer



Bulletin OEPP/EPPO Bulletin (2012) **42** (3), 367-381
ISSN 0250-8052. DOI: 10.1111/epp.2610

European and Mediterranean Plant Protection Organization
Organisation Européenne et Méditerranéenne pour la Protection des Plantes

PP 1/152(4)

Efficacy evaluation of plant protection products
Evaluation biologique des produits phytosanitaires

Design and analysis of efficacy evaluation trials

Specific scope

This standard is intended for use in association with EPPO Standards of set PP 1 *Standards for the efficacy evaluation of plant protection products* and provides detailed advice on the design and analysis of efficacy evaluation trials.

Specific approval and amendment

First approved in 1989–09.
First revision approved in 1998–09.
Second revision approved in 2006–09.
Revision mainly to reflect zonal assessment approved in 2012–09.

Introduction

This standard is intended to provide general background information on the design and analysis of efficacy evaluation trials. The EPPO Standards for the efficacy evaluation of plant protection products provide more detailed instructions on such trials for individual host/pest combinations. The set-up of a trial is first considered (experimental design, plot size and layout, role and location of untreated controls). The nature of observations to be made is then reviewed (types of variables, modes of observation). Finally, suggestions are made on the statistical analysis of the results of a trial and of a trial series (estimates of effects, choice of the statistical test, transformation of variables). Appendix 1 gives examples of scales used in the EPPO standards.

What follows is intended to give an outline of good statistical practice in the analysis of data. It is not, and cannot be, a prescription for all analyses, and cannot cover all situations. Practitioners should never underestimate the need for professional statistical advice. It is important for practitioners to understand the advice they receive. It is often better for them to perform a simple analysis that they can report and defend with confidence, than to accept advice that leads to an analysis that they may understand only partially. The bibliography at the end of these standards may be helpful. It gives several good texts that attempt to reveal the principles of good statistical practice, rather than to provide a series of statistical recipes to be followed blindly.

1. Experimental design

1.1 Experimental scope and objectives

Before the design of a trial is considered, its scope and objectives should be defined clearly, because these constrain the available choices of design. In practice, an iterative process is often used: scope and objectives are gradually adjusted to fit the experimental resources available. It is vital that the scope and objectives are updated to reflect decisions made during this process.

The scope of the trial reflects the range of practical outcomes that may result from the trial and which are relevant to its objectives. Part of the scope relates to the population which the trial is sampling. Another part determines the range of environmental conditions, crops, treatment chemicals, application methods and target pests which the trial is intended to test. The scope defines the context in which the experimental units and observations are studied.

The objectives of the trial should be in the form of questions about the treatments to which answers are desired. Typical answers will be 'yes' or 'no', a ranking of treatments or an estimate of a value.

The scope and objectives should form part of the trial protocol, as described in EPPO Standard PP 1/181 *Conduct and reporting of efficacy evaluation trials, including good experimental practice*. The planned experimental methods, design and analysis described below should also form part of the protocol.

EPPO Guideline PP 1/152(4)
Design and analysis of efficacy evaluation trials

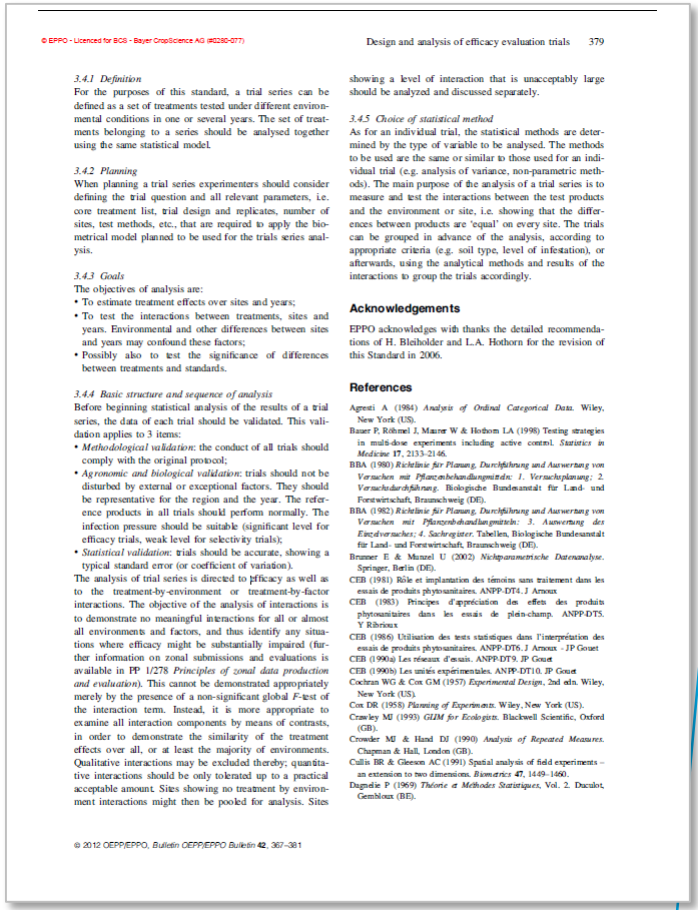
3.4.4 Basic structure and sequence of analysis

Before beginning statistical analysis of the results of a trial series, the data of each trial should be validated.

