Clippings from Electronic Notebooks

On the taking of averages: Variance and factor analyses compared

Date: Mon, 23 Mar 1992 02:38:44 EST
Sender: Q Methodology Network
<Q-Method@listserv.kent.edu>*
From: Steven R. Brown, Kent State University
<Sbrown@kent.edu>*
Subject: TAKING AVERAGES IN Q METHOD

In a recent post, I summarized Wolf's (1988) chapter on Q methodology, indicating what I thought were its strong points, and in a critical vein noted that Wolf "goes on to talk about calculating means and standard deviations for items and groups of items...." In response to this, Robert Mrtek commented as follows:

> Without defending Wolf, I think you should help me (and perhaps others) understand why you are critical about means taking.... If the Q set is structured to reflect (and then test) some *a priori* theory, the application of formal tests of significance to the means of grouped items which reflect the inherent structure of the Q set seems perfectly acceptable to me. Especially if a standard technique such as Analysis of Variance is used and the contrasts tested are orthogonal pre-planned comparisons.

The issue of taking the means of a set of items in a Q sort usually refers to the situation in which the Q sample (i.e., set of statements) is structured. To take a couple of illustrations:

- Gray, Koopman and Hunt (1991) created 30 statements each to represent the three phases of Bowlby's theory of loss — (x) urge to recover lost object, (y) disorganization, and (z) reorganization — for a total Q-sample size of N=90. Three scores were obtained for each person by summing the 30 Q-sort scores for each of the xyz Bowlby categories, and these scores were then treated as dependent variables and regressed against the independent variable of length of time since marital separation, as summed across all respondents.

---

* *Original heading altered to reflect current addresses.*

Operant Subjectivity, 1999(April), 22(3), 31-37.

- Proctor, Clarke and Mygdal (1989) administered N=28 one-word descriptors in a Q sort to n=196 students. Based on a prior norming study (n=215) — in which an R factor analysis had indicated three dimensions (capability, sensitivity, and authority) — category means were calculated for each respondent's self Q sort and ideal Q sort. The mean values were then submitted to a three-way analysis of variance (Year in school × Major × Self/ideal).

Studies such as the examples above are not really theoretical in any genuine sense, but merely hypothetico-deductive, *ad hoc*, and categorical. Proctor *et al.*, for example, simply open the door to a practically infinite string of possible variables to be entered into future ANOVAs, with no criterion for determining theoretical salience. A significant F-ratio from the ANOVA only explains the data at hand, therefore, and has no theoretical significance beyond that — hence its *ad hoc* character. In contrast, Newton's theory explained not only falling apples, but also the tides and the movements of planets. Stephenson's (1967) play theory also covered diverse activities. The capacity of these theories to account for more than simply the data that gave rise to them is what qualifies them as genuine rather than *ad hoc*.

Studies of the above kind also have very little to do with subjectivity. In Gray, for example, whatever Q factors might have existed in the data were obviously lost by summing over all persons. In both Gray and Proctor, whatever subjectivity may be at issue is trapped and compressed as a single-score dependent variable to be explained in terms of more fundamental (and objective) variables, such as length of time since divorce, year in school, etc.

Studies of this kind are typical R-methodological fare, and I would wager that they would be of no more interest to Robert Mrtek than to myself insofar as we might be in pursuit of insights into subjectivity. If I understand correctly, he probably would be interested in the Q sample structure itself and in explaining Q sorts in terms of that structure rather than in terms of external and objective variables such as length of time since divorce or year in school. Were he in a position to reanalyze Gray's data, he would likely wish to variance analyze each Q sort to determine which respondents were involved in trying to recover the lost object, which were in a state of disorganization and which were involved in the process of reorganization. He would also want to factor analyze the Q sorts to see what was actually operant (rather than simply sum across all responses).

