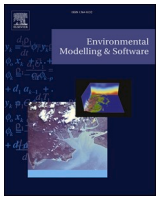
Environmental Modelling and Software 147 (2022) 105236

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft



EX-TRACT: An excel tool for the estimation of standard deviations from published articles

Marco Acutis [a], Tommaso Tadiello [a], Alessia Perego [a,*], Andrea Di Guardo [b], Calogero Schillaci [c,a], Elena Valkama [d]

[a] University of Milano, Dipartimento di Scienze Agrarie e Ambientali - Produzione, Territorio, Agroenergia. Via G. Celoria 2, 20133, Milan, Italy
[b] University of Milano Bicocca, Dipartimento di Scienze dell'Ambiente e della Terra. Piazza della Scienza 1– 20126 Milan, Italy
[c] Joint Research Centre, Sustainable Resources, Land Resources Unit, Via E. Fermi, 2749, I-21027, Ispra, Italy
[d] Natural Resources Institute Finland (Luke), Bioeconomy and environment, Sustainability Science and Indicators. Tietotie 4, 31600, Jokioinen, Finland

ARTICLE INFO

Keywords:
Data extraction
Standard deviation
Meta-analysis
Excel© tool
ANOVA

ABSTRACT

Meta-analysis, power analysis, and sensitivity analysis are widespread statistical techniques, which can be correctly performed only if variability statistics, such as standard deviation, are available; however, standard deviations are often missing in published articles. This work illustrates the functionality and the versatility of a newly developed Excel© tool for the standard deviation extraction from ANOVA and Multiple Comparison Test (MCT) results. The tool implements four methods, which can be alternatively applied according to the available statistics usually reported in ANOVA and/or MCT tables and graphs: 1) least significant difference (LSD), 2) significance level (p(F)), 3) letters for means separation assigned by MCT, 4) a range of significance level, indicated by "stars". The tool can be applied in one, two and three-way factorial experiments arranged in complete randomization, randomized block, split-plot or split-block. The performances of the different methods were tested in a case study about meta-analysis database preparation.

1. Introduction

In the scientific community, the information sharing is increasingly becoming important for enhancing knowledge integration. Open access journals and several international peer-review scientific publishers offer full access to scientific articles, so that there is full availability of data and statistical analysis results, which can be retrieved from such publications. This information can be employed for meta-analysis or for other data elaborations, such as global sensitivity analysis and the estimation of the number of replications that are needed to obtain a desired power in a new experiment (*i.e.,* power analysis). When data are retrieved from ANOVA experiments, the metric which is fundamental for performing the aforementioned types of analysis is the pooled standard deviation of the estimated means as it indicates the variability associated to the means (Koricheva et al., 2013). Note that this metric differs from the standard error associated to each mean because it is a single value that describes the pooled variability of all the means. In a one-way ANOVA experiment, the pooled standard deviation coincides with the residual standard deviation (Quinn and Keough, 2002).

Meta-analysis is a powerful statistical methodology for synthesizing research evidence across independent studies and for this reason it has increasingly gained importance in environmental and agricultural research (Philibert et al., 2012; Valkama et al., 2019). A key aspect of modern approaches to meta-analysis is weighting each study's effect size by the inverse of its variance (Borenstein, 2009; Weir et al., 2018) to avoid the bias produced by unweighted analyses (Koricheva et al., 2013). However, many published experiments failed to report sample sizes and variances-related statistics and this aspect often hampers to include these studies in a meta-analysis. With this regard, Valkama et al. (2019) reported that for a meta-analysis on nitrogen retention by buffer zones about 200 out of 246 analysed articles were not suitable for inclusion in the meta-analysis due to the lack of information about sample variances. In another meta-analysis on soil organic carbon changes due to cover cropping, 80 out of 131 articles did not report standard deviations of means (Jian et al., 2020).

Environmental and agricultural meta-analyses often attempt to include the possible largest number of articles by using alternative methods to overcome the lacking information about standard deviation.

* Corresponding author.
E-mail address: alessia.perego@unimi.it (A. Perego).

https://doi.org/10.1016/j.envsoft.2021.105236
Received 30 July 2021; Received in revised form 20 October 2021; Accepted 22 October 2021
Available online 3 November 2021

These alternative methods are often based on 1) performing unweighted meta-analysis (McDaniel et al., 2014; Jian et al., 2020), 2) replacing the missing standard deviation with a given percentage of the mean (frequently set to 10%) (Luo et al., 2010; Gattinger et al., 2012; García-Palacios et al., 2018), or 3) weighting the effect size by the sample size (Maillard and Angers, 2014; Morugán-Coronado et al., 2020). When available, information about data dispersion, such the percentage of the range of values or interquartile range, can be used to estimate standard deviation (Foscarini et al., 2010; Weir et al., 2019). The application of the first two methods is to be discouraged as the resulting meta-analysis is strongly biased due to the unvarying weight assignment. Similarly, the third method may be suitable only in the case of equal variances across the studies; however, variances are rarely equal across studies so that this method can potentially introduce serious bias (Koricheva et al., 2013; Gurevitch et al., 2018). In the meta-analysis, it was demonstrated by Hungate et al. (2009) that weighting by sample size is not a better solution than the unweighted analysis and both cause large distortion in the statistical tests.

Besides meta-analysis, global sensitivity analysis of models also requires the value of the standard deviation and the mean of the input factors as input information (Saltelli et al., 2004; Confalonieri et al., 2010; Quillet et al., 2013). Similarly, the value of the means standard deviation is fundamental for computing the replications number needed to set the power (*i.e.*, the probability to not committing a type II error) in an analysis of variance (ANOVA) or regression-based experiments (Ahrens and Pearson, 1974; Lenth, 2001; Dagnelie, 2013).

When variability metrics are reported in published articles, they are displayed in tables or graphs as error bars or written in the text. These metrics are usually referred to a specific treatment and they are reported in form of standard deviation, standard error of means or confidence intervals. These metrics are equivalent because all allow to obtain the standard deviation of a treatment through the use of the following identity:

$$s = s_{\bar{x}}\sqrt{N} \quad [1]$$

$$s = \frac{UCL - \overline{X}}{t\left(\frac{\alpha}{2}, N-1\right)}\sqrt{N} \quad [2]$$

where s = standard deviation, $s_{\bar{x}}$ = standard error of mean, N = sample size, UCL = upper confidence limit, $\overline{X}$ = mean value, $t_{\frac{\alpha}{2}}$ = Student-t value for a prefixed α level.

