DOCUMENT RESUME

ED 159 182 TM 007 359

AUTHOR Borich, Gary D.; And Others
TITLE An Index for Determining the Importance of a Region of Significance and Some Applications to Published ATI Research.
PUB DATE [75]
NOTE 16p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Analysis of Covariance; *Differences; Hypothesis Testing; Literature Reviews; Predictive Measurement; *Tests of Significance
IDENTIFIERS *Aptitude Treatment Interaction

ABSTRACT

The homogeneity of group regressions test and regions of significance test are two procedures which are frequently recommended for testing and aptitude-treatment or trait-treatment interaction (ATI) hypothesis. The former is used to determine if treatment group regressions are nonparallel while the latter is used to determine the range of aptitude values for which the heterogeneous group regessions are significantly different. While the above procedures have become somewhat commonplace among ATI researchers, quantitative indices which can be used to determine the practical importance of a region of significance are conspicuously absent from any discussion of ATI methodology. The purpose of this paper is to outline a quantitative procedure for determining the practical importance of a region of significance and to illustrate by reanalysing data from published ATI research studies that a failure to use the procedure could lead to trivial and/or erroneous research conclusions. (Author/CTM)

ED159182

TM

An Index for Determining the Importance of a Region of Significance and Some Applications to Published ATI Research

Gary D. Borich, Robert C. Godbout and Paul T. Marston

The University of Texas at Austin

R&D Center for Teacher Education
Education Annex 3.203
The University of Texas
Austin, Texas 78712

Running head: An Index for Determining the Importance of a Region of Significance

Abstract

The homogeneity of group regressions test and regions of significance test are two procedures which are frequently recommended for testing an aptitude-treatment or trait-treatment interaction hypothesis. The former is used to determine if treatment group regressions are nonparallel while the latter is used to determine the range of aptitude values for which the heterogeneous group regressions are significantly different. While the above procedures have become somewhat commonplace among ATI researchers, quantitative indices which can be used to determine the practical importance of a region of significance are conspicuously absent from any discussion of ATI methodology. The purpose of this paper is to outline a quantitative procedure for determining the practical importance of a region of significance and to illustrate by reanalyzing data from published ATI research studies that a failure to use the procedure could lead to trivial and/or erroneous research conclusions.

An Index for Determining the Importance of a Region of Significance and Some Applications to Published ATI Research

Gary D. Borich, Robert C. Godbout and Paul T. Marston

The University of Texas at Austin

The homogeneity of group regressions test (Edwards, 1971) and regions of significance test (Johnson & Neyman, 1936) are two procedures which are frequently recommended for testing an aptitude-treatment or trait-treatment interaction hypothesis. The former is used to determine if treatment group regressions (criterion regressed on aptitude) are nonparallel, i.e., heterogeneous, while the latter is used to determine the range of aptitude (trait) values for which the heterogeneous group regressions are significantly different. These procedures are illustrated in Figure 1.

Insert Figure 1 about here

While the above procedures have become somewhat commonplace among ATI researchers (Cronbach & Snow, 1974; Berliner & Cahen, 1973), quantitative indices which can be used to determine the practical importance of a region of significance (and, therefore, the ATI) are conspicuously absent from any discussion of ATI methodology. The purpose of this paper is to outline a quantitative procedure for determining the practical importance of a region of significance and to illustrate that a failure to use the procedure could lead to trivial and/or erroneous research conclusions. The first part of this paper will be devoted to the mechanics of determining the importance of a region of significance, the second part to the application of this technique to published research to illustrate its value in identifying trivial and/or erroneous research conclusions.

Importance of a Region of Significance in the Single Aptitude Case

The existence of a region of significance does not necessarily indicate the practical importance of that region. For example, if a region of significance contains no observed data points, then that region is of little importance. Furthermore, regions of significance are established on the basis of general relationships observed across the entire range of aptitude values. The Johnson-Neyman technique (1936), for example, defines a region of significance in terms of differences between group regressions (predicted values) and not on the basis of the observed data within that region. The actual pattern of observed results within a region of significance may be in conflict with the general predicted relationships and in this case the region would be of little importance. Figure 1 presents a simplified example of this situation.

Note in Figure 1 that the left region of significance (below point A) is evidenced because the Treatment 0 regression line (predicted scores) is significantly above the Treatment X regression line (predicted scores). The left region of significance would usually be taken as a region where Treatment 0 was superior. Note, however, that the observed data within this region indicate the exact <u>opposite</u> relationship. Any conclusion about the superiority of Treatment 0 within that region, then, is questionable.

