METHODOLOGY

Assessing Group Change Under Conditions of Anonymity and Overlapping Samples

Arthur G. Bedeian ▾ Hubert S. Feild

Background: A commonly used research design in the social sciences involves the matching of observations over 2 time periods (i.e., Time 1 → Time 2) to assess group change. Because coupled observations are usually correlated, a paired- or dependent-samples *t* test is generally recommended in such applications to determine if there has been a statistically significant change in mean scores across time. Consequently, it is typically believed that unless information for matching respondents' observations is available, researchers have no choice but to treat the observations as if they were independent.

Objectives: To demonstrate alternative statistical approaches for employing the paired samples *t* test when information for matching respondents' observations is unavailable and to illustrate the applicability of these alternatives to longitudinal designs in which respondents at Time 1 are partially replaced by new respondents at Time 2.

Method: Theoretical arguments and examples are employed to achieve the specified objectives.

Results/Discussion: Performing an independent-samples *t* test when a paired-samples *t* test is more appropriate will lead to a loss of statistical power and, thus, increase the likelihood of a Type II statistical error. The statistical approaches that are demonstrated allow researchers to account for pair wise dependency across observations and, therefore, to obtain a fairer test of group change in means.

Key Words: independent-samples *t* test • overlapping samples • paired-samples *t* test

Various research designs in the social sciences employ paired or matched observations for assessing group changes over time (i.e., Time 1 → Time 2). Pairing subjects and their responses from one time period to the next, data analysis is thus generally performed using a paired- or dependent-samples *t* test. It is not uncommon, however, for such longitudinal designs to involve overlapping samples in which only partial information on the pairing of subjects is available. In such situations, it is typically believed that researchers have no alternative but to treat the two samples as if they were independent. Assuming that the samples are from a normally distributed population, data analysis is often performed with an independent-samples *t* test. Zimmerman (1997) has observed, however, that significance tests of location, including the independent-samples *t* test, are inappropriate for such applications because repeated observv tions on the same respondents over period of time are usually correlati rather than independent. He furth demonstrates that performing an ind pendent-samples *t* test on correlat observations results in a loss of statis cal power and, thus, increases the lik lihood of a Type II statistical err(Indeed, when respondents are pair or matched in some way, even a corr lation of .10 between observations se ously distorts the significance level an independent-samples *t* test, resu ing in an overly conservative test group change between means (Zir merman, 1997).

The purpose of this article is twofol First, to suggest three alternative st tistical approaches for employing tl paired samples *t* test when explic information for matching respo dents' observations is unavailabl Second, to illustrate the applicabili of these alternatives to longitudin designs in which respondents at Tin 1 are partially replaced by ne respondents at Time 2 and when r information for estimating the degr

Arthur G. Bedeian, DBA, is Boyd Prof sor, Department of Management, Loui ana State University, Louisiana.

Hubert S. Feild, PhD, is Torchmark P fessor, Department of Manageme Auburn University, Alabama.

of pair wise dependency (i.e., correlation) between Time 1 — Time 2 observations is accessible.

Alternative Approaches

One alternative to performing an independent-samples t test for assessing group change when one is unable to match subjects' responses across time is to arrange Time 1 and Time 2 responses so as to maximize their negative pair wise relationship and then perform a paired-samples t test. This can be accomplished by arranging the Time 1 responses in ascending order and the Time 2 responses in descending order as shown in columns 1 and 2a of Table 1 and then performing a paired-samples t test, using the following formula:

$$t = D/(S^2_{m1}+S^2_{m2}-2r_{12}S_{m1}S_{m2})^{1/2} \quad (1)$$

where D is the mean difference in scores, S^2m1 and S^2m2 are the estimates of variance of the means for Time 1 and Time 2, r_{12} is the correlation between the two sets of observations, and degrees of freedom equals the number of respondents minus one. If a significant result is obtained, one can be very confident that there is a unique difference between the Time 1—Time 2 mean scores. In the present application (Table 1, comparing columns 1 and 2a), $t = 2.60$, $df = 9$, $p < .05$. This alternative is only slightly less conservative than setting the correlation between the Time 1 and Time 2 responses (i.e., r_{12} in Equation 1) equal to -1.0. It would be generally preferred, however, because assuming a perfect negative correlation between pairs of responses is unnecessarily conservative for most research designs.