- *Methodological validation* …
- *Agronomic and biological validation* …
- *Statistical validation*: trials should be accurate, showing a typical standard error (or **coefficient of variation**).

Questions

- What is a coefficient of variation?
- What values are typical?
- How can we validate trial data in practice?



3.4.1 Definition
For the purposes of this standard, a trial series can be defined as a set of treatments tested under different environmental conditions in one or several years. The set of treatments belonging to a series should be analysed together using the same statistical model.

3.4.2 Planning
When planning a trial series experimenters should consider defining the trial question and all relevant parameters, i.e. core treatment list, trial design and replicates, number of sites, test methods, etc., that are required to apply the biometrical model planned to be used for the trials series analysis.

3.4.3 Goals
The objectives of analysis are:
- To estimate treatment effects over sites and years;
- To test the interactions between treatments, sites and years. Environmental and other differences between sites and years may confound these factors;
- Possibly also to test the significance of differences between treatments and standards.

3.4.4 Basic structure and sequence of analysis
Before beginning statistical analysis of the results of a trial series, the data of each trial should be validated. This validation applies to 3 items:
- *Methodological validation*: the conduct of all trials should comply with the original protocol;
- *Agronomic and biological validation*: trials should not be disturbed by external or exceptional factors. They should be representative for the region and the year. The reference products in all trials should perform normally. The infection pressure should be suitable (significant level for efficacy trials, weak level for selectivity trials);
- *Statistical validation*: trials should be accurate, showing a typical standard error (or coefficient of variation).

The analysis of trial series is directed to efficacy as well as to the treatment-by-environment or treatment-by-factor interactions. The objective of the analysis of interactions is to demonstrate no meaningful interactions for all or almost all environments and factors, and thus identify any situations where efficacy might be substantially impaired (further information on zonal submissions and evaluations is available in PP 1/278 *Principles of zonal data production and evaluation*). This cannot be demonstrated appropriately merely by the presence of a non-significant global *F*-test of the interaction term. Instead, it is more appropriate to examine all interaction components by means of contrasts, in order to demonstrate the similarity of the treatment effects over all, or at least the majority of environments. Qualitative interactions may be excluded thereby; quantitative interactions should be only tolerated up to a practical acceptable amount. Sites showing no treatment by environment interactions might then be pooled for analysis. Sites showing a level of interaction that is unacceptably large should be analyzed and discussed separately.

3.4.5 Choice of statistical method
As for an individual trial, the statistical methods are determined by the type of variable to be analysed. The methods to be used are the same or similar to those used for an individual trial (e.g. analysis of variance, non-parametric methods). The main purpose of the analysis of a trial series is to measure and test the interactions between the test products and the environment or site, i.e. showing that the differences between products are 'equal' on every site. The trials can be grouped in advance of the analysis, according to appropriate criteria (e.g. soil type, level of infestation), or afterwards, using the analytical methods and results of the interactions to group the trials accordingly.

Acknowledgements

EPPO acknowledges with thanks the detailed recommendations of H. Bleiholder and L.A. Hothorn for the revision of this Standard in 2006.

References

Agresti A (1984) *Analysis of Ordinal Categorical Data*. Wiley, New York (US).
Bauer P, Röhmel J, Maurer W & Hothorn LA (1998) Testing strategies in multi-dose experiments including active control. *Statistics in Medicine* **17**, 2133-2146.
BBA (1980) *Richtlinie für Planung, Durchführung und Auswertung von Versuchen mit Pflanzenbehandlungsmitteln: 1. Versuchsplanung; 2. Versuchsdurchführung*. Biologische Bundesanstalt für Land- und Forstwirtschaft, Braunschweig (DE).
BBA (1982) *Richtlinie für Planung, Durchführung und Auswertung von Versuchen mit Pflanzenbehandlungsmitteln: 3. Auswertung des Einzelversuches; 4. Sachregister*. Tabellen, Biologische Bundesanstalt für Land- und Forstwirtschaft, Braunschweig (DE).
Brunner E & Munzel U (2002) *Nichtparametrische Datenanalyse*. Springer, Berlin (DE).
CEB (1981) Rôle et implantation des témoins sans traitement dans les essais de produits phytosanitaires. ANPP-DT4. J Arnoux
CEB (1983) Principes d'appréciation des effets des produits phytosanitaires dans les essais de plein-champ. ANPP-DT5. Y Ribrioux
CEB (1986) Utilisation des tests statistiques dans l'interprétation des essais de produits phytosanitaires. ANPP-DT6. J Arnoux - JP Gouet
CEB (1990a) Les réseaux d'essais. ANPP-DT9. JP Gouet
CEB (1990b) Les unités expérimentales. ANPP-DT10. JP Gouet
Cochran WG & Cox GM (1957) *Experimental Design*, 2nd edn. Wiley, New York (US).
Cox DR (1958) *Planning of Experiments*. Wiley, New York (US).
Crawley MJ (1993) *GLIM for Ecologists*. Blackwell Scientific, Oxford (GB).
Crowder MJ & Hand DJ (1990) *Analysis of Repeated Measures*. Chapman & Hall, London (GB).
Cullis BR & Gleeson AC (1991) Spatial analysis of field experiments – an extension to two dimensions. *Biometrics* **47**, 1449–1460.
Dagnelie P (1969) *Théorie et Méthodes Statistiques*, Vol. 2. Duculot, Gembloux (BE).