While plotted output provides general impressions about the importance of a region of significance, more objective measures of importance are often desirable. Two measures of the importance of a region of significance can be calculated--(1) the proportion of total observations within a region of significance and (2) an index of the overlap within a region.

<u>Proportion of total observations within a region of significance</u>. This index of importance is simply the number of observations falling within a region divided by the total number of observations. The greater this proportion, the greater the importance of the region of significance.

Index of overlap within a region of significance. Given a region of significance, we cannot be certain that any given S in the group predicted to be superior actually performed better than all the Ss in the other group. Some Group 1 Ss will perform better than Group 2 Ss even though the interaction and region of significance indicate that Group 2's treatment was superior to Group 1's treatment in that region. Figure 2 illustrates such overlap in a region of significance.

Insert Figure 2 about here

Consider the region of significance bounded by point A in Figure 2. Notice that, even though Treatment 0 is superior to Treatment X for the area which lies to the left of point A, some X Ss fall closer to the regression line for Treatment 0 than to the X regression line, and that some 0 Ss fall closer to the regression line for Treatment X than to the 0 regression line. We can expect such overlapping to occur even when regions of significance are defined at a high level of confidence.

An index of the extent of such overlapping is the percent of all subjects falling within a region of significance who actually demonstrate a criterion score inconsistent with their treatment group. The smaller the value of this index, the greater the importance of the region of significance. Such an index can be calculated by counting the number of subjects in the region of significance who, while assigned to the poorer treatment, actually performed above the midline between the regression lines for the two groups (i.e., a line equidistant from the two group regressions) and adding to this the number of subjects in the region of significance who, while assigned to the better treatment, actually performed below the midline between regressions.

The percentage of both types of deviations within a region is calculated by finding the midline between the group regressions and then determining whether each observation falls above or below this line. Let $Mpt(X_i)$ symbolize the midline criterion score for a predictor score of X_i. Note that $[Mpt(X_i), X_i]$ indicates the set of points falling on the midline. The midline between groups regressions is given by

$$Mpt(X_i) = \frac{\overline{Y}_1 + b_1(X_i - \overline{X}_1) - \overline{Y}_2 - b_2(X_i - \overline{X}_2)}{2} + \overline{Y}_2 + b_2(X_i - \overline{X}_2)$$

or, simplifying,

$$Mpt(X_i) = \frac{\overline{Y}_1 + b_1(X_i - \overline{X}_1) + \overline{Y}_2 + b_2(X_i - \overline{X}_2)}{2}$$

where $\overline{Y}_1$, $\overline{X}_1$ and b_1 represent the criterion mean score, aptitude mean score and regression coefficient (criterion on aptitude), respectively, for one treatment and $\overline{Y}_2$, $\overline{X}_2$ and b_2, these same values for the other treatment. For subject n with criterion score Y_n and aptitude score X_n, the distance from the midline is given by

$$D = Y_n - Mpt(X_n).$$

D will be zero when the observation falls on the midline, positive when it falls above it, and negative when it falls below it. D's for observations of the better treatment are expected to be positive and D's for observations of the poorer treatment are expected to be negative. Exceptions are considered "misses" and are tallied and reported as a percent of the total number of observations within the region. In Figure 2, two observations (O's) from the better treatment fell below the midline and two observations (X's) from the poorer treatment fell above it. Both types of "misses" constitute 28% of the observations that lay within the region of significance.

We, therefore, would report a 28% overlap for the region of significance bounded by the aptitude value A. A small amount of overlap indicates that the relationships among the data actually observed within the region are consistent with the predicted relationships used to establish the existence of that region of significance. A large amount of overlap indicates that the observed data contradict the validity of a region of significance. The greater the overlap, the less the importance of the region of significance.

It is important to note that a subject from Treatment 1 scoring closer to the regression line for Treatment 2 does not provide information as to whether that subject has been assigned to a treatment incorrectly. This becomes obvious when we consider a subject who is assigned to the better treatment within a region of significance but whose score falls, let us say, at or below the regression for the poorer treatment in this region. Such an S may be already performing the best that can be expected from either of the treatments and placing him in the opposing treatment might depress his criterion score below even its present level. The investigator cannot infer that the assignment of overlapping subjects to any other treatment would necessarily bring the data into better fit with the overall regression lines.

Importance of the Region of Significance in the Two-Aptitude Case

Regions of significance mathematically defined in a two-aptitude space also may have little or no practical importance. Heterogeneous group regressions for the two-aptitude case is illustrated in Figure 3.