When these metrics are not reported, a way to retrieve or estimate variability information is to use the data reported in the ANOVA tables. This analysis is the most frequently used statistical analysis in the environmental and agricultural research (Acutis et al., 2012). Although international journals (e.g., Nature Publishing Group, https://www.nature.com/documents/nr-reporting-summary-flat.pdf) ask authors to report the standard deviations of the means or other numerically equivalent statistics and authors can rely on the use of guides for the ANOVA result presentation (McIntosh, 2015), many publications still fail in reporting variability metrics as can be deduced from published meta-analyses (Mondal et al., 2020; Jian et al., 2020; Valkama et al., 2019; Haddaway et al., 2017). For instance, Haddaway et al. (2017 – additional file n.6) reported that the percentage of missing standard deviations was 40% and 58% in the two meta-analyses about soil organic concentration and soil organic carbon stock, respectively. This missing information can be estimated as the pooled error standard deviation (s_w), which can be considered as the standard deviation of the treatments means under the assumption of the homogeneity of variances. In ANOVA and in Multiple Comparison Tests (MCT), the main outcomes are computed as a function of s_w; however, only a few authors proposed a method to retrieve s_w from these outcomes (Thiessen Philbrook et al., 2007; Koricheva et al., 2013). As summarized by Weir et al. (2018), previous studies dealt with the extraction of standard deviation when it is missing. Some authors (e.g., Abrams et al., 2005; Sung et al., 2006) made available the code for potentially implementing the extraction of standard deviations, but a self-standing tool has not been developed yet. In addition, none of the listed studies offers the options to retrieve standard deviations from experiments analysed with ANOVA, especially in complex experimental designs. This latter point is of high importance in the environmental and agricultural fields of research. As ANOVA is the most utilized test in the statistical elaboration, there is the opportunity to valorise the hidden information in many published articles. None of the published approaches allow for extracting the standard deviations from the MCT.

The objective of this study is to provide an easy-to-use Excel© worksheet to compute s_w and/or its upper and lower limits using feasible methods to obtain s_w from the ANOVA and MCT outcomes. The software tool allows the user to obtain s_w from one-, two-, and three-way ANOVA factorial experiments in completely randomized design (CRD), completely randomized block design (CRBD), split-plot, split-block or split-split-plot design. The study also provides the theoretical background of each of the four implemented methods.

2. Material and methods

2.1. Methods for estimation of s_w and its uncertainty

The EX-TRACT tool offers four methods to estimate s_w according to the available information of ANOVA or MCT results reported in the published articles. The four methods were identified because they are the statistics most often employed and reported in published articles. The first three methods were chosen as they are based on all the possible statistics that are computed in experiments analysed with ANOVA and from which it is possible to extract s_w. The fourth method was chosen because MCTs results are also frequently reported in the published studies, and it has never used before to estimate s_w.

The four methods use the following input information: *i*) the LSD (Least Significant Difference) values; *ii*) the p(F) (*i.e.*, the type I probability error) values; *iii*) the letters assignment indicating differences among means based on the result of commonly used MCT such as LSD, Tukey (1953), Duncan (1955), SNK (Keuls, 1952), and REGWQ (Ryan, 1960; Einot and Gabriel, 1975; Welsch, 1977) tests; *iv*) "stars" used as a shorthand to indicate significance levels.

All these methods estimate an uncertainty range of s_w(*i.e.*, LL_{s_w} and UL_{s_w}, being the lower and upper limits of s_w, respectively), which includes its true value. The first two methods have an uncertainty that is only due to the rounding errors in the input data. The uncertainty of the other two methods is due to the fact that they do not rely on values (such as the LSD or the p(F) value) but on the classification of letter-labelled means and ranges of probabilities, respectively.

The four methods were developed for balanced experiments (*i.e.*, equal number of replications for each treatment) under the ANOVA assumption of homogeneity of variances. The application of these methods requires to indicate the experimental design, the number of levels for each factor, which is included in an experiment, and the number of replications. The type of experimental design is needed to retrieve the degrees of freedom of the error (i.e., the degrees of freedom within the groups, *DFW*) and the number of replications used to compute mean values. Moreover, all the methods, except the one using the LSD value, also require the values of the means of the source of variation to perform the computation of s_w.

2.1.1. From LSD (method i)

This method relies on the availability of the LSD value, which is commonly reported in tables and graphs. LSD is used as a threshold to define whether two means significantly differ or not.

For balanced experiments, LSD is defined as:

$$LSD = t_{\frac{\alpha}{2},DFW}\sqrt{\frac{2s_w^2}{n}} \quad [3]$$

where $t_{\frac{\alpha}{2},DFW}$ is the Student-t value for a selected significance level α, and n is the number of replications used to compute the means (Fisher, 1935).

Consequently,

$$s_w = \sqrt{\frac{LSD^2 n}{2\left(t_{\frac{\alpha}{2},DFW}\right)^2}} \quad [4]$$

The uncertainty range of s_w depends only on the number of the significant digits which are used in reporting the LSD value or derived from extracting the LSD value from a graph. Details about the uncertainty range calculation are given in Appendix A.

2.1.2. From p(F) (method ii)

The p(F) value defines the probability level of a false rejection of a null hypothesis for a specific effect, being the source of variation under examination.

The method to compute s_w from the p(F) value is because the F-statistic is the between-groups mean square (s_b^2) over the within-groups mean square (s_w^2) ratio. In a one-way balanced ANOVA,s_b^2 is computed as:

$$s_b^2 = n\sum_{i=1}^{N_t}\left(\overline{X}_i - \overline{\overline{X}}\right)^2 \quad [5]$$

where N_t is the number of treatments, $\overline{X}_i$ is the i_{th} treatment mean, and $\overline{\overline{X}}$ is the grand mean.

Then, s_w is computed as:

$$s_w = \sqrt{s_b^2 / invF_{(\alpha^*,DFB,DFW)}} \quad [6]$$

where *invF* is the inverse of the F cumulative distribution function, *DFB* is the between-groups degrees of freedom and α^* is the available value of p(F). Formulas for calculating s_b^2 for the main effects and interactions in multi-way experiments can be found in basic statistics textbooks (*e.g.*, Sokal and Rohlf, 2012).

The uncertainty in the estimate of $s_w$is due to the number of the significant digits used to report the means of treatments and the p(F) value. The procedure to assess the uncertainty of $s_w$estimation is reported in appendix B.

2.1.3. From Letters (method iii)

MCT results are commonly presented by labelling means in graphs and tables with the same letter if they are not significantly different (Quinn and Keough, 2002). This method can be applied if letter-labelled means are available in a published article. It computes two finite boundaries including the true value of s_w when at least two means are labelled with different letters and two means are labelled with the same letter. In the case of no significant differences between means (all the means share the same letter), only a minimum value of $s_w$can be defined, being the upper boundary equal to ∞. When all the means are classified as different (*i.e.*, all the means are labelled with different letters), only an upper boundary could be determined, being the lower boundary equal to zero.

This method implements five MCTs: LSD, Tukey, Duncan, Student-Newman-Keuls (SNK), Ryan-Einot-Gabriel-Welsh Studentized Range Q (REGWQ). It works in three steps for all the MCTs. In case of LSD, the procedure is as follows:

1) to seek for the minimum value of a difference between means that is declared significantly different (*i.e.*, to choose the minimum difference between two means that do not have any letter in common). Assuming this value as the LSD, s_w is calculated according to method i and represents the upper limit (UL_{s_w}) for s_w estimation. This is the highest value of s_w that is possible to assume.
2) to seek for the maximum difference that is declared not significantly different (*i.e.*, the maximum difference between two means that have at least one letter in common) and to use it as LSD value to compute s_w (according to the method i). This is the lowest value that is possible to assume, so it represents the lower limit (LL_{s_w}) for s_w;
3) to decide what s_w outcome is the most suitable. When the user intends to make a conservative choice, the user has to take the UL_{s_w} as the s_w value. Conversely, the mean between UL_{s_w} and LL_{s_w}, or the LL_{s_w} value can be taken for a more liberal choice.

