Organizational Research Methods

Decimal Dust, Significant Digits, and the Search for Stars

Arthur G. Bedeian, Michael C. Sturman and David L. Streiner

The online version of this article can be found at:

Published by:

⑤SAGE

On behalf of:

Ⓜ

The Research Methods Division of The Academy of Management

Additional services and information for *Organizational Research Methods* can be found at:

Email Alerts: http://orm.sagepub.com/cgi/alerts

Subscriptions: http://orm.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations http://orm.sagepub.com/cgi/content/refs/12/4/687

Decimal Dust, Significant Digits, and the Search for Stars

Arthur G. Bedeian
Louisiana State University
Michael C. Sturman
Cornell University
David L. Streiner
University of Toronto

The practice of rounding statistical results to two decimal places is one of a large number of heuristics followed in the social sciences. In evaluating this heuristic, the authors conducted simulations to investigate the precision of simple correlations. They considered a true correlation of .15 and ran simulations in which the sample sizes were 60, 100, 200, 500, 1,000, 10,000, and 100,000. They then looked at the digits in the correlations' first, second, and third decimal places to determine their reproducibility. They conclude that when $n < 500$, the habit of reporting a result to two decimal places seems unwarranted, and it never makes sense to report the third digit after the decimal place unless one has a sample size larger than 100,000. Similar results were found with rhos of .30, .50, and .70. The results offer an important qualification to what is otherwise a misleading practice.

Keywords: estimating and rounding numbers; significant digits; statistical precision; confidence intervals

I t has happened to us before, but recently it occurred again. This time, however, after years of wondering, we decided to delve a little deeper. We were running a standard set of Pearson product–moment correlations on a fresh data set to check for possible relationships among our study variables. We had complete data on 161 respondents and were particularly interested in the correlation between a specific predictor–outcome pair *x* and *y*, both assumed to be random and normally distributed variables. Running a standard software program, the correlation equalled .1547 ($p > .05$, two-tailed). This result left us puzzled, in that we had previously calculated the correlation between another predictor–outcome pair with the result that the correlation equalled .1552 ($p = .05$, two-tailed). On one hand, it thus seemed that whereas the first predictor–outcome pair was significantly correlated ($p = .05$), the other was not ($p > .05$), at least as judged by conventional standards (Sauley & Bedeian, 1989). On the other hand, was it possible that we would be actually more accurate in interpreting our results if we simply reported both correlations as equaling .15 and, thus, as significant?

687

In pondering these contrasting results, we came to wonder about the precision of our statistical methods and, in particular, Pearson's *r*. The *Publication Manual of the American Psychological Association* (American Psychological Association, 2001) advises authors to "express numerical values in the number of decimal places that the precision of measurement justifies" (p. 159), but it also recommends "in general, it is better to round to two decimal places" (p. 129). Common practice seems to follow the latter recommendation. Indeed, the Academy of Management (2007) style guide explicitly states, "Report only two decimal places for all statistics" (p. 473). Rounding to two decimal places is one of a large number of heuristics that are followed in the social sciences, often without question as to its logic or appropriateness (Vandenberg, 2006). We report correlations to two decimals, much like we accept an alpha of .05 as an appropriate cutoff for statistical significance, require coefficient alpha reliabilities of .70 and higher as evidence of internal consistency, report goodness-of-fit indices greater than .90 as evidence of well-fitting structural equation models, and look for eigenvalues greater than 1.00 in determining the number of factors for rotation and interpretation when conducting exploratory factor analyses (Lance, Butts, & Michels, 2006). Whereas we realize that the practice of reporting results to two decimal places may not appear at first blush to be of consequence, we suggest that on deeper reflection its appropriateness merits no less scrutiny. Indeed, one might even see the irony in that all of the preceding heuristics are reported to two decimal places. Many of us have, no doubt, had the experience of reviewing papers for journals and from students that report correlations to three, and even four, decimal places. It is commonly accepted that two decimal places are enough, but what "higher authority" underlies this conventional wisdom?