Insert Figure 3 about here

The importance of a region of significance in the two-aptitude case is (1) a positive function of the proportion of total observations which fall within that region and (2) a negative function of the amount of overlapping between the treatments within that region. Both of these indices of importance have been discussed with regard to the single-aptitude case. Generalization of the second (overlap) index to the two-aptitude case is as follows. With two aptitudes, a Treatment 1 observation evidences overlap if that observation falls closer to the Treatment 2 regression plane than the Treatment 1 regression plane and vice versa for a Treatment 2 observation. In other words, an observation is counted as overlapping if it lies on the "wrong" side of the midplane between the group regression planes. The midplane equation is

$$Mpt(A_{1i},A_{2i}) =$$
$$[\overline{Y}_1 + b_{11}(A_{1i} - \overline{A}_{11}) + b_{21}(A_{2i} - \overline{A}_{21}) + \overline{Y}_2 + b_{12}(A_{1i} - \overline{A}_{12}) + b_{22}(A_{2i} - \overline{A}_{22})]/2$$

where $Mpt(A_{1i},A_{2i})$ is the midplane criterion score for Aptitude 1 equal to A_{1i} and Aptitude 2 equal to A_{2i}; $\overline{Y}_1$ and $\overline{Y}_2$ are the criterion means for the two treatments; and b's are regression coefficients for aptitudes and treatments respectively; $\overline{A}_{11}$ and $\overline{A}_{21}$ are the Treatment 1 means on Aptitude 1 and Aptitude 2; and $\overline{A}_{12}$ and $\overline{A}_{22}$ are the Treatment 2 means on Aptitudes 1 and 2.

The overlap index is then the number of overlapping observations in a region divided by the total number of observations in that region. The midpoint of each region falling within the observed aptitude values is calculated. The significance of the difference (distance) between the corresponding region is a region of significance, while a nonsignificant difference at a midpoint indicates a region of nonsignificance.

Importance of the Region of Significance for Curvilinear Regressions

Recall considerations previously made with regard to the importance of a region of significance in the case of a linear relationship between aptitude and criterion. Such considerations also apply to regions of significance defined with regard to curvilinear regressions within treatments. The midline between curvilinear (quadratic) regression lines is given by the following equation:

$$Mpt(A_i) = [\overline{Y}_1 + b_{11}(A_i - \overline{A}_1) + b_{21}(A_i^2 - \overline{A}_1^2) + \overline{Y}_2 + b_{12}(A_i - \overline{A}_2) + b_{22}(A_i^2 - \overline{A}_2^2)]/2$$

where $Mpt(A_i)$ is the midline criterion score for the aptitude variable equal to A_i; $\overline{Y}_1$ and $\overline{Y}_2$ are the criterion means for the two treatments; $\overline{A}_1$ and $\overline{A}_2$ are the aptitude means for the two treatments; and the b's are regression coefficients for the aptitude, square of the aptitude, and treatments, respectively.

Applications of the Above Models to Published Research

Two of the above methods (one-aptitude linear and two-aptitude linear) for calculating size of and overlap in regions of significance have been applied to a sample of published and unpublished ATI studies to determine the importance--as opposed to simply the presence--of a region or regions of significance. Table 1 summarizes the results of this reanalyses (Borich, 1971; Carry, 1969; Eastman, 1972; Hughes, 1973; and Koran, 1969). The sample size for each study is shown in column 1. The percentage of the total cases lying in a region of significance are shown in columns 2 and 3. Finally, the percentage of overlap--those observations lying on the "wrong" side of the midline or midplane regression equation--are shown in columns 4 and 5.

The first thing to notice in the table is that the size of the sample bears little relation to the proportion of cases in the regions of significance. Borich's study had the fewest cases but shows the largest ratio of cases affected by the treatment (63 percent in the combined regions). In contrast, the largest study considered (Carry) had the smallest ratio of cases affected by the treatment with only 32 percent falling in the combined regions. Given that the overlap in the largest of these rather small regions was 37 percent, the logical conclusion is that the treatment in Carry's study had essentially no effect. The Eastman and Hughes studies did much better in terms of finding cases in the regions of significance. More than 40 percent of the sample fell in the largest region in each study. On the second test these two studies did not fare as well since they both had a relatively high degree of overlap in the largest region. About 50 percent overlap would be expected by chance so the 38 and 40 percent found by the two studies respectively do not represent a very clear-cut effect on the treatment groups. Rejecting the Eastman and Hughes studies leaves only the Koran and Borich results. Both show about 40 percent of the cases in the largest region of significance and a relatively low overlap within that region. Of course some caution must be observed with regard to the Borich findings since the small sample size makes the percentages quite unstable.

