Assessing Personality with Massively Missing Completely at Random data: An information Theoretic Approach

Part of an Experts Meeting on Personality Measurement: Edinburgh, Scotland, September 6-8

William Revelle[a] & David M. Condon[b] and Sonja Heintz[c]

[a]Department of Psychology, Northwestern University, Evanston, Illinois
[b]Department of Medical Social Sciences, Northwestern University Chicago Illinois
[c]Department of Psychology, University of Zurich

Slides at `http://personality-project.org/sapa`

Outline

Overview

1. Using big data techniques has changed the way we study personality. The increase in power due to very large samples allows the detection of small but meaningful effects – structural measures can have finer resolution than previously available and cross-validated predictive accuracy can be substantially enhanced.
2. But these techniques are not limited to big data. We discuss one such approach that can be used for both large ($N > 10^5$), medium ($N \approx 10^3 - 10^4$) as well as smallish (N=100-400) size samples.
3. We show how SAPA procedures improve scale reliability and validity over using short scales when giving random subsets of items selected from larger scales.
4. We encourage the use of item sampling (SAPA) techniques in many typical research applications, including ESM.

SAPA overview

1. At the sapa-project.org we use Synthetic Aperture Personality Assessment (SAPA) methods to assess 20K participants per month. This is just a technique of Massively Missing Completely at Random (MMCAR) data presentation. Each participant is given a random subset of items chosen from an item pool of more than 6600 items. These items, extended from the International Personality Item Pool (Goldberg, 1999) and the International Cognitive Ability Resource, assess temperament, cognitive ability, interests and attitudes as well as self reported behaviors.
2. Conventional psychometric techniques are used to identify homogeneous scales; empirical item selection procedures are use to develop optimal item composites to predict a wide range of criteria. Data analysis code is done using the *psych* package (Revelle, 2018) in R (R Core Team, 2018).

The basic problem: Fidelity versus bandwidth

1. Many personality traits, interests and cognitive abilities are multidimensional and have complex structure.
   - To measure these, we need to have the precision that comes with many participants.
   - But we also need the bandwidth that comes with many items.
   - But participants are reluctant to answer very many items.
2. This has led to the quandary of should you give many people a few items or a few people, many items?
3. Our answer is to do both, but with a *Massively Missing Completely At Random* (MMCAR) data structure.
4. We refer to this technique as *Synthetic Aperture Personality Assessment* (SAPA) to recognize the analogy to synthetic aperture radio astronomy (Revelle, Wilt & Rosenthal, 2010; Revelle, Condon, Wilt, French, Brown & Elleman, 2016)
5. This is functionally what Frederic Lord (1955, 1977) suggested 63 years ago. It is time to take him seriously.

Lord (1955) and matrix sampling

1. Given an N (subjects) by n (item) matrix, we can sample:
2. Type 1: Subjects – basic statistical theory
    - $\bar{x}$ and its standard error $\sqrt{\frac{\sigma^2}{N-1}}$
    - r_{xy} and its standard error $\sqrt{\frac{1-r^2}{N-2}}$
3. Type 2: Items – this is the basis of classical reliability theory especially domain sampling (Tryon, 1957, 1959)
    - $KR_{20} = \alpha = \lambda_3$ represent the correlation of a test with a test just like it sampled from a larger population of items
    - ω_h and ω_t similarly are estimates of what the general factor, ω_h, or total, ω_t, correlation would be with another representation in the domain
4. Type 12: Matrix sampling of subjects and items
    - Special case is balanced incomplete blocks (BIB).
    - General case is Missing Completely at Random (MCAR).

3 Methods of collecting 256 subject * items data

```
1) 8 x 32 complete                2) 32 x 8 complete        12) 32 x 32 MCAR p=.25
46213634521143453443645331212414  46323114                  ..3..2..6.....4.55.......44.....
21243623166421516154432261516513  25443314                  ...........4..6..45..3.4..6....1
51661351155165463622224435623344  43315423                  6..3.......6.1.....6.2.......5.6
11141343362332215612152135614522  26314145                  ....3522......5.3...3......5....
25353121264561433433232246526411  41435614                  ....3.2.2.......3..2......65..5.
61335154566424114612641225353516  42236153                  .....51....324.........23......5
24634342151536242425413513435116  62421344                  ....552............25...54.5....
11554654453123111162423325516334  35234443                  ...44.4.5....3..6...6........3..
                                  34514166                  ....61.523.2....2...........3...
Type 1 = sample subjects          63415154                  5.............42.4..6.5......61.
                                  44441342                  ....3....3.6..1.4...1..5......5.
                                  13514321                  1....54..........2.4.33..6......
                                  66365663                  4.....52..6.....44.3...........2
Type 2 = sample items             12264546                  ..44...1........1..42....5..1...
                                  31466135                  ..1..3.......2..3.521.......6...
                                  32645514                  .......3.142.........22......12.
                                  66151251                  .4...2..........3..162...4.....4
                                  14411441                  ..4..6..3.4...1....5.33.........
Type 12 sample items and subjects 62443636                  5..........243..5....41......1..
                                  33316236                  ..5..3..4...4.4..5..1.........4.
                                  63325425                  .....4.......3..5.2.....64.4..4.
                                  11531126                  ...1.1.2...6....4......55....2..
                                  61155546                  .....3..2..53.....2..2.3.3......
                                  33245361                  ......1...2..43...3.13........5.
                                  52241654                  ...2.........4..54...2.3..62....
                                  63212356                  22.......332..1.....5......6....
                                  24414663                  ...5..3.4.....3....5.241........
                                  63661414                  ......63.1.......6..5...4..2...5
                                  45555223                  ..2.4..5..........52.4.....44...
                                  14364433                  2.55.....2.....6.....6.....55...
                                  21461416                  ..5..........4....6341.4..2.....
                                  33232365                  ....55......5.......45....3..32.
```

3 Methods of collecting 256 subject * items data

1) complete (Ideal) 2) Sample people 3) Items

```
22552141414336514122645166143244    22552141414336514122645166143244 22552141
32144265454235634562343524256611    ................................ 32144265
43553143152141541641526114551151    ................................ 43553143
52654223445614444431162645313124    ................................ 52654223
62222255242315442652355414213325    ................................ 62222255
22125412454242154221456444214564    ................................ 22125412
65113311244511226522615346451412    ................................ 65113311
54436452425245244554632246526466    ................................ 54436452
55223643555215245514633426121226    55223643555215245514633426121226 55223643
35522554332664265346655451531612    ................................ 35522554
63261241341466311243222233323541    63261241341466311243222233323541 63261241
32224431433144451645255464435552    ................................ 32224431
11564655513111334341463561655541    11564655513111334341463561655541 11564655
24532624664444656366642463322555    ................................ 24532624
25516362264523255665245644125611    ................................ 25516362
32255635422342631523143414221354    ................................ 32255635
23244456631411361161615126144214    ................................ 23244456
34526633236542563633625123624421    ................................ 34526633
13451522616451531355135621451536    13451522616451531355135621451536 13451522
31625444241623135123121345134162    ................................ 31625444
44252526365556663522524162313453    ................................ 44252526
54361436651313615433261662235132    ................................ 54361436
46635454552135645224352362433436    46635454552135645224352362433436 46635454
26511624245416441145655363265265    ................................ 26511624
63512331235542645524352562623235    ................................ 63512331
11523665433656446452523322216333    11523665433656446452523322216333 11523665
56436532623253433145633663651242    ................................ 56436532
15136366233651513351113353151452    ................................ 15136366
46321152211446344326554442255226    ................................ 46321152
62156523111352364233551656146433    ................................ 62156523
65342552265235623363226156136333    6534255226523562336322615613633  65342552
55325212341345661654143661563533    ................................ 55325212
```

12 (Matrix) Sampling Methods of collecting 256 subject * items data

a) 32 x 16 balanced incomplete

```
................4122645166143244
................4562343524256611
................1641526114551151
................4431162645313124
................2652355414213325
........45424215........44214564
........24451122........46451412
........42524524........46526466
........55521524........26121226
........33266426........51531612
........3414663112432222........
........4331444516452554........
........5131113343414635........
........6644446563666424........
........2645232556652456........
32255635................14221354
23244456................26144214
34526633................23624421
13451522................21451536
31625444................45134162
44252526........35225241........
54361436........54332616........
46635454........52243523........
26511624........11456553........
63512331........55243525........
1152366543365644................
5643653262325343................
1513636623365151................
4632115221144634................
6215652311135236................
................................
................................
```

b) 32 x 8 SAPA p =.25

```
..55.....1......4...6..16.....4.
.2..4...45.....3........2.2....1
..5.....1......4...1..6...551...
..6..2.........4.......64...3.24
..22..5.......4.2..2...4..2.....
.2.2..1.4.....1..2......4.....6.
......1124......65...1...6......
.44..4.2......2.......2...52....
.5..3...5......4..1.6.......1..6
....25....2....6....65.......61.
....1...3.....311.4..22.........
...244......44....45...........2
.1.......1.11...4.4....5......4.
.....6.4.........3.66.2....2..5.
25...3.2.........6......4.12....
......3..22....3....1....42..3..
..2...5..31..1..........2....21.
...2.6...3...2.6.......12..2....
.....5...1...1.3135.1...........
...2..4........35.........13.16.
..2....6..5.......2....1....3.53
543.............5...2.....2.51..
4...54....2...6............3..36
....1..424....4.....6.......5.6.
........235...6..........262...5
.15....5..3....4.4........2..3..
........6.3......1.5.3..63...2..
..1...66.........35..1.35.......
......5221......4.......42..5...
.21........3..........1.5.1.6..3
...4....2.523...3...2........3..
55.2.....4.3.............1.6.5..
```