When Tukey test is used, the value of LSD is replaced by the so called "honest LSD" (*i.e.*, HSD). This value is computed as:

$$HSD = Q_{\alpha,k,D.F.E.}\sqrt{s_w^2/n} \quad [7]$$

where Q is the studentized range distribution and k is the number of all the means included in the experiment. Consequently:

$$s_w = \sqrt{n\frac{HSD^2}{Q_{\alpha,k,D.F.E.}^2}} \quad [8]$$

For the SNK, Duncan, and REGWQ tests, being step-down tests, s_w is computed from:

$$s_w = \sqrt{n\frac{C\_Val^2}{Q_{\alpha,k,D.F.E.}^2}} \quad [9]$$

where

$$C\_Val = Q_{\alpha^*,p,D.F.E.}\sqrt{s_w^2/n} \quad [10]$$

with p being the number of means whose range is to be sequentially tested and a^* the adjusted significance level for a test of the equality of p means (Day and Quinn, 1989). The a^* is differently defined for each test. For the SNK test, $\alpha^* = \alpha$; for the Duncan test, $\alpha^* = 1-(1-\alpha)^{p-1}$; for the REGWQ test, $\alpha^* = 1-(1-\alpha)^{p/k}$.

To calculate the Q value for the a^* probability for each MCT, we used the theoretical method proposed by Gleason (1998, 1999) and implemented in an Excel© code snippet proposed by Klasson (2018). An exact value for s_w could not be obtained with this method. Only the limits of the s_w range are computed with this method according to the difference between the means labelled with different and common letters. In particular, the uncertainty is reduced when the values of means are close enough to detect significant and non-significant differences within a narrow range.

2.1.4. From Stars (method iv)

Frequently, in scientific articles, simple indication like "*" or "**" are reported instead of the p(F) value. These symbols can have different correspondence to the significance level in different disciplines, journals, and statistical packages. Therefore, this method requires to know what level of significance corresponds to the symbol used. Three conventional significance codes associating p(F) and stars are proposed in the tool (Table 1), being the most common cases in scientific and technical literature. It is possible to obtain a bounded estimate of s_w only when the reported stars indicate a range of p(F); otherwise, one of the limits is zero or ∞ (Table 1).

Table 1
"From Stars" (method iv): lower (LL_{s_w}) and upper (UL_{s_w}) boundaries of the s_w estimation for each pair of "stars" and the associated p(F) value or range. Three different possible assignments of "stars" are implemented in the tool (i.e., significance code 1, 2, 3).

	Symbol	p(F)	LL_{s_w}	UL_{s_w}
Significance code 1:	Ns	>0.1	s_w for p(F) = 0.1	∞
	*	0.1 - 0.05	s_w for p(F) = 0.05	s_w for p(F) = 0.1
	**	<0.05	0	s_w for p(F) = 0.05
Significance code 2	Ns	>0.05	s_w for p(F) = 0.05	∞
	*	0.05 - 0.01	s_w for p(F) = 0.01	s_w for p(F) = 0.05
	**	<0.01	0	s_w for p(F) = 0.01
Significance code 3	Ns	>0.05	s_w for p(F) = 0.05	∞
	*	0.05 - 0.01	s_w for p(F) = 0.01	s_w for p(F) = 0.05
	**	0.01 - 0.001	s_w for p(F) = 0.001	s_w for p(F) = 0.01
	***	<0.001	0	s_w for p(F) = 0.001

2.2. Tool applicability

The EX-TRACT tool allows to operate with the most common experimental designs: one-, two- and three-way ANOVA in complete randomized design (CRD), complete randomized block design (CRBD), split-plot, split-split plot, or split-block design (Table 2). The tool allows to work with a maximum of 16 means (being the levels of the source of variation) with an unlimited number of replications. In factorial experiments (*i.e.*, where all possible main effects and interactions in multi-way experiments), to obtain the exact value or range of s_w, the EX-TRACT tool also requires the selection of the desired/available source of variation. The methods "from p(F)" and "from Stars" (method *ii* and *iv*) can be used for all experimental main factors and interactions, while methods "From LSD" and "From Letters" (method *i* and *iii*) can be used only in cases reported in Table 2.

Table 2
Type of experimental designs treated by the tool. The source of variation column indicates the cases treated in "From LSD" and "From Letters" (method i and iii). Note that the methods "From p(F)" and "From Stars" (method ii and iv) allow to estimate s_w for all sources of variation in all experimental designs.

Type of Experiment	Number of experimental factors	Source of variation allowed for *"From LSD"* and *"From Letters"* methods
Complete Randomized design (CRD)	1	A
	2	A, B, A × B
	3	A, B, C, A × B, A × C, B × C, A × B × C
Complete randomized block design (CRBD)	1	A
	2	A, B, A × B
	3	A, B, C, A × B, A × C, B × C, A × B × C
Split-Plot	2	A, B
Split-plot A x B main C sub[(1)]	3	A, B, C, A × B
Split-plot A main B x C sub[(2)]	3	A, B, C, B × C
Split-block	2	A, B
Split-split-plot	3	A, B, C

[1] Split-plot experiment with a factorial combination of two factors in the main-plots.
[2] Split-plot experiment with a factorial combination of two factors in the sub-plots.

2.3. EX-TRACT tool implementation

We chose to code the EX-TRACT tool in the Excel© environment following the "user-centered design (UCD)" paradigm (Barnum, 2011), which is an ISO standard (9241-210:2019), and it is based on a participatory approach. According to this, a panel of users (i.e., PhD students, junior and senior researchers) were involved in order to provide feedbacks in an iterative design process. The most reported requirement was the need of simple tool. With this regard, the Excel© software is a well-known environment and, therefore, users of different backgrounds can use the tool functionalities in an efficient way, without the need of knowing dedicated statistical packages. Such packages are more complex than a spreadsheet-based solution as they require coding skills and in-depth knowledge about the statistics theory behind each test. Moreover, the EX-TRACT tool allows for the visualization of input and output within the same page and for an automatic saving of the extraction results. The user interface of the EX-TRACT tool has a clear arrangement of the input and output data and includes a contextual help system. A detailed user manual is available in the Supplemental Material (S1). Finally, we have also provided a set of video tutorials to present the main features of the tool.

EX-TRACT results of the $s_w$extraction can be automatically saved in a sheet, which is structured as a database.

2.4. Evaluation of the tool utility: a case study

To evaluate the EX-TRACT tool capability in extracting s_w we applied it under the hypothesis of increasing the number of articles which can be used for carrying out a meta-analysis. Therefore, we created a database, which can be potentially used in a meta-analysis for detecting the effect of conservation agriculture on soil organic carbon stock.

According to Koricheva et al. (2013), a four-step procedure was adopted: (1) to perform a systematic search in scientific bibliography databases (*i.e.*, Scopus and WoS); (2) to select the studies that are potentially suitable for a meta-analysis, in which measurements were available from clearly defined controls and treatment groups; (3) to detect the standard deviation of control and treatment from a given study (note that an article may report more than one study); (4) to identify articles that were selected in the second step but did not report explicitly the standard deviation of control and treatment; in this case, it may possible to overcome this missing information by estimating s_w on the basis of the available ANOVA model features using the EX-TRACT tool.

