Optimal Data Analysis
Vol. 2, Release 2 (October 22, 2013), 94-97

2155-0182/10/$3.00

Ipsative Standardization is *Essential* in the Analysis of Serial Data

Paul R. Yarnold, Ph.D. and Robert C. Soltysik, M.S.
Optimal Data Analysis, LLC

An omnipresent experimental method in all quantitative scientific disciplines involves what is commonly called, for example, a time-series, repeated measures, clinical trial, test-retest, longitudinal, prospective, pre-post, AB, or, more generally, a serial design. In a serial design each observation is assessed on a measure on two or more test sessions spaced by a theoretically meaningful time span. This note presents a classic serial study with n=12 observations, each measured at the same four theoretically-significant times. Scatter plots illustrating test session and raw, normative, and ipsative standardized data demonstrate that ipsatively standardized data are clearly the most appropriate to statistically address fundamental questions that motivate such research.

There are two basic types of serial statistical designs: *N-of-1* single-case studies analyze data obtained for one observation, and *sample-based* investigations analyze data obtained for multiple observations.[1-6] Previous research examined and recommended using a "pre-processing" method for serial data, in which N-of-1 analysis is used to evaluate serial data of individual observations and assess the outcome (e.g., whether the case showed statistically significant improvement in the measure across the series), independently of data from other observations, before combining the observations for a subsequent sample-based investigation (e.g., factors predicting whether or not observations showed statistically significant improvement in the measure across the series).[7]

Research considered presently measured the level of a molecule in rat blood collected at four theoretically critical points to determine the effect of a pharmacological oncology treatment. The substantive focus of the data is irrelevant to the point of this note: it might as well have been a study of the weight of reality-show contestants at each of four scheduled "weigh-offs", or the number of typing errors made by students on four weekly quizzes. Conceptually the point of the design and the research is to assess if data changed across time in an organized, systematic manner, or in a disorganized, chaotic manner.

Data for this study, presented in Table 1 separately by observation (dummy-coded as 1-12) and testing period (dummy-coded as 1-4), include raw scores (*Raw*) and the corresponding normative (z_N) and ipsative (z_I) standard scores.[2] The computational formula for a standardized or z score is: (observation's score – mean score) / standard deviation (SD). For z_N, mean and SD are based on a sample of observations for whom

2155-0182/10/$3.00

a score was recorded: conceptually z_N measures the magnitude of any observation's score relative to the population of scores *for all observations*. In contrast, for z_I the mean and SD are based only on the data from the observation: conceptually z_I measures the magnitude of any observation's score relative to all scores in the population of scores *for the observation*.[2,3]

Table 1: Raw, z_N and z_I Data by Observation and Test Session

Observation	Test Session	*Raw* Score	z_N Score	z_I Score
1	1	124820	-0.950838618	1.3251421085
	2	81560	-0.879555332	0.2094230035
	3	47820	-0.837290734	-0.660765684
	4	39560	-0.510339148	-0.873799428
2	1	122040	-0.977369006	1.2652444035
	2	76500	-0.949258542	0.2672495973
	3	41980	-0.963754323	-0.489245368
	4	16700	-1.185151183	-1.043248633
3	1	422960	1.8944023744	1.3095198891
	2	271380	1.7352793294	0.2002261326
	3	172240	1.8569900367	-0.52530087
	4	109500	1.554243257	-0.984445152
4	1	117189	-1.023663579	1.1939232206
	2	84200	-0.84318844	0.3555004892
	3	53580	-0.712559523	-0.422713583
	4	25880	-0.914163673	-1.12671521
5	1	356400	1.2591999847	1.2884876019
	2	237660	1.2707749316	0.2582181214
	3	141380	1.1887252514	-0.577173032
	4	96160	1.1604553062	-0.969532691
6	1	352000	1.2172094421	1.2380738368
	2	252120	1.4699663193	0.3438745829
	3	150040	1.3762551619	-0.57002069
	4	100680	1.2938827079	-1.01192773
7	1	202400	-0.210469006	1.2579331645
	2	133800	-0.159931674	0.3196312993
	3	66546	-0.431784371	-0.600260153
	4	38980	-0.527460363	-0.977304311
8	1	157640	-0.637627343	1.2612165217
	2	107080	-0.528008706	0.2992023807
	3	62240	-0.525029613	-0.5539763
	4	38460	-0.542810418	-1.006442603
9	1	188540	-0.342739215	1.372310766
	2	100500	-0.61865043	0.0865315946
	3	54920	-0.683542193	-0.579140993
	4	34340	-0.664430085	-0.879701367
10	1	139300	-0.81265156	1.3883732411
	2	72740	-1.001053813	0.0305988813
	3	45280	-0.892293733	-0.529564639
	4	27640	-0.86220964	-0.889407483

