# Alternative Measures of Reliability: From $\alpha$ to $\omega$

Part of a symposium:
What Should We Do about Alpha II:
Alternatives to Alpha?
Society of Industrial and Organizational Psychology, Chicago.

William Revelle and David Condon

Northwestern University
Evanston, Illinois USA

Partially supported by a grant from the National Science Foundation: SMA-1419324
Slides at http://personality-project.org/sapa.html

April, 2018

# Outline

## What is reliability and why do we care?

1. Reliability is the correlation of a test with a test just like it.
   - It affects regression towards the mean ("reversion to mediocrity", Galton, 1886).
   - We use it to correct for attenuation (Spearman, 1904).
   - To estimate standard errors of measurement.
2. The other test can exist and we then have
   - Reliability over time
   - Reliability over alternative forms
3. If the other form does not exist, we find the correlation with a test "just like our test" that does not actually exist.
   - These reliabilities are based upon the internal structure of the test.
   - Simplest one to find has multiple names, most frequently called $\alpha$ (Cronbach, 1951) (also known as KR20, KR21 (Kuder & Richardson, 1937), $\lambda_3$ (Guttman, 1945).
   - These were shortcuts to the right answer that could be done with a desk calculator.
   - It is time to move on.

# Many ways of estimating reliabity: Most are not used
# All are easy to find

But what are they?

1. Multiple Occasions
   - 2 Occasions: Test retest reliability
   - Many Occasions: multilevel reliabilities (Shrout & Lane, 2012)
2. One Occasion
   - Greatest Lower Bound (Bentler, 2017)
   - $\omega_t$ (McDonald, 1999)
   - Best Split Half ($\lambda_4$) (Guttman, 1945)
   - Average Split half ($\approx \alpha = \lambda_3$) (Cronbach, 1951; Guttman, 1945)
   - $\omega_h$ (McDonald, 1999; Revelle & Zinbarg, 2009; Zinbarg, Revelle, Yovel & Li, 2005)
   - Worst Split Half ( $\beta$ ) (Revelle, 1979)

# How to find $\alpha$ or KR20: Use your Frieden calculator

$\alpha$ is a function of Total Test Variance ($V_X$), sum of item variances ($\Sigma v_i$ ) and number of items (n):

So, if you know how to add and subtract: $\alpha = \frac{V_X - \Sigma v_i}{V_x} \frac{n}{n-1}$

## How to do modern statistics: Use R

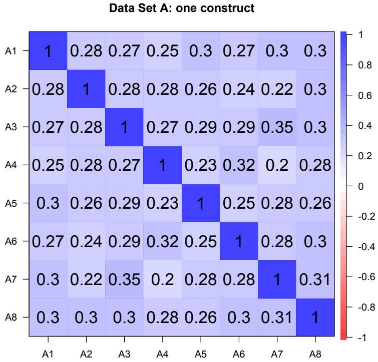
But we know more than addition and subtraction. We can do modern statistics and take advantage of computers.

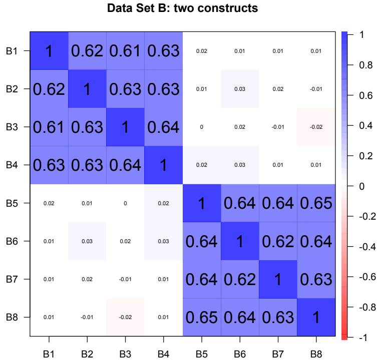
# Consider two data sets, A and B. They look similar.

```
headTail(A)
      A1  A2  A3  A4  A5  A6  A7  A8
1      2   2   3   2   1   2   2   2
2      3   3   4   3   4   4   3   2
3      2   4   3   2   3   2   2   1
4      4   2   2   3   3   2   2   1
...  ... ... ... ... ... ... ... ...
997    4   4   4   3   2   4   1   4
998    2   2   2   2   2   1   3   2
999    4   4   4   3   4   3   4   5
1000   5   3   2   4   3   4   3   3

describe(A,skew=FALSE)
   vars    n mean   sd min max range   se
A1    1 1000 3.03 1.03   1   6     5 0.03
A2    2 1000 2.96 1.09   1   6     5 0.03
A3    3 1000 3.01 1.06   1   6     5 0.03
A4    4 1000 3.03 1.05   1   6     5 0.03
A5    5 1000 3.03 1.00   1   6     5 0.03
A6    6 1000 3.05 1.03   1   6     5 0.03
A7    7 1000 3.02 1.02   1   6     5 0.03
A8    8 1000 3.00 1.05   1   6     5 0.03

 alpha(A)

Call: alpha(x = A)

  raw_alpha std.alpha G6(smc) average_r med_r
      0.75      0.75    0.73      0.28   .28
  95% confidence boundaries
lower alpha upper
0.73 0.75 0.78
```

```
 headTail(B)
      B1  B2  B3  B4  B5  B6  B7  B8
1      2   3   1   2   1   2   1   1
2      3   4   3   4   3   2   3   3
3      3   4   4   4   1   3   2   3
4      2   3   3   2   3   2   3   3
...  ... ... ... ... ... ... ... ...
997    5   5   4   3   2   3   3   3
998    1   1   2   2   4   4   4   5
999    4   3   5   4   3   2   3   3
1000   3   3   3   3   3   3   3   3

describe(B,skew=FALSE)
   vars    n mean   sd min max range   se
B1    1 1000 3.07 1.04   1   6     5 0.03
B2    2 1000 3.02 1.00   1   6     5 0.03
B3    3 1000 3.02 1.02   1   6     5 0.03
B4    4 1000 3.04 1.01   1   6     5 0.03
B5    5 1000 3.00 1.03   1   6     5 0.03
B6    6 1000 3.02 0.99   1   6     5 0.03
B7    7 1000 3.02 1.02   1   6     5 0.03
B8    8 1000 3.01 0.99   1   6     5 0.03

alpha(B)

Call: alpha(x = B)

  raw_alpha std.alpha G6(smc) average_r  med_r
      0.75      0.75    0.84       0.28    .03
 95% confidence boundaries
 lower alpha upper
0.73 0.75 0.78
```

