CHAPTER 11

Measures of Association and Item Analysis

OBJECTIVE

The main objective of this chapter is to review some special agreement coefficients that are used in item analysis, or used to quantify the extent of agreement among raters with respect to the rankings of subjects. We will discuss Cronbach's alpha, known to measure internal consistency of items in scale development. We will also discuss widely-used measures of association such as the Kendall's coefficient of concordance, the Spearman's correlation coefficient, among others. We will present the assumptions underlying these techniques as well as the practical situations in which their use is recommended.

CONTENTS

11.1 Overview

In this chapter, we will review Cronbach's alpha coefficient that is frequently used in item analysis, and will discuss a number of other measures of association aimed at evaluating the extent of agreement among raters with respect to the ranking of subjects. These special agreement coefficients do not fall into the category of chance-corrected measures such as Kappa, nor into the category of intraclass correlation coefficients. Nevertheless, they are often used by practitioners to address specific inter-rater reliability problems that cannot be resolved with the methods discussed in previous chapters.

Cronbach's alpha is discussed in section 11.2. Researchers involved in scale development will find this coefficient useful. It allows you to select an adequate set of questions for developing a scale needed to measure a particular construct of interest. In sections 11.3, 11.4, and 11.5 we present methods, which are recommended for evaluating the extent of agreement among raters when the subject's rank relative to other subjects is more relevant than its actual score. In many inter-rater reliability problems, the exact score assigned to subjects is not essential. What matters, is how each subject ranks with respect to other subject. We previously gave an example of government examiners scoring proposals submitted by contractors in response to a pre-solicitation notice. Although each individual proposal is scored, it is the ranking that determines the winner. Any inter-rater reliability experiment designed to improve the extent of agreement among government examiners for example, will produce data that should be analyzed with rank-based methods.

11.2 Cronbach's Alpha

Cronbach's Alpha[1] (α) is a measure of internal consistency that is popular in the field of psychometrics. This measure was originally developed in a context where a set of questions (also called items) are asked to a group of individuals with the objective of measuring a specific construct such as risk aversion, extraversion or introversion. The extent to which all questions contribute positively towards measuring the same concept is known as internal consistency. This is a key element for evaluating the quality of the overall score. Cronbach's alpha is one of the most widely-used measure of internal consistency.

Items that are internally consistent can be seen as raters that agree about the

[1] The term alpha (α) is used very often in the study of reliability as we saw in the past few chapters. We previously investigated Krippendorff's alpha as well as Aickin's alpha. Cronbach's alpha however, proposed by Cronbach (1951) does not have much in common with the previous alpha coefficients.

"true" value of the construct associated with the subjects that participated in the experiment. In that sense, Cronbach's alpha could be seen as an agreement coefficient, although some authors refer to it as a measure of association, or simply a measure of reliability. The terms agreement and association are generally used in this context in a lousy sense.

11.2.1 *Defining Cronbach's Alpha*

Table 11.1 contains the responses that 15 employees provided to 6 questions aimed at measuring the quality of leadership in the company they are working for. Each of these questions must be answered on a five-level Likert scale[2] defined as 1= "Strongly Disagree", 2="Disagree", 3="Undecided", 4="Agree", 5="Strongly Agree." The 5 questions are the following:

- **1A**: I know my organization's mission (what it is trying to accomplish).
- **1B**: I know my organization's vision (where it is trying to go in the future).
- **1C**: My senior (top) leaders use our organization's values to guide us.
- **1D**: My senior leaders create a work environment that helps me do my job.
- **1E**: My organization's leaders share information about the organization.
- **1F**: My organization asks what I think.

Let k be the number of items (or questions), s_i^2 the variance associated with item i, and s_{T}^2 the variance associated with the total (or sum) of all k item scores. Cronbach's alpha is mathematically defined as follows:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{1}{s_{\mathrm{T}}^2}\sum_{i=1}^{k} s_i^2\right) \tag{11.1}$$

Referring to Table 11.1, we have $k = 6$ (the number of questions), which leads to 6 item variances from s_1^2 to s_6^2. Each of the 6 item variances may be calculated using the data in each item's column. Then the total variance s_{T}^2 may be calculated using data in the Total column.

[2]Likert's scales are attributed to Rensis Likert (1931), and are further discussed by McIver and Carmines (1981)

11.2.2 *How Does Cronbach's Alpha Evaluate Internal Consistency*

A fundamental question to ask is whether the α coefficient as defined by equation 11.1 accomplishes its mission, which is to evaluate internal consistency. In our opinion, the answer is that this mission is well accomplished to a large extent. To see this, we need to re-write equation 11.1 as follows:

$$\alpha = \frac{k\overline{c}/\overline{v}}{1+(k-1)\overline{c}/\overline{v}}, \tag{11.2}$$

where $\overline{c}$ is the average of all $k(k-1)$ pairwise covariances associated with the k items under investigation, and $\overline{v}$ the average of all k item variances s_i^2 (i varying form 1 to k). The ratio $\overline{c}/\overline{v}$ (let's call it $\overline{r}^\star$) can be seen as a proxy measure of correlation between 2 arbitrary items taken from the set of k items. If the k items are internally consistent (meaning they are highly correlated), then you would expect $\overline{r}^\star$ to be close to 1. In the extreme situation of perfect correlation among the k items, the correlation coefficient $\overline{r}^\star$ will be 1, which in turn leads to an alpha coefficient of 1. If the k items are inconsistent (i.e. are weakly correlated) then $\overline{r}^\star$ is expected to be close to 0. If the k items are totally independent, then their correlation coefficient $\overline{r}^\star$ will be 0, leading to a 0 alpha coefficient. Because of all these properties, the alpha coefficient is seen as a measure of internal consistency that varies between 0 and 1 (i.e. between low consistency and high consistency).