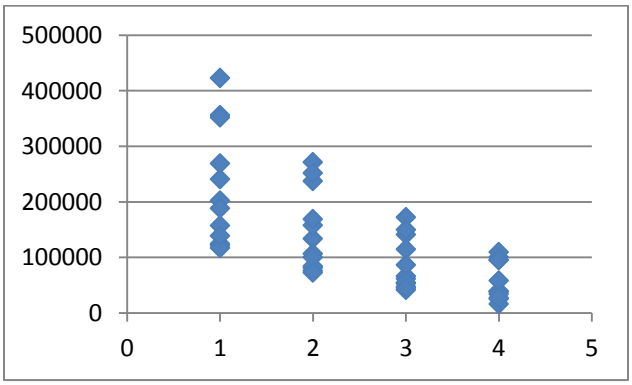
2155-0182/10/$3.00

11	1	240780	0.1558030453	1.3706827988
	2	158200	0.1761865732	0.090386616
	3	115000	0.61747363	-0.579373557
	4	95500	1.1409725441	-0.881695858
12	1	269380	0.4287415721	1.3028529781
	2	169180	0.3274397846	0.244473081
	3	86800	0.0068104107	-0.625679972
	4	58780	0.0570225026	-0.921646087

Exposition begins with a scatterplot that shows the *Raw* scores of all 12 observations on the four testing sessions (Figure 1).

Figure 1

Observations' *Raw* Scores
Across Four Serial Measurements



As seen, there is enormous variation in the *raw* scores among observations in every test period. Normative standardization may be done using mean (128300) and SD (93574) computed across sessions. Shown in Figure 2, this method produces a scatterplot closely resembling results obtained for *raw* data. Note that the scale of the vertical axis is now indexed in standard units.

Figure 2

Observations' z_N Scores (Computed using Grand Mean and SD) Across Four Serial Measurements

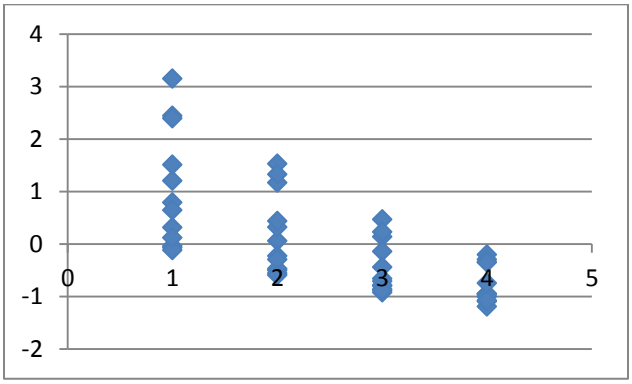
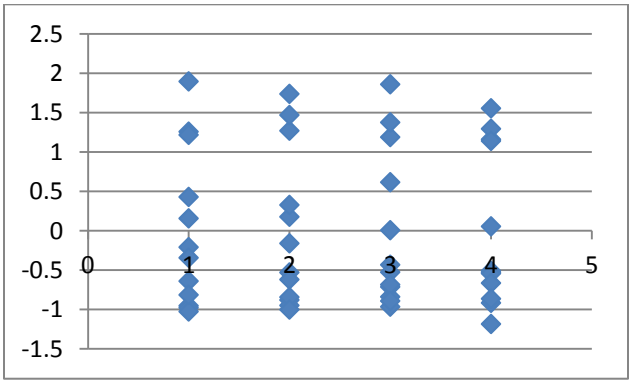


