FACTORS WHICH INFLUENCE THE RELIABILITY OF HUMAN JUDGMENTS OF STRESS

D. R. Brogan, University of North Carolina at Chapel Hill

Purpose of Investigation

It was desired to estimate the reliability with which persons could make judgments regarding the stress produced in them by various newspaper headlines. One reason for such estimation was that persons were to be selected from a North Carolina rural community to act as judges of stress for certain segments of the community. Naturally, persons selected to act as judges should be able to make reliable judgments regarding situations which produce stress. In addition, it was desired to investigate the factors which contribute to a person's having a high or low reliability in these judgments of stress. The factors reported on here are demographic characteristics of the persons.

Measuring Instrument

Two questionnaires were administered by personal interview to each subject in the sample in spring, 1970. There was a one month time span between the two interviews. The first questionnaire contained one hundred newspaper headlines. Some examples are "Eight per cent interest rate approved,""Impeach President Nixon," "A nuclear war threat is seen," "Franklinton mayor is returned to jail," "North Carolina schools plan various fee hikes." There was an equal representation of local, state, and national headlines in these 100 headlines.

Subjects were asked to rate each newspaper headline on a five point scale ranging from stressful (1) to reassuring (5), with the midpoint (3) used for a neutral reaction, i.e. neither stressful nor reassuring. In addition to the subject's responses to the 100 newspaper headlines, demographic data were collected on each subject.

The second questionnaire, administered one month later, contained fifty newspaper headlines from the original set of 100 on the first questionnaire. Keeping a fairly equal representation of national, state, and local headlines, the least ambiguous headlines in the first questionnaire were selected for the second questionnaire. The second questionnaire also contained questions about the stress produced by various life events such as a new birth in the family, a change in job, etc. The analysis of this life event data is reported elsewhere.[1]

Subjects

A previous study of the county had drawn a random sample of households, and then a random sample of one adult within each household.[2] The sample thus obtained was judged to be representative of the county, based on 1960 census data. The sample for this investigation was drawn by selecting a household adjacent to the household selected in the previous study, and then selecting at random one adult within each household. By this method, 111 adults were selected for this investigation.

In the second interview, 96 persons out of the original 111 were followed up, and these 96 persons constitute the sample for this investigation. The following are the demographic characteristics of this sample: 88% are rural residents, while the other 12% are residents of a small town. 63% are white; 51% are female. 73% have lived in the community for over 20 years, reflecting a very stable population. 10% are in the age range 20-29, 25% in the range 30-39, 27% in the range 40-49, 21% in the range 50-59, and 17% are 60 and above. 79% are married, 12% are widowed, 5% are single, and 4% are divorced or separated. 13% have some college education, 33% have a high school education, 38% have a junior high school education, and 16% have less than 6 years of formal schooling. 5% are employed in professional occupations, 10% in white collar occupations, 10% in skilled occupations, and 19% in semi-skilled occupations. 20% are farm laborers, and 32% are housewives. 23% have an annual income less than $3000, 33% have an annual income of $3000 to $6000, 28% have an annual income of $6000 to $9000, and the remaining 16% have an annual income of $9000 or over. 63% of the subjects are heads of households.

Measures of Reliability

The most commonly used measure of reliability is a correlation coefficient. Hence, a correlation coefficient between the responses on interview 1 and interview 2 was obtained for each of the 96 subjects by adding over the responses to the fifty newspaper headlines common to each interview. Letting X_{ijk} denote the response of person i (i=1,...,96) to newspaper headline j (j=1,...,50) at time k (k=1,2), then the measure of reliability COR for person i is given by

$$COR(i) = \frac{\sum_{j=1}^{50}(X_{ij1}-\bar{X}_{i.1})(X_{ij2}-\bar{X}_{i.2})}{\left[\sum_{j=1}^{50}(X_{ij1}-\bar{X}_{i.1})^2 \sum_{j=1}^{50}(X_{ij2}-\bar{X}_{i.2})^2\right]^{1/2}}$$

where

$$\bar{X}_{i.k} = \frac{1}{50}\sum_{j=1}^{50} X_{ijk}, \quad \text{for } k=1,2.$$

Note that this is not the structure of a correlation coefficient in the usual sense because, for a given person, the responses to the 50 newspaper headlines are probably not independent of one another. However, it is used here for a reliability measure in much the same way that it is used in cluster analysis as a measure of similarity between two units.[3] In the sample of 96 subjects, COR ranged from a low of .12 to a high of .99. Table 1 gives the frequency distribution of the values of COR. For this measure, a low score means low reliability and a high score means high reliability.