For the evaluation of methods performance, we used the two coefficients of variation of the s_w upper and lower limits ($CV_{lower}$and CV_{upper}), which were calculated as follows:

$$CV_{lower} = LL_{s_w} \Big/ \left(\overline{X}_{ctrl} + \overline{X}_{treat}\right) \Big/ 2\% \quad [11]$$

$$CV_{upper} = UL_{s_w} \Big/ \left(\overline{X}_{ctrl} + \overline{X}_{treat}\right) \Big/ 2\% \quad [12]$$

where LL_{s_w} an UL_{s_w} are the lower and upper limits of s_w, while $\overline{X}_{ctrl}$ and $\overline{X}_{treat}$ are the observed means of the control and the treatment, respectively.

To assess the standard deviation uncertainty when data are directly retrieved, we used the rounding error, while we used the digitizer error along with the rounding error when data were extracted from a graph.

3. Illustrative results and discussion

The main result of the present work is a tool that estimates s_w from ANOVA experiments and MCT. This tool implements and extends the s_w estimation approaches already proposed by Thiessen Philbrook et al. (2007) and Koricheva et al. (2013) to a wide range of experimental

designs (including 3-way ANOVA), which are typically used in the environmental sciences. Moreover, the tool implements new and original methods to estimate s_w on the basis on MCT results.

The following paragraphs describe the interface, the code validation, the performance evaluation of the four implemented methods, and the results of a case study.

3.1. Tool interface

As a result of the UCD approach, the developed Excel© spreadsheet shows an easy-to-use interface. In the EX-TRACT tool interface, input and output are displayed together in one page. Here, the user is guided in the data entry phase and supported by a help system. Drop-down lists offer the available choices for each input feature and an automatic check ensures the input data consistency.

The tool consists of six sheets: an introductory screen with general information and method selection (Fig. 1), four sheets for the extraction of s_w value and/or limits (*i.e.*, "From LSD", "From p(F)", "From Letters" and "From Stars"), and the Database sheet in which the article or experiment reference are automatically stored along with the associated output.

3.1.1. Input

The four s_w extraction methods require the following data as common input (Fig. 2): experiment type, source of variation, number of levels of each factor of the ANOVA experiment, and the number of replications. Note that the orange colour of the cell indicates a drop-down list of available options.

Besides the common input, each method requires specific inputs. The "From LSD" (method *i*), requires the value of the LSD and the probability level used to compute it. The "From p(F)" (method *ii*) requires the p(F) value, the mean values of the levels of the source of variation (MLs) under analysis, and the means value approximation (rounding). The "From Letters" (method *iii*) requires MLs, the associated letters, the type of MCT, and the significance level used to perform the MCT. The "From Stars" (method *iv*) requires MLs and the p(F) value or range which corresponds to the star significance codes adopted in the article.

Details about the specific input required for each method are reported in the user manual (Supplemental material, S1) and in the video tutorials available in the EX-TRACT YouTube channel (https://www.youtube.com/channel/UCKy_xnOpYHS0PcR5exhSv-A).

3.1.2. Output

Along with the estimated s_w and its range, the tool returns the degrees of freedom and the number of replications used to compute the means. The s_w is differently reported in the output section according to the different approaches described in paragraph 2.1. For LSD and p(F) methods (Fig. 3A) the output shows the estimated values of s_w with the LL_{s_w} and UL_{s_w} (in the interface named "Min" and "Max" respectively), while for "Letters" and "Star" only LL_{s_w} and UL_{s_w} are reported (Fig. 3B).

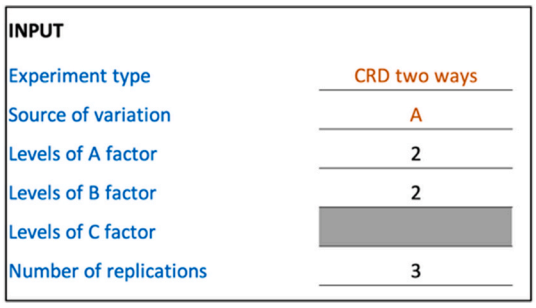


Fig. 2. Common input to the four methods to estimate standard deviation of means (s_w). This example shows a two-way CRBD experiment with two levels of the factors A and B, and three replications. The source of variation is set to the A factor. Regardless the method, the orange colour of a cell indicates a dropdown list of available options.

3.2. Tool validation and debugging

The correct coding of the algorithms in the tool was validated according to the following procedure:

1) generation of a set of synthetic experiments with s_w known according to the different experimental schemes handled by the tool (*e.g.*, RCDB, split-plot, etc) varying the number of factors (up to three) and levels;
2) performing an ANOVA analysis and different MCT for each experiment using the IBM-SPSS statistical package (version 26.1);
3) applying the EX-TRACT tool to compute the pooled s_w based on the data of the same experiments.

In the debugging phase, we performed a comparison between the tool output and the known value of s_w, which proved that the tool implementation was correct; more than 200 simulated experiments were analysed before the tool application to real cases.

3.3. Methods performances estimation

The $s_w$estimation from LSD (method *i*) is the most reliable method because of it is an exact computation of the standard deviations: the only uncertainty comes from the precision (in terms of the number of significant digits) used to express the LSD value. Most of the time the LSD value is reported with two or more digits: in this case, the uncertainty is for sure acceptable (see Appendix 1) for practical determination of standard deviation (for meta-analysis, sample size determination or sensitivity analysis).

As for the first method, the estimation of s_w computed from the p(F) (method *ii*) value follows an exact procedure (see Appendix 2 for the computation of uncertainties of the procedure). The p(F) and means values expressed with only one significant digit produce large

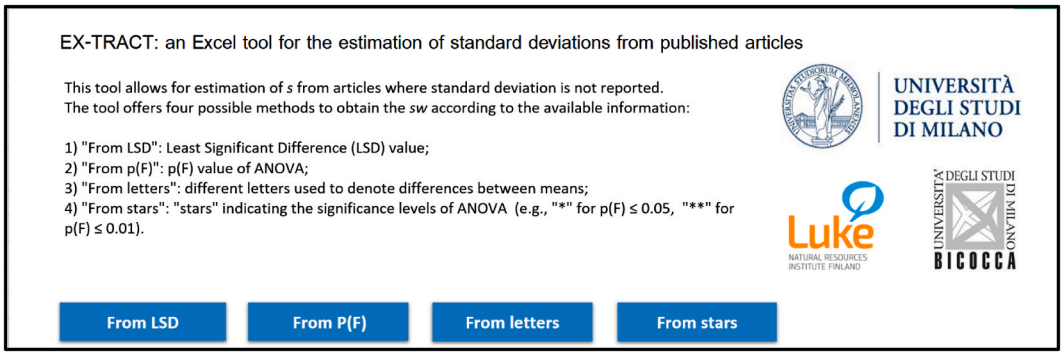


Fig. 1. The EX-TRACT tool introductory screen page with the selection of the methods to estimate the standard deviation of means.

