DOCUMENT RESUME

ED 151 689 CG 012 303

AUTHOR Schaffner, Paul E.; Darlington, Richard B.
TITLE A New Technique for Personality Scale Construction. Preliminary Findings.
PUB DATE Aug 77
NOTE 11p.; Paper presented at the Annual Convention of the American Psychological Association (San Francisco, California, August 26-30, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Behavior Rating Scales; *Diagnostic Tests; Measurement Instruments; *Personality Tests; *Predictive Validity; Psychological Tests; *Psychometrics; Research Tools; *Test Construction; *Test Validity
IDENTIFIERS Ridge Regression Analysis

ABSTRACT

Most methods of personality scale construction have clear statistical disadvantages. A hybrid method (Darlington and Bishop, 1966) was found to increase scale validity more than any other method, with large item pools. A simple modification of the Darlington-Bishop method (algebraically and conceptually similar to ridge regression, but independently derived) is proposed and examined; it is intended to increase scale validity by reducing validity shrinkage. In a double cross-validation design, MMPI items were used to predict diagnostic category (schizophrenic vs. other) for 200 hospitalized patients. With one anomalous exception, the proposed technique produced scales more valid than those from the original Darlington-Bishop technique. (Author)

ED151689

A New Technique for Personality Scale Construction: Preliminary Findings [1]

Paul E. Schaffner
Bowdoin College

and

Richard B. Darlington
Cornell University

Abstract

Most methods of personality scale construction have clear statistical disadvantages. A hybrid method (Darlington and Bishop, 1966) was found to increase scale validity more than any other method, with large item pools. A simple modification of the Darlington-Bishop method (algebraically and conceptually similar to ridge regression, but independently derived) is proposed and examined; it is intended to increase scale validity by reducing validity shrinkage. In a double cross-validation design, MMPI items were used to predict diagnostic category (schizophrenic vs. other) for 200 hospitalized patients. With one anomalous exception, the proposed tehcnique produced scales more valid than those from the original Darlington-Bishop technique.

Presented at the 85th Annual Convention of the American Psychological Association, August 1977

A New Technique for Personality Scale Construction: Preliminary Findings

The construction of personality and other inventory scales is an important topic whose statistical principles are not well understood. Three well-known techniques all have clear disadvantages. The use of clinical, prior, or intuitive weights takes no advantage of any data which may be available. Classical item analysis--in which a scale is constructed consisting of the items which correlate most highly with a criterion variable--make no use of any information about the intercorrelations among items. Multiple regression uses all of the relevant information available from sample data, but suffers from problems of excessive sampling error and validity shrinkage. This is especially the case when large numbers of predictors are available relative to the sample size. In fact, in many inventory scale construction problems, the number of items p may exceed the sample size n. For instance, a sample size of 200 or 300 is considered fairly large by most standards, but the best-known personality inventory--the MMPI--has 550 items. Multiple regression is of course impossible when the number of items exceeds the sample size, and very cumbersome or virtually impossible computationally when p is very large, regardless of n.

A fourth possible technique is stepwise multiple regression. It too uses all of the available relevant information, while greatly reducing the amount of computation required and providing tests of reasonable length. However, Darlington and Bishop (1966) have shown empirically the predictive inadequacy of this technique in comparison with classical item analysis. A major weakness of stepwise multiple regression in this context is the overestimation of ρ_{ct} by r_{ct} (where t is the scale or test and c is the criterion) at each iteration, resulting from sampling error.

Darlington and Bishop (1966) proposed and empirically investigated a 'hybrid' technique, and found that it produced scales more valid than those obtainable by other methods when large numbers of predictors are involved. Their technique begins with the construction of a provisional first-stage scale according to the item analysis approach, adopting the most valid individual items. (Unit weights are used exclusively in the Darlington-Bishop technique). Length of the first-stage scale can be determined post hoc if cross-validation is available, by choosing the set of items which yields the most valid test. In most applications of this technique, cross-validation would probably not be available; in these cases the researcher would select any number of items for the first-stage scale, according to whatever guidelines or restrictions s/he may be following. Partial correlations $r_{ic \cdot t}$ (where i represents the individual items) are then computed for the entire item pool using the test-construction sample data. Items with the highest partial correlation values--and therefore the greatest potential for increasing the validity of the first-stage scale--are then added to create a second-stage scale. This may involve the repetition of items already in the first-stage scale, or their effective deletion if repeated with negative unit weights. The second-stage scale is then used to compute a second set of partial correlations and the addition process is repeated. The optimal number of items to be added at each subsequent stage can be determined exactly post hoc if cross-validation is available. If not, Darlington and Bishop recommend the addition on each iteration of approximately 1/3 as many items as are included in the existing provisional scale (counting repeats and deletions). They also found that scale cross-validity does not generally rise after inclusion of the third-stage items.