With just two of the five studies reanalyzed providing important--as contrasted with significant--findings, the moral is clear. Only by applying measures that give the size of an ATI effect can the practical value of the study be determined. Admittedly, the criteria used to accept or reject a percentage must be somewhat arbitrary at this point. As more and more studies report these measures it will be possible to get a "feel" for the relative importance of these percentages much as we have a "feel" for the size of an important correlation.

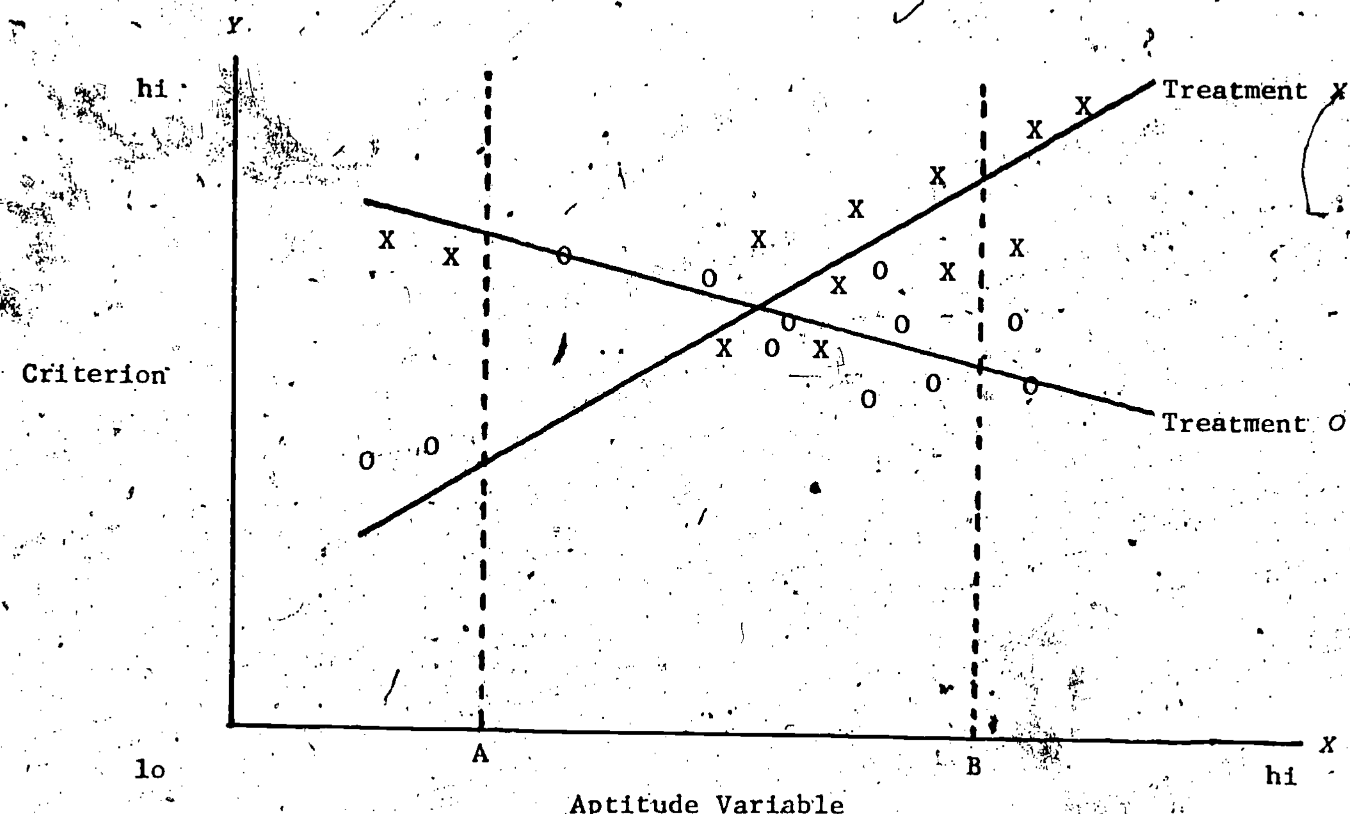

Figure 1. Heterogeneous group regressions. A region of significance is defined to the left of point A and to the right of point B.