© 2012 OEPP/EPPO, *Bulletin OEPP/EPPO Bulletin* **42**, 367–381

Coefficient of variation

The coefficient of variation (CV) is a **unitless** measure of **relative variability**. It is defined as the ratio of the standard deviation to the mean expressed as a percentage. (SAS 9.4 Help)

$$\widehat{cv} = \frac{s}{\bar{x}}$$

The coefficient of variation … provides a good yard-stick for appreciating the precision possible in an experiment, thereby aiding decisions on the sample size to be taken, and allowing **comparisons of variability** between experiments.

Ridgman W J (1975). Experimentation in Biology. An Introduction to Design and Analysis. Blackie, London.

The CV is sometimes used as a standard to gauge the **relative magnitude of error variation** compared with that from similar studies.

Littell R C, Stroup W W, Freund R J (2002). SAS for Linear Models. SAS Institute, Cary NC.

The CV on its own can be a poor indicator of quality. Trials that have been under disease, pest or other pressures and, as a consequence provide important genotype discrimination, often produce high CVs.

Kempton R A, Fox P N (eds.) (1997). Statistical Methods for Plant Variety Evaluation. Chapman & Hall, London.

What is a typical coefficient of variation?

Depends on:

// Site: intrinsic plot variability, application, accidents,

// Assessment: yield, disease, pest, weed species, aggregation.

// Method: measured, estimate, subjective rating, sample.

// Values: low yield, early epidemic => higher CV.

Published values can be found for yield in some crops.

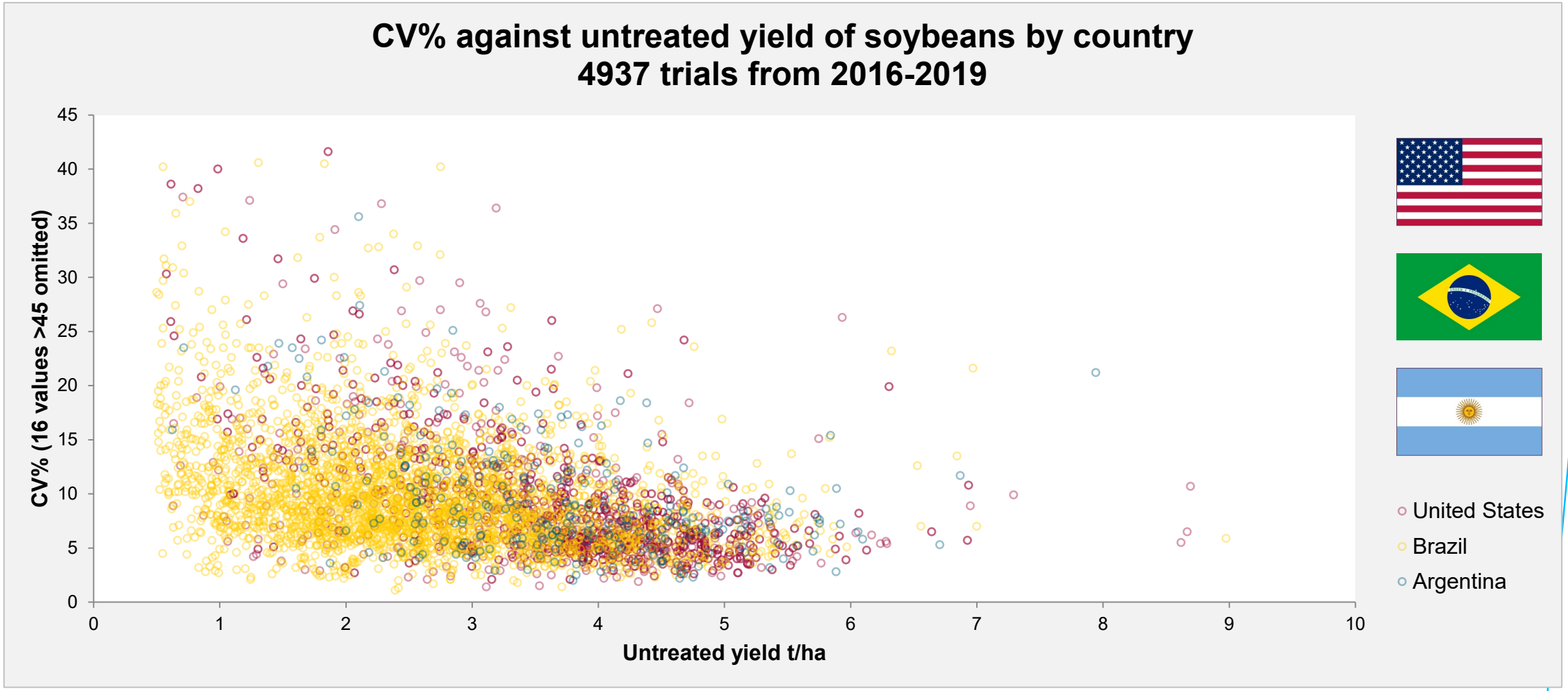
Not enough detail in most published reports.

