RELIABILITY ASSESSMENT OF A SURVEY INSTRUMENT

Ramesh M. Choudhari and Shobha R. Choudhari
South Carolina State University, Orangeburg, SC 29117

ABSTRACT

In a data oriented project, the reliability of results essentially depends on the reliability of the data collected, which depends on the reliability of the survey instrument used. So the reliability of the instrument becomes an important issue. The purpose of the paper is to explain the design and implementation of a system to assess the reliability of a survey instrument. Some of the popular software systems such as SPSS do have reliability procedures to compute reliability of an instrument. However, we have designed and implemented a system that is standalone, user friendly, and does not need specific technical knowledge of software to use. We have also achieved a high degree of automation for user convenience. This system would be useful and convenient in the assessment and construction of a survey instrument in research.

INTRODUCTION

For quantitative research, researchers develop a survey instrument consisting of items as a part of measurement procedure. These items need to have consistency of measurement for what it is supposed to be measuring. If the consistency of the items in a survey was questionable, then the results derived from that survey instrument would be questionable and so not reliable to use. This consistency is called reliability in quantitative research. Another way to look at reliability is to consider the correlation of an item or instrument with a hypothetical one which truly measures what it is supposed to. Since the true instrument is not available, we need to estimate reliability in different ways. There are four different ways to estimate reliability:

1. Internal Consistency Reliability: It is used to establish the consistency based on the correlation among the items of the survey instrument. The same instrument is administered to a group of subjects on one occasion to estimate reliability.

2. Parallel-Forms Reliability: It is estimated based on the two equivalent survey instruments of the scale constructed in the same way.

3. Test-Retest Reliability: It is estimated based on the correlation between two or more administrations of the same survey instruments on the same (or similar) sample at different times or locations, assuming there is no substantial change in the scale being measured between two administrations. The amount of time allowed between measures is critical, because longer the time interval, lower the correlation and so lower the reliability.

4. Inter-rater Reliability: It is estimated based on the correlation of scores between/among two or more raters/interviewers who rate the same survey instrument for the same sample. To get the expected outcome, the raters should be as blind as possible and should be randomly assigned.

Reliability is a property of the *scores of a measure* rather than the measure itself and is thus going to be *sample dependent.* Reliability analysis allows us to study the properties of measurement scales and the items that make them up. Data can be dichotomous, ordinal, or interval, but they should be coded numerically. Observations should be independent, and errors should be uncorrelated between items. Each pair of items should have a bivariate normal distribution. Scales should be additive, so that each item is linearly related to the total score. There are many other factors that could affect the reliability such as total number of questions in a survey and quality of the questions. More number of questions as well as effective questions will improve the reliability of the survey. In this paper, the researchers would like to focus on reliability of the data, and the computerized system developed to check the reliability of the survey instrument. The system we developed is a self-contained standalone system and does not need any statistical software to run this system. It is menu-driven and user does not need to know any commands. User also does not require any expertise of any statistical software to use this system. The system has a high degree of automation (only a few keystrokes) to perform a complete reliability analysis [1][11][15][18][21].

CONCEPTUAL MODELS

The conceptual models used to construct the reliability system are:

Alpha (Cronbach) Model: This is a model of internal consistency, based on the average inter-item correlation. The alpha reliability of the variable is derived by assuming each item represents a retest of a single item. For example, if there are five items, it's as if the five scores are the retest scores for one item. But the reliability is calculated in such a way that it represents the reliability of the *mean* of the items, not the reliability of any single item. So, for example, the alpha reliability of 10 items would be higher than that of 5 similar items. Alpha reliability should be regarded as a measure of internal consistency of the mean of the items at the time of administration of the questionnaire. It is not test-retest reliability. For that, the questionnaire has to be administered on two or more occasions. For dichotomous data, alpha coefficient is equivalent to the Kuder-Richardson 20 (KR20) coefficient.

Split-half Model: This model splits the scale into two parts and examines the correlation between the parts. It computes correlation between forms, Guttman split-half reliability, Spearman-Brown reliability (equal and unequal length), and coefficient alpha for each half.

Guttman Model: This model computes Guttman's lower bounds for true reliability. Reliability coefficients are lambda 1 through lambda 6.

Parallel Model: This model assumes that all items have equal variances and equal error variances across replications.

Strict Parallel Model: This model makes the assumptions of the parallel model and also assumes equal means across items.

Parallel and Strictly Parallel Model: This model computes test for goodness-of-fit of model, estimates of error variance, common variance, and true variance, estimated common inter-item correlation, estimated reliability, and unbiased estimate of reliability [9][10][11][16][23].