The second measure of reliability is the Fisher-z transformation of the correlation coefficient discussed above, i.e. FISH. For a given value of COR,

$$FISH = 1/2 \ln \frac{1+COR}{1-COR}.$$

The Fisher-z transformation was considered as a potential reliability measure because it is approximately normally distributed and, perhaps then, the distribution of the 96 values of FISH would be approximately normally distributed. In the sample of 96 subjects, the value of FISH ranged from .12 to 2.63. Table 1 gives the frequency distribution of the values of FISH. For this measure, a low score means low reliability and a high score means high reliability.

Correlation coefficients, although used extensively as reliability coefficients, aren't always the best indication of reliability since they only measure the linear relationship between two variables. For example, in this study, a subject could have answered each of the 50 newspaper headlines on the first interview with a judgment of 1 (stressful), and all 50 headlines on the second interview with a judgment of 5 (reassuring). The reliability, as measured by COR, would be 1.0. This doesn't sound intuitively reasonable, since this subject is making very different judgments on the same headline at the two interview times. One way to measure reliability without this drawback is to use a squared distance measure, also common to the field of cluster analysis.[3] The sum of squared deviations for each person i is given by

$$SSD(i) = \sum_{j=1}^{50}(X_{ij1}-X_{ij2})^2.$$

In the sample of 96 persons, this value ranged from a low of 3 to a high of 155. In order to use this measure so that a low value means low reliability, an adjusted squared deviation measure (ASD) is defined by

$$ASD = 1 - SSD/155.$$

This measure ranged from a low of 0.0 to a high of .98 in the sample of 96 persons. Table 1 gives the frequency distribution of ASD.

Table 2 gives the correlation coefficients between these three measures of reliability. In general, they are highly correlated with one another. Hence, it probably doesn't make too much difference which measure is used to select persons from the sample who would make "good," i.e., "reliable," judges. ASD is recommended, however, since it correlates highly with the other two and, in addition, has more of an intuitive appeal for this particular investigation.

Independent Variables

The independent variables which were used in an attempt to predict the degree of reliability are:

1) Residence (1=rural, 2=small town)

2) Race (1=white, 2=nonwhite)

3) Sex (1=male, 2=female)

4) Marital Status (1=not married, 2=married)

5) Head of Household (1=yes,2=no)

6) Occupation
   1=professional
   2=white collar
   3=skilled
   4=semi-skilled
   5=unskilled

7) Years Lived in Community
   1=$\leq$ 1 year
   2=1< years $\leq$ 3

3=3 < years ≤ 6
4=6 < years ≤ 10
5=10 < years ≤ 20
6=years > 20

8) Age (1=20-29, 2=30-39, 3=40-49, 4=50-59, 5=60 and above)

9) Education
1=some college or college graduate
2=high school graduate
3=less than high school, but more than junior high
4=less than 6 years or no schooling

10) Annual Family Income
1=income < 3000
2=3000 ≤ income < 6000
3=6000 ≤ income < 9000
4=9000 ≤ income < 15,000
5=income ≥ 15,000

Note that variables 1 thru 5 are dichotomous, nominal variables. Variables 6 thru 10 are categorized, ordinal variables.

Method of Analysis

The three dependent variables COR, ASD, and FISH were considered separately. For each dependent variable, the 10 independent variables were used in a stepwise multiple regression analysis in an effort to determine which variables influenced the measures of reliability.

Variables 7 thru 10 could be used as indicated above, or else the median value of the category could be assigned to each person within the category. Both methods were used here, and they gave similar results.