Type 12 sampling (matrix sampling)

1. Balanced incomplete blocks works but is hard if giving less than 50% coverage
    - 50% requires 6 blocks to be fully balanced (divide into 4ths and then present all pairs of the fourths)
    - AB, AC, AD, BC, BD, CD where A, B, C, and D are 1/4 of the total
    - Even then, items within blocks co-occur more than items beween blocks
    - 33% samples require 15 blocks, 25% 28 blocks
2. SAPA sampling (Massively Missing Completely at Random) allows any sampling rate.
3. BIB can be done with printed forms, MMCAR requires computer administration.
4. Possible to do FIML with BIB design, need to do pairwise complete for SAPA.

Why we care: Breadth vs. depth of measurement

1. Factor structure of domains needs multiple constructs to define structure.
2. Each construct needs multiple items to be measured reliably.
3. This leads to an explosion of potential items.
4. But, people are willing to answer only a limited number of items.
5. This leads to the use of short and shorter forms (the NEO-PI-R (Costa & McCrae, 1992) with 300, the IPIP (Goldberg, 1999) Big 5 with 100, the BFI (John, Donahue & Kentle, 1991) with 44 items, the BFI2 (Soto & John, 2017) with 60, the 30 item 'Short Five' (Konstabel, Lönnqvist, Leikas, Velàzquez, H, Verkasalo, & et al., 2017), the TIPI (Gosling, Rentfrow & Swann, 2003) with 10 and the 10 item BFI (Rammstedt & John, 2007)) to include as part of other surveys.
6. Unfortunately, with this reduction of items, breadth of substantive content is lost. We offer an alternative procedure.

Example studies with subject/item tradeoffs

1. The Potter-Gosling internet project (outofservice.com) has given over 10,000,000 tests since 1997. Originally the 44 items of the Big Five Inventory (BFI) (John et al., 1991) although they are now giving the BFI2 (Soto & John, 2017).
2. The Stillwell-Kosinski (mypersonality.org) Facebook application (no longer in service) gave 7,765 people the IPIP version of the NEO-PI-R with facets (300 items), 1,108,472 the IPIP NEO-PI R domains (100 items), and 3,646,237 brief (20 item) surveys. Cross linked to likes and Facebook pages (Kosinski, Matz, Gosling, Popov & Stillwell, 2015; Youyou, Kosinski & Stillwell, 2015).
3. Johnson reports two data sets: 300 IPIP-NEO items for 145,388 participants and 120 IPIP-NEO items for 410,376 participants (Johnson, 2014).
4. Smaller scale studies include the BBC data set of the 44 items of the BFI on 386,375 and the initial report on the BFI-2 (Soto & John, 2017) with several thousand subjects with 60 items.

Exceptions to the shorter and shorter inventory trend

1. Lew Goldberg and his colleagues at the University of Oregon developed the Eugene-Springfield sample (Goldberg & Saucier, 2016) which has given several thousand items to $\approx 1,000$ predominantly white middle class participants over 10 years. This sample has been the basis of the development and validation of the International Personality Item Pool (see ipip.ori.org).
2. In fact, many of the subsequent attempts at personality scale development have used the Eugene-Springfield sample, e.g., the BFI (John et al., 1991), and the Big Five Aspect Scales (BFAS) of DeYoung, Quilty & Peterson (2007).

The Eugene Springfield sample and the International Personality Item Pool

1. Unfortunately, many of the items that have come out of the E-S sample were prematurely selected to represent the Big 5. That is, even though meant to capture the many dimensions of the lexicon, the adjectival descriptors used had been trimmed to those matching the 5 factors that have been known since the 1950's (Kelly & Fiske, 1950, 1951; Tupes & Christal, 1961; Norman, 1963).
2. Because of the ease of use and the openness of the IPIP, most of the short forms followed the Big Five structure that came out of the E-S sample.

SAPA techniques can work for you

1. At the Personality Project (Revelle et al., 2010, 2016) (now at sapa-project.org) we have taken the opposite direction and have given more and more items including measures of temperament, ability, and interests and we are now developing item statistics on more than 6,600 items (Condon & Revelle, 2017) for almost 500,000 participants (but using SAPA procedures).
2. We have reported computer simulations of our procedures but now we want to demonstrate with real data the amazing power of massively missing data.
3. In particular, we want to show that the techniques can work on relatively small samples (as small as $\approx 100 - 400$ as well as the the larger ones we have been working with.
4. We reported some of this before at ECP-19 (Revelle & Condon, 2018) and are now pushing even further,

Trading items for people: Studies, Items, People, Items x People

Table: Data sets vary in their sampling strategy and the Potter-Gosling and Stillwell/Kosinski data sets seem to have more data than the others

Study	N	Items (n)	Items/ Person (k)	Items* People
Potter-Gosling	10^7	44	44	$4.4 * 10^8$
Stillwell-Kosinski	$4.5 * 10^6$	20-300	20-300	$1.7 * 10^8$
Johnson	$4.1 * 10^5$	120	120	$4.9 * 10^7$
Johnson	$1.4 * 10^5$	300	300	$4.3 * 10^7$
SAPA* (2010-2017)	$2.5 * 10^5$	2000	100-150	$2.5 * 10^7$
SAPA+ (2017-2018)	$2 * 10^5$	6,600	100-150	$2 * 10^7$
SAPA- (pre 2010)	$6.3 * 10^4$	500	100-150	$9.4 * 10^6$
Eugene-Springfield	10^3	3,000	3,000	$3 * 10^6$

But given basic statistical theory, is it worth while to increase the sample size so much? What is the effect of giving more items at the cost of reducing the sample size?

Consider the amount of *information* which varies by number of correlations $\frac{n*(n-1)}{2}$ and 1/(standard error of the r) $\approx \sqrt{N}$.

Trading items for people: Studies: Items, People, Items x People and Information

Information varies by the number of correlations $(n * (n-1)/2)$ weighted by their standard errors which vary by $\sqrt{N}$

Table: Data sets vary in their sampling strategy and the seemingly smaller sets, by giving many more items actually have more total information

Study	N	Items (n)	Items/ Person (k)	Items* People	Information
SAPA+	$2 * 10^5$	6,600	100-150	$2 * 10^7$	$2.2 * 10^8$
E-S	10^3	3,000?	3,000	$3 * 10^6$	$1.4 * 10^8$
S-Ki	$4.5 * 10^6$	20-300	20-300	$1.7 * 10^8$	$9.5 * 10^7$
SAPA*	$2.5 * 10^5$	1-2,000	100-150	$2.5 * 10^7$	$7.5 * 10^7$
Johnson	$1.4 * 10^5$	300	300	$4.3 * 10^7$	$1.7 * 10^7$
SAPA (pre 2010)	$4.3 * 10^3$	500	100-150	$2.5 * 10^7$	$9.3 * 10^6$
Johnson	$4.1 * 10^5$	120	120	$4.9 * 10^7$	$4.6 * 10^6$
P-G	10^7	44	44	$4.4 * 10^8$	$3.0 * 10^6$

SAPA pairwise:

+SAPA 2017-2018 (100) * SAPA 2013-2017 (1,400) -SAPA 2010-2013 (1100)

Many items versus many people

1. Not only do want many people, we also want many items.
2. Resolution (fidelity) goes up with sample size, N, (standard errors are a function of $\sqrt{N}$)

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N-1}} \qquad \sigma_r = \sqrt{\frac{1-r^2}{N-2}}$$

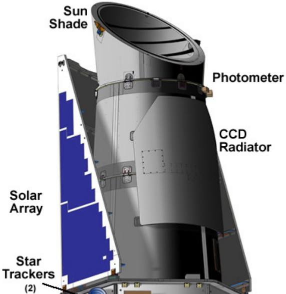
3. Also increases as number of items, n, measuring each construct (reliability as well as signal/noise ratio varies as number of items and average correlation of the items)

$$\lambda_3 = \alpha = \frac{n\bar{r}}{1+(n-1)\bar{r}} \qquad s/n = \frac{n\bar{r}}{(1-n\bar{r})}$$

4. Breadth of constructs (band width) measured goes up by number of items (n).
5. Thus, we need to increase N as well as n. But how?