I hasten to add that I would be the last to argue against ever proceeding in this fashion *in principle*. I have done it myself on many occasions, and Stephenson's *The Study of Behavior* (1953) is replete with examples and guidelines for this kind of tack. What I found objectionable in Wolf's presentation was the presumption that analysis of this genre should be conventionally carried out, as if it were part of a "normal science" of subjectivity.

But to return to Robert Mrtek's initial query: What would be wrong with this? If the Q sample is structured, why not analyze the Q sorts in terms of that structure (e.g., via variance analysis)? The answer is that if this is all that is done, opportunities for important discoveries can be lost and erroneous conclusions can be reached. An example may help clarify what is at stake.

Suppose we were interested in knowing why some people favor certain numbers (e.g., when they bet on the lottery), and let's say that we speculate that it has something to do with whether numbers are odd or even. For a quick experiment, we take the numbers from 0 to 9 and use these as our Q sample — i.e., we place the number "0" on a card, the number "1" on another card, and so forth, and ask friends of ours to Q sort them from those numbers they prefer most (+2) to least (-2). The Q sample is structured, half even and half odd numbers; each Q sort can therefore be variance analyzed. Five such Q sorts were rendered (by myself), and the results of the factor and variance analyses were as shown in Table 1.

Table 1

	Factor Loadings				ANOVA	Means	
Q Sort	*A*	*B*	*C*	*D*	*p(F)*<	*odd*	*even*
1	(62)	08	-10	07	0.02	-0.80	0.80
2	18	(95)	-19	09	0.02	-0.80	0.80
3	19	(75)	-19	03	0.02	-0.80	0.80
4	-21	-15	(65)	10	0.30	0.40	-0.40
5	25	12	11	(78)	0.30	-0.40	0.40

Significant loadings in parentheses ($p<0.05$).

Note first, at the right end of the table, that even numbers received significantly higher scores ($p<0.02$) in Q sorts 1-3, but that theoretical expectations were not met with respect to Q sorts 4 and 5; i.e., subjects 4 and 5 did not give significantly higher scores to even numbers or odd numbers. (The means in the table are the average Q sort scores for even numbers 02468 vs. odd numbers 13579 on a scale ranging from +2 to -2.) Note also that Q sorts 1-3 (despite the fact that they all gave higher scores to even numbers) are not on the same factor, a fact which an investigator restricted to means or ANOVA could never be aware of.

From the standpoint of variance analysis, Q sorts 4 and 5 are a mystery: Hypothesis-testing methods can only tell us what the Q sorts are *not* doing — they are not discriminating between odd and even numbers — but they cannot tell us what the Q sorts *are* doing. Factor analysis shows these two Q sorts on factors C and D, so we can at least examine the factor scores in hopes of detecting a pattern.

The factor scores are as shown in Table 2. Factor A is defined solely by Q sort no. 1 (hence the factor is simply that Q sort), and the preference for even numbers is clear. Q sort no. 2 (factor B) was formed from no. 1 by taking those even numbers from the center and placing them toward the +2 end (and odd numbers from the center and placing them toward the -2 end). No. 2 is consequently uncorrelated with no. 1 (thus they are on separate factors), even though both of them have significantly distinguished even from odd numbers.

Table 2

QSort 1 / Factor A				
-2	-1	0	+1	+2
1	3	6	2	0
	5	7	4	
		8		
		9		

QSorts 2, 3 / Factor B				
-2	-1	0	+1	+2
7	5	0	4	6
	9	1	8	
		2		
		3		

QSort 4 / Factor C				
-2	-1	0	+1	+2
0	2	6	8	9
	1	5	7	
		4		
		3		

QSort 5 / Factor D				
-2	-1	0	+1	+2
1	4	2	0	8
	7	5	3	
		6		
		9		

Q sort No. 3 also defines factor B, but this Q sort was constructed in a wholly different way. First, the numbers were all squared ($6^2 = 36$, $4^2 = 16$, etc.). Then, the squared numbers were ranked in terms of the extent to which their squares were associated with the number 6 (my favorite number) — hence 6, 8, and 4 were given high scores since their squares (36, 64, 16) all contained a 6.

