DOCUMENT RESUME

ED 048 346 TM 000 420

AUTHOR Chen, Martin K.
TITLE Extension of Guilford's Rating Adjustment Technique to Situations Where Not All Raters Rate All Ratees.
PUB DATE Feb 71
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association, New York, New York, February 4-7, 1971

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS Analysis of Variance, Classroom Observation Techniques, *Error Patterns, Mathematical Models, *Rating Scales, *Reliability, *Research Methodology, *Statistical Analysis

ABSTRACT

Generally, ratings have notoriously low inter-rater reliabilities. Because of differences in orientation, background, and expectations, ratings are seldom made from the same point of reference; thus, many types of error mask the true rating variance. Guilford's technique identified most types of constant error by analysis of variance and then eliminated by an adjustment technique. This, however, does not remove all possible errors; it only "cleans up" the ratings to the extent these known errors no longer contribute to the error variance. The technique, moreover, is not applicable where not all raters have ratings for all ratees on all traits. It is proposed that if the number of missing values is relatively small, one of several methods be used to estimate the missing data before an analysis of variance is performed. Depending on the validity of the statistical assumptions made in the estimation, these methods are capable of producing reasonable estimates for the missing data. Three possible methods are illustrated with data taken from Guilford's work. The accuracy of the methods is compared before performing the analysis of variance and making the required adjustments. Further, the inter-rater reliabilities of the adjusted data and of the unadjusted data are estimated, using the Spearman-Brown formula by analysis of variance. A comparison of the two reliability coefficients reflects the degree to which the adjustment has been useful. (Author)

Paper presented at American Educational Research Association
annual meeting in
New York, February 4-7, 1971

Extension of Guilford's Rating Adjustment Technique to Situations Where Not All Raters Rate All Ratees

Martin K. Chen
National Center for Health Services Research and Development
Rockville, Maryland 20852

Generally, ratings have notoriously low inter-rater reliabilities. Because of differences in orientation, background and expectations, it is seldom that the raters rate the ratees from the same point or points of reference. The result is that the ratings are composed of many types of error that mask the true rating variance.

Guilford (1954) has developed a technique by which most types of constant error are identified by analysis of variance and then eliminated by an adjustment technique. The main effects and interaction effects of the analysis of variance are linked with leniency error, halo error, and rater-trait interaction error. Leniency error corresponds with significant main rater effect due to the fact that some raters tend to over-rate or under-rate all of the ratees. Halo effect corresponds with significant rater-ratee interaction effect due to the fact that some raters tend to over-rate or under-rate certain ratees in all the traits. The rater-trait interaction effect is due to the fact that some raters tend to over-rate or under-rate certain traits of all of the ratees.

Following identification of these errors, Guilford then systematically

proceeds to eliminate them by making proper adjustments which free the ratings of these errors. This technique does not remove all the possible errors from the ratings; it only "cleans up" the ratings to the extent these known errors are no longer contributing to the error variance.

Guilford's technique, however, is not applicable in situations where not all the raters have ratings for all the ratees on all the traits. It is proposed that in cases where the number of missing values is relatively small, say less than 10 per cent, one of several methods be used to estimate the missing data before an analysis of variance is performed on the data. Depending on the validity of the statistical assumptions that are made in the estimation, these methods are capable of producing "reasonable" estimates for the missing data, reasonable in the sense that when they are included in the analysis of variance schema, the results of the analysis approximate those that would obtain had there been no missing values.

When there are more than one observations in each cell, the obvious estimate is the mean of the other observations in that cell. This procedure, however, is not applicable to the rating situation where each rater rates each ratee on each trait only once, resulting in a single observation in each cell in a three-factor (rater, ratee, and trait) analysis layout. In this situation estimates of missing values in the empty cells must be based on the remainder of the cells with known values.

There are two conditions in which missing values may occur. One condition

is that one or more raters may not know one or more ratees well enough to provide reasonable or valid ratings. Or a rater may know a ratee well enough to rate him on certain traits that he has had opportunities to observe, but not on other traits which he has had no opportunities to observe.

Another condition relates to the necessity of eliminating ratings that are obviously out of line with the rest of the ratings. This is a tricky problem that only the experimenter himself with his unique knowledge of the experimental situation can cope with. Yates (1933) provides some guidelines which may be helpful. Although his guidelines are intended for agricultural researchers, they are equally applicable to social sciences research. If, in the opinion of Yates, the experimenter knows that some external influences have affected a particular observation to the degree that it becomes outstanding, then this observation should be rejected and its value estimated. On the other hand, an outstanding observation per se is not enough reason for rejection if it is known that whatever factors that affected that particular observation also affected other observations to unknown degrees.

The utility of the estimation procedures may be questioned on ground that analysis of variance techniques are available for factorial designs with missing values. This argument, however, does not hold in the case of ratings where it is desired to know the inter-rater reliabilities of the ratings on one or more traits. Furthermore, as Yates (1933) points out,

appropriate estimates of missing values included in the treatment means make these means efficient estimates of treatment differences, free from bias due to other effects.