A short diversion: the history of optical telescopes

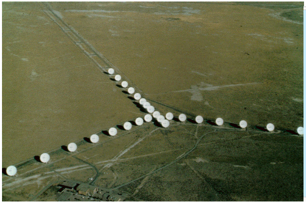
Resolution varies by aperture diameter (bigger is better)

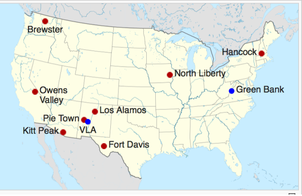


A short diversion: history of radio telescopes

Resolution varies by aperture diameter (bigger is still better)

Aperture can be synthetically increased across multiple telescopes or even multiple observatories





Can we increase N (subjects) and n (items) at the same time?

1. Frederic Lord (1955) introduced the concept of sampling people as well as items.
2. Apply basic sampling theory to include not just people (well known) but also to sample items within a domain (less well known).
3. Basic principle of Item Response Theory and tailored tests.
4. Used by Educational Testing Service (ETS) to pilot items.
5. Used by Programme for International Student Assessment (PISA) in incomplete block design (Anderson, Lin, Treagust, Ross & Yore, 2007).
6. Discussed at IMPS (2017) meeting by Rutkowski and Matta in the missing data symposium.
7. Can we use this procedure for the study of individual differences without being a large company?
8. Yes, apply the techniques of radio astronomy to combine measures synthetically and take advantage of the web.

Subjects are expensive, so are items

1. In a survey such as Amazon's Mechanical Turk (MTURK), we would need to pay by the person and by the item.
2. Volunteer subjects are not very willing to answer many items.
3. Why give each person the same items? Sample items, as we sample people.
4. Synthetically combine data across subjects and across items. This will imply a missing data structure which is
   - Missing Completely At Random (MCAR), or even more descriptively:
   - Massively Missing Completely at Random (MMCAR) (we sometimes have 99% missing data although our median is only 93% missing!)
5. This is the essence of Synthetic Aperture Personality Assessment (SAPA) (Condon & Revelle, 2014; Condon, 2014; Revelle et al., 2016, 2010).
6. This is a much higher rate of missingness than discussed in the balanced incomplete block design of NAEPS or PISA.

3 Methods of collecting 256 subject * items data

a) 8 x 32 complete

```
46213634521143453443645331212414
21243623166421516154432261516513
51661351155165463622224435623344
11141343362332215612152135614522
25353121264561433433232246526411
61335154566424114612641225353516
24634342151536242425413513435116
11554654453123111162423325516334
```

b) 32 x 8 complete

```
46323114
25443314
43315423
26314145
41435614
42236153
62421344
35234443
34514166
63415154
44441342
13514321
66365663
12264546
31466135
32645514
66151251
14411441
62443636
33316236
63325425
11531126
61155546
33245361
52241654
63212356
24414663
63661414
45555223
14364433
21461416
33232365
```

c) 32 x 32 MCAR p=.25

```
..3..2..6.....4.55.......44.....
...........4..6..45..3.4..6....1
6..3.......6.1.....6.2.......5.6
....3522......5.3...3......5....
....3.2.2.......3..2......65..5.
.....51....324.........23......5
....552............25...54.5....
...44.4.5....3..6...6........3..
....61.523.2....2...........3...
5.............42.4..6.5......61.
....3....3.6..1.4...1..5......5.
1....54..........2.4.33..6......
4.....52..6.....44.3...........2
..44...1........1..42....5..1...
..1..3.......2..3.521.......6...
.......3.142.........22......12.
.4...2..........3..162...4.....4
..4..6..3.4...1....5.33.........
5..........243..5....41......1..
..5..3..4...4.4..5..1.........4.
.....4.......3..5.2.....64.4..4.
...1.1.2...6....4......55....2..
.....3..2..53.....2..2.3.3......
......1...2..43...3.13........5.
...2.........4..54...2.3..62....
22.......332..1.....5......6....
...5..3.4.....3....5.241........
......63.1.......6...5..4..2...5
..2.4..5..........52.4.....44...
2.55.....2.....6.....6.....55...
..5..........4....6341.4..2.....
....55......5.......45....3..32.
```

Synthetic Aperture Personality Assessment

1. Give each participant a random sample of *pn* items taken from a larger pool of n items. p_i might be anywhere from .01 to 1.
2. Find covariances based upon "pairwise complete data". Each pair appears with probability $p_i p_j$ with a median of .01.
3. Find scales based upon basic covariance algebra.
    - Let the raw data be the matrix ${}_N\boldsymbol{X}_n$ with N observations converted to deviation scores.
    - Then the item variance covariance matrix is ${}_n\boldsymbol{C}_n = \boldsymbol{X'X}N^{-1}$
    - and scale scores, ${}_N\boldsymbol{S}_s$ are found by $\boldsymbol{S} = {}_N\boldsymbol{X}_{pp}\boldsymbol{K}_s$.
    - ${}_n\boldsymbol{K}_s$ is a keying matrix, with $k_{ij} = 1$ if $item_i$ is to be scored in the positive direction for scale j, 0 if it is not to be scored, and -1 if it is to be scored in the negative direction.
    - In this case, the covariance between scales, ${}_s\boldsymbol{C}_s = {}_s\boldsymbol{S_N}'{}_N\boldsymbol{S}_s N^{-1} =$

$$ {}_s\boldsymbol{C}_s = (\boldsymbol{XK})'(\boldsymbol{XK})N^{-1} = \boldsymbol{K'X'XK}N^{-1} = \boldsymbol{K'}{}_n\boldsymbol{C}_n\boldsymbol{K}. \quad (1) $$

4. That is, we can find the correlations/covariances between scales from the item covariances, not the raw items.

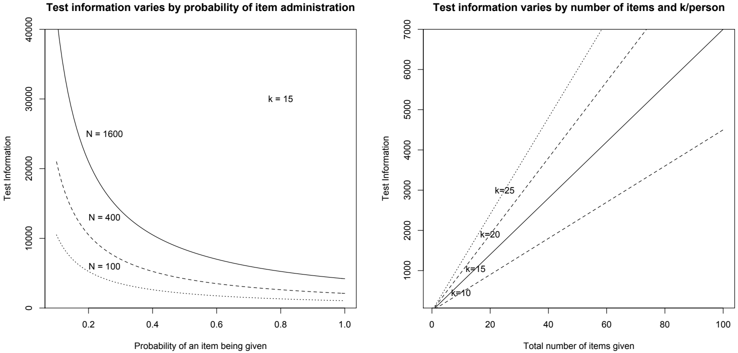
Total information

1. The information in a correlation varies by its standard error $\sigma_r = \sqrt{\frac{1-r^2}{N-2}}$
2. In SAPA, k items/person are randomly selected with probability p from a larger number, n (k = pn).
3. Thus,the number of subjects per item is pN.
4. The total number of correlations is just $\frac{n*(n-1)}{2}$ and the number of subjects per correlation is p^2N.
5. Total information is number of correlations * $\sqrt{p^2N} = \frac{n*(n-1)}{2}\sqrt{p^2N} = \frac{(k/p)((k/p)-1)}{2} * \sqrt{p^2N} = \frac{k*(k-1)\sqrt{N}}{2*p}$.
6. For the "normal case" where $p = 1$, the information is just what we expect–a quadratic function of k: $I_{kN} = \frac{k*(k-1)\sqrt{N}}{2}$.
7. But the more interesting case (the SAPA case) is for $p < 1$ the information is a hyperbolic function of p: $I_{pkN} = \frac{k*(k-1)\sqrt{N}}{2*p}$ but a linear function of the total number of items given (n= k/p) $I_{pkN} = \frac{n*(k-1)}{2} * \sqrt{N}$

Total information varies by the number of items (n) and the probability of sampling (p) and total sample size (N)