Cannot compare with our assessments.

Solution: Use our own historical data.

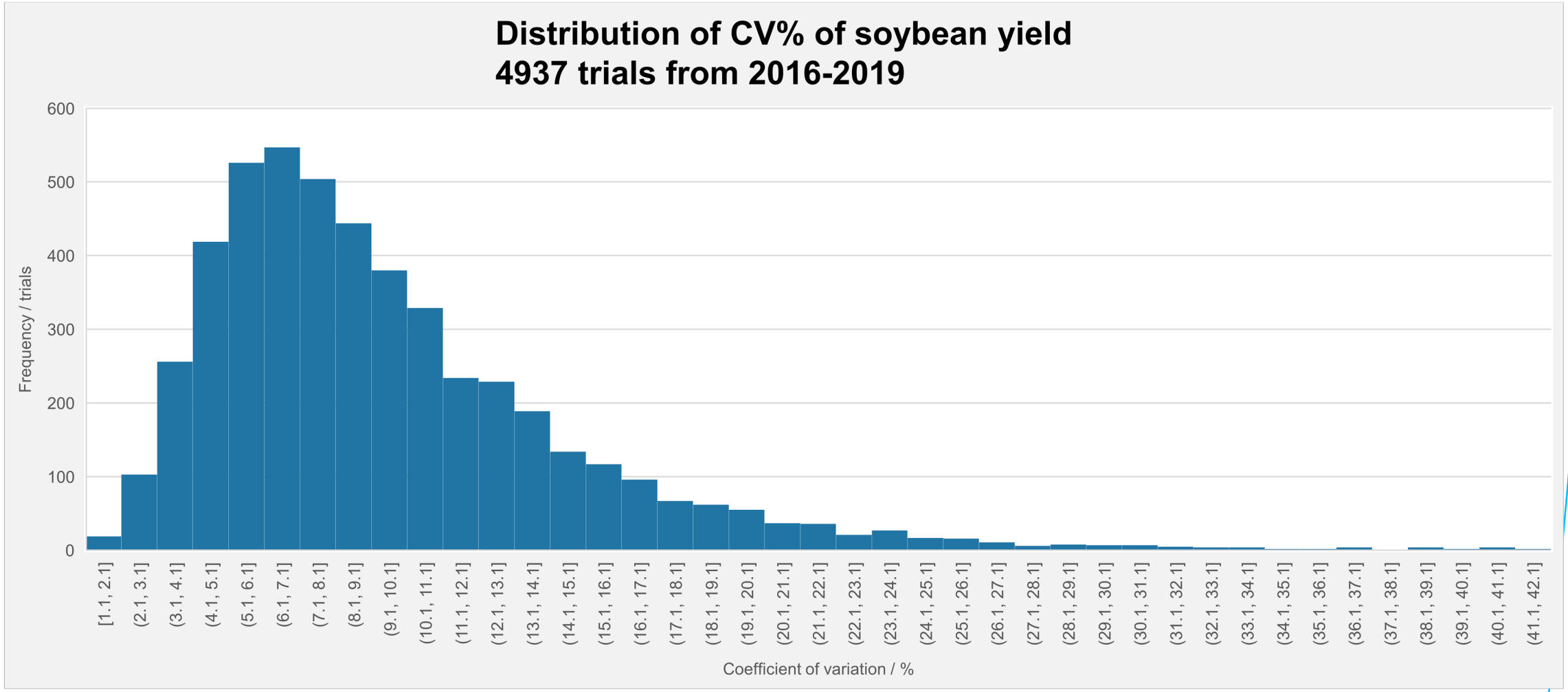
Coefficient of variation and yield

CV can be country-dependent and related to yield.



