Australian Journal of Crop Science AJCS
AJCS 5(13):1876-1881 (2011) ISSN:1835-2707

Research Note

Reporting numbers in agriculture and biology: Don't overdo the digits

Marcin Kozak[1], Ricardo Antunes Azevedo[2], Justyna Jupowicz-Kozak[3], Wojtek Krzanowski[4]

[1]Marcin Kozak, University of Information Technology and Management in Rzeszow, Poland
[2]Departamento de Genética, Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, SP 13418-900, Brazil
[3]Department of Botany, Faculty of Agriculture and Biology, Warsaw University of Life Sciences, Nowoursynowska 159, 02-787 Warsaw, Poland
[4]School of Engineering, Mathematics and Physical Sciences, University of Exeter, Harrison Building, North Park Road, Exeter EX4 4QF, UK

***Corresponding author: nyggus@gmail.com**

Abstract

Texts in agricultural and biological sciences are full of numbers, but far too often these numbers are not presented well. The way that numbers are reported does matter, as it can either help or hinder the reader in understanding the message. In particular, how many digits are chosen for each number is important. This choice should be governed by the so-called effective digits, that is, those digits that vary in the given set of data. We give various examples to demonstrate this facet, and offer some guidelines for the selection of number of digits to be used. Our hope is that authors of research articles will use these guidelines in order to make their reports more readable and understandable.

Keywords: data, effective digits, statistics, tables.

Introduction

The agricultural and biological literature is full of numbers. This is because most experimental data are quantitative in nature, and numbers provide an important means of interpreting biological phenomena. If these numbers are poorly presented then important biological conclusions may be obscured. Sometimes graphical methods are needed to present data appropriately, in which case it is crucial that graphs follow standard rules for their construction and presentation [see Kozak (2010) for basic rules, with agricultural and biological examples]. In many other examples, however, it is more advisable and advantageous for data to be presented as raw or summary numbers, within either sentences or tables, or within a hybrid of the two known as a text-table (Tufte, 2001; Kozak, 2009). However, the way that these numbers are presented needs some analogous rules or guidelines, as otherwise the reader might have problems in interpreting the phenomena that they address. In this paper we consider just one particular aspect of presenting numbers: the appropriate number of digits to use.

Sometimes exact values have to be presented for particular purposes, as for example in official statistics reports where precise data reporting can be more important than data interpretation (Kozak and Krzanowski, 2010). However, using numbers to report scientific research is somewhat different. For science data, interpretation is the most important aspect, and this will often require sacrificing some precision in the reported numbers if maximum impact is to be achieved. For example, when comparing two cultivars evaluated for yield potential, it is not necessary to go to four decimal places of accuracy and write that cultivar A had mean grain yield of 5.3761 t ha^{-1} while cultivar B of 4.9987 t ha^{-1} It will be perfectly adequate for the purposes to say that cultivar A had mean grain yield of 5.38 t ha^{-1} while cultivar B of 5.00 t ha^{-1}. Or even, that cultivar A had mean grain yield of 5.4 t ha^{-1} while cultivar B of 5.0 t ha^{-1}.

Thus, when setting out to report on scientific experimentation, a decision must be made regarding the number of digits to retain in the numbers that are presented. Ehrenberg (1977) suggests using two significant or effective digits for reporting numerical values. The term "effective digits" seems to be more appropriate here, because "significant digits" carries a specifically technical meaning – see, e.g., Hagy and Bayless (2010). Effective digits are those digits that vary in the data set under consideration. Thus, returning to the numbers in the example at the end of the previous paragraph, the pair 5.38, 5.00 uses two effective digits while the pair 5.4, 5.0 uses one effective digit. Indeed, in many instances, three effective digits will suffice, while more than three digits may confuse the reader. One important point here is that far too often the reported "precision" of numbers is spurious rather than true (MacDonald-Ross, 1977). For example, presenting the average of a set of numbers to six decimal places (as might be output by a computer or pocket calculator) is pointless if each individual value has actually been measured to only one decimal place of accuracy. Indeed, in standard field experiments, a precision of 0.1 kg in the value of 5.3761 t ha^{-1} wheat mean grain yield is not considered essential. Koschat (2005) suggests that data should be presented with a maximum of 6 digits, which in the case of big numbers can be accomplished by adjusting the scale (e.g. from kg to t) and rounding. According to him, "Usually the left-most digits of a number are more important than the digits to the right. Retaining too many digits hinders the reader from paying attention to the more important digits."