Significant Digits and Confidence Intervals

In the physical sciences, rules for determining the appropriate number of significant digits in a calculation are a standard topic discussed in introductory textbooks (e.g., Nowlin, 2006-2007). As explained in such discussions, the significance of a digit has to do with whether it represents a true measurement. It is also typically noted that whereas no measurements are exact, responsible reporting requires presenting results in a manner such that the smallest placeholder that can vary conveys a meaningful value. Yet in the organizational sciences, this issue seems rarely considered. Instead, results are reported to two decimal places because, well, it seems that this is just the way it is done. If a study reports a correlation between two variables of .15, one would hope that the 5 is not simply a random number. Stated directly, each number reported as a measurement should not degrade the precision of a result.[1] The 5 should, in some meaningful way, be more likely to represent a true measurement than any other trailing digit, such as 4 or 6.

Concern over the precision of correlations presented as point estimates has led some methodologists to recommend the use of confidence intervals, which provide a range of plausible parameter values and, thus, are viewed as more informative (Zou, 2007). For a given investigation, confidence intervals reveal both the magnitude and the precision of an estimated effect. More important, with respect to the use of simple correlations, confidence intervals also provide a basis for establishing the accuracy and, thus, appropriateness of trailing digits. Indeed, a confidence interval can be used as a significance test, in that if it

does not include the null value (usually zero), a sample correlation is said to be statistically significant.

On the basis of our original example, we computed a traditional 95% confidence interval (i.e., 1.96 σ_p, and again we realize we just arbitrarily reported using two significant digits) for a true correlation of .15 (with an *n* from our original example of 161).

Given $r = .15$ and $n = 161$,

$$\text{Prob}(-.007 \text{ to } .315) = .95$$

The high degree of uncertainty suggested by the width of this confidence interval provides a formal expression of probable sample error, in that, as with other sample statistics, the standard error of a correlation coefficient is inversely related to sample size. A reduction in a confidence interval is, thus, possible by increasing sample size. This said, however, because any further reduction varies as $1/\sqrt{n}$, obtaining smaller intervals for a given confidence level can require a dramatic increase in the number of participants necessary to enhance the power of appropriate statistical tests (Klugh, 1986, p. 188).

A Simulation

The issue at hand, however, is not that a correlation is "just" not quite significant at $p < .05$. With any statistical test, there will be thresholds for which some values fall barely to one side or the other. Of greater importance, and often unconsidered, is the meaning and precision of the trailing digits associated with a point estimate. Does the 5 in the .15 of our example mean anything, or is it simply a random number?

In pursuing this line of thought, we conducted a simulation to investigate the precision of Pearson correlations. We wrote a computer program that generated two values (*X* and *Y*), based on a true relationship of .15. The procedure for the simulation was as follows: *X* was generated first as a random normally distributed number with a mean of 0 and standard deviation of 1. *Y* was then created as $[(.15) * X + \sqrt{(1 - .15^2)} *$ error], where error was an independent random normally distributed number ($M = 0$, $SD = 1$). The process was repeated for the various sample sizes, and the entire process was repeated to obtain 10,000 correlations for each sample size condition. Simulations were conducted for sample sizes of 60, 100, 200, 500, 1,000, 10,000, and 100,000. After estimating the correlations to four decimal places, we then looked at the digits in the correlations' first, second, and third decimal places to determine their reproducibility. Table 1 presents results examining the digit patterns for the observed correlations, without any rounding. In contrast, Table 2 presents the distribution of digits after rounding; in each set of three columns, the first column reports the distribution of digits had we rounded the separate correlations to a single decimal place, the second column had we rounded each correlation to two decimal places, and the third column had we rounded each correlation to three decimal places.

For both tables, we tested whether the digits in each distribution appeared random. That is, we used a chi-square test to determine whether the distribution of the frequency of digits for each distribution was significantly different than what would be expected due to chance if the distribution were uniform. If we were unable to conclude that a distribution was not uniform,

Table 1
Number of Times Out of 10,000 of the Occurrence of the Specified Digit for $r = .15$