For k items/subject and N subjects, if every item is given with probability p, the information in the test is

$$I_{pkN} = \frac{k*(k-1)\sqrt{N}}{2*p} = \frac{n*(k-1)}{2} * \sqrt{N}$$



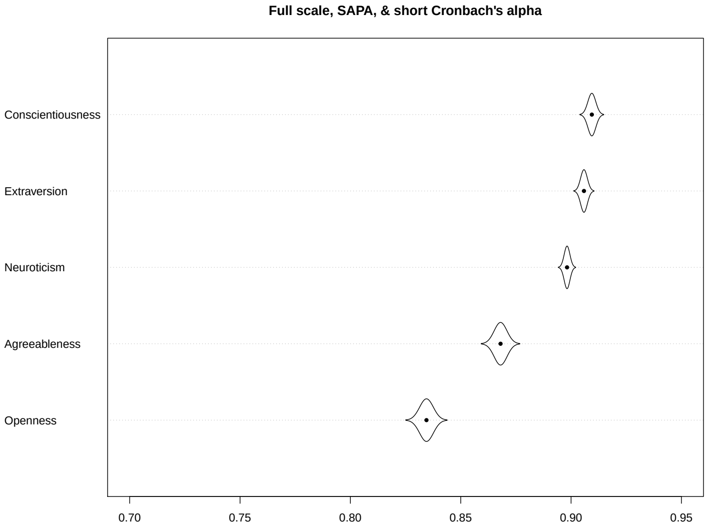
Demonstrations of SAPA with real data

1. At ECP (Revelle & Condon, 2018) we showed simulations of SAPA techniques using 70 items from the SPI-135 (Condon, 2017)
    - We demonstrated full recovery of structure for short (15 item) "sapaized" scales sampled from the SPI-70.
    - Reported success at 2,000 subjects, hinted at it working for 1,000
    - The question was: "how low can we go?"
2. Now we show simulations of SAPA technique using John Johnson's NEO-IPIP-300 data (145,388 complete participants) from Johnson (2014) from which we derived the 120-item IPIP-NEO
3. We show results for 15 item "sapaized" scales from the full 120 for 2,000 subjects and for 30 item "sapaized" scales for 400 subjects.

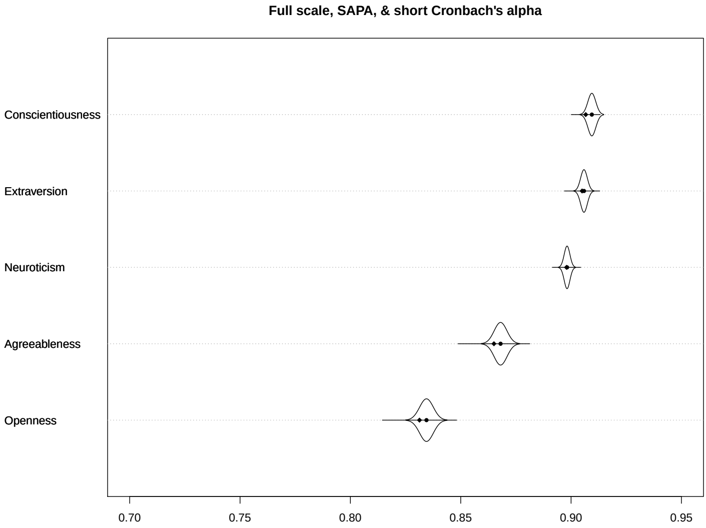
Simulation technique

1. From a complete data set do 20 non-overlapping samples ("K-fold") of size N=2000 or 400 for n=120 items
2. Compare 3 types of models
3. The structure of all items (i.e., k = 120)
4. The structure of sapaized items (k =15/30)
    - For each of N subjects, make all except k=15/30 missing completely at random
    - Thus $p = k/n = 15/120 = .125$ or $p = 30/120 = .25$
5. The structure of short scales (Big 5 from 15/30 items)
6. We now compare the full data, "sapaized" data, and the short scales.
7. For each graph we show cats eyes for the complete data, error bars for sapaized and short scales.

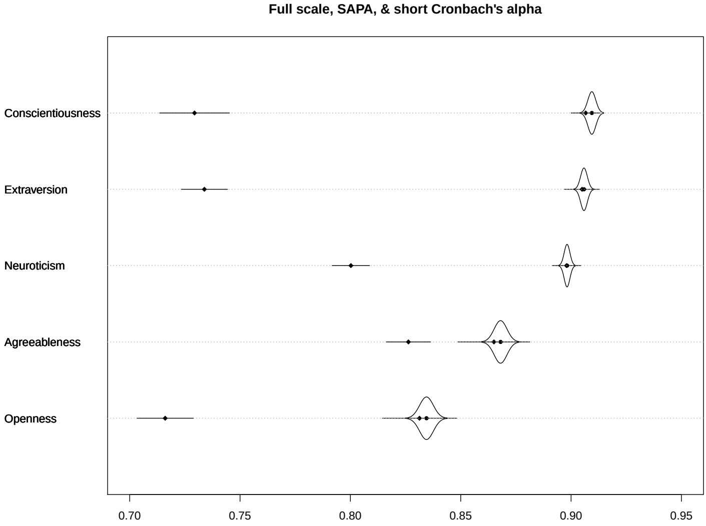
α for 2000 subjects sample from Johnson 120 item IPIP, full scales



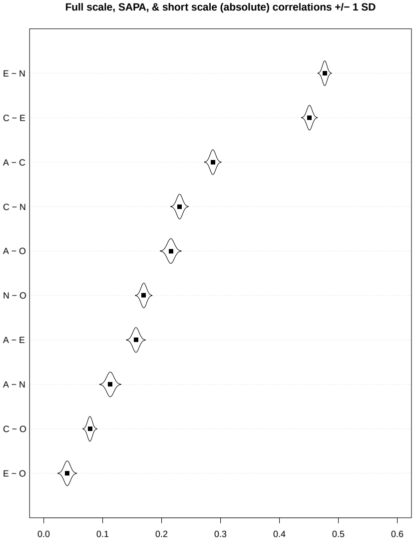
α for 2000 from Johnson 120 item IPIP, sapa sampling 15 items



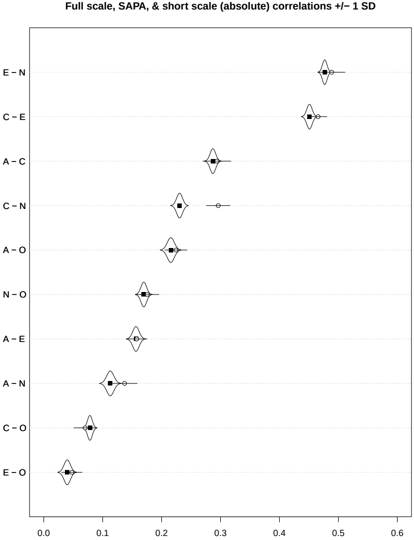
α N=2000 from Johnson 120 item IPIP, 15 item sapa + short scales



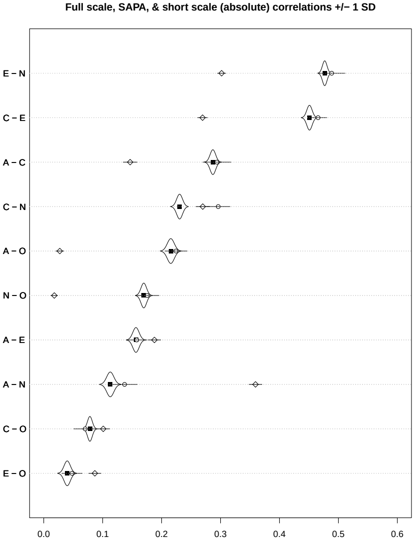
N = 2000 subjects Johnson 120 item IPIP, full scales



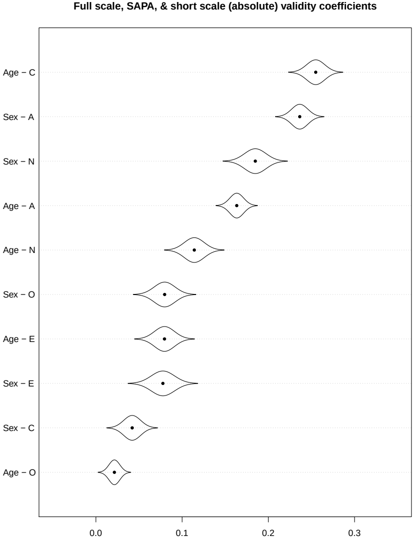
N=2000 Johnson 120 item IPIP, sapa sampling 15 items



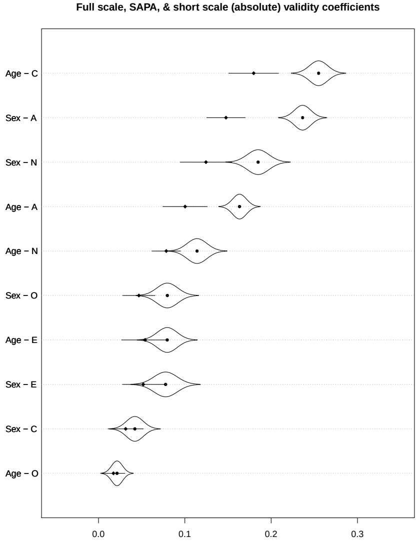
N=2000 120 item IPIP, 15 item sapa and short scales