This aspect has been emphasized also by other authors (e.g., Wainer, 1992; Ryder, 1995; Lang, 2004). Despite these warnings, however, the scientific literature is still full of badly presented numbers, ones that are difficult to read and compare.

The simple message of this paper is therefore to warn against overdoing the number of digits when reporting numerical values if the aim of the report is to ensure that their text is comprehensible. We do this by first giving some specific illustrative examples, and then providing some general guidelines.

Examples

Table 1 gives some examples of published sentences that are difficult to assimilate easily because too many digits have been used to present numerical values. Alongside each sentence is a revised text that we believe to be more comprehensible. At first glance the difference might not be that obvious, but if the values in the original text are compared first, and then the same is done for the values in the revised text, the difference is more evident.

Table 2 is a simplified version of a table from Cha-um et al. (2010), while Table 3 is a more readable revised version. We decided to present three effective digits for the means and two effective digits for the standard errors in most of the columns, but four effective digits are given for the means of TC as there are some very large values.

One situation where decimal digits are often overused is in reporting correlation coefficients. For example, Zhang et al. (2010) gave correlation coefficients to four decimal digits (e.g., 0.2719). In this instance, there was no need to provide such accuracy. Actually, even three decimal digits (e.g., Irzykowska and Bocianowski, 2008; Liu et al., 2010; Reig et al., 2010) are too many, and instead two decimals would generally appear to be sufficient (e.g., in or Mohammady et al., 2009). This guideline will most likely apply also to other similar coefficients ranging from –1 to 1, as for example for the Yule coefficient that represents species co-occurrence (e.g., Kozak and Lewandowski, 2010). For path coefficients, which are seldom smaller than –1 or bigger than 1, it also suffices to provide two decimal digits (e.g., Kozak et al., 2007, 2008; Scheneiter et al., 2009).

Another common coefficient, the coefficient of variation, usually does not need more than two digits if presented as a percentage, or two decimal digits otherwise (e.g. Gupta et al., 2009); two decimal digits for percentage coefficients of variation (e.g., Ginigaddara and Ranamukhaarachchi, 2009; Tavares et al., 2010), and even three (e.g., Oselebe and Tenkouano, 2009), are not needed.

Appropriate presentation of *P*-values is also important. A general guideline would be to use three decimal digits: two are too few (there is quite a difference between 0.046 and 0.054, and with two decimal digits both would become 0.05). There is no need to use four decimal digits (instead of $P = 0.0134$ it will suffice to write $P = 0.013$, especially as *P*-values are very variable – Cumming, 2008). Thus the smallest reported *P*-value should be $P < 0.001$; there even seems to be no need to distinguish it from $P < 0.0001$ (Lang, 2004).

Guidelines

The above numerical examples show that authors' choice of the appropriate number of digits is of considerable importance. Too few digits will make comparison of values difficult, but on the other hand too many digits might confuse the reader. We therefore now present some guidelines that we hope will be of use in general.

1. *Obey the conventions of the context*

Sometimes you will need high accuracy for a reported number if this is the requirement in the scientific community at which your work is directed. You will clearly need to adhere to these requirements, so only follow the guidelines below if there are no such conventions or constraints.