Coefficient of variation distribution

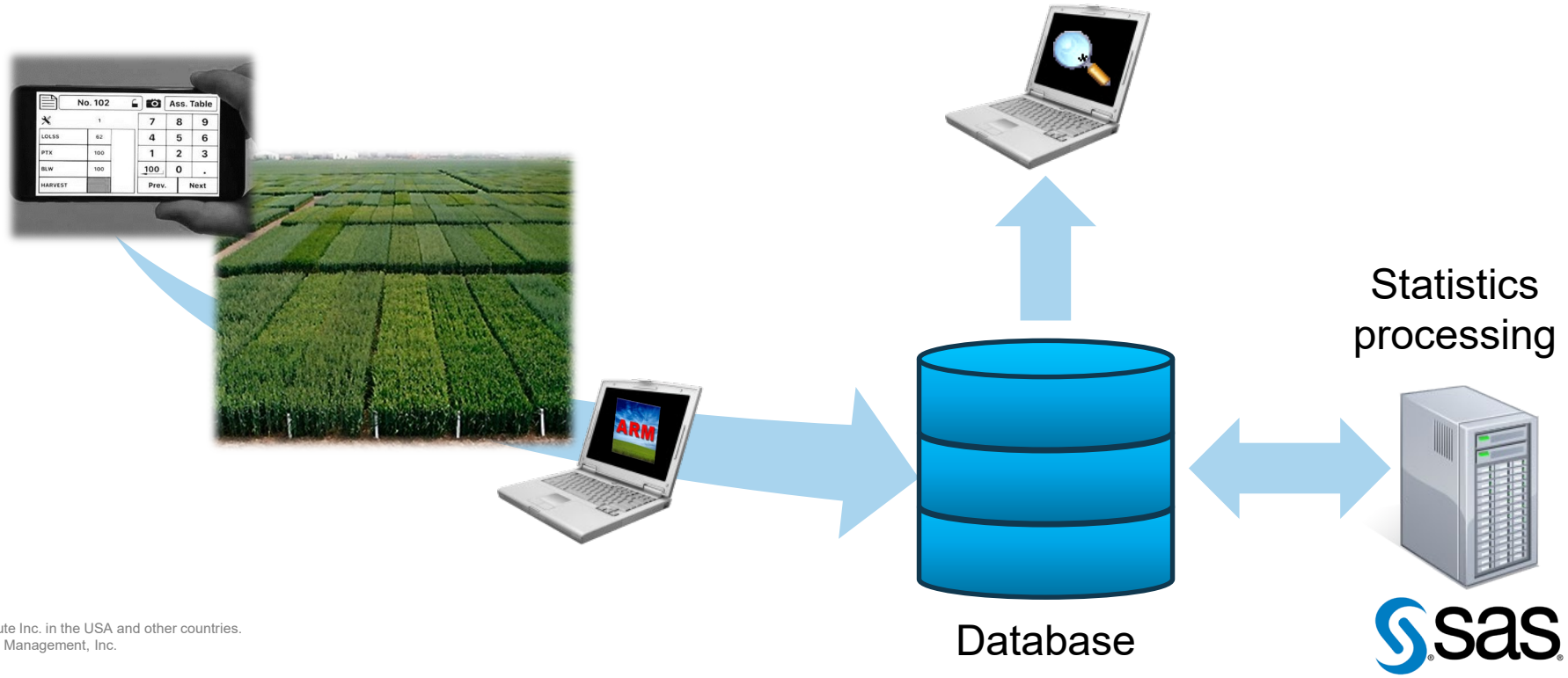
A histogram is more useful for validation.



Scout field trials system

The Scout system manages the whole process of planning, running, analysing, reporting and summarising field trials.

Scout



Scout statistics for single trials calculates CV

Meets EPPO standard.

Automatic based on rules.

Rules ensure consistency.

Uses SAS PROC GLM.

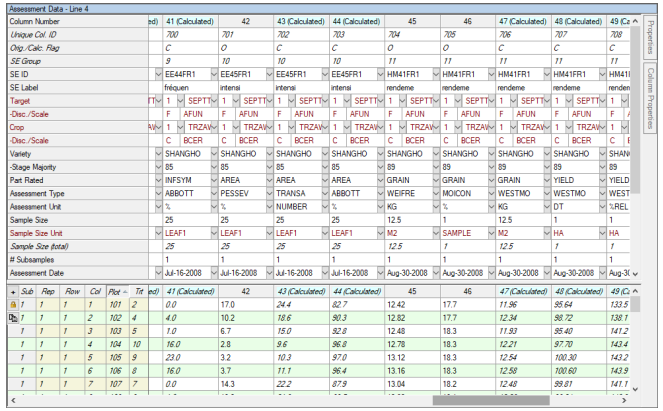
Traditional transforms.

Saves statistics to database.

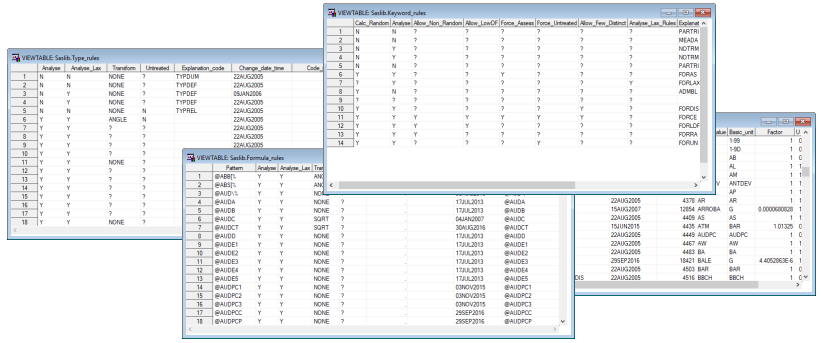
Includes coefficient of variation.

Used for registration reports.