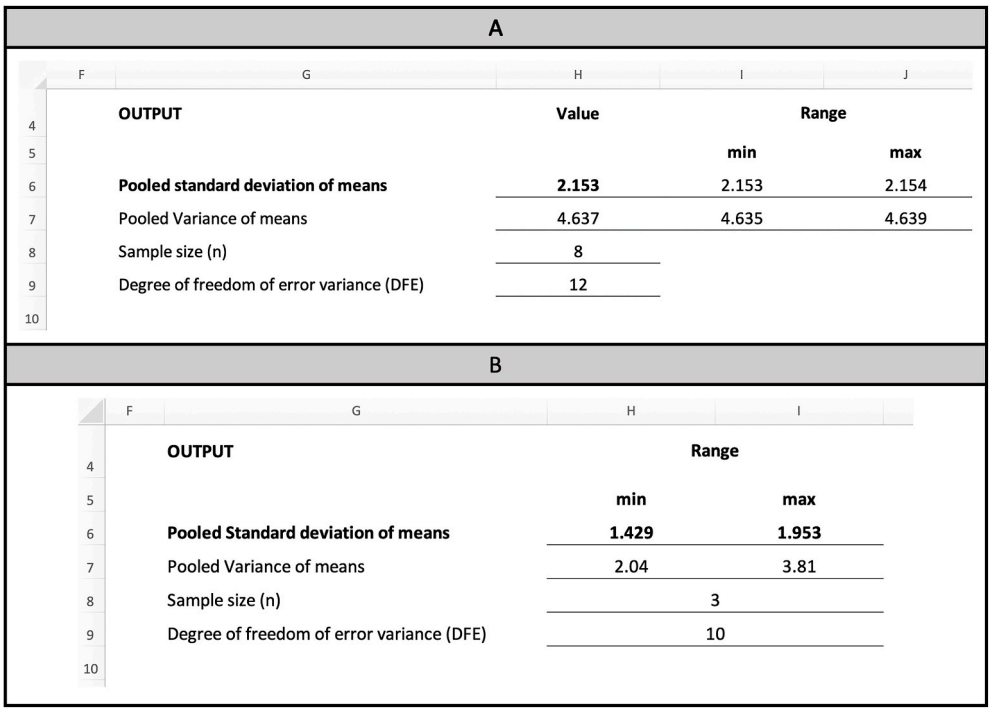


Fig. 3. A) Output for LSD (method i) and p(F) (method ii), B) Output for "Letters"(method iii) and "Stars" (method vi).

uncertainty in s_w estimation, while when they are expressed with two significant digits uncertainty is low in most of the cases, as already stated by Thiessen Philbrook et al. (2007) in medical field.

Obtaining the estimation of s_w from "Letters" (method *iii*) is an interesting perspective because MCT comparisons and the consequent letters assignment are widespread in scientific publications, even if the uncertainty of estimation is unpredictable, depending mainly on the distribution of the treatments means. The user is then free to alternatively keep the central value of the s_w range or the upper limit in the case a conservative approach is required.

Two cases can also arise while using this method: *a)* all the means are declared as different so that the lower limit of s_w results equal to zero and *b)* all the means are declared not different, so that the upper limit of s_w results equal to ∞. In the former case the conservative approach suggests using the upper limit of s_w. In the latter case, the user can follow a liberal approach and keep the lower limit of s_w. To our knowledge this is the first time that a method to estimate s_w from the multiple comparison letters attribution is proposed.

The method *iv* "Stars" is the option that yields the highest uncertainty for s_w estimation due to the limited input information (*i.e.*, p(F) below a threshold or included between two values). As for this method, the greatest input information is given when stars indicate a p(F) included within two boundaries (*e.g.*, when "*" indicates 0.01 <p(F) < 0.05) which allows the method to compute two defined limits (both different from 0 and ∞). In the other cases this method can compute only one $s_w$boundary. Similarly, to the "Letters" method, the uncertainty of the estimation is unpredictable, being dependent on the values of the means and significance levels available from the ANOVA analysis.

3.4. Tool utility: a case study

The case study concerned the database creation for a meta-analysis regarding the effects of conservation agriculture on soil fertility (Table 3). To populate the database, standard deviations of the means were directly retrieved from published articles or estimated by the EX-TRACT tool. The final database consisted of 43 studies published in literature that meet the meta-analytic requirements listed by Koricheva et al. (2013). Studies from 1 to 18 reported the standard deviations of treatments in the text, in tables, or as graphical bars. Studies from 19 to 43 required EX-TRACT to compute thes_w. In the latter case, the "Method" column defines the specific tool method utilized (Table 3). Means of control and treatment, s_w data (the exact value or the LL_{s_w} and UL_{s_w}), and coefficient of variation (CV) of the two boundaries are also displayed in Table 3.

This example showed that EX-TRACT tool allowed to obtain the minimum number of studies (*i.e.,* 30–40) required to perform a meta-analysis (Valentine et al., 2010).

The global influence of a single data in a meta-analysis depends on the ratio of s_w over mean (*i.e.,* the coefficient of variation CV); the power analysis also relies on the use of CV.

As shown in Table 3, CV% had a considerable variation across the 43 experiments, (from 0.5% to 34.9%). The lowest is the CV% range (CV % (UL_{s_w}) - CV % (LL_{s_w})) the most reliable the s_w estimate. The extent of the CV range depends on the method of extraction. Direct and LSD methods allow to obtain narrow CV ranges for which the maximum value of range was 0.04% and 0.1%, respectively. When standard deviation is estimated from a graph, the uncertainty is still low with a maximum range value of 0.33%.

As for the p(F) method, only one experiment was found; however, the result of p(F) indicates a low CV range (0.26%), and this agrees with the fact that is an exact procedure, leading to a $s_w$value with limited uncertainty.

With the s_w extraction from "Letters", it was possible to estimate the uncertainty (*i.e.*, the CV range) in seven out of ten studies, since in three cases one of the limits was zero or ∞. This method performed worse than in the LSD and p(F) methods, having a mean CV range of 4.21%.

With the s_w extraction from "Star", it was possible to estimate the uncertainty in three out of six studies, since in three cases one of the limits was zero or ∞. Due to the limited input information used with the "Star" method, the mean CV range (9.41%) was much higher than the other methods.

Given that the uncertainty obtained using "Letters" and "Star" was larger, the conservative approach suggests using the upper limit of the estimated s_w. Although this conservative approach reduces the weight of

Table 3
Pooled standard deviation of means (s_w) of the studies included in the case study (meta-analysis about the effect of conservation agriculture on soil organic carbon). The table reports in the first 18 rows the articles that had the standard deviation being directly reported (in text, in tables, or as graphical bars); in this case, this statistic is the standard deviation of the single treatment. The observations from 19 to 43 are those that required the EX-TRACT tool to compute the s_w. In the latter case the "Method" column defines the specific tool method utilized. Means of control and treatment, the s_w data (the exact value or the LL_{s_w} *and* UL_{s_w}), and the coefficient of variation of the two boundaries are also displayed.