| Digit | $n = 60$ | | | $n = 100$ | | | $n = 200$ | | | $n = 500$ | | | $n = 1{,}000$ | | | $n = 10{,}000$ | | | $n = 100{,}000$ | | |
	1st digit	2nd digit	3rd digit	1st digit	2nd digit	3rd digit	1st digit	2nd digit	3rd digit	1st digit	2nd digit	3rd digit	1st digit	2nd digit	3rd digit	1st digit	2nd digit	3rd digit	1st digit	2nd digit	3rd digit
0	3,199	1,144	1,044	3,007	1,154	1,017	2,384	1,020	1,024	1,252	1,021	1,025	555	736	1,034	0	0	1,025	0	0	1,526
1	3,282	1,071	1,021	3,854	1,082	1,033	5,254	1,024	984	7,452	984	961	8,893	863	961	10,000	3	1,026	10,000	0	1,126
2	2,378	1,061	1,003	2,504	1,067	967	2,235	1,058	1,003	1,293	993	1,031	552	971	1,041	0	191	1,045	0	0	998
3	933	1,055	1,010	598	1,013	1,006	125	1,040	969	3	1,013	982	0	1,205	996	0	1,320	987	0	0	836
4	191	1,053	981	35	1,024	1,016	2	958	1,030	0	1,046	997	0	1,209	1,006	0	3,438	966	0	4,985	550
5	15	1,011	996	2	962	1,000	0	1,034	984	0	1,042	1,016	0	1,309	981	0	3,521	1,017	0	5,015	467
6	2	911	1,026	0	1,001	962	0	957	996	0	996	999	0	1,175	999	0	1,312	977	0	0	671
7	0	910	972	0	965	1,007	0	976	1,018	0	979	963	0	1,006	1,006	0	207	990	0	0	1,006
8	0	930	970	0	904	1,022	0	986	1,009	0	960	1,012	0	782	988	0	8	965	0	0	1,340
9	0	854	977	0	828	970	0	947	983	0	966	1,014	0	744	988	0	0	1,002	0	0	1,480
p	<.001	<.001 *	.765	<.001	<.001 *	.776	<.001	.112	.932	<.001	.533	.798	<.001	<.001	.826	<.001	<.001	.658	<.001	<.001	<.001

Note: Table reports the frequency of each digit, from 10,000 simulations of the specified parameters. The p values indicate the results of a chi-square test, which examined whether the distribution of digits is significantly different from a uniform distribution. An asterisk indicates that the distribution of the frequency of digits is significantly different than what would be expected due to chance if the distribution were uniform; however, in these indicated cases the distribution has the incorrect modal response.

Table 2
Number of Times Out of 10,000 of the Occurrence of the Specified Digit, Rounding to the Respective Decimal Place for $r = .15$

Digit	$n = 60$			$n = 100$			$n = 200$			$n = 500$			$n = 1,000$			$n = 10,000$			$n = 100,000$		
	1st digit	2nd digit	3rd digit	1st digit	2nd digit	3rd digit	1st digit	2nd digit	3rd digit	1st digit	2nd digit	3rd digit	1st digit	2nd digit	3rd digit	1st digit	2nd digit	3rd digit	1st digit	2nd digit	3rd digit
0	1,581	1,012	1,002	1,345	1,016	996	744	994	1,006	115	1,021	1,044	4	740	1,037	0	0	1,057	0	0	1,576
1	3,344	1,094	1,029	3,613	1,118	1,028	4,229	1,029	993	4,889	970	974	4,987	793	993	4,952	1	1,028	4,985	0	1,327
2	2,872	1,097	1,022	3,462	1,076	1,011	4,314	1,021	1,004	4,887	1,019	1,001	5,001	866	969	5,048	52	1,012	5,015	0	1,052
3	1,681	1,061	1,024	1,396	1,013	990	702	1,029	943	109	997	1,003	8	1,136	1,035	0	541	1,020	0	0	935
4	442	1,016	975	179	1,058	1,034	11	1,008	1,030	0	1,039	986	0	1,185	1,009	0	2,433	953	0	218	689
5	76	1,047	995	5	983	965	0	988	1,013	0	1,049	1,015	0	1,309	997	0	3,930	1,007	0	9,585	441
6	3	958	999	0	985	1,014	0	1,018	979	0	1,028	1,030	0	1,234	983	0	2,428	1,046	0	197	582
7	1	932	1,029	0	953	998	0	951	1,002	0	973	963	0	1,082	990	0	561	930	0	0	805
8	0	911	934	0	942	995	0	976	1,030	0	955	991	0	914	1,002	0	53	957	0	0	1,177
9	0	872	991	0	856	969	0	986	1,000	0	949	993	0	741	985	0	1	990	0	0	1,416
p	<.001	<.001 *	.551	<.001	<.001 *	.869	<.001	.753	.763	<.001	.227	.794	<.001	<.001	.890	<.001	<.001	.071	<.001	<.001	<.001

Note: Table reports the frequency of each digit, from 10,000 simulations of the specified parameters. The p values indicate the results of a chi-square test, which examined whether the distribution of digits is significantly different from a uniform distribution. An asterisk indicates that the distribution of the frequency of digits is significantly different than what would be expected due to chance if the distribution were uniform; however, in these indicated cases the distribution has the incorrect modal response.