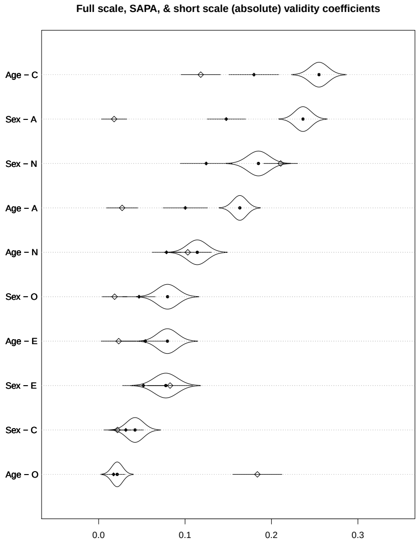
N=2000 from Johnson 120 item IPIP, full scales: Validity



N=2000 Johnson 120 item IPIP, sapa sampling 15 items: Validity



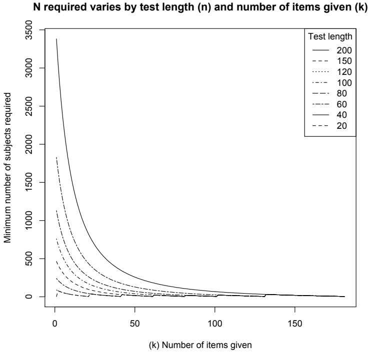
N=2000 120 item IPIP, sapa 15 items + Short scales: Validity

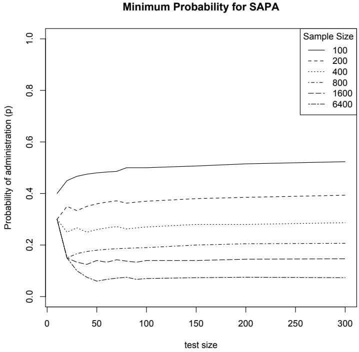


Can we apply SAPA to smaller samples: e.g. How low can we go?

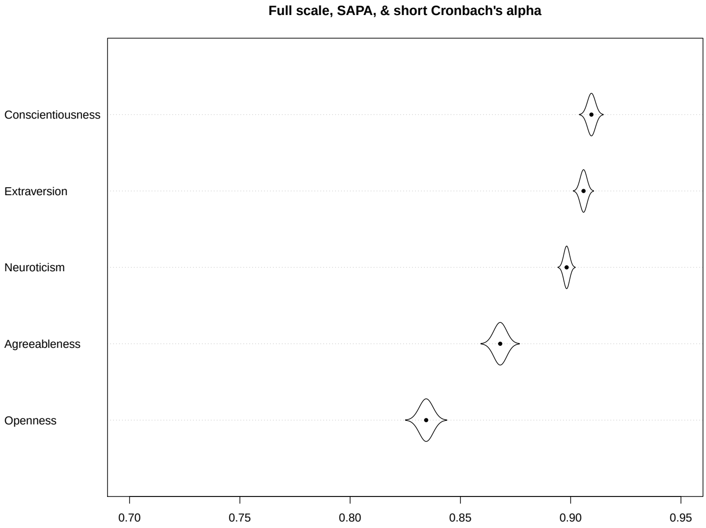
1. It seems as if the limit of the procedure is based upon the likelihood of random samples not producing empty correlation cells.
2. If we sample items with probability p, then the expected number of observations per pairwise correlation is p^2N.
3. But the standard deviation of a binomial with probability P = $\sqrt{\frac{PQ}{N}}$ where Q = 1-P.
4. We need to make sure that $p^2N - 3 * \sqrt{\frac{p^2*(1-p^2)}{N}} > 2$
5. Functionally, this means that $p^2N > 25$.
6. Thus, we can calculate the number of items (k) we can sample from the total number of items (n) for sample size (N).
7. We show this graphically

Probability of sampling is limited by sample size

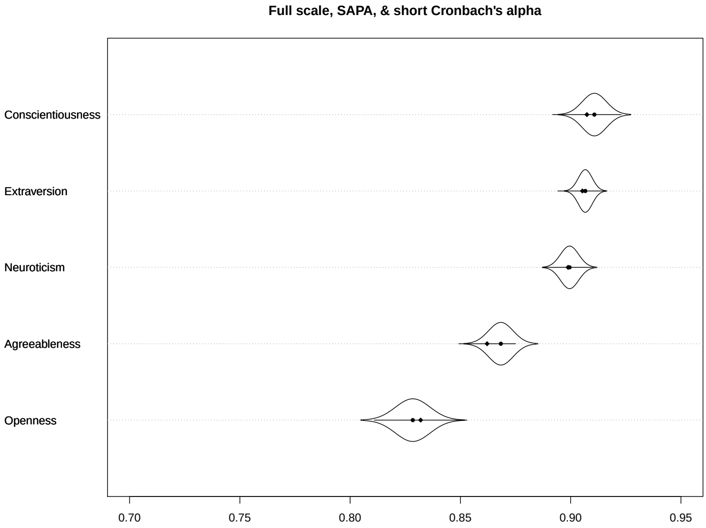

N required varies by test length (n) and number of items given (k)


Minimum Probability for SAPA

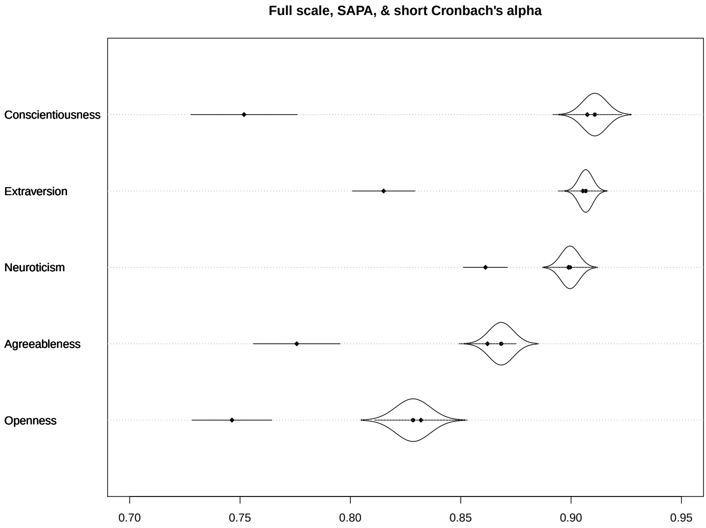
α N=400 Johnson 120 item IPIP, full scales



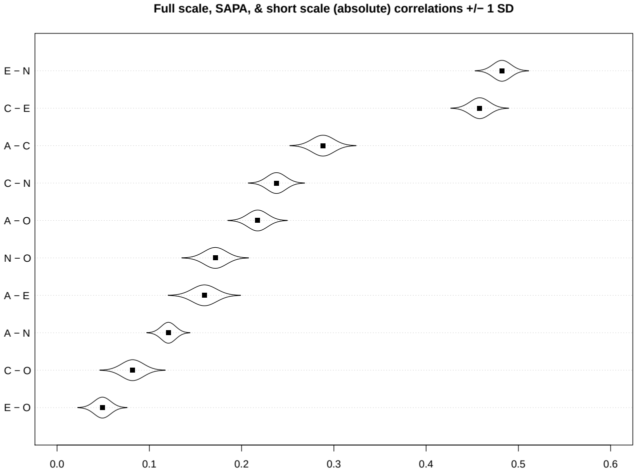
α N=400 Johnson 120 item IPIP, sapa sampling 30 items at a time



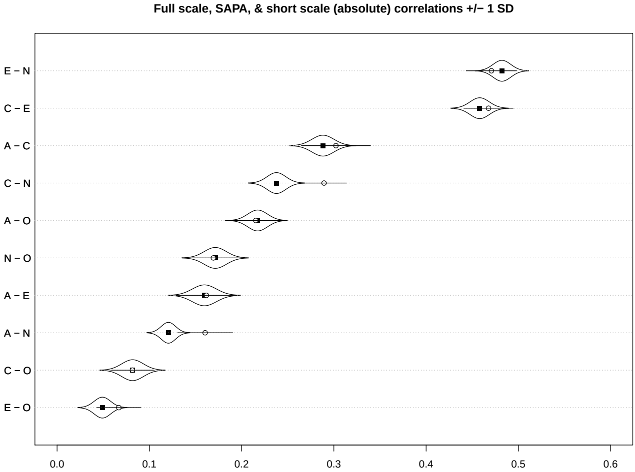
α N = 400 Johnson 120 item IPIP, 30 item sapa and short scales