Study number	Method	Control Mean	Treatment Mean	UM	s_w [(1)]	LL_{s_w} [(1)]	UL_{s_w} [(1)]	CV % (LL_{s_w})	CV % (UL_{s_w})
1	direct	33.23	35.55	t ha^{-1}	**3.27**			9.05	9.08
2	direct	54.61	70.57	t ha^{-1}	**5.62**			8.31	8.33
3	direct	38.00	40.94	t ha^{-1}	**1.48**			3.34	3.36
4	direct	29.13	32.59	t ha^{-1}	**2.41**			7.44	7.47
5	direct	44.72	43.32	t ha^{-1}	**12.20**			25.65	25.68
6	direct	23.10	21.68	t ha^{-1}	**2.33**			9.62	9.66
7	direct	68.06	74.71	t ha^{-1}	**5.74**			7.19	7.20
8	direct	48.95	49.43	t ha^{-1}	**3.00**			5.44	5.46
9	direct	51.50	54.64	t ha^{-1}	**2.88**			5.16	5.18
10	direct	68.10	66.70	t ha^{-1}	**1.77**			2.34	2.35
11	direct	32.39	40.28	t ha^{-1}	**1.88**			4.61	4.64
12	from graph	22.26	28.23	t ha^{-1}	**3.42**			12.30 [(2)]	12.80 [(2)]
13	from graph	22.67	26.97	t ha^{-1}	**2.49**			9.48 [(2)]	9.87 [(2)]
14	from graph	26.83	32.26	t ha^{-1}	**3.18**			9.44 [(2)]	9.82 [(2)]
15	from graph	28.92	34.35	t ha^{-1}	**1.81**			5.03 [(2)]	5.23 [(2)]
16	from graph	43.52	47.33	t ha^{-1}	**4.62**			9.74 [(2)]	10.14 [(2)]
17	from graph	21.91	21.04	t ha^{-1}	**0.48**			1.94 [(2)]	2.02 [(2)]
18	from graph	3.52	3.92	t ha^{-1}	**0.34**			7.99 [(2)]	8.31 [(2)]
19	LSD	10.63	14.21	t ha^{-1}	**0.53**	0.41	0.76	3.30	6.16
20	LSD	55.0	58.8	t ha^{-1}	**1.80**	1.79	1.80	3.15	3.16
21	LSD	8.26	13.44	t ha^{-1}	**1.36**	1.35	1.36	12.48	12.52
22	LSD	5.96	19.38	t ha^{-1}	**1.45**	1.44	1.45	11.40	11.44
23	LSD	15.67	17.3	t ha^{-1}	**1.05**	1.04	1.05	6.33	6.36
24	LSD	9.20	12.89	t ha^{-1}	**2.62**	2.61	2.63	23.66	23.84
25	LSD	3.61	6.05	t ha^{-1}	**0.42**	0.42	0.42	8.60	8.69
26	LSD	0.83	3.13	%	**0.69**	0.68	0.69	34.59	34.88
27	p(F)	21.3	21.8	t ha^{-1}	**3.18**	3.15	3.21	14.62	14.88
28	Letter	0.86	0.98	%		0.01	**0.02**	1.49	2.24
29	Letter	51.1	46.6	t ha^{-1}		0.24	**2.00**	0.49	4.10
30	Letter	36.7	41.1	t ha^{-1}		**2.57**	∞	6.61	
31	Letter	15.73	16.82	t ha^{-1}		0.08	**0.47**	0.49	2.86
32	Letter	50.57	54.29	t ha^{-1}		1.62	**4.41**	3.09	8.41
33	Letter	27.76	31.08	t ha^{-1}		0.00	**1.45**		4.92
34	Letter	1.04	2.46	%		0.06	**0.19**	3.43	10.68
35	Letter	13.8	15.8	t ha^{-1}		0.00	**0.53**		3.59
36	Letter	76	81	t ha^{-1}		11.91	**15.11**	15.17	19.25
37	Letter	0.65	1.62	%		0.04	**0.11**	3.88	9.97
38	Stars	3.37	6.49	t ha^{-1}		0.39	**0.89**	7.83	18.06
39	Stars	4.46	8.26	t ha^{-1}		0.92	**1.69**	14.47	26.57
40	Stars	13.6	22.5	t ha^{-1}		0.00	**4.86**		26.90
41	Stars	13.20	15.34	t ha^{-1}		**0.94**	∞	6.61	
42	Stars	44.7	47.1	t ha^{-1}		**0.92**	∞	2.01	
43	Stars	1.37	1.81	%		0.04	**0.13**	2.34	8.22

(1) Bolded values were utilized to compute the weight of the effect size in the meta-analysis.
(2) Uncertainties of SD extraction in agreement with Drevon et al. (2017).

a specific study in the meta-analysis, it allows to include the study in the analysis. This approach is recommended by Weir et al. (2018), who studied the effect of different approach of standard deviation on the reliability of meta-analysis results. The same authors recommended to use several methods, which led to standard deviation values comparable to those estimated in our study with "Letter" and "Stars" methods. As a final consideration, we encourage the user to apply the best method, being LSD and p(F) methods when the required input information is available.

4. Conclusions

Standard deviation is a strictly required statistic for proper data processing, such as meta-analysis, sensitivity analysis, and power analysis. The present work illustrates the functionality of an Excel© tool (EX-TRACT), which allows to estimate the pooled standard deviation of existing experimental datasets when this value is not directly reported. This tool is capable to extract standard deviation from ANOVA and post-ANOVA multiple comparison tests, being the most common test in the environmental research, by applying four methods, which can be chosen according to the availability of one of the following information: LSD values, p(F) values, letters used to classify means (five multiple comparison tests are managed by the tool), and stars attribution to indicate significance level.

The flexibility of the EX-TRACT tool is also ensured by the wide range of experiment types (one-, two-, three-ways) and design schemes (CRD, CRDB, Split-plot, Strip-block), in which standard deviation can be extracted.

The LSD and p(F) methods are capable to estimate standard deviation with a defined level of uncertainty, which is comparable to the one directly reported in tables or graphs. Although the other two methods, "Letters" and "Stars", are associated with a higher degree of uncertainty, they offer the possibility to derive standard deviation from those datasets, which would be otherwise neglected.

The application to a real case demonstrates that the tool allows to double the number of studies that can be included in a meta-analysis. Therefore, the EX-TRACT tool offers an operational facilitation in meta-analytic research, valorising the hidden information in published

articles, increasing the number of suitable studies, and avoiding questionable procedure of study weighting.

Software availability

Name of Software: EX-TRACT.
Developers: Andrea Di Guardo (andrea.diguardo@unimib.it), Marco Acutis (marco.acutis@unimi.it)
Year first available: 2021.
Language: VBA for Excel version 2016, 2019 and 365.
Availability: https://doi.org/10.6084/m9.figshare.14987130.
Supported systems: Microsoft Windows, macOS.
Licence: CC BY 4.0.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is a part of the project SOMMIT: Sustainable management of soil organic matter to mitigate trade-offs between C sequestration and nitrous oxide, methane, and nitrate losses that received funding from the European Joint Programme SOIL (grant agreement ID: 862695). This study is also funded by the European Union's Horizon 2020 Framework Programme for Research and Innovation (H2020-RUR-2017-2) as part of the LANDSUPPORT project (grant agreement No. 774234), which aims at developing a decision support system for optimizing soil management in Europe. The work is also financially supported by the Doctoral School of Agriculture, Environment, and Bioenergy and of the University of Milan.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envsoft.2021.105236.