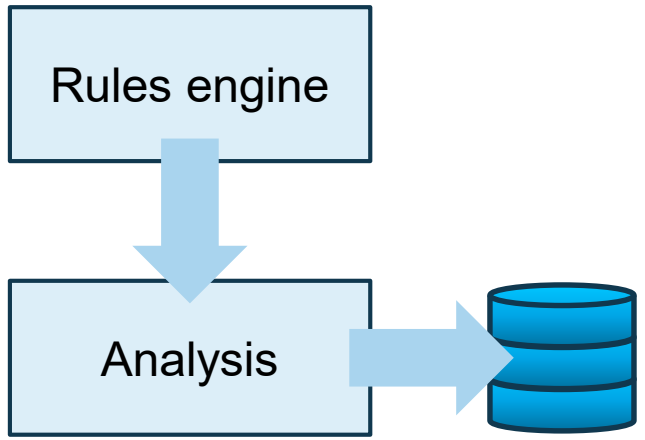
Available for query and summary.



Assessments

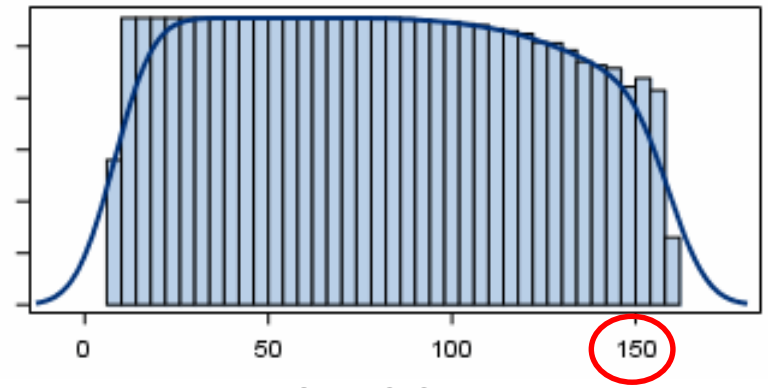


Rules tables

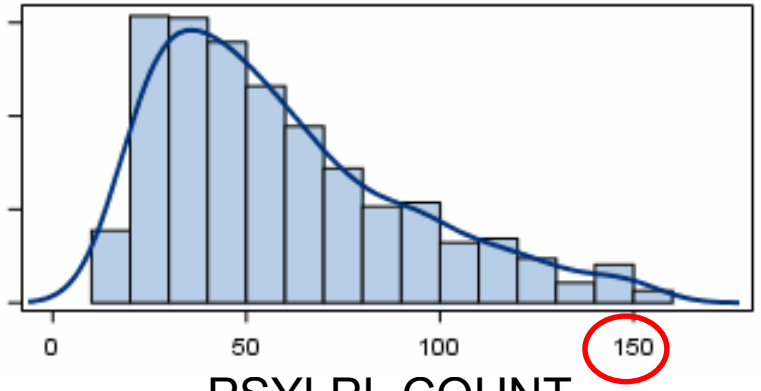


CV ranges vary depend on crop/target and assessment type

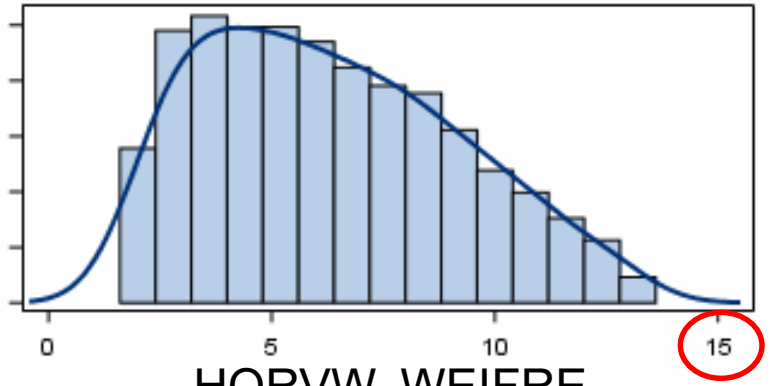
Examples of reference distributions



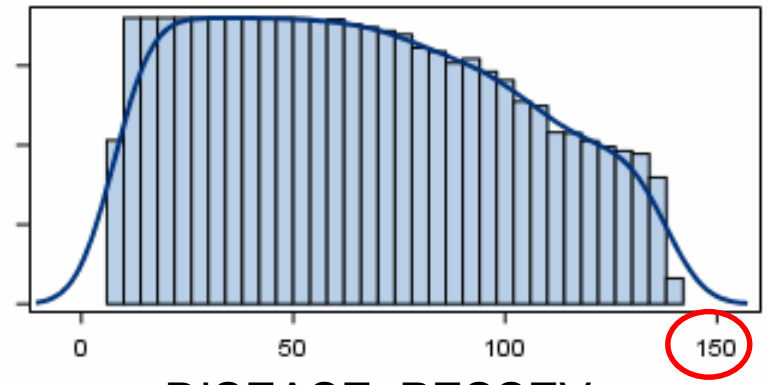
PEST, COUNT



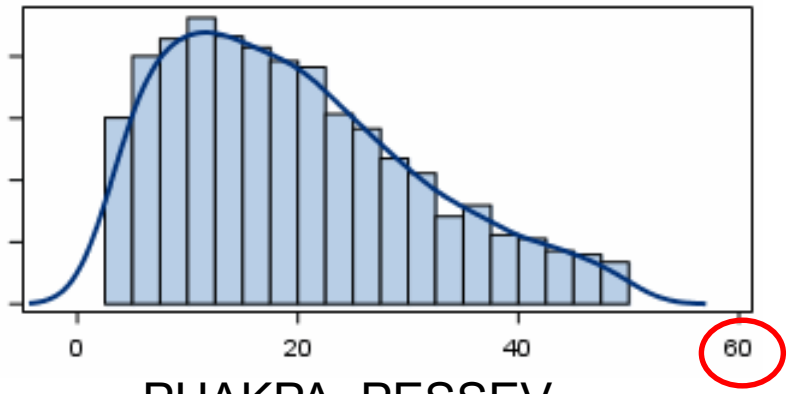
PSYLPI, COUNT



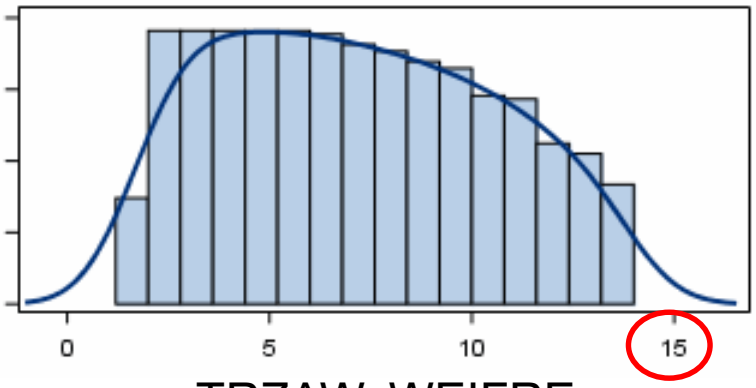
HORVW, WEIFRE



DISEASE, PESSEV



PHAKPA, PESSEV



TRZAW, WEIFRE

CV inspector tool