691

such that the probability of occurrence was the same for each digit, this would suggest that the decimal place occupied by the digit conveys no information. We also examined each distribution to determine whether the "correct" digit was the modal response.

Referring to Table 2, when $n = 60$ and $n = 100$, the second decimal place is an imprecise estimate of the true value 5. Although a chi-square test reveals that the distributions of digits in the second decimal place are not random (i.e., we reject the null hypothesis that the distribution of digits comes from a uniform distribution at $p < .001$), the mode does not peak at the anticipated (i.e., "correct") number. That is, when the true correlation is .15, instead of 5 being the most common digit, the most common digit is 2 when $n = 60$ and 1 when $n = 100$. Stated differently, the second digit was precise (i.e., 5) only 10.47% of the time, whereas we would expect it to be precise 10% of the time, if only by chance (additionally, the observed frequency of 5 being observed was not greater than we would expect due to chance, based on a t test, at $p = .13$). When the correlation is reported to three decimal places, the third digit is essentially random (a chi-square test does not reject the null hypothesis that the third digit comes from a uniform distribution, at $p = 0.55$).

Without rounding (see Table 1), when the sample size is 200, 500, or 1,000, the first digit is correctly identified as 1 more than 50% of the time. It is, however, only when $n = 1,000$ (or greater) that the distribution of the second digit is nonrandom and most commonly correctly identified as 5. The third digit, however, remains essentially random with (Table 2) or without (Table 1) rounding. Even when $n = 10,000$, we cannot reject the null hypothesis that the distribution of digits in the third decimal place is uniform ($p = .658$ when not rounding and $p = .071$ when rounding).

In our simulation, it is only when $n = 100,000$ that we truly have precision to two decimal places and some confidence in the third. The first digit was correct 100% of the time (see Table 1), and the second digit had a true value (5) 96% of the time when reduced to two digits (see Table 2). Finally, with a sample size of 100,000, a test of the distribution of digits in the third decimal place did reject the null hypothesis that the distribution of digits was uniform when both rounded and not rounded (both ps $< .001$) and the modal response was the correct value of 0.

Additional simulations were run, repeating the procedures described above, but for rhos of .30, .50, and .70. We ran these simulations to see whether there were notable differences in precision when using higher rhos, because higher rhos should have less variance in their observed correlations (Zho, 2007). The results (available on request), however, present largely the same pattern. For both $\rho = .30$ and $\rho = .50$, the second digit of the correlation is random or does not have the appropriate modal response until the sample size is 500 or greater; for $\rho = .70$, the second digit is not accurate until the sample size is 100 or greater. For all three additional simulations, the third digit appears random until $n = 100,000$. A summary of the necessary sample sizes from our simulations to achieve precision in each respective decimal place is reported in Table 3.

Precision Fallacy

The lesson that seems evident from these results is to avoid overinterpreting data by reporting more decimal places than the data will support. Good (1968) has observed that

Table 3
Precision of Correlations at Various Rhos and Sample Sizes.

ρ	Meaningful to 1 decimal place	Meaningful to 2 decimal places	Meaningful to 3 decimal places
.150	60	1,000	100,000
.300	60	500	100,000
.500	60	500	100,000
.700	60	100	100,000

Note: Table reports the minimum sample size, from 10,000 simulations of the specified parameters, necessary for (a) the distribution of each respective digit to be from a nonuniform distribution and (b) the modal digit to be the correct digit as specified by the true correlation (ρ).

"inexperienced statisticians often overestimate the degree of precision and objectivity that can be attained" in data analysis (p. 293). He referred to this tendency as the precision fallacy, wherein "when we know a machine or formal system . . . can produce an exact answer to a question, we are tempted to provide an answer and inquire no further" (p. 293). McCall and Bobko (1990) labeled this *pseudoprecision* (p. 389). They viewed such "precision" as a contextual effect and wondered what impact the computation of results to, say, six decimal places would have on researchers' tendency to overinterpret data. In this connection, Cohen (1990) has advised that despite there being "computer programs that report by default four, five, or even more decimal places for all numerical results," as social scientists (in contrast to atomic scientists), we should know better because, as demonstrated in the present analysis, "these superfluous decimal places are no better than random numbers" (p. 1305).

