Empirical Estimates of Reliability

I. Cronbach's Alpha

Preliminary steps, raw, standardized, KR20

II. Special Cases

III. Example

average inter-item correlation, item total correlation, alpha if item deleted

IV. Cronbach's Alpha: Some Properties

I. Cronbach's Alpha

Preliminary steps

- Generate descriptive statistics, including means, standard deviations (and/or variances, skewness and kurtosis)
- Obtain frequency tables and histograms
- Check for errors in entry, coding, etc.
- Variables do not need to be normally distributed, but when they are highly skewed or kurtotic or they respondents have not used the full range of values, you may want to consider the wording of that item.
- Check correlations to confirm scoring direction is correct and potentially eliminate items that are supposed to correlate that do not

I. Cronbach's Alpha

Cronbach's alpha (Cronbach, 1951) is an estimate of internal reliability (sometimes called the "consistency coefficient")

Conceptually based on the proportion of true score to total observed score variance

$$R_{xx} = \frac{s_t^2}{s_t^2 + s_e^2} = \frac{s_t^2}{s_o^2}$$

I. Cronbach's Alpha

If we can estimate the proportion of the observed score variance that is due to measurement error, then we can estimate reliability

$$\text{Proportion error} = 1 - \frac{s_e^2}{s_o^2}$$

Cronbach's alpha (α) raw score form is:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum s_i^2}{s_X^2}\right)$$

k = number of items, s_i^2 is the variance for each item, and s_x^2 is the variance for the composite scale score (as a sum of the items)

I. Cronbach's Alpha

The domain sampling model conceptualizes the items (testlets or subtests) as retests, so that the average correlation between these subtests is a measure of reliability

Cronbach's alpha in the standardized form is:

$$\alpha = \frac{k\overline{r}_{ii'}}{1+(k-1)\overline{r}_{ii'}}$$

$\overline{r}_{ii'}$ is the average correlation among all pairs of items, and k is the number of items

I. Cronbach's Alpha

The standardized coefficient alpha is the alpha for the set of items after they have been standardized (converted to *z*-scores) and will be equal or higher than the raw score version

Raw score alpha assumes the variances of the of the items are equal, and if they are not, the raw score estimate will be smaller than the standardized estimate

Usually similar, but when items are on very different scales (e.g., some 5-point and some 9-point scales), the difference may be larger

I. Cronbach's Alpha

Composites scores calculated by the sum or mean tend to weight items with larger variances more heavily

Standardizing items before computing the composite will equally weight them, because variances are all equal to 1

In most applications, researchers do not bother to do standardize items, sometimes because the original metric is lost (e.g., average of items on a 7-point no longer between 1 and 7, but are z-score values instead)

I. Cronbach's Alpha

What is an acceptable alpha? Exceeding .70 is widely mentioned as a cutoff for acceptable reliability, but what is "acceptable" or "good" depends heavily of the consequences of using a measure with some certain level of reliability.

Many scales with an alpha of .70 can be improved, however.

And this value has been grossly over applied and over stated.

I. Cronbach's Alpha

The .70 criteria is commonly attributed to Nunnally (1978), a highly regarded psychometrician, but using.70 as a standardwas clearly not his intention:

> what a satisfactory level of reliability is depends on how a measure is being used. In the early stages of research . . . one saves time and energy by working with instruments that have only modest reliability, for which purpose reliabilities of .70 or higher will suffice. . . . In contrast to the standards in basic research, in many applied settings a reliability of .80 is not nearly high enough. In basic research, the concern is with the size of correlations and with the differences in means for different experimental treatments, for which purposes a reliability of .80 for the different measures is adequate. In many applied problems, a great deal hinges on the exact score made by a person on a test. . . . In such instances it is frightening to think that any measurement error is permitted. Even with a reliability of .90, the standard error of measurement is almost one-third as large as the standard deviation of the test scores. In those applied settings where important decisions are made with respect to specific test scores, a reliability of .90 is the minimum that should be tolerated, and a reliability of .95 should be considered the desirable standard. (pp. 245-246)

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill. Quote also given by: Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria what did they really say?. *Organizational research methods*, *9*(2), 202-220.