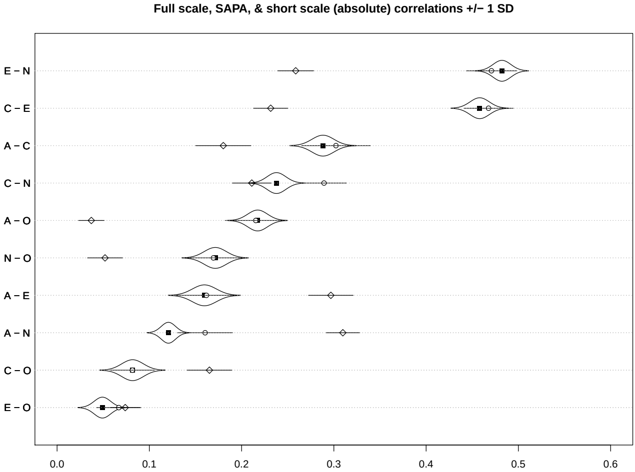
N=400 subjects Johnson 120 item IPIP, full scales: correlations



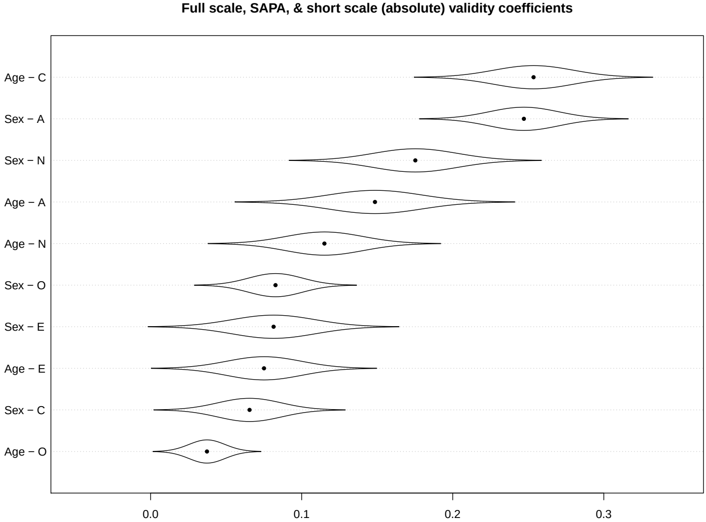
N=400 Johnson 120 item IPIP, sapa sampling 30 items: correlations



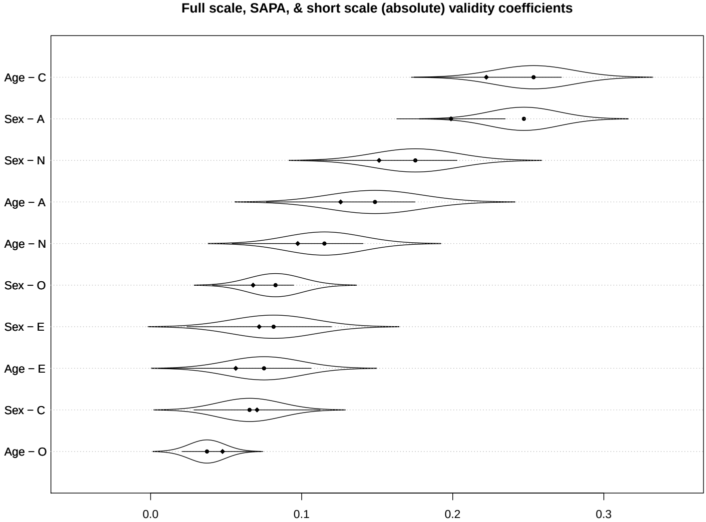
N=400 120 item IPIP, sapa sampling 30 at +short scales



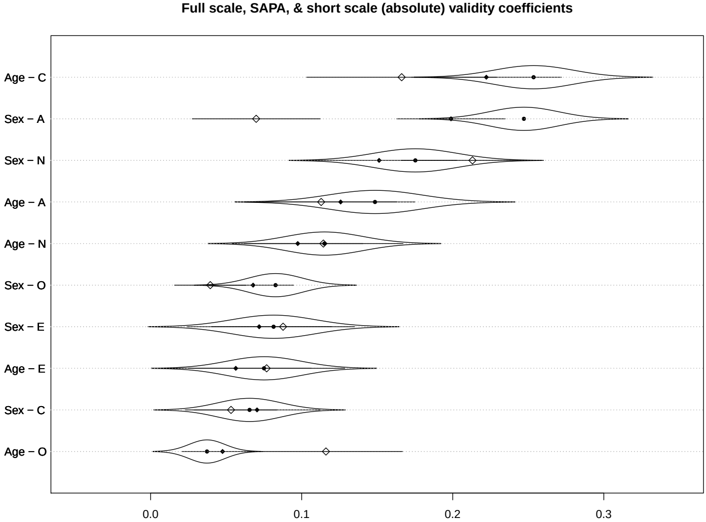
N=400 from Johnson 120 item IPIP, full scales: Validity



N=400 Johnson 120 item IPIP, sapa sampling 30 items: Validity



N=400 120 item IPIP, sapa 30 items + Short scales: Validity



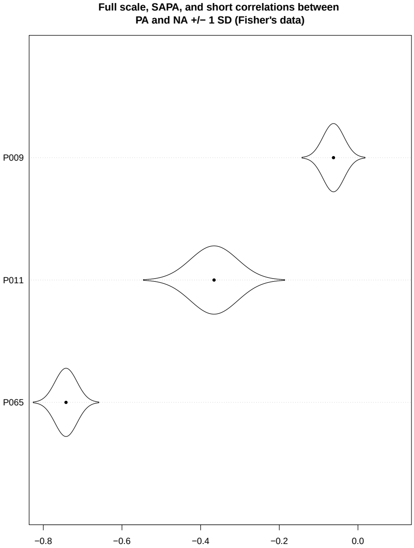
Some comments

1. These simulations were done with 400-2,000 simulated (but real) subjects
2. With p=.25, the number of pairwise correlations was $p^2 N = .04 * 400 \approx 25/pair$ and yet the results are very stable!
3. With such missingness, the correlation matrices are improper, but the `fa` function will give a minres solution anyway.
4. Even better is to use *minchi* fitting to take into account the variation of sample sizes.

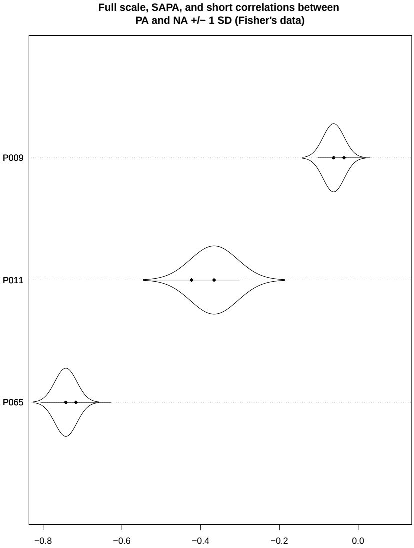
Experience Sampling Methods

1. We can also apply this technique to ESM data
2. Two data sets were examined: Fisher (2015) and Wilt, Funkhouser & Revelle (2011)
3. We were particularly interested in the correlation within subjects between Positive and Negative Affect
4. Rather than do k-fold sampling, we did boot strap resampling of the entire data or with 50% sapaized samples

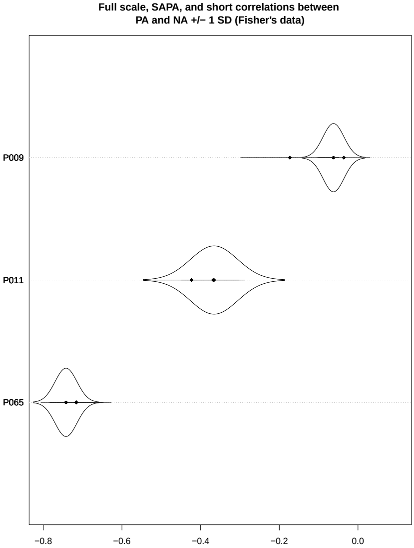
Aaron Fisher data - as reported



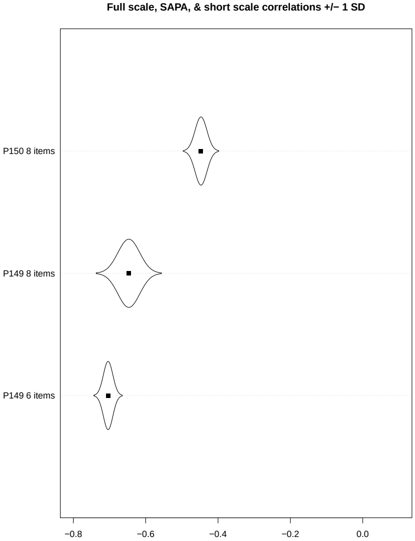
Aaron Fisher data – comparing full versus sapaized 50%