RELABILITY SYSTEM

Based on the models explained above, the researchers developed the computerized reliability system. The system has several modules to collect the information about a data file to estimate reliability for the survey instrument on which data was collected. A data file consists of data lines and a data line consists of data-fields. Data-fields represent items in a survey form and association between data-fields and items has to be left to right and one to one. The model and implementation of the reliability system is explained below.

SYSTEM MODEL

The system model is essentially composed of three main modules: User Interface Module, Processing Module, and Output Module. A procedural architecture of the system is shown in Figure 1.

User Interface Module

The user interface module contains the mouse and keyboard event handlers to collect information from the users regarding the total number of data-fields in the data-file and locations of a data file and an output file.

Processing Module

The processing module consists of several submodules, each one responsible for processing specific tasks, such as accessing the values of each data-field, processing the data, computing several types of statistical analysis, estimating reliability of the instrument based on different reliability models.

Output Module

The output module is responsible for saving the results created by the processing module in an output file and displaying the results [3][4][8][14] [22].

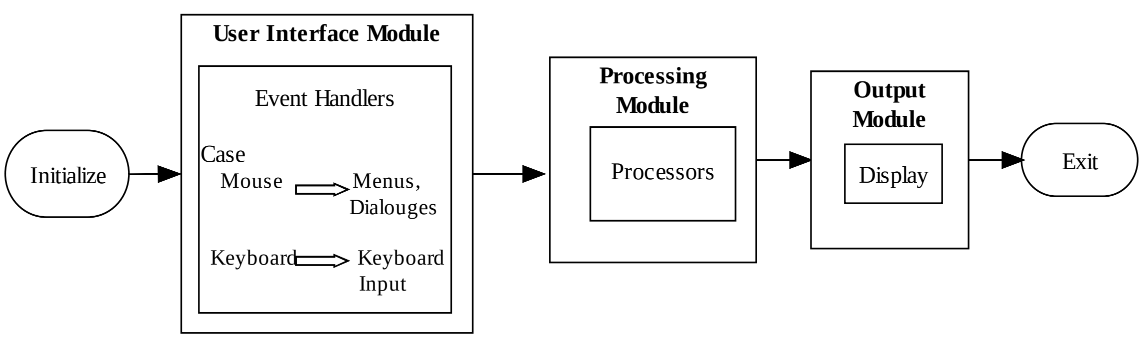

Figure 1. Procedural Architecture of the System

IMPLEMENTATION

A prototype of the system is implemented in the Microsoft Windows environment. The menu driven graphic user interface is implemented using the current GUI techniques. User dialogues are implemented to collect user input as needed.

The heart of the system is the processing module that is implemented by the main processor. The processor consists of several subprocessors that are responsible for specific tasks such as processing user inputs, processing data, and performing necessary computations by implementing the conceptual models on the data.

In this process, several types of descriptive and inferential statistics are generated. Descriptive statistics on items such as individual item means and standard deviations, as well as, correlation coefficients are generated. In addition, item-total statistics are also generated, where the Cronbach Alpha coefficient of reliability is computed with the scale mean, scale variance, and corrected item-total correlation for each case in which an individual item is deleted. This will aid in estimating the relative contribution of each item to the overall instrument. Cronbach Alpha and Standard Alpha reliability coefficients are computed to estimate the reliability of the instrument. Spearman-Brown and Guttman coefficients can be generated for the split-half model in addition to the alpha coefficient for each part. If the total number of items in the sample is odd, the system will divide the sample into two parts with the first part having one more item. The system saves all the generated results in the output file.

The output module is implemented by the display procedures using data aware controls. It displays the results of the reliability analysis and the results will be saved in a text file that can be opened by any text editor or world processor [2][4][5] [6] [12] [13][19][20][22].

TESTING

The testing of this system was done on more than one data set. A data file with at least two subjects and at least four items is required to cover all the categories of reliability analysis generated by this system. A data set used for testing in the following example contains 18 items and 268 subjects.

This menu driven system collects the necessary input from user through menus and user dialogues. The output file that contains results of the reliability analysis on the data file after running the reliability system on the test data is shown below (on the next page) in Figure 2. The output generated by the system consists of several sections such as correlation matrix, item statistics, item-total statistics, and various reliability coefficients. Results of each section start on a new page in the original output though here the output is provided without page breaks.