2. *Three effective digits will usually suffice*

For example, in these three numbers:

123.4
123.6
122.1

the two last digits vary in the data, and so they all are needed. However, in these numbers:

12356.4
12342.6
1222.1

the five last digits vary in the data, and so not all of them are required. These numbers:

12400
12300
1200

which have two effective digits, might suffice to interpret the data, depending on the context. With three effective digits the set of numbers would become

12360
12340
1220

So, how do we decide between two or three? Ehrenberg (1977) calls for two effective digits, but this will clearly depend on the context and the desired accuracy of numbers in that context. For example, as mentioned above, correlation coefficients practically never require three effective digits. A general guideline might therefore be: if there are no special features in the data that require greater accuracy, two effective digits should suffice. However, three effective digits are often acceptable in agricultural and biological sciences. For example, here:

314.3
295.7
287.3

there are four effective digits, which seems too many. Using three of them:

314
296
287

seems fine, but reducing to two:

310
300
290

seems to be too drastic (although ease of comparison of these numbers is evident). The absolute differences between all corresponding pairs of these numbers are presented in Table

Table 1. Examples of overusing digits when presenting numbers. In each case, the revised version simplifies the sentence by making it easier to read and most importantly, by facilitating the comparison of values.

Source	Original text	Revised text	Comment
Selvaraj et al. 2010	"Jaccard's coefficient of similarity ranged from 0.470 to 0.839 with a mean of 0.640. Most of the pair-wise similarity values fell into the range of 0.601-0.700. "	"Jaccard's coefficient of similarity ranged from 0.47 to 0.84 with a mean of 0.64. Most of the pair-wise similarity values fell into the range of 0.60-0.70. "	No need to provide three decimal numbers for Jaccard's coefficient: two give the same information and are easier to read and compare.
Shahinul Islam 2010	"This effect of the media for another treatment apparently was caused by a much higher embryo formation from anthers cultured on the induction medium compared with pre-culture and washing medium (325.50 vs 207.83 in Table 1)."	"This effect of the media for another treatment apparently was caused by a much higher embryo formation from anthers cultured on the induction medium compared with pre-culture and washing medium (325 vs 208 in Table 1)."	Decimal digits are completely unnecessary to compare the values of 325.50 and 207.83 (which use five effective digits). Much easier to read are values of 325 and 208 (using three effective digits).
Nääs et al. 2010	"Five birds were randomly selected to be weighed every day, and the mean body weight at 42 day was 2981.5 ± 86.7 g."	"Five birds were randomly selected to be weighed every day, and the mean body weight at 42 day was 2.98 ± 0.09 kg."	No need here to give the decimal digit for the mean and standard deviation (although the authors do not state what 86.7 stands for). In fact, since this is quite a general piece of information, here even 3.0 ± 0.1 kg might suffice.
Costa et al. 2010	"The values of Taylor's Power Law's b were greater than the unity, 1.4376 for nymphs and 1.4760 for adults in Area 1, and 1.2977 for nymphs and 1.2504 for adults in Area 2, showing aggregated distribution for both nymphs and adults of *D. citri*, with values of coefficient of determination (R^2) ranging from 0.9580 to 0.9853..."	"The values of Taylor's Power Law's b were greater than the unity, 1.44 for nymphs and 1.48 for adults in Area 1, and 1.30 for nymphs and 1.25 for adults in Area 2, showing aggregated distribution for both nymphs and adults of *D. citri*, with values of coefficient of determination (R^2) ranging from 96 to 99%..."	The original sentence is full of numbers, which makes the sentence difficult to read. Even worse, the numbers consist of 5 digits, of which 4 are decimal ones, which makes the sentence a big challenge. Coefficients of determination do not need so many digits, and in fact giving them in % is better, because the reader would him or herself transform the values to percents; and in addition, we save three digits. The revised sentence is much easier to read.
Zhang et al. 2010	"The relative contribution of epistasic effect ranges from 6.23 % to 37.73%."	"The relative contribution of epistasic effect ranges from 6 to 38%."	In almost all situations there is no need to provide two decimal digits for percents, especially those ranging from 6 to 38. We can omit the first "%".
Sharief et al. 2009	"The highest mean performance were detected in the cross Gm. 3 x Gm. 1021 (15.090t/ha) followed by the cross Gm.1 x Gm. 1021 (14.640t/ha) and the cross Gm. 8 x Gm. 1021(13.731t/ha), respectively."	"The highest mean performance were detected in the cross Gm. 3 x Gm. 1021 (15.1 t/ha) followed by the cross Gm.1 x Gm. 1021 (14.6 t/ha) and the cross Gm. 8 x Gm. 1021 (13.7 t/ha), respectively."	No need to give three decimal digits (so, kg per ha) for maize yield. Comparing these three numbers based on five effective digits: 15.090, 14.640 and 13.731, is far more difficult than comparing these three numbers based on three effective digits: 15.0, 14.6 and 13.7.