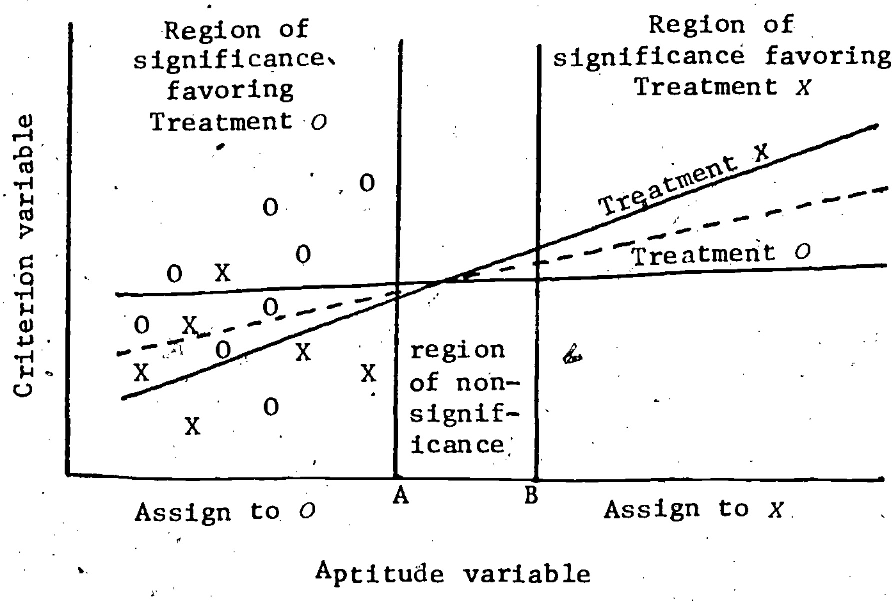

Figure 2. Overlap within a region of significance.

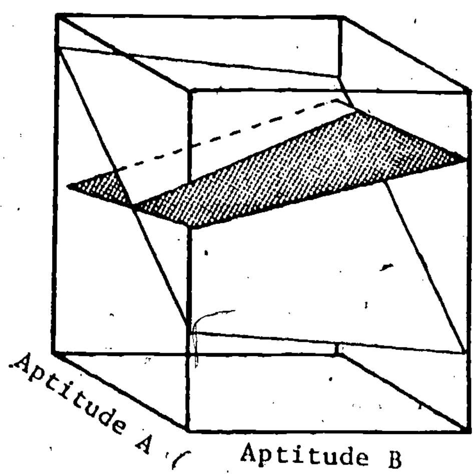

Figure 3. Heterogeneous group regressions for the two-aptitude case.

Table 1

	N	Percent in Regions		Percent Overlap	
Borich (1971)	30	40	23	8	29
Carry (1969)	181	21	11	37	30
Eastman (1972)	80	43	0	38	–
Hughes (1973)	123	41	*	40	*
Koran (1969)	76	0	47	–	31

Note: Hughes data consisted of one predictor variable, all other studies had two predictors.

*Only a single region of significance was definable. A single variable, of course, does not preclude two regions.

–Index not applicable since region had no cases.

References

Behr, M. J. Interactions between "structure-of-intellect" factors and two methods of presenting concepts of modular arithmetic--A summary paper. Journal of Research in Mathematics Education, 1970, 1, 29-42.

Berliner, D. C., & Cahen, L. S. Trait-treatment interaction and learning. In F. Kerlinger (Ed.), Review of Research in Education, Peacock Press, 1973.

Borich, G. D. Final report: Learning and transfer in concept attainment as a function of concept rule, cue saliency, and stimulus variety. Bloomington, Indiana: Indiana University, 1971.

Carry, L. R. Interaction of visualization and general reasoning abilities with instructional treatment in algebra. Unpublished doctoral dissertation, Stanford University, 1969.

Cronbach, L. J., & Snow, R. E. Aptitudes and instructional methods. (preliminary edition) School of Education, Stanford University, 1974.

Eastman, P. M. The interaction of spatial visualization and general reasoning abilities with instructional treatment in quadratic inequalities: A follow-up study. Unpublished doctoral dissertation. The University of Texas at Austin, 1972.

Edwards, A. L. Experimental design in psychological research. New York: Holt, Rinehart and Winston, Inc., 1971.

Hernandez, N. G. A model of classroom discourse for use in conducting aptitude-treatment interaction studies. Journal for Research in Mathematics Education, 1973, 4, 161-169.

Hughes, D. C. An experimental investigation of the effects of pupil responding and teacher reacting on pupil achievement. American Educational Research Journal, 1973, 10, 21-37.

Koran, M. R. The effects of individual differences on observational learning in the acquisition of a teaching skill. Unpublished doctoral dissertation, Stanford University, 1969.