Q sort no. 3 was contrived with no anticipation that it would relate to anything. The fact that it resulted in an ordering of items which produced a significant F-ratio, however (and produced a significant loading on factor B as well), demonstrates the limitations of both ANOVA and factor analysis, and emphasizes the importance of an interview after the Q sort has been completed. Only if we had interviewed no. 3 would we have discovered that he relates to numbers in terms of their squares, and that his favorite number was the key to understanding his ranking.

Two people can define the same factor for different reasons, as in Q sorts 2 and 3. The methodological point is that it is quite possible for individuals who are factor-analytic lookalikes (such as Q sorts 2 and 3 above) to appear quite different from one another at deeper levels. As Brown and Mathieson

(1990), for example, showed with respect to poetic understanding: Both sophisticates and novices defined the same factor, the former because they had reasoned their way to a positive view of the poem in question, and the novices because they did not know enough to be critical and so adopted a positive view by default. Stephenson (1953) has said, "A theory can be rejected even when every tested proposition supports it, as shown by F-tests and the like" (p. 80). It is likely that situations such as this are among those he had in mind: Q sort no. 3 produced a significant F-ratio for distinguishing odd and even numbers, yet an explanation in these terms would be off the mark.

Q sorts 4 and 5 defined factors C and D, respectively. Before continuing with the reading below, however, the reader is encouraged to return to the factor scores in the previous table and to savor what is involved inferentially by trying to divine the principle(s) involved in C and D. Recall that Q sorts 1-3 distinguished even from odd numbers; since nos. 4 and 5 did not, and since 4 and 5 are orthogonal to factors A and B, their operations are quite likely based on some principle quite different from odd vs. even.

Factor C should be readily obvious: Hypothetical respondent no. 4 has expressed a preference for large numbers (+2) over small ones (-2), hence the 9 through 0 sequence. This preference ordering is no respecter of the odd-even distinction, as the insignificant F-ratio for Q sort no. 4 attests. Were odd/even the only known distinction, factor C would rank as an important discovery, and analysis of variance could be used to test for it *in the future*. ANOVA would be of practically no use in detecting it in the first place, however, since *ANOVA is a procedure for verification, not discovery.*

Incidentally, a person who was asked to go through this exercise did not immediately see the high-to-low principle in factor C, but instead saw that the highest number 9 (+2) was balanced by the lowest number 0 (-2); that 8 was balanced by 1 (at +1 and -1), and so on, as if a back-and-forth principle were at issue. This idea was abandoned for the more parsimonious high-to-low principle once the latter was recognized. This is a good example of the way in which abductory logic operates, and how factor analysis provides its underpinning. This has always been known in Q methodology and is now gaining recognition for factor analysis more generally (see Baird 1992).

But factor D is an even better example. One person who pondered the factor scores initially suspected the number 3 to be important: Beginning under -2, the difference between number 1 and number 4 is 3, between 4 and 7 is 3, and the same between 2 and 5, 6 and 9, and 0 and 3, which takes us from -2 up through +1; however, the principle falters since the difference between number 3 and number 8 (under +3) is 5. Clearly, odd vs. even is not at issue either, as the insignificant F-ratio attests. Nor is high-to-low involved, which is why Q sorts 4 and 5 are uncorrelated. What operant function does factor D document? Or is it simply random noise?

Ideas have momentum, and since numerical principles were at issue in factors A, B, and C, it is an easy matter to assume that something numerical (like the number 3) will be at issue in D also. But form rather than number is the operative principle in factor D: Number 8 (under +2) is a curvy number, whereas number 1 (under -2) is a straight line; 0 and 3 (score +1) are also curvy, and 4 and 7 (score -1) are composed of straight lines. As in the case of factor C, factor D represents a principle of numerical preference which is operant and which factor analysis will reveal, but which will remain obscure through the application of variance analysis.