Table 2. Table from Cha-um et al. (2010) (simplified): "Chlorophyll a (Chla), chlorophyll b (Chlb), total chlorophyll (TC) and total carotenoids (Cx+c) of Phalaenopsis acclimatized *in vitro* under different temperatures and relative humidity for 30 days and subsequently transferred to in vivo for 14 days. Errors of mean are represented by ± SD."

Temp. (ºC)	RH (%)	Chl_a (μg g^{-1} FW)	Chl_b (μg g^{-1} FW)	TC (μg g^{-1} FW)	C_{x+c} (μg g^{-1} FW)
	60±5	829.5±13.6ab	223.1±2.8ab	1052.6±16.3a	266.2±4.3a
15 ± 2	80±5	309.3±18.3cd	112.1±1.3ab	421.4±2.0bc	103.0±1.4bc
	95±5	241.0±4.2d	65.6±2.0ab	306.6±2.2bc	74.8±1.4c
	60±5	861.5±8.4a	276.2±3.9a	1137.7±12.2a	274.2±3.4a
25 ± 2	80±5	558.2±12.6bc	164.9±6.6ab	723.1±19.2ab	184.9±4.5ab
	95±5	349.5±8.69cd	105.5±1.3ab	455.0±9.9bc	117.8±2.4bc
	60±5	451.0±4.0cd	147.5±14.7ab	598.5±5.5bc	130.7±1.8bc
35 ± 2	80±5	262.7±2.9d	77.6±1.7ab	340.3±17.5bc	64.4±2.5c
	95±5	186.4±3.83d	38.5±5.9b	224.9±4.2c	53.8±8.2c

Different letters in each column show significant difference at $p \leq 0.01$ (**) by Turkey's Honestly Significant different test (Tukey's HSD). ["Turkey's" is original]

4. We see that the differences between the original (four-effective-digit) numbers are quite precisely represented by the differences between the three-effective-digit numbers, but not by the differences between the two-effective-digit numbers.

One additional consideration when selecting the number of effective digits is the amount of variability in the numbers. Here:

301.3
398.7
353.3

there are three effective digits, but because the variability in these three numbers is quite large, two effective digits might suffice:

301
399
353

If we consider, however, the following numbers that have three effective digits:

347.3
349.1
352.8

and their approximation with two effective digits:

347
349
353

we see that without knowledge of the context for the values it is difficult to decide whether two or three digits are needed. If these differences, although seemingly small, turn out to be quite big given the data context, then three effective digits will be required; otherwise two will suffice to show that differences are small.

3. ***Take into account any results of statistical analysis***

The decision between two or three effective digits is also very important when presenting numbers that have been subjected to statistical analysis. For example, suppose that a least significant difference has been computed for a set of mean values and a decision is being made regarding the rounding of the numbers for reporting purposes. It is then crucial to ensure that significance levels for the pair-wise differences between the means are the same for the rounded values that they were for the original values. Noting how greatly rounding affected the pair-wise differences in the small example above, it is evident that in such a case one must be very careful. So, if one attempts to use two effective digits and the aim is to make calculations based on the numbers, one needs to be assured that

(i) the results of any calculations and their associated data interpretation are not affected too much by the rounding to two effective digits;

(ii) the chosen accuracy suffices for the specific context; and

(iii) there is no discrepancy between the presented data and the statistical analysis based on the original data.

4. ***Remember that too many digits can give a spurious appearance of precision***

This usually occurs when derived numerical estimates such as means are obtained with greater accuracy than that of the original values.

5. ***Bear in mind that digits can hinder the reading and comparing of numbers***

Too many digits make it difficult to read and compare values and so, if it is at all possible, consider adjusting the scale (e.g., from kg/ha to t/ha). Rounding errors can also influence the conversion of numbers from one reference measurement system to another (e.g., temperature measures from Fahrenheit to Celsius).