Results

Only the results obtained by using the independent variables as indicated previously will be discussed, since the results obtained by using the midpoint of the categories for variables 7 thru 10 were very similar.

1. Regression Analysis of COR

The correlations between COR and each of the independent variables ranged, in absolute value, from .001 to .18. COR was most highly correlated with race (r=.18), head of household (r=-.13), and income (r=-.13). All of these correlations are, of course, very small. These three variables enter the stepwise regression equation in the above order and give a multiple R of .22 (multiple R^2=.05). If all independent variables are allowed into the equation, multiple R increases only to .24 (multiple R^2=.06). Hence, none of the independent variables really explain a significant amount of the variability in COR. (None of the F tests for a regression effect were significant at α=.05.)

2. Regression analysis of FISH

The correlation of FISH with the 10 independent variables ranged, in absolute value, from .06 to .23. The highest correlations were with race (r=.23), income (r=-.16), sex (r=-.11), and head of household (r=-.11). The first two variables entering the stepwise regression equation were race and sex. This yielded a multiple R of .26 (multiple R^2=.07). The F ratio, with (2,93) df, was 3.421, which is statistically significant at α=.05. However, the multiple R^2 is still comparatively small.

3. Regression analysis of ASD

The correlation of ASD with the 10 independent variables ranged, in absolute value, from .001 to .135. The highest correlations were with head of household (r=-.135) and race (r=.131). Using these two variables, multiple R was .18(multiple R^2=.03), and it was nonsignificant at α=.05.

Discussion

Only the reliability measure FISH showed a statistically significant relationship with any of the independent variables, although the multiple R was still quite small. Of interest are the correlation coefficients of highest value. For example, race had the highest correlation with the dependent variables FISH and COR and the second highest correlation with ASD. In all three cases, the correlation coefficient indicated that blacks have higher reliability coefficients than whites, although the magnitude of this relationship was statistically significant only for the reliability coefficient FISH. Head of household was another variable showing some of the highest correlations with the dependent variables. All three correlation coefficients indicated that persons who are head of households have higher reliability coefficients, although, again, the particular correlation coefficients are not statistically significant at α=.05.

In general, one can conclude that none of the independent variables are useful in predicting the degree of reliability of the judges.

FOOTNOTES

[1]Miller, F. T., W. K. Bentz, and D. R. Brogan (1971). "Perception of Life Crisis Events: A Comparative Study of Rural and Urban Americans," submitted to Cross-Cultural Psychology.

[2]Edgerton, J. Wilbert, Willard Bentz, and William Hollister (1970). "Demographic Factors and Responses to Stress Among Rural People," Amer. J. Public Health, 60, 1065-1071.

[3]Sokal, R. R. and P. H. Sneath (1963). The Principles of Numerical Taxonomy. W. H. Freeman, San Francisco, California.

Table 1

Frequency Distribution of COR, ASD, and FISH

COR		ASD		FISH	
Range	Frequency	Range	Frequency	Range	Frequency
.0-.10	0	.00-.10	4	.00-.20	1
.11-.20	1	.11-.20	8	.21-.40	2
.21-.30	1	.21-.30	5	.41-.60	8
.31-.40	1	.31-.40	18	.61-.80	23
.41-.50	6	.41-.50	18	.81-1.00	36
.51-.60	11	.51-.60	12	1.01-1.20	8
.61-.70	27	.61-.70	13	1.21-1.40	9
.71-.80	27	.71-.80	8	1.41-1.60	3
.81-.90	16	.81-.90	5	1.61-1.80	0
.91-1.00	6	.91-1.00	5	1.81-2.00	3
				2.01-2.20	0
Minimum .12		Minimum 0.0		2.21-2.40	1
Maximum .99		Maximum .98		2.41-2.60	1
Median .71		Median .46		2.61-2.80	1
				Minimum .12	
				Maximum 2.63	
				Median .88	

Table 2

Correlation Coefficients Between COR, ASD, and FISH

	COR	ASD
ASD	.859	-
FISH	.886	.840