Given that some of the same individuals at Time 1 are also surveyed at Time 2, it might be argued that the relationship between the two sets of observations is not negative. Thus, a second alternative would be to calculate an independent-samples t value, but use the degrees of freedom appropriate for a paired-samples t test. This can be accomplished by arranging the Time 1 and Time 2 responses so that there is no pair wise relationship (i.e., dependency) and then performing a paired-samples t test. One such arrangement is shown in columns 1 and 2b in Table 1, where the Time 1 responses have been ordered from the lowest to the highest and the Time 2 responses have been arrayed in such a way as to achieve the lowest correlation possible with their Time 1 counterparts. In Table 1, the desired matching of Time 1 and Time 2 (see column 2b) responses was accomplished by hand. For larger data sets, the necessary matching would be more efficiently realized by randomly sorting the Time 2 responses. With most longitudinal data, this second alternative would also tend to be conservative, although somewhat less than the first alternative. In the present application (Table 1, comparing columns 1 and 2b), $t = 3.65$, $df = 9$, $p < .01$).

A final alternative requires a familiarity with both the survey instrument being employed for data collection and the variables being studied. If it is known, for example, from previous work that the estimated reliability of measurement device X is .70, and a uniform change in a focal variable is expected across an intervention's targeted subjects, one may simply substitute X's estimated reliability (i.e., r_{xx}) for r_{12} in Equation 1. Because reliability sets an upper limit on validity (Ghiselli, Campbell, & Zedeck, 1981), this is a reasonable option. Following classical test theory, the reliability coefficient a researcher substitutes in exercising this alternative may be based on any of three empirical approaches (Pedhazur & Schmelkin, 1991): (a) the correlation between observations on the same instrument given at dif-

A final alternative requires a familiarity with both the survey instrument being employed for data collection and the variables being studied.

TABLE 1. Examples of Time 1 and Time 2 Responses

	Time 1 Observations	Possible Arrangements of Time 2 Observations		
Respondent	*(1)*	*(2a)*[a]	*(2b)*[b]	*(2c)*[c]
1	50	66	60	56
2	51	64	61	59
3	52	62	59	58
4	53	61	54	54
5	54	60	64	62
6	55	59	66	59
7	56	59	56	64
8	57	58	58	66
9	58	56	62	61
10	59	54	59	60
M (SD)	54.50 (3.03)	59.90 (3.57)	59.90 (3.57)	59.90 (3.57)
r_{xx}	—	−0.98[d]	0.01[d]	0.60[e]

[a]Time 2 observations arranged in descending order.
[b]Time 2 observations arranged so that there is no relationship with Time 1 observations.
[c]A third possible arrangement of Time 1 and Time 2 observations.
[d]Correlation between Time 1 and Time 2 observations.
[e]Estimated reliability of Time 1/Time 2 measurement instrument.

ferent times (the test-retest approach), (b) the correlation between comparable forms of the same instrument (the equivalent forms approach), and (c) the correlation between comparable parts of same instrument (the internal-consistency approach). If differential change were expected, then the substituted coefficient (however derived) could be set to less than the estimated reliability. To illustrate this final alternative, column 2c in Table 1 presents a third possible arrangement of Time 2 responses. Comparing column 1 and 2c, but substituting a reliability coefficient of .60 (rather than .70) in Equation 1 for r_{12}, $t = 5.71$, $df = 9$, $p < .001$.

