Example of Cronbach's alpha calculation

Steve Simon

3/30/2019

```
suppressWarnings(suppressMessages(library(psych)))
```

My Clinical Research Methods class seemed a bit confused about my discussion of Cronbach's alpha (hereafter just called alpha). It's either because alpha is a confusing statistic or because Steve Simon is a confusing teacher. Probably a bit of both.

So let me take a bit of time to elaborate a bit more on this statistic.

Alpha only makes sense for composite measures. A composite measure is a single measurment that is computed as a sum of several individual items. What alpha does is to measure the internal consistency of the individual items.

The data set I will use to illustrate alpha is found in the R package. The name of the data set is "attitude" and it was included as an example of multiple linear regression in a classic textbook, Chatterjee and Price (1977), Regression Analysis by Example.

The R package has further details about this data set in html format (https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/attitude.html).

I will ignore the first column, an overall rating, and compute a composite measure of

- complaints: Handling of employee complaints
- privileges: Does not allow special privileges
- learning: Opportunity to learn
- raises: Raises based on performance
- critical: Too critical
- advance: Advancement

The first column is an overall rating. I will ignore this column from the perspective of reliability for reasons that I do not want to get into. I would note, however, that the overall rating could be a criterion, and if it correlated well with the composite scale of the remaining six items, that would be one way to establish validity.

Each of these items ranges from 0 to 100 and represents the number of "positive" responses in a survey of employees in a given department. Now, you should be a bit concerned about the fifth item, critical. Is this coded in a way that large values mean that most of the employees in a department thought their manager was too critical? Or do large values mean a "postive" response meaning that most employees disagreed that their manager was too critical?

Here is a listing of the entire data set (excluding the first column).

```
attitude[, -1] # exclude the first column
```

```
##    complaints privileges learning raises critical advance
## 1          51         30       39     61       92      45
## 2          64         51       54     63       73      47
## 3          70         68       69     76       86      48
## 4          63         45       47     54       84      35
## 5          78         56       66     71       83      47
## 6          55         49       44     54       49      34
## 7          67         42       56     66       68      35
## 8          75         50       55     70       66      41
## 9          82         72       67     71       83      31
## 10         61         45       47     62       80      41
## 11         53         53       58     58       67      34
## 12         60         47       39     59       74      41
## 13         62         57       42     55       63      25
## 14         83         83       45     59       77      35
## 15         77         54       72     79       77      46
## 16         90         50       72     60       54      36
## 17         85         64       69     79       79      63
## 18         60         65       75     55       80      60
## 19         70         46       57     75       85      46
## 20         58         68       54     64       78      52
## 21         40         33       34     43       64      33
## 22         61         52       62     66       80      41
## 23         66         52       50     63       80      37
## 24         37         42       58     50       57      49
## 25         54         42       48     66       75      33
## 26         77         66       63     88       76      72
## 27         75         58       74     80       78      49
## 28         57         44       45     51       83      38
## 29         85         71       71     77       74      55
## 30         82         39       59     64       78      39
```

Before you compute a composite score, look at the correlaton matrix of the six individual items. I am rounding the correlations to a single decimal.

```
round(cor(attitude[, -1]), digits=1)
```

```
##            complaints privileges learning raises critical advance
## complaints        1.0        0.6      0.6    0.7      0.2     0.2
## privileges        0.6        1.0      0.5    0.4      0.1     0.3
## learning          0.6        0.5      1.0    0.6      0.1     0.5
## raises            0.7        0.4      0.6    1.0      0.4     0.6
## critical          0.2        0.1      0.1    0.4      1.0     0.3
## advance           0.2        0.3      0.5    0.6      0.3     1.0
```

Notice that all of the correlations are positive. That includes "critical" which provides some reassurance that large values on this item represent a good thing.