It is important to point out that although including the estimates in the analysis of variance schema enables one to perform the analysis of variance on orthogonal data, these estimates add no new information to the experiment. For this reason, the degrees of freedom in the error estimates must be reduced by the number of missing values, regardless of which estimation procedure is used.

In this paper three estimation procedures are briefly discussed and employed to estimate "missing" values in Guilford's (1954) data. Since the "missing" values are actually known, it is possible to compare the estimates with the observed values and to compare the estimates themselves. The closeness of the estimates to the observed data provides a measure of the accuracy of the estimates.

Winer (1962) describes two methods for estimating single-valued missing cells. The simpliest method, referred to as Winer-1 in this paper, is to estimate the missing value by the deviation of the sum of the means of the particular row and column (in which the missing value lies) from the grand mean. Smybolically, $X'_{ij} = \bar{A}'_i + \bar{B}'_j - \bar{G}'$, where X'_{ij} is the estimated value, $\bar{A}'_i$ the row mean, $\bar{B}'_j$ the column mean, and $\bar{G}'$ the grand mean. This method assumes that there is no two-factor interaction effect, since

this effect is not in the model.

The second method, termed Winer-2, is based on the thesis that proper estimates when included in the analysis of variance schema should preserve the profile of simple effects which would obtain if there were no missing values. The missing value is thus proportional to the neighboring cell in the next column as the two adjacent cells above and below the missing value are proportional to <u>their</u> neighboring cells in the next column. The missing value is obtained by solving a direct proportional equation, in which the two known proportions are averaged. Symbolically, this could be represented thus:

$$\frac{M_{i,j}}{X_{i,j+1}} = \frac{\frac{X_{i-1,j}}{X_{i-1,j+1}} + \frac{X_{i+1,j}}{X_{i+1,j+1}}}{2}$$

Where $M_{i,j}$ is the missing value to be estimated, $X_{i+1,j}$ is the value below the missing value, $X_{i-1,j}$ the value above the missing value, $X_{i-1,j+1}$ and $X_{i+1,j+1}$ are respectively adjacent to the two values in the next column. You will notice that this equation is not applicable if the missing value happens to be in the first or last row of the schema. Since Winer does not discuss this situation, I merely used one proportion rather than the average of the two proportions in such cases. I am aware of the fact this procedure will subject the estimate to greater error, but as I will attempt to demonstrate later in this paper, this procedure is not recommended even if the missing value is not in the first or last row.

The third method is that proposed by Yates (1933). This method produces estimates that, when included in the analysis of variance layout, make the error variance the smallest of all possible error variances when other estimates are used. The formula for computing the missing value is $X' = \frac{pP + qQ - T}{(p - 1)(q - 1)}$.

Where p is the number of treatments, q the number of blocks, P and Q the treatment and block totals respectively, from which the value is missing, and T the grand total.

If there are more than one missing value, T is augmented by the sum of rough estimates of all but one of the missing values, which could be the sum of the means of treatments exclusive of the missing values or (M - 1) times the grand mean, M being the number of missing values. Once a missing value is estimated by this formula, then that estimate is added to T in place of the rough estimate for that value. When all the missing values are estimated, the second iteration begins with new T values. What makes this formula attractive is that regardless of the initial values of the rough estimates, convergence occurs very rapidly, usually at the third or fourth iteration. I have written a computer program in PL/1 that does the job fairly efficiently with two missing values. If you are doing work in this area, I will be glad to send you a source deck, provided you have

a functional PL/1 compiler on your system.

Table 1 is reproduced from Guilford (1954). The numbers with a slash across them are those that are supposed to be missing and estimated by the three methods. Table 2 gives the observed values together with the estimated values by the three methods for each of the five traits. You will notice that in some instances some of the estimated values come fairly close to the observed values, but in many cases they differ from the observed values considerably. On the average, Yates' procedure appears to produce more accurate estimates than the other two, the worst being Winer's second method.

The important question is, however, not whether or not the estimation procedure is capable of producing accurate estimates of missing data, but rather whether or not the estimated values, when included in the analysis of variance schema, will change the decisions that are normally made in hypothesis testing with complete data. The answer to this question is furnished by a detailed comparison of Tables 3 and 4. Table 3 is reproduced from Guildord. The three analyses of variance are based on his original, complete data. Table 4 summarizes the results of the analyses based on data that include estimated values by Yates' procedure. You will notice that with the exception of Part II, in which the interaction is significant at the .01 level in Table 3 and at the .05 level in Table 4, the two tables are identical in terms of the decisions that have to be made. These results

are in accord with Federer's (1955, p. 125) observations that the validity of the analysis of variance procedure is not disturbed if the proportion of estimated missing values is not large.