Appendix A. Uncertainty calculation in the estimation of s_w when LSD value is available (method *i*)

a) Let LSD_APX be the approximation value of LSD that is automatically computed by the software, based on the number of significant digits of LSD (*e.g.*, 0.5)
b) Let min_LSD = LSD - LSD_APX such as LSD = *round*(LSD,-APX)
c) Let max_LSD = LSD + LSD_APX such as LSD = *round*(LSD,+APX)

The results of b) and c) are then plugged into the equation [4] to obtain the lower and the upper limits of s_w.

Note that in this case the ratio ($R = ((LL_{s_w} - LL_{s_w})/s_w)*100$) depends only on the number of significant digits used to report LSD (or from reading the length of LSD if the value is obtained from a graph) and from the LSD value itself. Defining LSD* as the significant digits of the LSD written in the significant normalized scientific notation (Fleisch and Kregenow, 2014) without the decimal point, the R value is given by 100/LSD*. The tool considers the final zeroes of a number (integer or decimal) as significant.

For instance, if the value of the LSD is 1 (the normalized scientific notation is $1*10^0$ and so LSD* = 1), the R is 100%. If the LSD is 2.2 (the normalized scientific notation is $2.2*10^0$ and LSD* = 22), the R is 4.5%. If LSD is 99.9 (the normalized scientific notation is $9.99*10^1$ and LSD* = 999), the R is 0.001%.

Appendix B. Uncertainty calculation in the estimation of s_w when p(F) value is available

Let APX be the approximation of mean values set by user (*e.g.*, 0.005).
For each $\overline{X}_i$,

a) let min_ $\overline{X}_i$ be the smallest value such as the rounding of $\overline{X}_i$ is obtained using the approximation value APX, *i.e.*, *round*(min_ $\overline{X}_i$, -APX) = $\overline{X}_i$
b) let max_ $\overline{X}_i$ be the greatest value such as the rounding of $\overline{X}_i$ is obtained using the approximation value APX, *i.e.*, *round*(max_$\overline{X}_i$, +APX) = $\overline{X}_i$

and let.

a) PfAPX be the approximation value of p(F) that is automatically computed by the software, based on the number of significant digits of p(F)
b) min_p(F) be the smallest values such as the rounding of p(F) is obtained using the approximation value PfAPX, *i.e.*, *round*(p(F), -PfAPX) = p(F)
c) max_p(F) be the greatest values such as the rounding of p(F) is obtained using the approximation value PfAPX, *i.e.*, *round*(p(F), +PfAPX) = p(F)

Then, the minimum value of s_w can be estimated by replacing each $\overline{X}_i > \overline{\overline{X}}$ with min_ $\overline{X}_i$ and each $\overline{X}_i < \overline{\overline{X}}$ with max_ $\overline{X}_i$ in formula [5] and using p(F)_min as p(F) value in the formula [6]. Similarly, the maximum value of s_w can be estimated replacing each $\overline{X}_i > \overline{\overline{X}}$ with max_ $\overline{X}_i$ and each $\overline{X}_i < \overline{\overline{X}}$ with min_$\overline{X}_i$ in formula [5] and using p(F)_min as p(F) value in formula [6].

References

Abrams, K.R., Gillies, C.L., Lambert, P.C., 2005. Meta-analysis of heterogeneously reported trials assessing change from baseline. Stat. Med. 24, 3823–3844. https://doi.org/10.1002/sim.2423.

Acutis, M., Scaglia, B., Confalonieri, R., 2012. Perfunctory analysis of variance in agronomy, and its consequences in experimental results interpretation. Eur. J. Agron. 43, 129–135. https://doi.org/10.1016/J.EJA.2012.06.006.

Ahrens, H., Pearson, E., 1974. Biometrika Tables for Statisticians, 2. Cambridge University Press, London. https://doi.org/10.1002/bimj.19740160412, 1972. XVIII, 385 S. Biom. J. 16, 291–291.