II. Special Cases

The Kuder-Richardson 20 (KR20; Kuder & Richardson, 1937) is a special case of Cronbach's alpha when the items are binary (e.g., yes/no or correct/incorrect)

Predated Cronbach's alpha

It is equivalent to the raw score form of Cronbach's alpha, so computation of α for a set of binary items will give the same result as the KR_{20}

II. Special Cases

You may also hear about the KR_{21}, a variant if one assumes equal difficulty of all items (Kuder & Richardson, 1937)—we will discuss item difficulty later in the course

II. Special Cases

Hoyt's Method (Hoyt, 1941) also predated Cronbach's coefficient with a conceptual formula and method based on analysis of variance (ANOVA) which is also equivalent to α

ANOVA partitions variance into between- and within-group variance, which can be translated into the Classical Test Theory ideas of true and error score variance

$$R_{xx} = Hoyt's = \frac{MS_B - MS_W}{MS_B} = \frac{s_o^2 - s_e^2}{s_o^2}$$

III. Example

Data from the Late Life Study of Social Exchanges (LLSSE), a national longitudinal study of social relationships funded by NIH

Ages 65-90 years of age

N = 916

Extensive interview of social relationships, health, and mental health

Example uses items from the Positive and Negative Social Exchanges (PANSE) measure

III. Example

Negative Social Exchange Items

Factor	In the past month, how often did the people you know. . . (*0 = never, 1 = not very often, 2 = sometimes, 3 = often, 4 = very often*)
Unwanted advice or intrusion	. . .give you unwanted advice?
	. . .question or doubt your decisions?
	. . .interfere or meddle in your personal matters?
Failure to provide help	. . .let you down when you needed help?
	. . .ask you for too much help?
	. . .fail to give you assistance that you were counting on?
Unsympathetic or insensitive behavior	. . .do things that were thoughtless or inconsiderate?
	. . .act angry or upset with you?
	. . .act unsympathetic or critical about your personal concerns?
Rejection or neglect	. . .leave you out of activities you would have enjoyed?
	. . .forget or ignore you?
	. . .fail to spend enough time with you?

Newsom, J.T., Mahan, T.L., Rook, K.S., & Krause, N. (2008). Stable negative social exchanges and health. *Health Psychology, 27*, 78-86.

III. Example

SPSS reliability command output (Inter-Item Correlations)

Inter-Item Correlation Matrix

	w1unw1 unwanted advice	w1unw2 question your decisions	w1unw3 interfere in personal matters	w1dwn1 let you down	w1dwn2 ask for too much help	w1dwn3 fail to give assistance	w1out1 leave you out	w1out2 forget/ignore you	w1out3 fail spend enough time w/ you	w1fai1 do thoughtless things	w1fai2 act angry or upset with you	w1fai3 act unsympathetic or critical
w1unw1 unwanted advice	1.000	.760	.690	.554	.528	.437	.527	.492	.437	.504	.496	.466
w1unw2 question your decisions	.760	1.000	.812	.549	.591	.502	.503	.530	.482	.476	.531	.513
w1unw3 interfere in personal matters	.690	.812	1.000	.538	.567	.531	.466	.506	.498	.469	.515	.547
w1dwn1 let you down	.554	.549	.538	1.000	.757	.799	.555	.565	.479	.547	.527	.538
w1dwn2 ask for too much help	.528	.591	.567	.757	1.000	.709	.555	.590	.506	.520	.551	.549
w1dwn3 fail to give assistance	.437	.502	.531	.799	.709	1.000	.429	.480	.444	.435	.497	.545
w1out1 leave you out	.527	.503	.466	.555	.555	.429	1.000	.837	.782	.640	.551	.536
w1out2 forget/ignore you	.492	.530	.506	.565	.590	.480	.837	1.000	.825	.625	.560	.577
w1out3 fail spend enough time w/ you	.437	.482	.498	.479	.506	.444	.782	.825	1.000	.570	.501	.536
w1fai1 do thoughtless things	.504	.476	.469	.547	.520	.435	.640	.625	.570	1.000	.720	.693
w1fai2 act angry or upset with you	.496	.531	.515	.527	.551	.497	.551	.560	.501	.720	1.000	.820
w1fai3 act unsympathetic or critical	.466	.513	.547	.538	.549	.545	.536	.577	.536	.693	.820	1.000