The taking of averages is unavoidable in science; what is critical is how those averages are to be taken. A factor, for example, is also an average — i.e., the factor array is a merger ("average") of the Q sorts defining the factor. But factors in Q methodology conform to Zizek's (1913) "postulate of the greatest possible homogeneity of series," which states that "the average shall refer to a complex of causes as nearly unified as possible, since only in this way will it possess a definitely intelligible content..." (p. 65). (Unlike logical categories such as college major (Proctor 1989) or number of months of marital separation (Gray 1991), Q factors are demonstrably homogeneous — by operation.) Zizek continues:

> If masses of items, which have evidently been variously influenced by quite independent causes, are taken together in a series the average so computed has little scientific value, since it does not express the activity of a unified complex of natural or social causes and is, as a rule, poorly adapted to purposes of comparison. (p. 65)

Although Zizek is not using the term "items" in the same sense as Q statements, the postulate applies here as well. Simply because a group of statements has been declared homogeneous on categorical grounds (i.e., as meaning "thus and so" in general) provides no guarantee that they will be so viewed in the singular situation of Q sorting, for as Stephenson (1953) has said, "we fully expect (and indeed hope for it) that the statements will 'mean' very different things for different persons in different interactional settings, or for the same person in different settings" (p. 144). And to the extent that different meanings are at issue, norming should be avoided, by variance-analytic or any other means.

The issue ultimately is not one of averaging or not averaging, but of when (and how) to ask nature a question, and when (and how) to listen to the answer; and in this regard some of the very best advice comes from Robert Mrtek's own field of medicine. Writing at the time of the American Civil War, French physiologist Claude Bernard (1865/1927), like Zizek a half century later, was warning against "the use of averages which, in medicine and physiology, leads, so to speak, necessarily to error" (p. 134); but Bernard's more fundamental contribution was to remind us of the two operations combined in experimentation — of premeditating and observing. Premeditation of necessity involves *a priori* ideas of the kind expressed formally in hypotheses to be tested: These are the probes that put questions to nature. But when nature replies, Bernard says that observers must divest themselves of preconceived ideas so as to be able to hear more clearly: "the observer's mind must be passive, that is, must hold its peace; it listens to nature and writes at nature's dictation" (p. 22).

In many respects, Q-sample structures, P sets, and conditions of instruction represent formalized probes into nature, and these are often framed in the language of variance analysis; nature, however, speaks most clearly in the language of factor analysis, and in terms of *natural* (i.e., operant) rather than *logical* categories. It is here that discoveries are made. What was missing in Wolf's (1988) account, therefore, was recognition of this aspect of Q methodology, which, after all, is apt to overshadow all else in a science of subjectivity.

References

Baird, D. (1992). *Inductive logic*. Englewood Cliffs, NJ: Prentice Hall.

Bernard, C. (1927). *An introduction to the study of experimental medicine*. (H.C. Green, Trans.). New York: Macmillan. (Original work published 1865)

Brown, S.R. & Mathieson, M. (1990). The operantcy of practical criticism. *Electronic Journal of Communication/La Revue Electronique de Communication*, 1(1). Available: http://www.cios.org/www/ejc/v1n190.htm.

Gray, C., Koopman, E. & Hunt, J. (1991). The emotional phases of marital separation: An empirical investigation. *American Journal of Orthopsychiatry*, 61, 138-143.

Proctor, T.J., Clarke, C.M. & Mygdal, W.K. (1989). Teacher education students' perceptions of self and the ideal teacher. *Educational Research Quarterly*, 13(3), 44-52.

Stephenson, W. (1953). *The study of behavior*. Chicago: University of Chicago Press.

Stephenson, W. (1967). *The play theory of mass communication*. Chicago: University of Chicago Press.

Wolf, R.M. (1988). Q-methodology. In J.P. Keeves (Ed.), *Educational research, methodology, and measurement: An international handbook* (732-734). Oxford: Pergamon.

Zizek, F. (1913). *Statistical averages: A methodological study*. (W.M. Persons, Trans.). New York: Henry Holt.