Barnum, C.M., 2011. Usability Testing Essentials: Ready, Set, test!. Morgan Kaufmann Publishers, Burlington.
Borenstein, M., 2009. Introduction to Meta-Analysis. John Wiley & Sons, Chichester, U.K.
Confalonieri, R., Bellocchi, G., Tarantola, S., Acutis, M., Donatelli, M., Genovese, G., 2010. Sensitivity analysis of the rice model WARM in Europe: exploring the effects of different locations, climates and methods of analysis on model sensitivity to crop parameters. Environ. Model. Software 25, 479–488. https://doi.org/10.1016/j.envsoft.2009.10.005.
Dagnelie, P., 2013. Statistique théorique et appliquée. Tome 1. Statistique descriptive et bases de l'inférence statistique. Bruxelles, De Boeck.
Day, R.W., Quinn, G.P., 1989. Comparisons of treatments after an analysis of variance in ecology. Ecol. Monogr. 59, 433–463. https://doi.org/10.2307/1943075.
Drevon, D., Fursa, S.R., Malcolm, A.L., 2017. Intercoder reliability and validity of WebPlotDigitizer in extracting graphed data. Behav. Modif. 41, 323–339. https://doi.org/10.1177/0145445516673998.
Duncan, D.B., 1955. Multiple range and multiple F tests. Biometrics 11, 1–4. https://doi.org/10.2307/3001478.
Einot, I., Gabriel, K.R., 1975. A study of the powers of several methods of multiple comparisons. J. Am. Stat. Assoc. 70, 574–583. https://doi.org/10.1080/01621459.1975.10482474.
Fisher, R.A., 1935. Design of Experiments. Oliver & Boyd, London.
Fleisch, D., Kregenow, J., 2014. A Student's Guide to the Mathematics of Astronomy. Cambridge University Press, Cambridge. https://doi.org/10.1017/cbo9781139542388.
Foscarini, F., Bellocchi, G., Confalonieri, R., Savini, C., Van den Eede, G., 2010. Sensitivity analysis in fuzzy systems: integration of SimLab and DANA. Environ. Model. Software 25, 1256–1260. https://doi.org/10.1016/j.envsoft.2010.03.024.
García-Palacios, P., Gattinger, A., Bracht-Jørgensen, H., Brussaard, L., Carvalho, F., Castro, H., Clément, J.-C., De Deyn, G., D'Hertefeldt, T., Foulquier, A., Hedlund, K., Lavorel, S., Legay, N., Lori, M., Mäder, P., Martínez-García, L.B., Martins da Silva, P., Muller, A., Nascimento, E., Reis, F., Symanczik, S., Paulo Sousa, J., Milla, R., 2018. Crop traits drive soil carbon sequestration under organic farming. J. Appl. Ecol. 55, 2496–2505. https://doi.org/10.1111/1365-2664.13113.
Gattinger, A., Muller, A., Haeni, M., Skinner, C., Fliessbach, A., Buchmann, N., Mader, P., Stolze, M., Smith, P., Scialabba, N.E.-H., Niggli, U., 2012. Enhanced top soil carbon stocks under organic farming. Proc. Natl. Acad. Sci. U.S.A. 109, 18226–18231. https://doi.org/10.1073/pnas.1209429109.
Gleason, J.R., 1998. dm64: quantiles of the studentized range distribution. Stata Technical Bulletin 46, 6–10.
Gleason, J.R., 1999. An accurate, non-iterative approximation for studentized range quantiles. Comput. Stat. Data Anal. 31, 147–158. https://doi.org/10.1016/s0167-9473(99)00002-x.
Gurevitch, J., Koricheva, J., Nakagawa, S., Stewart, G., 2018. Meta-analysis and the science of research synthesis. Nature 555, 175–182. https://doi.org/10.1038/nature25753.
Haddaway, N.R., Hedlund, K., Jackson, L.E., Kätterer, T., Lugato, E., Thomsen, I.K., Jørgensen, H.B., Isberg, P.E., 2017. How does tillage intensity affect soil organic carbon? A systematic review. Environ. Evid. 6, 1–48.
Hungate, B.A., van Groenigen, K.J., Six, J., Jastrow, J.D., Luo, Y., de Graaff, M.A., van Kessel, C., Osenberg, C.W., 2009. Assessing the effect of elevated carbon dioxide on soil carbon: a comparison of four meta-analyses. Global Change Biol. 15, 2020–2034. https://doi.org/10.1111/j.1365-2486.2009.01866.x.
Jian, J., Du, X., Reiter, M.S., Stewart, R.D., 2020. A meta-analysis of global cropland soil carbon changes due to cover cropping. Soil Biol. Biochem. 143, 107735 https://doi.org/10.1016/j.soilbio.2020.107735.
Keuls, M., 1952. The use of the "studentized range" in connection with an analysis of variance. Euphytica 1, 112–122. https://doi.org/10.1007/bf01908269.
Klasson, K.T., 2018. QXLA: adding upper quantiles for the studentized range to Excel for multiple comparison procedures. J. Stat. Software 85, 9. https://doi.org/10.18637/jss.v085.c01.
Koricheva, J., Gurevitch, J., Mengersen, K., 2013. Handbook of Meta-Analysis in Ecology and Evolution. Princeton University Press, Princeton.
Lenth, R.V., 2001. Some practical guidelines for effective sample size determination. Am. Statistician 55, 187–193. https://doi.org/10.1198/000313001317098149.
Luo, Z., Wang, E., Sun, O.J., 2010. Can no-tillage stimulate carbon sequestration in agricultural soils? A meta-analysis of paired experiments. Agric. Ecosyst. Environ. 139, 224–231. https://doi.org/10.1016/j.agee.2010.08.006.
Maillard, É., Angers, D.A., 2014. Animal manure application and soil organic carbon stocks: a meta-analysis. Global Change Biol. 20, 666–679. https://doi.org/10.1111/gcb.12438.
McDaniel, M.D., Tiemann, L.K., Grandy, A.S., 2014. Does agricultural crop diversity enhance soil microbial biomass and organic matter dynamics? A meta-analysis. Ecol. Appl. 24, 560–570. https://doi.org/10.1890/13-0616.1.
McIntosh, M.S., 2015. Can analysis of variance Be more significant? Agron. J 107, 706–717. https://doi.org/10.2134/agronj14.0177.
Mondal, S., Chakraborty, D., Bandyopadhyay, K., Aggarwal, P., Rana, D.S., 2020. A global analysis of the impact of zero-tillage on soil physical condition, organic carbon content, and plant root response. Land Degrad. Dev. 31, 557–567. https://doi.org/10.1002/ldr.3470.
Morugán-Coronado, A., Linares, C., Gómez-López, M.D., Faz, Á., Zornoza, R., 2020. The impact of intercropping, tillage and fertilizer type on soil and crop yield in fruit orchards under Mediterranean conditions: a meta-analysis of field studies. Agric. Syst. 178, 102736 https://doi.org/10.1016/j.agsy.2019.102736.
Philibert, A., Loyce, C., Makowski, D., 2012. Assessment of the quality of meta-analysis in agronomy. Agric. Ecosyst. Environ. 148, 72–82. https://doi.org/10.1016/J.AGEE.2011.12.003.
Quillet, A., Garneau, M., Frolking, S., 2013. Sobol' sensitivity analysis of the Holocene Peat Model: what drives carbon accumulation in peatlands? J. Geophys. Res. Biogeosci. 118, 203–214. https://doi.org/10.1029/2012JG002092.
Quinn, G.P., Keough, M.J., 2002. Experimental Design and Data Analysis for Biologists. Cambridge University press, Cambridge.
Ryan, T.A., 1960. Significance tests for multiple comparison of proportions, variances, and other statistics. Psychol. Bull. 57, 318–328. https://doi.org/10.1037/h0044320.
Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M., 2004. Sensitivity Analysis in Practice. Wiley, New York.
Sokal, R.R., Rohlf, F.J., 2012. Biometry: the Principles and Practice of Statistics in Biological Research. In: Extensively Rev, fourth ed. W.H. Freeman, New York.
Sung, L., Beyene, J., Hayden, J., Nathan, P.C., Lange, B., Tomlinson, G.A.A., 2006. Bayesian meta-analysis of prophylactic granulocyte colony-stimulating factor and granulocyte-macrophage colony-stimulating factor in children with cancer. Am. J. Epidemiol. 163, 811–817. https://doi.org/10.1093/aje/kwj122.
Thiessen Philbrook, H., Barrowman, N., Garg, A.X., 2007. Imputing variance estimates do not alter the conclusions of a meta-analysis with continuous outcomes: a case study of changes in renal function after living kidney donation. J. Clin. Epidemiol. 60, 228–240. https://doi.org/10.1016/j.jclinepi.2006.06.018.
Tukey, J.W., 1953. Some selected quick and easy methods of statistical analysis. Trans. N.Y. Acad. Sci. Series 2 (16), 88–97.
Valentine, J.C., Pigott, T.D., Rothstein, H.R., 2010. How many studies do you need?: a primer on statistical power for meta-analysis. J. Educ. Behav. Stat. 35, 215–247. https://doi.org/10.3102/1076998609346961.
Valkama, E., Usva, K., Saarinen, M., Uusi-Kämppä, J., 2019. A meta-analysis on nitrogen retention by buffer zones. J. Environ. Qual. 48, 270–279. https://doi.org/10.2134/jeq2018.03.0120.
Weir, C.J., Butcher, I., Assi, V., Lewis, S.C., Murray, G.D., Langhorne, P., Brady, M.C., 2018. Dealing with missing standard deviation and mean values in meta-analysis of continuous outcomes: a systematic review. BMC Med. Res. Methodol. 18, 25. https://doi.org/10.1186/s12874-018-0483-0.
Weir, C.J., Assi, V., Na, L., Lewis, S.C., Murray, G.D., Langhorne, P., Brady, M.C., 2019. Unreported summary statistics in trial publications and risk of bias in stroke rehabilitation systematic reviews: an international survey of review authors and examination of practical solutions. J. Stroke Med. 2, 136–142.
Welsch, R.E., 1977. Stepwise multiple comparison procedures. J. Am. Stat. Assoc. 72, 566–575. https://doi.org/10.2307/2286218.