Table 11.1: Leadership Data from 15 Respondents

	Questions/Items						
Respondent	1A	1B	1C	1D	1E	1F	Total
1	5	5	3	3	2	2	20
2	4	4	3	5	4	4	24
3	4	5	4	4	4	4	25
4	5	3	4	4	2	2	20
5	4	3	5	4	3	3	22
6	5	2	1	1	1	1	11
7	5	5	4	4	4	4	26
8	4	4	3	2	1	1	15
9	5	5	1	1	1	1	14
10	5	4	1	2	2	1	15
11	4	4	3	4	3	2	20
12	3	2	3	4	3	4	19
13	5	5	5	1	4	1	21
14	5	5	3	2	4	4	23
15	5	4	3	2	4	2	20

Example 11.1

To illustrate the calculation of Cronbach's alpha, let us consider the data in Table 11.1. As previously mentioned, the number of items being studied is $k = 6$. The 6 item-level variances are given by $s_1^2 = 0.4095$, $s_2^2 = 1.1429$, $s_3^2 = 1.6381$, $s_4^2 = 1.8381$, $s_5^2 = 1.4571$, and $s_6^2 = 1.6857$. The variance associated with the total score is $s_{\text{T}}^2 = 18.3810$. We calculated these variances using Excel and the appropriate variance function. Since the sum of the item-level variances is 8.1714, the alpha coefficient is calculated as follows:

$$\alpha = (6/(6-1)) \times (1 - 8.1714/18.3810) = 0.6665.$$

Is there a threshold that α must exceed before we can conclude that the items are internally consistent ? The answer is that there is no official and widely-accepted threshold. A rule of thumb that has been advocated in the literature (c.f. Nunnaly, 1978) is to require α to equal 0.70 or exceed it before the items are considered internally consistent.

There are a few important comments about Cronbach's alpha that are worth mentioning:

- The α coefficient as defined by equation 11.1 is expected to always fall between 0 and 1. In reality, that will not always be the case, especially when the number subjects participating in the experiment is small. The alpha coefficient could indeed take a negative value. The cause of this odd situation is some negative between-item covariances with a large absolute value. The only thing that is known with certainty is α always being below 1.
- Cronbach's alpha is the generalization of the Kuder-Richardson Formula 20 (often referred to in the literature as KR-20), which was proposed for dichotomous items by Kuder and Richardson (1937). Although the KR-20 is still being mentioned in the literature, it no longer plays an important role in the study of reliability other than being a special case of Cronbach's alpha.
- **Standardized Cronbach's alpha**

  When the k items under investigation use different measurement units, summing the variances s_i^2 will be problematic since they would also be expressed in different units. The same issue can be raised regarding the calculation of the total variance s_{T}^2.

  This problem has been resolved by using the *Standardized Cronbach's Alpha* coefficient, which is defined as follows:

$$\alpha_{\text{S}} = \frac{k\overline{r}}{1 + (k-1)\overline{r}}, \tag{11.3}$$

  where $\overline{r}$ is the average of all distinct $k(k-1)/2$ pairwise correlation coefficients between the k items. Note that the correlation coefficient between 2 items can always be calculated whether they are both expressed with the same units or with different units.

To illustrate the calculation of the standardized Cronbach's alpha, let us consider the leadership data of Table 11.1. For this dataset, Cronbach's alpha was estimated to be 0.6665.

Example 11.2

Since the leadership test is based on 6 items (or questions), there are $6 \times (6-1)/2 = 15$ different pairwise correlation coefficients that can be calculated. All 15 correlation coefficients are shown in Table 11.2. The average correlation coefficient is given by $\overline{r} = 0.1838$, which leads to a standardized alpha of,

$$\alpha_{\text{S}} = \frac{6 \times 0.1838}{1 + (6-1) \times 0.1838} = 0.5747.$$

Table 11.2: Item Pairwise Correlation Coefficients from Table 11.1 Data

	1A	1B	1C	1D	1E
1B	0.4176				
1C	-0.2209	0.1566			
1D	-0.5708	-0.1478	0.4583		
1E	-0.1295	0.3321	0.6103	0.4190	
1F	-0.4470	0.0515	0.4126	0.7223	0.6927

It appears that the standardized Cronbach's alpha is smaller that the raw coefficient. This is not an indication of the superiority of the raw alpha coefficient. These 2 coefficients are based on 2 different computation techniques, and each needs specific benchmarks for interpreting the extent of internal consistency.

► Length of the Test

The test length measured by the number of items, affects the magnitude of Cronbach's alpha. That is, alpha increases as the number k of items goes up. Consequently, a good test must have an adequate number of items in order to achieve a reasonable internal consistency as evaluated with Cronbach's alpha.

Moreover, many test developers want the magnitude of alpha to be an attribute of the test. This will be the case only if alpha is calculated based on a sample of test takers that is representative of the universe they were selected from. Even in this case, alpha would be an attribute of the test and the target population of potential test takers. However, if the sample of test takers is not representative of any target population, then alpha becomes an attribute of both the test and the specific sample used, with limited value beyond.