This scale construction technique can be understood as a compromise between classical item analysis and stepwise regression, combining the advantages of each. By starting with what would be the final scale under classical item-analysis procedures, the researcher can do no worse than to match the effectiveness of that technique. In subsequent stages this technique does make use of all available relevant information about the items (including item intercorrelations) by adding new items according to their partial correlation values.

The problem of validity shrinkage due to overestimation of $\bar{r}_{ct}$ still exists in the Darlington-Bishop technique, though lessened by the great reduction in number of iterations performed. The present study was undertaken to test the effectiveness of a modification of the Darlington-Bishop technique which may raise its effectiveness further by reducing validity shrinkage due to overestimation of ρ_{ct}. The modification consists of reducing the value of r_{ct} obtained at each stage in the scale construction process. A positive constant less than 1.00 is multiplied by r_{ct} and the resulting value is inserted in place of r_{ct} in the numerator of the partial correlation $r_{ic \cdot t}$. This has the effect of increasing the importance of item validities relative to item intercorrelations in determining which items will be added to the existing provisional scale. The correction factor proposed for the present study is:

$$1 - \frac{1}{N \cdot \overline{r_{ic}^{2}}} \qquad [1]$$

which, when multiplied b r_{ct}, can be shown to produce a more accurate estimate of ρ_{ct} than r_{ct} itself.

Examination of the formula for the correction factor reveals at least three interesting properties. First, the size of the reduction of r_{ct} is negatively related to the overall predictive power of the item pool. Second, the size of the reduction of r_{ct} is negatively related to sample size n. Finally, the correction factor is mechanically simple to use. Only one computation is needed for all operations on any given data set and criterion, and only readily-available statistics are involved in its computation.

The original Darlington-Bishop technique and classical item analysis can be reformulated in terms of the value each implicitly assigns to this correction factor. Item analysis uses a value of 0.0 in all cases, so that potential additions are evaluated only in terms of their predictive validities. The Darlington-Bishop technique always assumes a value of 1.00. The modification proposed here can thus be interpreted as an intermediate approach in relation to these other two techniques. It would seem plausible that other values for the correction factor (between o.0 and 1.0) could also increase scale validity over the levels obtained by adoption of either extreme value (i.e., item analysis or the Darlington-Bishop method).

A certain modification of the Darlington-Bishop technique is also suggested by the theory of ridge regression (Hoerl & Kennard, 1970; Price, 1977). We shall give only the justification of the ridge regression approach here, since it is simpler. It can be shown algebraically that redge regression is equivalent to the following procedure: multiply all observed correlations in a data set, both item validities and item intercorrelations, by a constant $1/(1+k)$; then run the regression in the normal way on these adjusted correlations. It can be shown that the procedure described in the preceding paragraphs is algebraically almost equivalent to that procedure. Thus the proposed modification can be considered as suggested by the theory of ridge regression. With very large values of p, ridge regression becomes computationally prohibitive, whereas the technique described above does not.

Method

Subjects and item pool: MMPI responses of 96 diagnosed schizophrenics and 104 non-schizophrenic psychiatric patients were used in the study. These data were taken from a larger sample made available to the investigators by Albert Rosen, which sample was also used in one of the studies reported by Darlington and Bishop. Two subsamples of 100 individuals were created, each including 48 diagnosed schizophrenics and 52 non-schizophrenic patients. For budgetary reasons, the first 150 of the 550 MMPI items were adopted as the item pool for this study. This procedure is likely to weaken rather than strengthen the effectiveness of the technique under study, as the problem of validity shrinkage increases with larger item pools. Parallel analyses were also undertaken using only the first 90 MMPI items.