If you notice that some of the correlations appear to be negative or very close to zero, that is a warning signal that maybe your composite will be a poor composite. It will end up being a mixture of apples and oranges. The zero and negatively correlated items will weaken the overall signal.

```
composite <-
  attitude$complaints +
  attitude$privileges +
  attitude$learning +
  attitude$raises +
  attitude$critical +
  attitude$advance
print(composite)
```

```
##  [1] 318 352 417 328 401 285 334 357 406 336 323 320 304 382 405 362 439
## [18] 395 379 374 247 362 348 293 318 442 414 318 433 361
```

```
summary(composite)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   247.0   320.8   359.0   358.4   399.5   442.0
```

The best department manager had a score of 442 and the worst department manager had a score of 247. The average rating was 358.4333333.

The simplest approach for assessing reliability is to split the composite into two halves and then corelating those two halves.

```
first_half <-
  attitude[ , 2] +
  attitude[ , 4] +
  attitude[ , 6]
second_half <-
  attitude[ , 3] +
  attitude[ , 5] +
  attitude[ , 7]
split_half_correlation <- cor(first_half, second_half)
print(round(split_half_correlation, digits=2))
```

```
## [1] 0.75
```

The correlation here is pretty good. You would like to see a correlation of 0.7 or greater. That's an arbitrary cut-off, so don't give up hope if your split-half correlation is 0.64.

Most researchers consider the split-half correlation to be an underestimate of the internal consistency, because shortening a composite measure by 50% will hurt reliability.

There is an adjustment, Spearman-Brown, that compensates for this. There is a general formula for the Spearman-Brown adjustment, but for split-half correlations, it simplifies to

$$r^* = \frac{2r}{1+r}$$

In our example, that would be

```
spearman_brown_adjustment <-
   2*split_half_correlation/(1+split_half_correlation)
print(round(spearman_brown_adjustment, digits=2))
```

```
## [1] 0.86
```

Now, the choice of how to split up your composite is very arbitrary, so very few researchers rely on the split-half correlation.

Alpha is another measure of internal consistency and very similar in spirit to the split-half correlation.

To compute alpha, you compare the variances of the individual items to the variance of the composite score.

$$\alpha = \frac{I}{I-1}\left(1 - \frac{\Sigma S_i^2}{S^2}\right)$$

where S_i^2 is the variance of the individual item and S^2 is the variance of the composite.

You can think of alpha as a measure of the synergy of the individual items. The whole is greater than the sum of the parts, so the saying goes.

Another way of thinking about it is that if the individual items are positively correlated, a high score on one item will carry over into the other items. This amplifies the signal you get when you compute the sum of the items.

The calculations are a bit tedious, but I'm happy to do it. Tedious is my middle name.

```
v2 <- var(attitude$complaints)
v3 <- var(attitude$privileges)
v4 <- var(attitude$learning)
v5 <- var(attitude$raises)
v6 <- var(attitude$critical)
v7 <- var(attitude$advance)
v <- var(composite)
print(round(v2, 1))
```

```
## [1] 177.3
```

```
print(round(v3, 1))
```

```
## [1] 149.7
```

```
print(round(v4, 1))
```

```
## [1] 137.8
```

```
print(round(v5, 1))
```

```
## [1] 108.1
```

```
print(round(v6, 1))
```

```
## [1] 97.9
```

```
print(round(v7, 1))
```

```
## [1] 105.9
```

```
print(round(v2+v3+v4+v5+v6+v7, 1))
```

```
## [1] 776.6
```

```
print(round(v, 1))
```

```
## [1] 2388.1
```

Notice that the sum of the individual item variances is a lot smaller than the variance of the composite. That's a good sign.

```
I <- 6
alpha <- I/(I-1)*(1-(v2+v3+v4+v5+v6+v7)/v)
print(round(alpha, digits=3))
```

```
## [1] 0.81
```