6. ***Relate the numbers to the precision of instruments and precision of measurements***

Numbers must not be reported with greater precision than the documented precision of the instruments (or the methods) used to obtain the various measurements.

Table 3. Revised Table 2. Note how much easier it is to compare the values in columns, and how much easier they are to read. At no cost (because there is no need to present as precise numbers as in Table 2), space is saved and the reading and interpretation of data is made much easier.

Temp. (ºC)	RH (%)	Chl_a (μg g^{-1} FW)			Chl_b (μg g^{-1} FW)			TC (μg g^{-1} FW)			C_{x+c} (μg g^{-1} FW)		
15 ± 2	60 ± 5	829	±14	ab	223	± 2.8	ab	1053	±16	a	266	± 4.3	a
	80 ± 5	309	±18	cd	112	± 1.3	ab	421	± 2	bc	103	± 1.4	bc
	95 ± 5	241	± 4	d	66	± 2.0	ab	307	± 2	bc	75	± 1.4	c
25 ± 2	60 ± 5	862	± 8	a	276	± 3.9	a	1138	±12	a	274	± 3.4	a
	80 ± 5	558	±12	bc	165	± 6.6	ab	723	±19	ab	185	± 4.5	ab
	95 ± 5	350	± 8	cd	106	± 1.3	ab	455	±10	bc	118	± 2.4	bc
35 ± 2	60 ± 5	451	± 4	cd	148	±14.7	ab	599	± 6	bc	131	± 1.8	bc
	80 ± 5	263	± 3	d	78	± 1.7	ab	340	±18	bc	64	± 2.5	c
	95 ± 5	186	± 4	d	39	± 5.9	b	225	± 4	c	54	± 8.2	c

Different letters in each column show significant difference at $p \leq 0.01$ (**) by Turkey's Honestly Significant different test (Tukey's HSD).

Table 4. Differences between numbers 314.3, 295.7 and 287.3 presented in three ways: with four (first numbers in cells), three (second numbers in cells) and two (third numbers in cells) effective digits. Differences based on four effective digit numbers are exact, and the worth of each rounding is judged by the closeness of its differences to the corresponding exact ones. Clearly the numbers with three effective digits give quite close differences (first two values in each cell), but the numbers with two effective digits do not (first and third values in each cell).

	314.3 / 314 / 310	295.7 / 296 / 300
295.7 / 296 / 300	18.6 / 18 / 10	0.0 / 0 / 0
287.3 / 287 / 290	27.0 / 27 / 20	8.4 / 9 / 10

Conclusion

We hope that the examples and guidelines given in this paper are of help to anyone who wants to decide how many digits to use when reporting numbers, and we finish with a thought. Perhaps journals should also start adding more strict or specific rules in the Instructions for Authors regarding ways of presenting numbers?

References

Cha-um S, Ulziibat B, Kirdmanee C (2010) Effects of temperature and relative humidity during in vitro acclimatization, on physiological changes and growth characters of *Phalaenopsis* adapted to *in vivo*. Aust J Crop Sci 4(9):750-756

Costa MC, Barbosa JC, Yamamoto PT, Leal RM (2010) Spatial distribution of *Diaphorina citri Kuwayama* (Hemiptera: Psyllidae) in citrus orchards. Sci Agr 67:546-554

Cumming G (2008) Replication and *p* intervals: *p* values predict the future only vaguely, but confidence intervals do much better. Persp Psychol Sci 3:286-300.

Ehrenberg ASC (1977) Rudiments of numeracy. J Royal Stat Soc A 140:277-297

Ginigaddara GAS, Ranamukhaarachchi SL (2009) Effect of conventional, SRI and modified water management on growth, yield and water productivity of direct-seeded and transplanted rice in central Thailand. Aust J Crop Sci 3(5):278-286

Gupta S, Kozak M, Sahay G, Durrai AA, Mitra J, Verma MR, Pattanayak A, Thongbam PD, Das A (2009) Genetic Parameters of Selection and Stability and Identification of Divergent Parents for Hybridization in Rice Bean [*Vigna umbellata* Thunb. (Ohwi and Ohashi)] in Mid Altitudes of Meghalaya. J Agr Sci 147:581-588.