An Extension

A consideration of the third alternative described above suggests an extension to situations involving longitudinal designs in which respondents at Time 1 are partially replaced by new respondents at Time 2, and no information for estimating the dependency between Time 1 and Time 2 observations is available. Consider two samples (N_1 and N_2) with an overlap of L repeat respondents. Using the difference in sample means as an estimate of the difference in population means, the variance of the difference in sample means is

$$Var\ (\chi_1 - \chi_2) = \sigma^2 \left(\frac{1}{N_1} + \frac{1}{N_2} - \frac{2L\rho}{N_1 N_2}\right) \quad (2)$$

where ρ is the sample correlation for identified pairs of Time 1 and Time 2 responses. Applying this modification to the denominator in Equation 1 yields the following equivalent expression:

$$t = D/\sigma^2 \left(\frac{1}{N_1} + \frac{1}{N_2} - \frac{2L\rho}{N_1 N_2}\right)^{1/2} \quad (3)$$

Formula 3 can be applied directly if L and ρ are known. Mann and Martin (1999) have discussed possible methods for estimating L in situations where it might normally be unknown due to concerns over confidentiality. In an application that Mann and Martin described in detail, respondents were asked at both Time 1 and Time 2 to indicate their Social Security numbers on separate sheets of paper. This information was then used to match Time 1 and Time 2 responses so that both L and ρ could be estimated, with ρ being the sample correlation for the identified pairs of responses. A similar method for matching anonymous pre- and post-test respondents using an anonymous sticker identification system is described by McGloin, Holcomb, and Main (1996).

Such methods, of course, are inapplicable in situations where a researcher initially has no intention of resampling at a later date and, thus, sees no need to request identifying codes at Time 1 or in situations where a decision is made by a different researcher to conduct a followup study. The alternatives previously described, however, offer three solutions to such dilemmas. Given that L can be estimated by assuming that the proportion of overlap between the Time1 and Time 2 respondents is equivalent to the proportion of known overlap between the Time 1 and Time 2 population of subjects, ρ may be estimated using either the first or second alternatives described above. That is, by calculating the sample correlation between Time 1 and Time 2 responses after simply arranging the Time 1 responses and the Time 2 responses (as shown in columns 1 and 2a in Table 1) so as to maximize their negative pair wise relationship or, less conservatively, by arranging the Time 1 and Time 2 responses (as shown in columns 1 and 2b in Table 1) so that there is no pair wise dependency. As a third basis for estimating the sample correlation, a researcher might choose to follow the final alternative described above and substitute the estimated reliability of one's survey instrument for ρ.

The point of the preceding discussion is to inform researchers that, contrary to having no choice other than using an independent-samples t test in the absence of pair wise information, alternatives do exist. Moreover, when faced with a situation where there is replacement in sampling units from one period to another, it is still possible to account for pair wise dependency and conduct a paired-samples t test. Whereas the proffered alternatives may not offer exact solutions, they do represent reasonable resolutions that allow researchers greater latitude than heretofore acknowledged.

Contrary to having no choice other than using an independent-samples t test in the absence of pair wise information, alternatives do exist

Accepted for publication January 3, 2001:

The authors thank Barbara L. Mann, Associate Professor, Department of Mathematic and Statistics, Wright State University, an Donald W. Zimmerman, Professor Emeritus Department of Psychology, Carleton University, for the helpful vetting on an earlier draft

Address reprint requests to Arthur G Bedeian, PhD, Department of Management Louisiana State University, Baton Rouge, L 70803-6312. (e-mail: abede@lsu.edu).

References

Ghiselli, E. E., Campbell, J. P., & Zedeck S. (1981). *Measurement theory for th behavioral sciences.* San Francisco Freeman.

Mann, B. L., & Martin, P. A. (1999 Overlapping samples in organization research. *Nursing Research, 48*, 23(233.

McGloin, J., Holcomb, S., & Main, D. (1996). Matching anonymous pr posttests using subject-generated info mation. *Evaluation Review, 20*, 72 736.

Pedhazur, E. J., & Schmelkin, L. P. (199 *Measurement, design, and analysis: integrated approach.* Hillsdale, N Lawrence Erlbaum Associates.

Zimmerman, D. W. (1997). A note interpretation of the paired-sample test. *Journal of Educational and Beh ioral Statistics, 22*, 349-360.