Several sources note that alpha is the average of every possible split-half correlation (after a Spearman-Brown adjustment). That's not too difficult to show, although tedious as all the earlier

calculations. There is a slight discrepancy, which I chalk up to rounding.

```
r234 <- cor(
  attitude[ , 2]+attitude[ , 3]+attitude[ , 4],
  attitude[ , 5]+attitude[ , 6]+attitude[ , 7])
r235 <- cor(
  attitude[ , 2]+attitude[ , 3]+attitude[ , 5],
  attitude[ , 4]+attitude[ , 6]+attitude[ , 7])
r236 <- cor(
  attitude[ , 2]+attitude[ , 3]+attitude[ , 6],
  attitude[ , 4]+attitude[ , 5]+attitude[ , 7])
r237 <- cor(
  attitude[ , 2]+attitude[ , 3]+attitude[ , 7],
  attitude[ , 4]+attitude[ , 5]+attitude[ , 6])
r245 <- cor(
  attitude[ , 2]+attitude[ , 4]+attitude[ , 5],
  attitude[ , 3]+attitude[ , 6]+attitude[ , 7])
r246 <- cor(
  attitude[ , 2]+attitude[ , 4]+attitude[ , 6],
  attitude[ , 3]+attitude[ , 5]+attitude[ , 7])
r247 <- cor(
  attitude[ , 2]+attitude[ , 4]+attitude[ , 7],
  attitude[ , 3]+attitude[ , 5]+attitude[ , 6])
r256 <- cor(
  attitude[ , 2]+attitude[ , 5]+attitude[ , 6],
  attitude[ , 3]+attitude[ , 4]+attitude[ , 7])
r257 <- cor(
  attitude[ , 2]+attitude[ , 5]+attitude[ , 7],
  attitude[ , 3]+attitude[ , 4]+attitude[ , 6])
r267 <- cor(
  attitude[ , 2]+attitude[ , 6]+attitude[ , 7],
  attitude[ , 3]+attitude[ , 4]+attitude[ , 5])
all_halves <- c(
  r234, r235, r236, r237,
  r245, r246, r247,
  r256, r257,
  r267)
print(round(2*all_halves/(1+all_halves), digits=2))
```

```
##  [1] 0.72 0.78 0.79 0.86 0.78 0.86 0.85 0.79 0.87 0.87
```

```
avg_all_halves <- mean(all_halves)
print(round(2*avg_all_halves/(1+avg_all_halves), digits=3))
```

```
## [1] 0.819
```

There were some questions about Kuder-Richardson 20 (KR-20) as well. It works pretty much the same way as alpha.

You use KR-20 for a composite variable that is composed of binary items. You can create a set of binary items from the previous data set by enforcing a pass-fail criteria on the ratings. A supervisor gets a passing grade on a particular item if 60% or more of his employees rate him positively on a particular item.

Note: I normally do not recommend taking nice continuous scales and creating binary variables. You are throwing away information. I'm doing this only to illustrate the calculation of KR-20.

```
binary <- NULL
binary$complaints <- as.numeric(attitude$complaints >= 60)
binary$privileges <- as.numeric(attitude$privileges >= 60)
binary$learning <- as.numeric(attitude$learning >= 60)
binary$raises <- as.numeric(attitude$raises >= 60)
binary$critical <- as.numeric(attitude$critical >= 60)
binary$advance <- as.numeric(attitude$advance >= 60)
binary
```

```
## $complaints
##  [1] 0 1 1 1 1 0 1 1 1 1 0 1 1 1 1 1 1 1 1 0 0 1 1 0 0 1 1 0 1 1
##
## $privileges
##  [1] 0 0 1 0 0 0 0 0 1 0 0 0 0 1 0 0 1 1 0 1 0 0 0 0 0 1 0 0 1 0
##
## $learning
##  [1] 0 0 1 0 1 0 0 0 1 0 0 0 0 0 1 1 1 0 0 0 1 0 0 0 1 1 0 1 0
##
## $raises
##  [1] 1 1 1 0 1 0 1 1 1 1 0 0 0 0 1 1 1 0 1 1 0 1 1 0 1 1 1 0 1 1
##
## $critical
##  [1] 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 1 0 1 1 1 1 1
##
## $advance
##  [1] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0
```

```
binary_composite <-
  binary$complaints +
  binary$privileges +
  binary$learning +
  binary$raises +
  binary$critical +
  binary$advance
binary_composite
```

```
##  [1] 2 3 5 2 4 0 3 3 5 3 1 2 2 3 4 3 6 5 3 3 1 4 3 0 2 6 4 1 5 3
```