III. Example

SPSS reliability command output (Summary Item Statistics)

Summary Item Statistics

	Mean	Minimum	Maximum	Range	Maximum / Minimum	Variance	N of Items
Item Means	.546	.416	.761	.345	1.829	.011	12
Inter-Item Correlations	.566	.429	.837	.407	1.949	.011	12

III. Example

SPSS reliability command output (Reliability Statistics)

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.940	.940	12

III. Example

SPSS reliability command output (Item-Total Statistics)

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
w1unw1 unwanted advice	5.7871	150.887	.686	.642	.936
w1unw2 question your decisions	5.8155	147.438	.729	.750	.935
w1unw3 interfere in personal matters	6.1015	150.386	.719	.707	.935
w1dwn1 let you down	6.0207	148.431	.753	.752	.934
w1dwn2 ask for too much help	5.9989	147.984	.754	.662	.934
w1dwn3 fail to give assistance	6.0360	149.112	.676	.700	.937
w1out1 leave you out	6.1321	150.191	.745	.762	.934
w1out2 forget/ignore you	6.1255	148.197	.771	.794	.933
w1out3 fail spend enough time w/ you	6.0349	149.870	.703	.726	.935
w1fai1 do thoughtless things	5.9891	149.906	.723	.638	.935
w1fai2 act angry or upset with you	5.9891	148.385	.732	.729	.934
w1fai3 act unsympathetic or critical	5.9978	148.780	.740	.728	.934

IV. Cronbach's Alpha: Some Properties

- Cronbach's alpha is an estimate of internal reliability or consistency and does not indicate stability over time necessarily
- Alpha is a *lower bound estimate* of reliability, and actual reliability may be higher
- Alpha is equal to the estimate of reliability from all possible split halves
- Alpha assumes unidimensionality—if the measure really assesses more than one hypothetical construct (or factor), the estimate may be incorrect (lower than for each factor)

IV. Cronbach's Alpha: Some Properties

- A more heterogeneous group will have a higher alpha than a more homogeneous group, all other things equal
- Speeded tests may inflate alpha (Lord & Novick, 1968), related to the homogeneity phenomenon above
- Test length affects alpha—longer tests are more reliable

  Consider a single-item test vs. multiple item test

  Think about domain sampling—larger sample of items should be a better estimate of the population of items

IV. Cronbach's Alpha: Some Properties

Spearman-Brown prophecy formula

$$R_{xx-revised} = \frac{nR_{xx-original}}{1+(n-1)R_{xx-original}}$$

n is the factor by which the size is increased

If length is increased from a 10-item test is increased to 20 items (with the same average inter-item correlation), n = 2, because the length is increased by a factor of 2

IV. Cronbach's Alpha: Some Properties

Spearman-Brown prophecy formula

If length is increased from a 10-item test is increased to 20 items (with the same average inter-item correlation), n = 2, because the length is increased by a factor of 2. Assume the original reliability $R_{xx\text{-original}}$ is .6.