RELIABILITY ANALYSIS

CORRELATION MATRIX

ITEMS	1	2	3	4	5	6	7	8	9
1	1.0000								
2	0.2509	1.0000							
3	0.7117	0.2183	1.0000						
4	0.5385	0.3338	0.5346	1.0000					
5	0.1953	0.3092	0.2714	0.4434	1.0000				
6	0.2870	0.3850	0.2882	0.3854	0.3919	1.0000			
7	0.3869	0.3849	0.3368	0.3118	0.5427	0.5465	1.0000		
8	0.2196	0.4313	0.1408	0.3554	0.5577	0.3460	0.5523	1.0000	
9	0.2868	0.3599	0.1700	0.4589	0.6311	0.3234	0.5611	0.6941	1.0000
10	0.2697	0.1268	0.1333	0.1564	0.2639	0.1758	0.1584	0.2533	0.2919
11	0.5469	0.3210	0.4826	0.6226	0.5204	0.3435	0.5107	0.5352	0.6210
12	0.6230	0.3276	0.4948	0.4373	0.2732	0.4131	0.4757	0.3326	0.2805
13	0.3992	0.2895	0.4558	0.5180	0.4130	0.3806	0.4123	0.3215	0.3509
14	0.2376	0.4029	0.1733	0.3715	0.2632	0.4617	0.4215	0.3752	0.3878
15	0.4952	0.3007	0.3449	0.4533	0.5432	0.5132	0.5484	0.5374	0.5765
16	0.5270	0.2998	0.4456	0.3981	0.2563	0.2949	0.4541	0.4015	0.3548
17	0.5096	0.3422	0.4546	0.4309	0.4453	0.4536	0.6373	0.4256	0.4216
18	0.5282	0.2615	0.4550	0.5063	0.4929	0.2986	0.4220	0.4145	0.4485

ITEMS	10	11	12	13	14	15	16	17	18
10	1.0000								
11	0.3227	1.0000							
12	0.2956	0.5068	1.0000						
13	0.0208	0.5635	0.3622	1.0000					
14	0.1153	0.2788	0.2463	0.2074	1.0000				
15	0.5472	0.5618	0.5208	0.2986	0.4092	1.0000			
16	0.3414	0.5223	0.5136	0.3876	0.2816	0.5364	1.0000		
17	0.1199	0.5952	0.5031	0.5445	0.2655	0.3972	0.5462	1.0000	
18	0.4086	0.6429	0.5377	0.3611	0.3043	0.6459	0.5908	0.5067	1.0000

ITEM STATISTICS

TOTAL NUMBER OF ITEMS = 18
TOTAL NUMBER OF CASES = 268

NO	MEAN	STD DEV
1	1.7500	0.9367
2	2.9030	1.1075
3	1.8955	1.0113
4	2.7239	1.2141
5	2.2687	1.1262
6	3.0933	1.2158
7	2.2052	1.0308
8	2.6530	1.2701
9	2.3694	1.2931
10	3.2052	1.0488
11	2.4478	1.1551

Figure 2. Results on Reliability Analysis of a Survey Instrument

12	2.2164	1.1006
13	1.6903	0.8640
14	3.8209	1.1107
15	2.7500	1.3101
16	2.1716	1.0206
17	2.2537	1.0435
18	2.4552	1.1619

	MEAN	MINIMUM	MAXIMUM	RANGE	MAX/MIN	VARIANCE
ITEM MEANS	2.4930	1.6903	3.8209	2.1306	2.2605	0.2874
ITEM VARIANCES	1.2510	0.7464	1.7163	0.9699	2.2994	0.0718

ITEM-TOTAL STATISTICS

STATISTICS FOR SCALE	MEAN	VARIANCE	STD DEV	NO OF VARIABLE
	44.8731	175.6168	13.2520	18

NO	SCALE MEAN IF ITEM DELETED	SCALE VARIANCE IF ITEM DELETED	CORRECTED ITEM-TOTAL CORRELATION	ALPHA IF ITEM DELETED
1	43.1231	160.1084	0.6172	0.9189
2	41.9701	161.0777	0.4736	0.9221
3	42.9776	161.0482	0.5277	0.9207
4	42.1493	154.6892	0.6441	0.9180
5	42.6045	156.9216	0.6176	0.9186
6	41.7799	157.1461	0.5575	0.9202
7	42.6679	156.7170	0.6911	0.9170
8	42.2201	154.2997	0.6244	0.9186
9	42.5037	152.9850	0.6553	0.9177
10	41.6679	164.8668	0.3583	0.9245
11	42.4254	152.4401	0.7658	0.9149
12	42.6567	157.0128	0.6306	0.9183
13	43.1828	162.6443	0.5548	0.9203
14	41.0522	161.2482	0.4656	0.9223
15	42.1231	149.9361	0.7469	0.9151
16	42.7015	158.2926	0.6340	0.9183
17	42.6194	156.9258	0.6733	0.9174
18	42.4179	154.0269	0.7018	0.9165