Hagy M, Bayless M (2010) Significant figures calculator http://calculator.sig-figs.com/

Irzykowska L, Bocianowski J (2008) Genetic variation, pathogenicity and mycelial growth rate differentiation between *Gaeumannomyces graminis* var. *tritici* isolates derived from winter and spring wheat. Ann Appl Biol 152:369–375

Kozak M (2009) Text-table: an undervalued and underused tool for communicating information. Eur Sci Ed 35(4):103-105

Kozak M (2010) Basic principles of graphing data. Scientia Agricola 67:483-494

Kozak M, Bocianowski J, Rybiński W (2008) Selection of promising genotypes based on path and cluster analyses. J Agr Sci 146:85-92

Kozak M, Krzanowski WJ (2010) Effective presentation of data. Eur Sci Ed 36(2):41-42

Kozak M, Lewandowski M (2010) *Eriophyoid mites* (Acari: Eriophyoidea) on coniferous trees: is the occurrence of one species associated with the other? Exp Appl Acarol 50:115-122

Kozak M, Singh PK, Verma MR, Hore DK (2007) Causal mechanism for determination of grain yield and milling quality of lowland rice. Field Crops Res 102:178-184

Lang T (2004) Twenty statistical errors even you can find in medical research articles. Croat Med J 45(4): 361-70

Liu WS, Dong M, Song ZP, Wei W (2009) Genetic diversity pattern of *Stipa purpurea* populations in the hinterland of Qinghai-Tibet plateau. Ann Appl Biol 154: 57–65

Macdonald-Ross M (1977) How numbers are shown. AV Commun Rev 25:359-409

Mohammady S, Arminian A, Khazaei H, Kozak M (2009) Does water-use efficiency explain the relationship between carbon isotope discrimination and wheat grain yield? Acta Agr Scand, Sect B – Plant Soil Sci 59:385-388

Nääs IA, Romanini CEB, Neves DP, Nascimento GD, Vercellino RA (2010) Broiler surface temperature distribution of 42 day old chickens. Sci Agr 67:497-502

Oselebe HO, Tenkouano A (2009) Ploidy versus gender effects on inheritance of quantitative traits in *Musa* species. Aust J Crop Sci 3(6):367-373

Reig C, Mesejo C, Martínez-Fuentes A, Agustí M (2010) Climatic factors affecting the occurrence of cloudy stain on the fruit skin of 'Triumph' Japanese persimmon. Ann Appl Biol 156:421–429

Ryder K (1995) Guidelines for the presentation of numerical tables. Res Vet Sci 58(1):1-4

Scheneiter O, Rosso B, Corletto M (2009) Attributes related to seasonal herbage growth in white clover. Sci Agr 66:20-27

Selvaraj I, Nagarajan P, Thiyagarajan K, Bharathi M (2010). Predicting the relationship between molecular marker heterozygosity and hybrid performance using RAPD markers in rice (*Oryza sativa* L.). Afr J Biotech 9(45):7654-7659

Shahinul Islam S M (2010) The effect of colchicine pretreatment on isolated microspore culture of wheat (*Triticum aestivum* L.). Aust J Crop Sci 4:660-665

Sharief AE, El-Kalla SE, Gado HE, Abo-Yousef HAE (2009) Heterosis in yellow maize. Aust J Crop Sci 3(3):146-154

Tavares CRD, Bendassolli JA, Ribeiro DN, Rossete ALRM, Prestes CV, Tavares GA (2010) ^{15}N-labeled glyphosate synthesis and its practical effectiveness. Sci Agr 67:96-101

Tufte ER (2001) The visual display of quantitative information. Graphics Press LLC, Cheshire, 2001

Wainer H (1992) Understanding Graphs and Tables. Educ Res 21(1):14-23.

Zhang ZB, Xu P, Jia JZ, Zhou RH (2010) Quantitative trait loci for leaf chlorophyll fluorescence traits in wheat. Aust J Crop Sci 4(8):571-579