# But they are actually quite different in their internal structure.



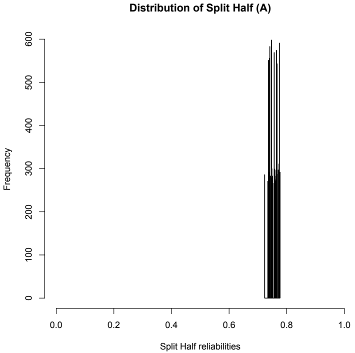



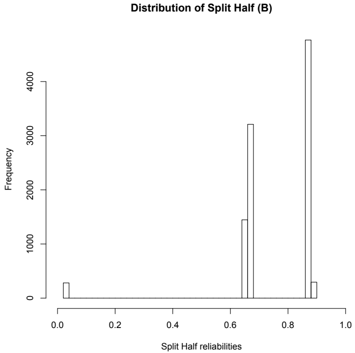

$\alpha = .75$
$\omega_h = .70$

$\alpha = .75$
$\omega_h = .03$

What is this thing called $\omega_h$?

# Distribution of all Split Half reliabilities differ




```
Split half reliabilities
Call: splitHalf(r = A, raw = TRUE)

Maximum split half reliability  =  0.78
Guttman lambda 6                =  0.73
Average split half reliability  =  0.75
Guttman lambda 3 (alpha)        =  0.75
Minimum split half reliability  =  0.72

 Quantiles of split half reliability
    2.5% 50% 97.5%
   0.72 0.76 0.78
```

```
Split half reliabilities
Call: splitHalf(r = B, raw = TRUE)

Maximum split half reliability =  0.88
Guttman lambda 6               =  0.84
Average split half reliability =  0.75
Guttman lambda 3 (alpha)       =  0.75
Minimum split half reliability =  0.03

 Quantiles of split half reliability
    2.5% 50% 97.5%
  0.03 0.87 0.88
```

## Modeling the items and their correlations

1. An item represents at least four sources of variance
   - General factor variance ($\boldsymbol{g}$) which is common to all items on the test
   - Group factor variance ($\boldsymbol{f}$) which is common to some of the items on the test
   - Specific item variance ($\boldsymbol{s}$) reliable variance specific to an item
   - Residual (error) variance ($\boldsymbol{e}$), shared with nothing
2. Formally (in matrix notation) an item can be thought of as

$$\boldsymbol{x} = \boldsymbol{cg} + \boldsymbol{Af} + \boldsymbol{Ds} + \boldsymbol{e}.$$

   And a test ($\boldsymbol{X}$) made up of these items has a variance-covariance matrix ($\boldsymbol{C} = \boldsymbol{X'X}/N$) . The total test variance is just the sum of the individual elements of $\boldsymbol{C}$ and is

$$V_X = \mathbf{1C1'} = \mathbf{1'cc'1} + \mathbf{1'AA'1} + \mathbf{1'DD'1} + \mathbf{1'ee'1}$$

3. Our challenge: Estimate the variance associated with each of these sources.

## This leads to several different estimates of reliability

1. Greatest Lowest Bound is total reliable variance (if we can estimate the specific somehow ( ala Bentler (2017); Wood, Harms, Lowman & DeSimone (2017)):

$$glb = \frac{\mathbf{1'cc1 + 1'AA'1 + 1'DD'1}}{V_X}. \tag{1}$$

2. Total common variance for each variable is general + group = communality

$$h_i^2 = c_i^2 + \Sigma A_{ij}^2.$$

and therefore

$$\omega_t = \frac{\mathbf{1'cc'1 + 1'AA'1}}{V_x} = 1 - \frac{\Sigma(1 - h_i^2)}{V_x} = 1 - \frac{\Sigma u_i^2}{V_x}. \tag{2}$$