Figure 3 shows z_N scores computed with the mean and SD for each session (which are presented in Table 1). Note that this method *increases* the relative dispersion between the observations as a result of equating the SD—which is now 1 for all four test sessions. And, this method also makes a statistical comparison between test sessions pointless as a consequence of equating the means—which are now 0 for all four test sessions.

2155-0182/10/$3.00

Figure 3

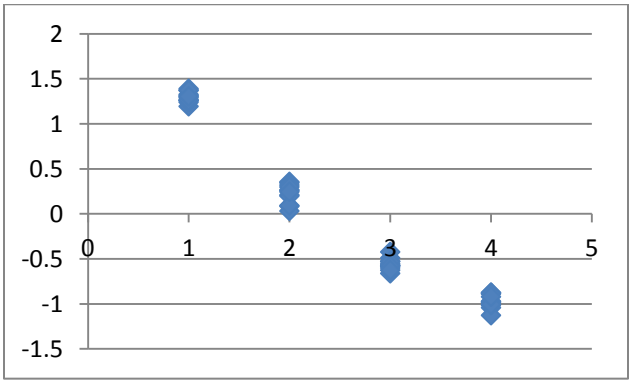
Observations' z_N (Computed using Session Means and SDs) Across Four Serial Measurements



Finally, Figure 4 presents the scatterplot showing observations' z_I scores measured across four test sessions, which clearly shows ipsative standardization of the *raw* data eliminated the variability between observations attributable to "base-rate" differences in their individual mean and variability parameters. However, when data for individual observations are viewed in the context of their own specific base-rate—that is, in terms of their own mean and SD, then as is seen in Figure 4, the molecules in the rat blood appear to have behaved notably uniformly.

Figure 4

Observations' z_I Scores Across Four Serial Measurements



References

[1]Yarnold PR, Soltysik RC (2005). *Optimal data analysis: Guidebook with software for Windows*. Washington, D.C.: APA Books.

[2]Yarnold PR (1992). Statistical analysis for single-case designs. In: FB Bryant, L Heath, E Posavac, J Edwards, E Henderson, Y Suarez-Balcazar, S Tindale (Eds.), *Social Psychological Applications to Social Issues, Volume 2: Methodological Issues in Applied Social Research*. New York, NY: Plenum, pp. 177-197.

[3]Yarnold PR (1988). Classical test theory methods for repeated-measures *N=1* research designs. *Educational and Psychological Measurement*, *48*, 913-919.

[4]Yarnold PR, Soltysik RC, Collinge W (2013). Modeling individual reactivity in serial designs: An example involving changes in weather and physical symptoms in fibromyalgia. *Optimal Data Analysis*, *2*, 37-42.

[5]Mueser KT, Yarnold PR, Foy DW (1991). Statistical analysis for single-case designs: Evaluating outcomes of imaginal exposure treatment of chronic PTSD. *Behavior Modification*, *15*, 134-155.

[6]Soltysik RC, Yarnold PR (2010). The use of unconfounded climatic data improves atmospheric prediction. *Optimal Data Analysis*, *1*, 67-100.

[7]Yarnold PR, Feinglass J, Martin GJ, McCarthy WJ (1999). Comparing three pre-processing strategies for longitudinal data for individual patients: An example in functional outcomes research. *Evaluation and the Health Professions*, *22*, 254-277.

Author Notes

E-mail: Journal@OptimalDataAnalysis.com.

ODA Blog: http://odajournal.wordpress.com/