Having demonstrated the utility of Yates' procedure, I will now turn to Guilford's procedure of making adjustments on the ratings to increase their reliability. I will not go into the detail of his adjustment procedure, since it is clearly described in his book. Guilford reasons that by eliminating certain types of error, the ratings will have greater reliability. Table 5 furnishes empiric evidence of the correctness of his statement in a rather dramatic way. While the reliabilities could have been computed with data that include estimated values, Table 5 is based on Guilford's original data and his adjusted data. In all the five traits the improvement in reliability is considerable, in one case the new coefficient being more than double the coefficient based on unadjusted data. Since unreliability of measures attenuates their correlation with other measures and tends to inflate the standard error of difference in means, Guilford's technique should be a boon to researchers interested in measurement validity and significant differences.

References

1. Guilford, J. P. Psychometric Methods. New York: McGraw-Hill, 1954.

2. Yates, F. The analysis of replicated experiments when the field results are incomplete. Empire Journal of Experimental Agriculture, I, 129-142, 1933.

3. Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

4. Federer, W. T. Experimental design. New York: McMillan, 1955.

Table 1 *

Ratings of Seven Ratees in Five Traits, as Given by Three Raters

Rater / Ratee	Trait A			Trait B			Trait C			Trait D			Trait E		
	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
1	⅘	6	5	5	⅘	5	3	4	5	⅘	6	7	3	3	3
2	9	8	7	7	7	7	5	5	5	8	7	7	5	2	5
3	3	4	3	3	5	5	⅗	3	5	7	6	5	1	6	5
4	7	5	⅘	3	6	3	1	4	5	3	5	3	3	5	1
5	9	2	9	7	4	7	7	3	7	8	2	7	5	⅘	7
6	3	4	3	5	4	⅘	3	6	⅗	5	4	⅘	1	2	3
7	7	3	7	7	3	7	5	5	7	5	5	5	5	4	7

*Table reproduced from Guilford (1954, p.282). Slashed numbers are supposed to be "missing" values that are to be estimated.

Table 2

Estimates of Two Missing Values with Corresponding Observed Values for Five Traits

	Observed		Yates	Winer-1	Winer-2
Trait A	1	5	6.72	6.36	6.75
	2	5	6.27	6.20	6.43
Trait B	1	5	4.33	4.57	5.01
	2	3	5.06	4.90	7.10
Trait C	1	3	3.01	6.98	6.80
	2	3	5.83	4.64	8.00
Trait D	1	5	7.46	6.96	6.85
	2	5	4.38	4.19	6.20
Trait E	1	3	6.33	6.03	8.33
	2	7	4.97	4.87	6.00

Table 3 *

Summary of Analysis of Variance of Ratings of Seven Ratees in Five Traits as Given by Three Raters

I. Ignoring Ratee Differences

Source	SS	DF	MS	F	P
Between Raters (R)	9.05	2	4.52	1.35	NS
Between Traits (T)	46.53	4	11.63	3.47	.05
Interaction (R X T)	12.96	8	1.62	.48	NS
Within	301.71	90	3.35		
Total	370.25	104			

II. Ignoring Differences between Traits

Source	SS	DF	MS	F	P
Between Raters (R)	9.05	2	4.52	2.26	NS
Between Ratees (I)	94.92	6	15.82	7.91	.01
Interaction (R X I)	98.68	12	8.22	4.11	.01
Within	167.60	84	2.00		
Total	370.25	104			

III. Ignoring Rater Differences

Source	SS	DF	MS	F	P
Between Ratees (I)	94.92	6	15.82	6.25	.01
Between Traits (T)	46.53	4	11.63	4.60	.01
Interaction (I X T)	51.47	24	2.14	.85	NS
Within	177.33	70			
Total	370.25	104			

*Reproduced from Guilford (1954, p.283).

Table 4

Summary of Analysis of Variance of Ratings of Seven Ratees in Five Traits as Given by Three Raters, with Estimated Missing Values

I. Ignoring Ratee Differences

Source	SS	DF	MS	F	P
Between Raters (R)	9.75	2	4.88	1.34	NS
Between Traits (T)	47.80	4	11.95	3.28	.05
Interaction (R X T)	21.53	8	2.69	0.74	NS
Within	290.72	80	3.63		
Total	369.80	94			

II. Ignoring Differences between Traits

Source	SS	DF	MS	F	P
Between Raters (R)	9.75	2	4.88	1.80	NS
Between Ratees (I)	86.76	6	14.46	5.36	.01
Interaction (R X I)	73.25	12	6.10	2.26	.05
Within	200.05	74	2.70		
Total	369.81	94			

III. Ignoring Rater Differences

Source	SS	DF	MS	F	P
Between Ratees (I)	86.76	6	14.46	5.26	.01
Between Traits (T)	47.80	4	11.95	4.34	.01
Interaction (I X T)	70.20	24	2.93	1.06	NS
Within	165.05	60	2.75		
Total	369.81	94			

Table 5

Spearman-Brown Reliabilities of Original Ratings and of Adjusted Ratings Computed by Analysis of Variance

	Original Ratings	Adjusted Ratings
Trait A	r = .58	r = .91
Trait B	r = .45	r = .85
Trait C	r = .45	r = .83
Trait D	r = .50	r = .79
Trait E	r = .27	r = .72