RELIABILITY COEFFICIENTS

CRONBACH ALPHA COEFFICIENT = 0.9231
STANDARDIZED ALPHA = 0.9232

Figure 2-(cont'd). Results on Reliability Analysis of a Survey Instrument

STATISTICS FOR SPLIT-HALF METHODS

NUMBER OF ITEMS IN PART1 =	9
NUMBER OF ITEMS IN PART2 =	9
MEAN FOR PART 1 =	21.8619
MEAN FOR PART 2 =	23.0112
VARIANCE FOR PART 1 =	48.3966
VARIANCE FOR PART 2 =	46.7602
CORRELATION BETWEEN THE TWO PARTS =	0.8457
EQUAL LENGTH SPEARMAN-BROWN COEFFICIENT =	0.9164
UNEQUAL LENGTH SPEARMAN-BROWN COEFFICIENT =	0.9164
GUTTMAN SPLIT HALF COEFFICIENT =	0.9163
ALPHA COEFFICIENT FOR PART1 =	0.8531
ALPHA COEFFICIENT FOR PART2 =	0.8646

Figure 2-(cont'd). Results on Reliability Analysis of a Survey Instrument

CONCLUSION

In experimental or applied research projects, data becomes an important entity. The collection of data depends on a survey instrument that researchers use. The reliability of the results depends on the reliability of the data collected. The reliability of data in turn depends on the reliability of the instrument used. So the reliability of an instrument becomes an important issue. The purpose of our system is to assess the reliability of the instrument before it can be used. It also gives opportunity to correct and enhance the quality of the instrument before its use. The reliability system is necessary for researchers and analysts since reliable data is important to get the consistent and reliable results from the data. We hope, in this respect, this system would be an important and useful tool for researchers in developing reliable survey instruments for their research.

REFERENCES

[1] Armor, D. J. (1974). Theta reliability and factor scaling. Pp. 17-50 in H. Costner, ed., *Sociological methodology*. San Francisco: Jossey-Bass.

[2] Bell Doug. *Software Engineering: A Programming Approach*. Addison Wesley, 2000.

[3] Booch Grady. *Object Oriented Analysis and Design with Applications*. The Benjamin/Cummings Publishing Company, Inc., 1993.

[4] Borland Software Corporation. *Borland Delphi for Windows*. Scotts Valley, CA.

[5] Budd Timothy. *An Introduction to Object-Oriented Programming*. 3rd Edition. Addison Wesley Publishing, 2001.

[6] Cantu Marco. *Mastering Delphi 7*. SYBEX Inc., 2004.

[7] Coad Peter and Yourdon Edward. *Object Oriented Analysis*. Prentice Hall, 1990.

[8] Dersham Herbert and Jipping Michael. *Programming Languages: Structures of Models*. 2nd Edition. PWS Publishing Company, Inc., 1995.
[9] Ebel, Robert L. (1951). Estimation of the reliability of ratings. *Psychometrika* 16: 407-424.
[10] Fleiss, J. L., Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33: 613-619.
[11] Hopkins, W. G. (2000). Summarizing Data: Precision of Measurement. In: A new view of statistics. Internet Society for Sport Science: http://www.sportsci.org/resource/stats/precision.html.
[12] Jalote Pankaj. *An Integrated Approach to Software Engineering*. 3rd Edition. Springer, 2005.
[13] Koffman Elliot. *Turbo Pascal. 5th Edition*. Addison Wesley Publishing, 1995.
[14] Kovach Warren. *Delphi 3: User Interface Design*. Prentice Hall, Europe, 1998.
[15] Litwin, Mark S. How to Assess and Interpret Survey Psychometrics. The Survey Kit Series, Sage Publications. 2nd Edition. June 2003.
[16] McKelvie, S. J. (1992). Does memory contaminate test-retest reliability? *Journal of Gen Psychology* 119(1):59-72. This article reports that reliability estimates under test-retest designs are not inflated due to memory effects.
[17] Rachele Warren. *Learn Object Pascal with Delphi*. Wordware Publishing, Inc., 2001.
[18] Reliability Analysis. http://www2.chass.ncsu.edu/garson/pa765/reliab.htm.
[19] SPSS Inc., *SPSS Statistical Algorithms*, SPSS Inc., Chicago, IL, 1985, pp. 176-189.
[20] Thurrott Paul, Brent Gary, Bagdazian Richard, and Tendon Steve. *Delphi 3 SuperBible*. Watt Group Press, 1996.
[21] Walter, S. D.; Eliasziw, M. ; and Donner, A. (1998). Sample size and optimal designs for reliability studies. *Statistics in Medicine* 17: 101-110.
[22] Williams Shirley and Walmsley Sue. *Discover Delphi Programming Principles Explained.* Addison Wesley Publishing, 1998.
[23] Zumbo, B. D.; Gadermann, A. M.; & Zeisser, C.. (2007). Ordinal versions of coefficients alpha and theta for likert rating scales. *Journal of Modern Applied Statistical Methods*, 6, 21-29.