3. Total due to just the general ($\omega_h$) ($\omega$ hierarchical to reflect the way we estimate it)

$$\omega_h = \frac{\mathbf{1'cc'1}}{V_X}. \tag{3}$$

## Use R for statistics

1. R is the open source collaborative enterprise of 20-30 statisticians that allows thousands of developers to contribute modern statistics.
2. Download R from the Comprehensive R Archive Network https://cran.r-project.org
3. Install it (current version is 3.4.4) and start it.
4. Install useful packages (choose from 12,484) e.g., install the package that downloads "task views".

R code
```
install.packages("ctv")
library("ctv")
install.views("Psychometrics")
```

5. Or just install psych

R code
```
install.packages("psych", dependencies=TRUE)
```

6. Use them and cite them (R Core Team, 2018), (Revelle, 2018)

## Using the *psych* package in R

R code

```
library(psych)

mydata <-  read.file()
#or

mydata <-
     read.clipboard.tab()

describe(mydata)

omega(mydata)

alpha(mydata)

splitHalf(mydata)
```

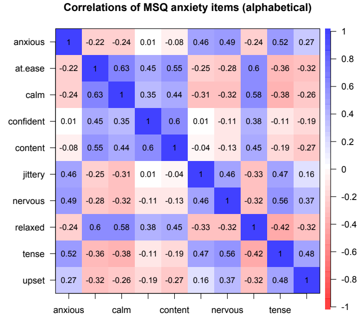
1. Make the psych package active (just need to do this once per session)
2. Read your data from your disk (or a remote server). Will read .sav, .dat, .csv, .txt files.
3. Or, read an Excel file from the clipboard
4. Find the basic descriptive statistics for your data
5. Find and display the omega statistics
6. Or just find alpha (Do you really want to do that?)
7. Show the split half information

## An example of 10 anxiety items

1. Items taken from the Motivational State Questionnaire (Revelle & Anderson, 1997) given over many years at the PMC lab (N=3032).
2. These items were chosen because they were duplicated in another test at the same time (ala Wood et al., 2017) and are useful examples of various reliability coefficients.
3. We first show the correlation plot (`corPlot`), and then reorganize it for a more clear structure (`fa` and `mat.sort` )
4. Then find $\omega_t$, $\alpha$, and $\omega_h$ using `omega`.
5. We show the commands because each step is just one line of code.

# The structure of 10 anxiety items from the MSQ

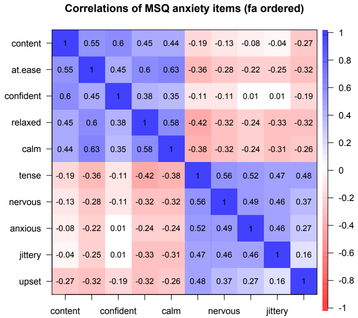
Alphabetical ordering of 10 MSQ anxiety items. Hard to see structure.




R code

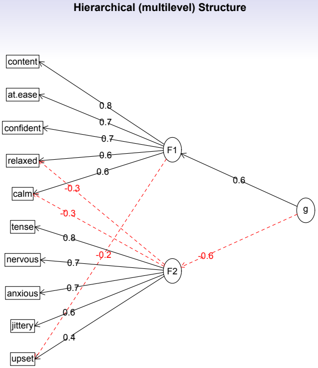
```
R <- lowerCor(msq1[msqitems])
corPlot(R,numbers = TRUE)
```

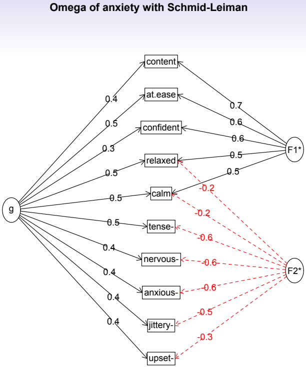
“Alabama need not come first”. Sort by something meaningful.




R code

```
f2 <- fa(R,2)  #extract 2 factors
sorted <- mat.sort(R,f2)
corPlot(sorted,numbers=TRUE)
```




1. Extract 2 factors (default is 3).
2. Correlate them.
3. Factor the correlations.
4. Show the diagram.

1. Apply the Schmid & Leiman (1957) transformation.
2. Find the g loadings and the group factor loadings.
3. Show the diagram.