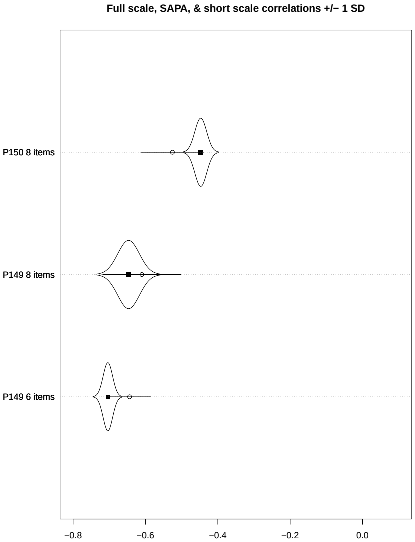
Aaron Fisher data – comparing full versus sapaized 50%



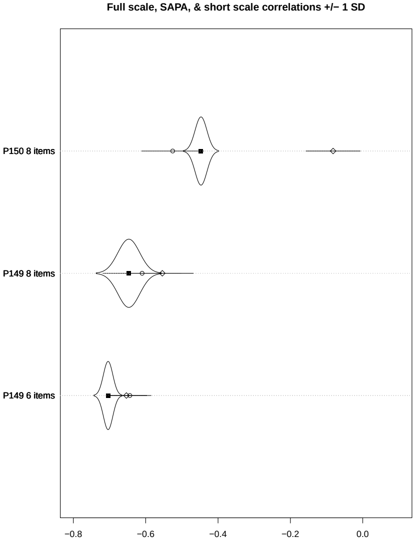
Wilt-Revelle data - as reported



Wilt-Revelle data – comparing full versus sapaized 50%



Wilt-Revelle data – comparing full versus sapaized 50%



SAPA or MMCAR procedures are very powerful

1. At the SAPA-project we estimate difficulty parameters and covariance structures for 1,000s of items even though only 100-150 items are answered per subject.
2. More importantly, the same procedures can be used for people with smaller sample sizes with fewer items (e.g. MTurk research).
3. Structure of ability measures using the open source ability test from the International Cognitive Ability Resource (ICAR) http://icar-project.com
4. Data sharing: https://dataverse.harvard.edu/dataverse/SAPA-ProjectCode/manuscript/
5. SPI development (Condon, 2017): https://sapa-project.org/research/SPI/SPIdevelopment.pdf
6. SPI scales, norms, IRT parameters: https://sapa-project.org/research/SPI
7. Today's slides at http://personality-project.org/sapa
8. Join the ICAR and SAPA projects.

R code

The next few slides show the R code used for these analyses

R code

```
sapa <- read.file()  #this searched for and loaded the master data file
icar.dictionary <- read.file()  #this searches for and loads a dictionary file

sapa <- SAPAdata18aug2010thru7feb2017  #change the name to make the following analyses cleare
 #the raw sapa has alphanumeric codes for some fields. Convert to numeric
sapa <- char2numeric(sapa)
 icar <- sapa[rownames(icar.dictionary)] #Identify the ICAR items
 icar.keys <- ItemLists[417:421]  #the keys are loaded as part of the sapa read.file line
icar  #show the item numbers of ICAR

icar.16 <- c("q_12007","q_12033" ,"q_12034","q_12058","q_12045", "q_12046","q_12047",
"q_12055", "q_11003", "q_11004","q_11006" ,"q_11008","q_12004","q_12016","
q_12017", "q_12019")

#now do random resamples, 100 times each of various fractions of the icar.16 data
 icar.16.sapa.05 <-fa.sapa(sapa[icar.16],4,n.iter=100,frac=.05)
  icar.16.sapa.1 <-fa.sapa(sapa[icar.16],4,n.iter=100,frac=.1)
 icar.16.sapa.2  <-fa.sapa(sapa[icar.16],4,n.iter=100,frac=.2)
 icar.16.sapa.4  <-fa.sapa(sapa[icar.16],4,n.iter=100,frac=.4)
```

More R

R code

```
error.dots(icar.16.sapa.05,head=40,tail=40,sort=FALSE,main="ICAR 16 4 Factors  pairwise = 12
error.dots(icar.16.sapa.1,head=40,tail=40,sort=FALSE,main="ICAR 16 4 Factors pairwise = 2479"
error.dots(icar.16.sapa.2,head=40,tail=40,sort=FALSE,main="ICAR 16 4 Factors pairwise = 4950"
 error.dots(icar.16.sapa.4,head=40,tail=40,sort=FALSE,main="ICAR 16 4 Factors 4 pairwise = 9


 #now show the factor intercorrelations
 names(icar.16.sapa.05$cis$means.rot) <- label.icar.rot
names(icar.16.sapa.1$cis$means.rot) <- label.icar.rot
 names(icar.16.sapa.2$cis$means.rot) <- label.icar.rot
 names(icar.16.sapa.4$cis$means.rot) <- label.icar.rot

  error.dots(icar.16.sapa.05$cis$means.rot,se=icar.16.sapa.05$cis$sds.rot,sort=FALSE,xlim=c(
  error.dots(icar.16.sapa.1$cis$means.rot,se=icar.16.sapa.1$cis$sds.rot,sort=FALSE,xlim=c(0,
  error.dots(icar.16.sapa.2$cis$means.rot,se=icar.16.sapa.2$cis$sds.rot,sort=FALSE,xlim=c(0,
  error.dots(icar.16.sapa.4$cis$means.rot,se=icar.16.sapa.4$cis$sds.rot,sort=FALSE,xlim=c(0,

  op <- par(mfrow=c(2,2))
   error.dots(icar.16.sapa.05$cis$means.rot,se=icar.16.sapa.05$cis$sds.rot,sort=FALSE,xlim=c
  error.dots(icar.16.sapa.1$cis$means.rot,se=icar.16.sapa.1$cis$sds.rot,sort=FALSE,xlim=c(0,
  error.dots(icar.16.sapa.2$cis$means.rot,se=icar.16.sapa.2$cis$sds.rot,sort=FALSE,xlim=c(0,
  error.dots(icar.16.sapa.4$cis$means.rot,se=icar.16.sapa.4$cis$sds.rot,sort=FALSE,xlim=c(0,
```

More R

R code

```
op <- par(mfrow=c(1,1)
icar.16.sapa.05$cis$mean.pair   #1238.375
icar.16.sapa.1$cis$mean.pair    # 2478.613
icar.16.sapa.2$cis$mean.pair  #  4950.1
icar.16.sapa.4$cis$mean.pair  # 9906.598

#now, just take out a general factor
 icar.16.sapa.1.05 <-fa.sapa(sapa[icar.16],n.iter=100,frac=.05)
  icar.16.sapa.1.1 <-fa.sapa(sapa[icar.16],n.iter=100,frac=.1)
  icar.16.sapa.1.2  <-fa.sapa(sapa[icar.16],n.iter=100,frac=.2)
   icar.16.sapa.1.4  <-fa.sapa(sapa[icar.16],n.iter=100,frac=.4)

error.dots(icar.16.sapa.1.05,head=40,tail=40,sort=FALSE,main="ICAR 16 1 Factor   5% sample",
error.dots(icar.16.sapa.1.1,head=40,tail=40,sort=FALSE,main="ICAR 16 1Factor 10% sample" ,xl
error.dots(icar.16.sapa.1.2,head=40,tail=40,sort=FALSE,main="ICAR 16 1 Factor 20% sample",xl
 error.dots(icar.16.sapa.1.4,head=40,tail=40,sort=FALSE,main="ICAR 16 1 Factor 40% sample",x


error.dots(icar.16.sapa.1.05,head=40,tail=40,sort=FALSE,main="ICAR 16 1 Factor   5% ",xlim=c
error.dots(icar.16.sapa.1.1,head=40,tail=40,sort=FALSE,main="ICAR 16 1Factor 10% " ,xlim=c(0
error.dots(icar.16.sapa.1.2,head=40,tail=40,sort=FALSE,main="ICAR 16 1 Factor 20% ",xlim=c(0
 error.dots(icar.16.sapa.1.4,head=40,tail=40,sort=FALSE,main="ICAR 16 1 Factor 40% ",xlim=c(

icar.keys <- keys.list[398:402]


R3Diq.sapa  <-fa.sapa(sapa[icar.keys[[4]]],1,n.iter=100,frac=.05)
R3Diq.sapa.1  <-fa.sapa(sapa[icar.keys[[4]]],1,n.iter=100,frac=.1)
R3Diq.sapa.2  <-fa.sapa(sapa[icar.keys[[4]]],1,n.iter=100,frac=.2)
R3Diq.sapa.4  <-fa.sapa(sapa[icar.keys[[4]]],1,n.iter=100,frac=.4)

R3Diq.sapa$cis$mean.pair
```