Use key criteria including crop, target, assess type, assess unit to group CV values

Inspects Excel tables generated by Scout.

// Works with a standard table template.

// Recognises common field labels.

// Finds assessment descriptions automatically.

// Can enter own field labels or assessment descriptions.

Buttons step over sheets and assessments.

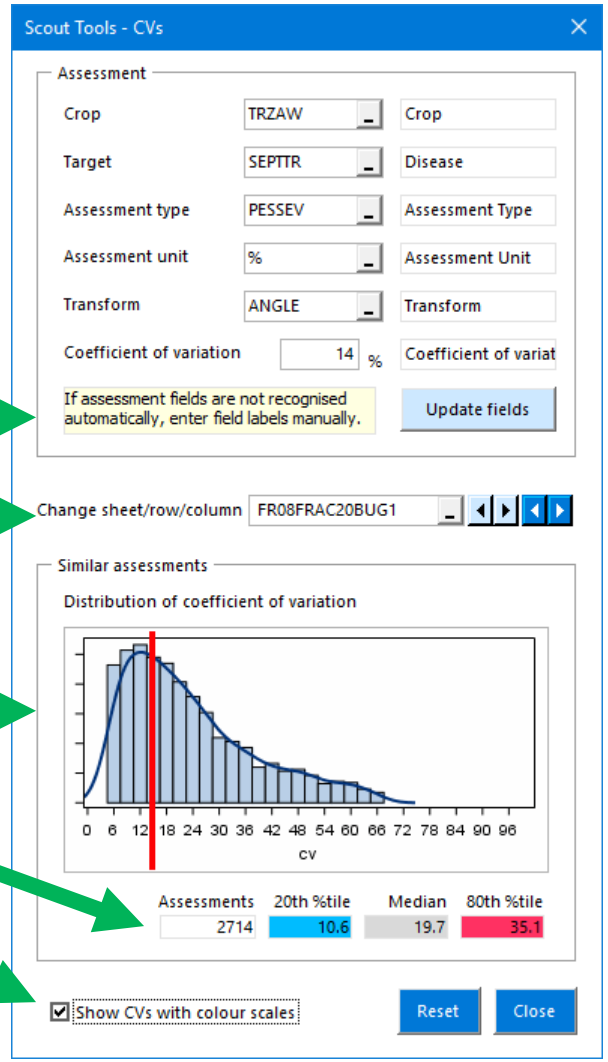
Compares CV with similar assessments.

// Detailed: Histogram with current CV marked.

// Statistics: 20th percentile, median, 80th percentile.

// Overview: Format the Excel table with colour scales.

Live demonstration.