$$R_{xx-revised} = \frac{nR_{xx-original}}{1+(n-1)R_{xx-original}}$$

$$= \frac{2(.6)}{1+(2-1).6} = \frac{1.2}{1.6} = .75$$

IV. Cronbach's Alpha: Some Properties

Spearman-Brown prophecy formula (using average inter-item correlation)

$$R_{XX} = \frac{k\bar{r}_{ii'}}{1+(k-1)\bar{r}_{ii'}}$$

k is number of items, and $\bar{r}_{ii'}$ is the average inter-item correlation

IV. Cronbach's Alpha: Some Properties

Spearman-Brown prophecy formula (using average inter-item correlation)

$$R_{XX} = \frac{k\bar{r}_{ii'}}{1+(k-1)\bar{r}_{ii'}}$$

$$= \frac{5(.4)}{1+(5-1).4} = \frac{2}{2.6} = .77$$

IV. Cronbach's Alpha: Some Properties

Spearman-Brown prophecy formula (using average inter-item correlation)

$$R_{XX} = \frac{k\bar{r}_{ii'}}{1+(k-1)\bar{r}_{ii'}}$$

$$= \frac{20(.4)}{1+(20-1).4} = \frac{8}{8.6} = .93$$

IV. Cronbach's Alpha: Some Properties

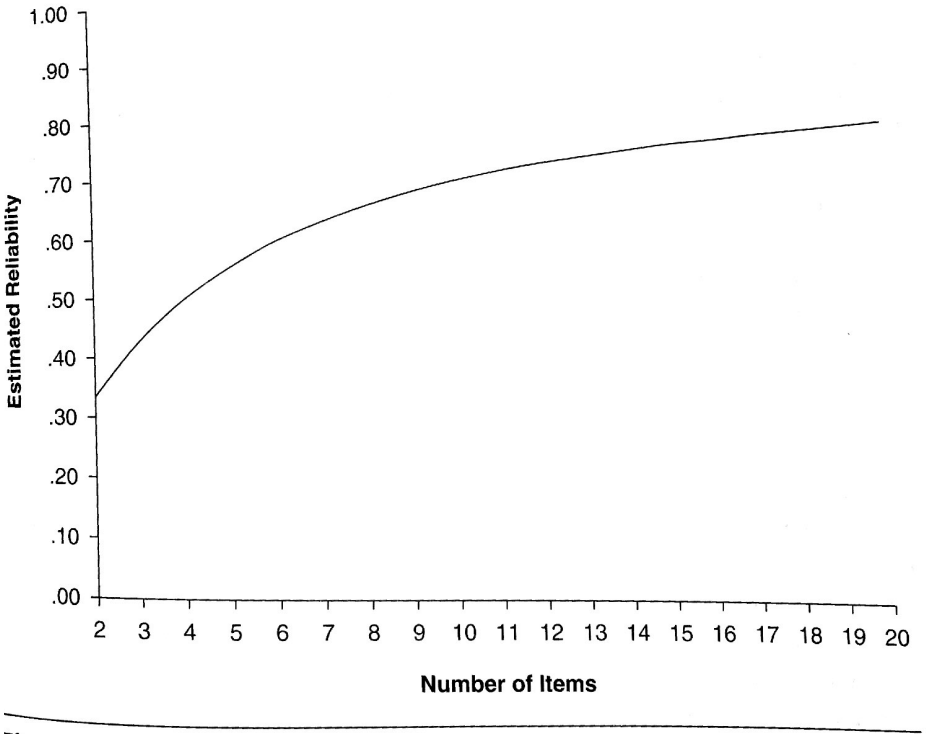

Figure 6.3 The Association Between Number of Items and Reliability (for a Test With an Average Interitem Correlation of .30)

Furr & Bacharach (2014, p. 151)

IV. Cronbach's Alpha: Some Properties

- Does not indicate that alpha is “biased” by the number of items, but it may be difficult to reach acceptable reliability with short scales even if inter-item correlation is fairly high
- Longer scales may still have high reliability even though some items are not so good
- Good idea to also look at average inter-item correlation and item-total statistics because of the sensitivity to length