R code

```
 #this next one fails
#R3Diq.sapa.01  <-fa.sapa(sapa[icar.keys[[4]]],1,n.iter=100,frac=.01)
# error.dots(R3Diq.sapa..01,head=40,tail=40,sort=FALSE,main="3D rotation items  -- 1% sample
  op <- par(mfrow=c(2,2))
 error.dots(R3Diq.sapa,head=40,tail=40,sort=FALSE,main="3D rotation items  pairwise = 200",x
 error.dots(R3Diq.sapa.1,head=40,tail=40,sort=FALSE,main="3D rotation items  pairwise = 399",
 error.dots(R3Diq.sapa.2,head=40,tail=40,sort=FALSE,main="3D rotation items   pairwise = 798
  error.dots(R3Diq.sapa.4,head=40,tail=40,sort=FALSE,main="3D rotation items  pairwise=1596",

  op <- par(mfrow=c(1,1)


  samp.size <- data.frame(fraction=fraction,sample = fraction * 255348,icar16=fraction * 743

  #now, some omega comparisons
  om16 <- omega(sapa[icar.16],4)   #the complete sample
   omega.diagram(om16)

# a 5% sample
  sapa.5.samp <- sapa[sample(1:255348,12767,replace=TRUE),icar.16]
  om.samp.5 <- omega(sapa.5.samp,4)
  omega.diagram(om.samp.5)

  mean(count.pairwise(sapa.5.samp,diagonal=FALSE),na.rm=TRUE)
  cp <- count.pairwise(sapa.5.samp)
> mean(diag(cp))
[1] 3718.562

corPlot(factor.congruence(om16,om.samp.5),numbers=TRUE,gr=gr,main="Factor Congruence: total s

sapa.02.samp <- sapa[sample(1:255348,5000,replace=TRUE),icar.16]
om.02 <- omega(sapa.02.samp,4)
 corPlot(factor.congruence(om16,om.02),numbers=TRUE,gr=gr,main="Factor Congruence: total samp
 mean(count.pairwise(sapa.02.samp,diagonal=FALSE),na.rm=TRUE)
```

R code

```
#now some interesting simulations


sim.1 <- sim.irt(24,200000,low=-1,high=1,a=3)
simp <- sim.1$items
 filter <- matrix(NA,nrow=200000,ncol=24)
 filter <- sample(1:24,24*200000,replace=TRUE)

 filter<- matrix(filter,ncol=24)
simp[filter > 12 ] <- NA.  #rhis is a 50% sample
simp <- matrix(simp,ncol=24)

sim.fa.008 <- fa.sapa(simp,frac=.008,n.iter=100)
sim.fa.0016<- fa.sapa(simp,frac=.0016,n.iter=100)
 sim.fa.016<- fa.sapa(simp,frac=.016,n.iter=100)
sim.fa.032<- fa.sapa(simp,frac=.032,n.iter=100).
sim.fa.032$cis$mean.pair    #[1] 1600.788

#Now, do this again for a 25% sample
simp.25 <- sim.1$items
simp.25[filter > 6 ] <- NA.  #rhis is a 50% sample
simp,25 <- matrix(simp.25,ncol=24)
```

Error dots for factor loadings

NOte that the error bars are smaller for the 25 versus 50 % samples

R code

```
op <- par(mfrow=c(2,2))
error.dots(sim.fa.004,sort=FALSE,xlim=c(0,1),main="200 pairwise, 50
error.dots(sim.fa.008,sort=FALSE,xlim=c(0,1),main="400 pairwise, 50
error.dots(sim.fa.016,sort=FALSE,xlim=c(0,1),main="800 pairwise, 50
error.dots(sim.fa.032,sort=FALSE,xlim=c(0,1),main="1600 pairwise, 5

error.dots(sim.fa.25.016,sort=FALSE,xlim=c(0,1),main="200 pairwise,
error.dots(sim.fa.25.032,sort=FALSE,xlim=c(0,1),main="400 pairwise,
error.dots(sim.fa.25.064,sort=FALSE,xlim=c(0,1),main="800 pairwise,
error.dots(sim.fa.25.128,sort=FALSE,xlim=c(0,1),main="1600 pairwise
op <- par(mfrow=c(1,1))
```

Anderson, J., Lin, H., Treagust, D., Ross, S., & Yore, L. (2007). Using large-scale assessment datasets for research in science and mathematics education: Programme for International Student Assessment (PISA). *International Journal of Science and Mathematics Education*, *5*(4), 591–614.

Condon, D. M. (2014). *An organizational framework for the psychological individual differences: Integrating the affective, cognitive, and conative domains*. PhD thesis, Northwestern University.

Condon, D. M. (2017). The SAPA Personality Inventory: An empirically-derived, hierarchically-organized self-report personality assessment model. Technical report, Northwestern University.

Condon, D. M. & Revelle, W. (2014). The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. *Intelligence*, *43*, 52–64.

Condon, D. M. & Revelle, W. (2017). The times they are a-changin' (in personality assessment). In *European Conference on Psychological Assessment*.

Costa, P. T. & McCrae, R. R. (1992). *NEO PI-R professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the big five. *Journal of Personality and Social Psychology*, *93*(5), 880–896.

Fisher, A. J. (2015). Toward a dynamic model of psychological assessment: Implications for personalized care. *Journal of Consulting and Clinical Psychology*, *83*(4), 825 – 836.

Goldberg, L. R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe*, volume 7 (pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.

Goldberg, L. R. & Saucier, G. (2016). The Eugene-Springfield Community Sample:

Information available from the research participants. Technical Report 56-1, Oregon Research Institute, Eugene, Oregon.

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality*, *37*(6), 504 – 528.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory-Versions 4a and 54*. Berkeley, CA: University of California, Berkeley, Institue of Personality and Social Research.

Johnson, J. A. (2014). Measuring thirty facets of the five factor model with a 120-item public domain inventory: Development of the ipip-neo-120. *Journal of Research in Personality*, *51*, 78 – 89.

Kelly, E. L. & Fiske, D. W. (1950). The prediction of success in the VA training program in clinical psychology. *American Psychologist*, *5*(8), 395 – 406.

Kelly, E. L. & Fiske, D. W. (1951). *The prediction of performance in clinical psychology*. Ann Arbor, Michigan: University of Michigan Press.

Konstabel, K., Lönnqvist, J.-E., Leikas, S., Velàzquez, R. G., H, H. Q., Verkasalo, M., , & et al. (2017). Measuring single constructs by single items: Constructing an even shorter version of the "short five" personality inventory. *PLoS ONE*, *12*(8), e0182714.

Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, *70*(6), 543.

Lord, F. M. (1955). Estimating test reliability. *Educational and Psychological Measurement*, *15*, 325–336.

Lord, F. M. (1977). Some item analysis and test theory for a system of computer-assisted test construction for individualized instruction. *Applied Psychological Measurement*, *1*(3), 447–455.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factors structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, *66*(6), 574–583.

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

Rammstedt, B. & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of Research in Personality*, *41*(1), 203 – 212.

Revelle, W. (2018). *psych: Procedures for Personality and Psychological Research*. https://cran.r-project.org/web/packages=psych: Northwestern University, Evanston. R package version 1.8.4.

Revelle, W. & Condon, D. M. (2018). Using sapa to study the structure of 6600 personality and ability items. In *Part of symposium: Measuring personality: What's new and how does it* *help personality psychology*, Zadar, Croatia. European Conference on Personality.

Revelle, W., Condon, D. M., Wilt, J., French, J. A., Brown, A., & Elleman, L. G. (2016). Web and phone based data collection using planned missing designs. In N. G. Fielding, R. M. Lee, & G. Blank (Eds.), *SAGE Handbook of Online Research Methods* (2nd ed.). chapter 37, (pp. 578–595). Sage Publications, Inc.

Revelle, W., Wilt, J., & Rosenthal, A. (2010). Individual differences in cognition: New methods for examining the personality-cognition link. In A. Gruszka, G. Matthews, & B. Szymura (Eds.), *Handbook of Individual Differences in Cognition: Attention, Memory and Executive Control* chapter 2, (pp. 27–49). New York, N.Y.: Springer.

Soto, C. J. & John, O. P. (2017). The next big five inventory (bfi-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of personality and social psychology*, *113*(1), 117–143.

Tryon, R. (1957). Communality of a variable: Formulation by cluster analysis. *Psychometrika*, *22*(3), 241–260.

Tryon, R. (1959). Domain sampling formulation of cluster and factor analysis. *Psychometrika*, *24*(2), 113–135.

Tupes, E. C. & Christal, R. E. (1961). Recurrent personality factors based on trait ratings. Technical Report 61-97, USAF ASD Technical Report, Lackland Air Force Base (reprinted in Journal of Personality (1992) 60: 225–251).

Wilt, J., Funkhouser, K., & Revelle, W. (2011). The dynamic relationships of affective synchrony to perceptions of situations. *Journal of Research in Personality*, *45*, 309–321.

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, *112*(4), 1036–1040.