CV inspector dashboard

Allows direct database connection

Look up a TPT ID

fa19

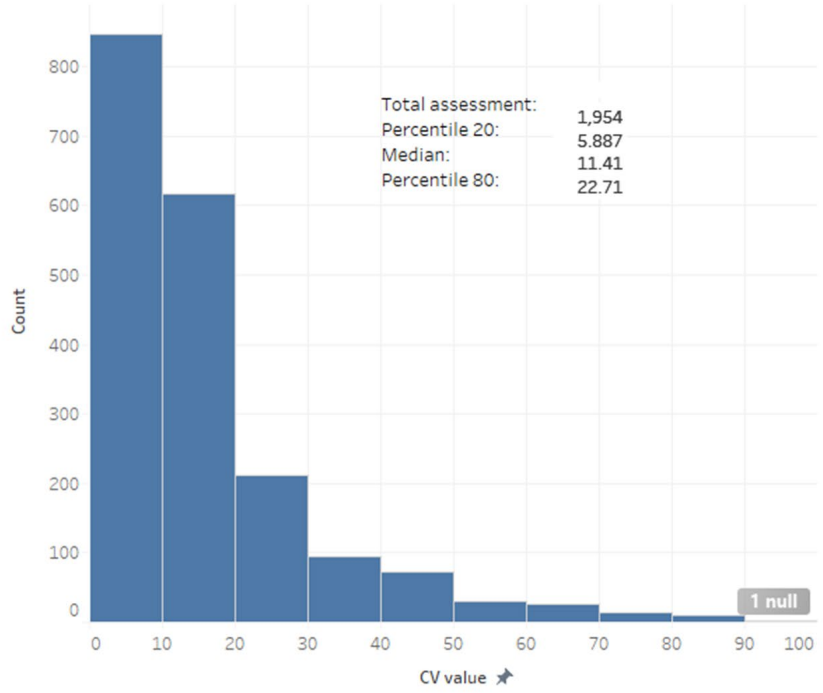
trial	column	crop	target	assess_type	assess_unit	transfor..	
FA19ARGOZIGM01	1	GLXMA	SEPTGL	PESINC	%	ANGLE	12.2
	3	GLXMA	SEPTGL	PESSEV	%	ANGLE	16.2
	4	GLXMA	SEPTGL	ABBOTT	%	ANGLE	16.2
	6	GLXMA	SEPTGL	PESSEV	%	ANGLE	14.5
	7	GLXMA	SEPTGL	ABBOTT	%	ANGLE	14.5
	8	GLXMA	SEPTGL	DEFOLI	%	ANGLE	3.4
	12	GLXMA	SEPTGL	PESSEV	%	ANGLE	22.5
	13	GLXMA	SEPTGL	ABBOTT	%	ANGLE	22.5
	14	GLXMA	SEPTGL	PESSEV	%	ANGLE	16.6
	15	GLXMA	SEPTGL	ABBOTT	%	ANGLE	16.6
	16	GLXMA	SEPTGL	PESSEV	AUDPC	NONE	22.5
	17	GLXMA	SEPTGL	PESSEV	SAUDPC	NONE	22.5
	18	GLXMA	SEPTGL	ABBOTT	%	NONE	22.5
FA19ARGOZISA01	2	GLXMA	SEPTGL	PESSEV	%	ANGLE	11.2
	3	GLXMA	SEPTGL	ABBOTT	%	ANGLE	11.2
	4	GLXMA	SEPTGL	PESSEV	%	ANGLE	11.4
	5	GLXMA	SEPTGL	ABBOTT	%	ANGLE	11.4
	6	GLXMA	SEPTGL	PESSEV	%	ANGLE	13.0
	7	GLXMA	SEPTGL	ABBOTT	%	ANGLE	13.0
	8	GLXMA	SEPTGL	DEFOLI	%	ANGLE	14.6
	14	GLXMA	SEPTGL	PESSEV	AUDPC	NONE	16.0
	15	GLXMA	SEPTGL	PESSEV	SAUDPC	NONE	16.0
	16	GLXMA	SEPTGL	ABBOTT	%	NONE	16.0
FA19ARGOZISM01	5	GLXMA	SEPTGL	PESSEV	%	ANGLE	10.1
	6	GLXMA	SEPTGL	ABBOTT	%	ANGLE	10.1
	8	GLXMA	SEPTGL	PESSEV	%	ANGLE	16.1
	9	GLXMA	SEPTGL	ABBOTT	%	ANGLE	16.1
	10	GLXMA	SEPTGL	DEFOLI	%	ANGLE	9.7

The view is filtered on trial, which keeps 235 of 109,359 members.
CV value for each assessment is found in last column.

Select your criteria for CV distribution

crop	target	assess_type	assess_unit	transformation
glxma	septgl	(All)	(All)	(All)

CV value distribution



The data is filtered on crop, target, assess_type, assess_unit and transformation. The crop filter keeps LYPES. The target filter keeps XANTSP. The assess_type filter keeps 228 of 308 members. The assess_unit filter keeps 84 of 80 members. The transformation filter keeps 7 of 7 members.

CV values from historical trials as a quality measure: a tool

CV-Werte von historischen Versuchen als Qualitätsmesslatte für Bewertungen: ein Werkzeug

Discussion Questions

Discussion and questions

Notes from the meeting

- Do we use all assessments?
  - Yes, from trials in the last ten years.
- How do we advise colleagues to use CV values? Are there fixed upper limits?
  - We recommend using CV values with biological expertise. There are no fixed upper limits, but with yield we exclude unusually high values.
- As there is a relationship between CV and mean, would it be possible also to include the mean?
  - In principle yes. A graphic would be possible but units and sample basis would give problems.
- A later discussion about the trend for higher yields to have lower CV postulated the possible applicability of Taylor's Power Law.

Thank you!

Peter Lancashire
Yinfei Li