R code

```
om <- omega(R,2)
omega.diagram(om,sl=FALSE)
```

# Calculating multiple measures of internal consistency reliability.

| | Formula | Calculation | Reliability measure |
|---|---|---|---|
| Total variance = $V_X = \Sigma(R_{ij}) = 39.80$ | | | |
| Total reliable item variance = $\Sigma r_{ii} = 6.71$ | $glb = \frac{V_X - tr(R) + \Sigma(r_{ii})}{V_X}$ | $\frac{39.80-10+6.71}{38.90}$ | = .917 |
| r best split (A= 1, 4, 6, 7, 10 vs B = 2,3,5, 8, 9) = .834 | $\lambda_4$ = best split half = $\frac{2r_{ab}}{1+r_{ab}}$ | $\frac{2*834}{1+.834}$ | = .909 |
| Total common variance = $\Sigma h_i^2 = 5.27$ | $\omega_t = \frac{V_X - tr(R) + \Sigma h_i^2}{V_X}$ | $\frac{39.80-10+5.27}{39.80}$ | = .881 |
| Total squared multiple correlations $\Sigma(SMC) = 4.36$ | $\lambda_6 = \frac{V_X - tr(R) + \Sigma(SMC)}{V_X}$ | $\frac{39.80-10+4.36}{39.80}$ | = .858 |
| | $\alpha = \frac{n}{n-1}\frac{V_X - tr(R)}{V_X}$ | $\frac{10}{9}\frac{39.80-10}{39.80}$ | = .832 |
| Average correlation = $\frac{V_X - tr(V_X)}{n*(n-1)} = 0.331$ | $\alpha = \frac{n\bar{r}}{1+(n-1)\bar{r}}$ | $\frac{10*.331}{1+9*.331}$ | = .832 |
| r worst split (A = 1-5 vs. B= 6-10) = .397 | $\beta$ = worst split half = $\frac{2r_{ab}}{1+r_{ab}}$ | $\frac{2*.397}{1+.397}$ | = .569 |
| Sum of g loadings = 4.84 (bi-factor) | $\omega_g = \frac{(\Sigma g_i)^2}{V_X}$ | $\frac{4.84^2}{39.80}$ | = .589 |
| Sum of g loadings = 4.21 (Schmid-Leiman) | $\omega_h = \frac{(\Sigma g_i)^2}{V_X}$ | $\frac{4.21^2}{39.80}$ | = .446 |

## Reliability needs to step into the modern era

Reliability is too important to leave to the calculator era. We need to do modern statistics. This is not hard. Use R.

| Types of Reliability | | Estimation functions |
|---|---|---|
| Test-retest | | `testRetest` |
| | Many occasions | `multilevel.reliability` |
| Parallel form approach | | |
| | Parallel tests | `scoreItems` |
| | Duplicated tests | `testRetest` |
| Internal consistency | | |
| | greatest split half ($\lambda_4$) | `splitHalf` |
| | $\omega_t$ | `omega` |
| | SMC adjusted ($\lambda_6$) | `splitHalf` |
| | $\alpha$ ($\lambda_3$) | `alpha` |
| | average split half | `splitHalf` |
| | $\omega_g$ | `omega` |
| | smallest split half | `splitHalf` |
| | worst split half ($\beta$) | `iclust` |
| Other forms of reliability | | `ICC`, `cohen.kappa` |

1. Use R
2. Use the *psych* package
3. Report 21st century statistics
4. Slides at `http://personality-project.org/sapa.html`
5. Questions?

Bentler, P. M. (2017). Specificity-enhanced reliability coefficients. *Psychological Methods*, *22*(3), 527 – 540.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland*, *15*, 246–263.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255–282.

Kuder, G. & Richardson, M. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151–160.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, N.J.: L. Erlbaum Associates.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Revelle, W. (1979). Hierarchical cluster-analysis and the internal structure of tests. *Multivariate Behavioral Research*, *14*(1), 57–74.

Revelle, W. (2018). *psych: Procedures for Personality and Psychological Research*. https://cran.r-project.org/web/packages=psych: Northwestern University, Evanston. R package version 1.8.2.

Revelle, W. & Anderson, K. J. (1997). Personality, motivation and cognitive performance. Final report to the Army Research Institute on contract MDA 903-93-K-0008, Northwestern University:.

Revelle, W. & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega and the glb: comments on Sijtsma. *Psychometrika*, *74*(1), 145–154.

Schmid, J. J. & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*(1), 83–90.

Shrout, P. E. & Lane, S. P. (2012). Psychometrics. In *Handbook of research methods for studying daily life*. Guilford Press.

Spearman, C. (1904). "General Intelligence," objectively determined and measured. *American Journal of Psychology*, *15*(2), 201–292.

Wood, D., Harms, P. D., Lowman, G. H., & DeSimone, J. A. (2017). Response speed and response consistency as mutually validating indicators of data quality in online samples. *Social Psychological and Personality Science*, *8*(4), 454–464.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's $\alpha$, Revelle's $\beta$, and McDonald's $\omega_H$: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*(1), 123–133.