Calculate the probability of a 1 for each item.

```
p2 <- sum(binary$complaints==1)/30
p3 <- sum(binary$privileges==1)/30
p4 <- sum(binary$learning==1)/30
p5 <- sum(binary$raises==1)/30
p6 <- sum(binary$critical==1)/30
p7 <- sum(binary$advance==1)/30
q2 <- 1 - p2
q3 <- 1 - p3
q4 <- 1 - p4
q5 <- 1 - p5
q6 <- 1 - p6
q7 <- 1 - p7
print(round(p2, digits=2))
```

```
## [1] 0.73
```

```
print(round(p3, digits=2))
```

```
## [1] 0.27
```

```
print(round(p4, digits=2))
```

```
## [1] 0.37
```

```
print(round(p5, digits=2))
```

```
## [1] 0.67
```

```
print(round(p6, digits=2))
```

```
## [1] 0.9
```

```
print(round(p7, digits=2))
```

```
## [1] 0.1
```

The formula for KR-20 is very similar to the formula for Cronbach's alpha.

$$KR - 20 = \frac{I}{I-1}\left(1 - \frac{\Sigma p_i q_i}{S^2}\right)$$

where $\Sigma p_i q_i$ is the sum of the variances of each individual item and S^2 is the variance of the composite. The similarity becomes even more striking if you think about the variance of a Bernoulli random variable. I won't clutter this handout up with those sort of things, though.

```
v <- var(binary_composite)
I <- 6
kr <- I/(I-1)*(1-(p2*q2+p3*q3+p4*q4+p5*q5+p6*q6+p7*q7)/v)
print(round(kr, digits=2))
```

```
## [1] 0.71
```

The value for the KR-20 coefficient is larger than 0.7, which is what most researchers like to see.

When should you use Cronbach's alpha and KR-20

There is a fair amount of controversy over these two statistics. First, they are not a measure of uni-dimensionality. To establish unidimensionality, you would need to run some sort of factor analysis or structural equations model.

What does uni-dimensionality mean, though? And why it is important? Uni-dimensionality means that all the items in the composite score are measuring more or less the same thing. Even with sophisticated models, you would have a hard time establishing this.

You'd like to have all the items measuring the same thing, because if a few of the items are measuring something else, a second dimension, then you are diluting the signal associated with most of items.

What Cronbach's alpha and KR-20 are measuring is internal consistency. Do the individual items in your composite travel together–large values of one item associated with large values of the other items and small values of one item associated with small values of the other items.

In my opinion, this is a small and unimportant feature of a composite scale. But other researchers will insist on this before they will publish your work. So you need to know about this.

Keep in mind that Cronbach's alpha and KR-20 only work for composite measures. If you are measuring pain using the Wong-Baker FACES pain rating scale (https://wongbakerfaces.org/), that's a single number with six possible values. Because it is single item so there is nothing internal to be consistent about. If you still need to assess reliability, and that would probably mean something like test-retest reliability.

As another example, the Clark scale (https://www.cancerresearchuk.org/about-cancer/melanoma /stages-types/clark-breslow-staging))is a measure used in melanoma cases. It is a single item with five levels describing how deeply a melanoma has penetrated the skin. Again, you can't use alpha because it is a single item scale. Since this is a physician report, your main concern about reliability would be assuring that two different physicians would, more often than not, use the same level when they are looking at the same melanoma. The intra-class correlation coefficient is your best bet here.

If you have a composite measure, then you can certainly consider alpha and KR-20. Use alpha when

the individual items have more than two levels and use KR-20 when the individual items are binary.