The tendency of social scientists to fall victim to the precision fallacy and overinterpret their data may, in part, be attributed to the fact that as Tukey (1969) has noted, philosophers tell us that data analysis should be "unequivocal and without error" (p. 85). Because absolute precision is unattainable in statistical analysis, however, there is nearly always the need to balance accepted method with judgment. Experience has shown that "there is no royal road to statistical induction, that the informed judgment of the investigator is the crucial element in the interpretation of data" (Cohen, 1990, p. 1305). Our opening example highlights this point by illustrating how, other things being equal, declaring the calculated value from a statistical test to be significant can be a somewhat arbitrary decision.

In this respect, given that the *Publication Manual of the American Psychological Association* (American Psychological Association, 2001) recommends that "in general, it is better to round to two decimal places" (p. 129), our results offer an important qualification to what is otherwise a misleading guideline. When *n* is less than 500, the habit of reporting a result to two decimal places seems unwarranted; the second digit appears, according to our results, random. Concomitantly, it never makes sense, unless one has a sample size greater than 100,000, to report results beyond the first two leading digits. Readers who object to "throwing away" decimals might bear in mind Cohen's (1990) conclusion that beyond being useless, uncertain digits are actually worse than random numbers, in that "the clutter they create . . . serves to distract the eye and mind from the necessary comparisons among meaningful leading digits." To this, Cohen (1990) added, in such situations, "less is indeed

more'' (p. 1305). We suspect, however, that all too often researchers are readily mislead into thinking that ''more is more'' in their search for stars (i.e., asterisks indicating significance) in the decimal dust of their data.

Note

1. To be ''not random,'' we mean that the 5 in the correlation .15 should convey more information than had we simply selected a random digit from 0 to 9. As explained anon, to represent this idea we examined the distribution of the frequency of digits for various correlations and tested to determine whether we could reject the null hypothesis that the distribution was uniform (i.e., that the likelihood of any single digit in the distribution was equally probable). Where we could not reject this hypothesis (at $p < .05$), we concluded that the decimal place occupied by a digit conveyed no information.

References

Academy of Management. (2007). Style guide for authors. *Academy of Management Journal*, *50*, 472-475.

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304-1312.

Good, I. J. (1968). Statistical fallacies. In D. L. Sills (Ed.), *International encyclopedia of the social sciences* (Vol. 5, pp. 292-301). New York: Macmillan.

Klugh, H. E. (1986). *Statistics: The essentials for research* (3rd ed.). Hillsdale, NJ: Erlbaum.

Lance, C. E., Butts, M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods*, *9*, 202 -220.

McCall, M. W., Jr., & Bobko, P. (1990). Research methods in the service of discovery. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1, pp. 380-418). Palo Alto, CA: Consulting Psychologist Press.

Nowlin, D. (2006-2007). Precision: The neglected part of the measurement standard. *Mathematics Teacher, 100,* 356-361.

Sauley, K. S., & Bedeian, A. G. (1989). .05: A case of the tail wagging the distribution.*Journal of Management*, *15*, 335-344.

Tukey, J. W. (1969). Analyzing data: Sanctification or detective work? *American Psychologist*, *24*, 83-91.

Vandenberg, R. J. (2006). Statistical and methodological myths and urban legends: Where, pray tell, did they get this idea? *Organizational Research Methods*, *9*, 194-244.

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, *12*, 399-413.

Arthur G. Bedeian is a Boyd Professor of Management at Louisiana State University. He is a past president of the Academy of Management, a former dean of the academy's Fellows Group, and a recipient of the academy's Distinguished Service Award, the Ronald G. Greenwood Lifetime Achievement Award, and the Richard M. Hodgetts Service Award.

Michael C. Sturman is an associate professor and the Blanchard Professor of Human Resource at Cornell University's School of Hotel Administration. He received his PhD in 1997 from Cornell University's School of Industrial and Labor Relations. His research examines individual job performance over time and the consequences of compensation decisions.

David L. Streiner is the director of the Kunin-Lunenfeld Applied Research Unit and assistant vice president, research, at the Baycrest Centre and a professor in the Department of Psychiatry at the University of Toronto. He is the author of two books on statistics, one on epidemiology and one on measurement theory.