Design: A double cross-validation design was adopted, with scales constructed in each subsample cross-validated in the other subsample. Parallel tests were conducted on the complete 150-item pool and the first 90 items of the pool. The original study design called for three different scale-construction techniques to be used on each item pool with each subsample: classical item analysis, the Darlington-Bishop technique, and the modification of the latter obtained by multiplying r_{ct} values by [1]. For the two subsamples the actual correction factor values obtained were .4254 and .3303. A fourth 'technique' was subsequently 'invented,' using a correction factor value of .7000 in order to sample the full range of possible values more evenly and completely.

Procedure: Product-moment correlations were used as indices of validity. Unit weights (positive and negative) were used exclusively in constructing each scale. Cross-validation of all potential scales was undertaken so as to achieve optimum performances for each technique. Up to three iterations were performed on each data set for each technique, except classical item analysis for which the first-stage test is certain to be optimal by that method. The scale construction steps for all four techniques are described above.

Results and Discussion

The results of each of sixteen scale constructions (four techniques applied to four different data sets) are presented in Table 1. In the data set comprised by subsample A with a 90-item pool, the scale produced by classical item analysis contained four items. None of the three iterative procedures was able to increase the validity of that four-item scale. This finding is somewhat surprising, in that personality scales using MMPI or similar items are generally much longer. The other 12 completed scales included from 12 to 34 items, more in keeping with common experience in this area. In the other three data sets, the Darlington-Bishop technique yielded more valid scales (compared with item analysis) twice. The technique proposed in this paper performed better than classical item analysis, better than the Darlington-Bishop technique, and better than or equal to the intermediate-correction-factor-value 'technique' in all three cases.

The percentage increases in scale validity obtained by the proposed modification technique are not large in this study. However, the percentage increases obtained here by the use of the original Darlington-Bishop technique are consistently smaller (0% to 2%) than those obtained in the 1966 study (1% to 22%) where all 550 MMPI items were used. Thus there is reason to believe that the proposed modification also would have greater beneficial impact under more typical personality scale construction conditions. Further research is now underway to more thoroughly test the properties of this technique. Its simplicity in use, its conceptual and algebraic similarity to ridge regression, its performance in the present study, and the previously-established value of the original Darlington-Bishop technique, together suggest that this modification will prove to be a useful supplement to existing inventory scale construction technology.

References

Darlington, R.B., and Bishop, C.H. Increasing test validity by considering interitem correlations. Journal of Applied Psychology, 50(4), 1966, 322-330.

Hoerl, A.E., and Kennard, R.W. Ridge regression: biased estimation for nonorthogonal problems. Technometrica, 12, 1970, 69-82.

Price, . Ridge regression: applications to non-experimental data. Psychological Bulletin, 1977, in press.

Footnotes

[1] This research was aided by a Grant-in-Aid of Research from Sigma Xi, the Scientific Research Society of North America, to the first author.

Table 1.

Performances of Four Different Scale Construction Techniques Applied to Four Data Sets

Number of items in item pool	Subsample	Characteristics of original scale		Value of correction factor	Number of productive iterations performed	Number of items in final scale, by weightings					Cross-validity of final scale	Percentage improvement in predictive power	
		Number of items	cross-validity			1	2	3	4	Tot.		R	R^2
150	A n=100	12	.226	(.0000)	-------	12	0	0	0	12	.226	---	---
				.4254	2	11	4	2	0	17	.238	5.31	10.90
				.7000	0	12	0	0	0	12	.226	0.00	0.00
				1.000	1	18	4	0	0	18	.230	1.77	3.57
	B n=100	33	.337	(.0000)	------	33	0	0	0	33	.337	----	----
				.3303	2	31	1	2	0	34	.361	7.12	14.75
				.7000	1	33	1	0	0	34	.343	1.78	3.59
				1.000	1	30	4	0	0	34	.344	2.08	4.20
90	A n=100	04	.207	(.0000)	------	4	0	0	0	4	.207	----	----
				.4254	0	4	0	0	0	4	.207	0.00	0.00
				.7000	0	4	0	0	0	4	.207	0.00	0.00
				1.000	0	4	0	0	0	4	.207	0.00	0.00
	B n=100	32	.251	(.0000)	------	32	0	0	0	32	.251	----	----
				.3303	3	31	0	0	1	32	.274	9.16	19.17
				.7000	3	31	0	0	1	32	.274	9.16	19.17
				1.000	0	32	0	0	0	32	.251	0.00	0.00