Overview
000

Using Qualtrics
00

Analyzing your data
000000

More analysis
0000000

Advanced analysis
000000
0

Psychology 360: Personality Research
Analyzing the data for the final projects

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

NORTHWESTERN
UNIVERSITY

November, 2022

Overview
ooo

Using Qualtrics
oo

Analyzing your data
oooooo

More analysis
ooooooo

Advanced analysis
oooooo
o

Outline

The basic steps

1. Install R on your computer (just do this once) (go to
   https://cran.r-project.org
2. Install the *psych* and *psychTools* (just do this once) (install
   packages option)
3. Make *psych* active (library(psych))
4. Get the data in Qualtrics ready to be read
5. Read the data into R using read.file
6. Find basic descriptive statistics
7. Score the scales
8. Do regressions or correlations to test your hypotheses

Read the How To's

1. An introduction (vignette) of the *psych* package
2. An overview (vignette) of the *psych* package
3. Installing R and some useful packages
4. Using R and the *psych* package to find *omega$_h$* and ω_t.
5. Using R and the *psych* for factor analysis and principal components analysis.
6. Using the scoreItems function to find scale scores and scale statistics (this document).
7. Using mediate and setCor to do mediation, moderation and regression analysis

Once R is installed and psych is installed

```
                         ┌─── R code ───┐
library(psych)   #necessary whenever you want to run functions
                 in the psych package
my.data <- read.clipboard()   #normal free field with spaces in between values
#or
my.data <- read.clipboard.csv()  #if the data are in csv format
my.keys <- make.keys(my.data,list(scale1 = cs(Q1,Q2, -Q6),
          scale2 = cs((Q3,Q4,Q5)  #etc
dim(my.data)    #how many subjects (rows) and variables (colums)
describe(my.data)
my.scales <- scoreItems(my.keys,my.data)
my.scales
```

Getting the data from Qualtrics

1. Download the data from Qualtrics as either an Excel file or as a csv (comma separated values).
2. A little bit of preprocessing is necessary
   - Prepare a line before your data with short abbreviations for each item
   - This can either be an abbreviation of the item (no more that about 10 characters) or
   - Some clever name such as Q1, Q2, Q3, ...
3. Select this row and all of the data (ignore the first few columns with id number and names, ip addresses, etc)
4. Copy this selection to your clipboard
5. `my.data <- read.clipboard.csv()`
6. Or paste your data into a new csv file so you can automate the process
7. Consider the example data in a file small.msq

Reading from a remote file

```R code
fn <- "https://personality-project.org/r/psych/HowTo/scoring.tutorial/small.msc
my.data <- read.file(fn)
dim(my.data)
describe(my.data)
```

```
fn <- "https://personality-project.org/r/psych/HowTo/scoring.tutorial/small.msq.txt"
my.data <- read.file(fn)
Data from the .txt file https://personality-project.org/r/psych/HowTo/scoring.tutorial/small.msq.txt has b
> dim(my.data)
[1] 200  14
> describe(my.data)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
active	1	200	0.86	0.84	1.0	0.78	1.48	0	3	3	0.62	-0.45	0.06
alert	2	199	0.97	0.85	1.0	0.90	1.48	0	3	3	0.54	-0.42	0.06
aroused	3	200	0.52	0.76	0.0	0.38	0.00	0	3	3	1.30	0.92	0.05
sleepy	4	198	1.60	1.08	2.0	1.62	1.48	0	3	3	-0.05	-1.31	0.08
tired	5	200	1.78	1.01	2.0	1.86	1.48	0	3	3	-0.30	-1.04	0.07
drowsy	6	200	1.50	1.08	1.0	1.49	1.48	0	3	3	0.02	-1.28	0.08
anxious	7	200	0.60	0.83	0.0	0.45	0.00	0	3	3	1.31	1.05	0.06
jittery	8	200	0.66	0.87	0.0	0.50	0.00	0	3	3	1.22	0.63	0.06
nervous	9	200	0.44	0.77	0.0	0.25	0.00	0	3	3	1.81	2.63	0.05
calm	10	200	1.55	0.91	2.0	1.57	1.48	0	3	3	-0.06	-0.80	0.06
relaxed	11	199	1.67	0.86	2.0	1.71	1.48	0	3	3	-0.20	-0.63	0.06
at.ease	12	198	1.40	0.96	1.0	1.38	1.48	0	3	3	0.08	-0.96	0.07
gender	13	136	1.54	0.50	2.0	1.55	0.00	1	2	1	-0.18	-1.98	0.04
drug	14	136	1.50	0.50	1.5	1.50	0.74	1	2	1	0.00	-2.01	0.04

Overview
○○○

Using Qualtrics
○○

Analyzing your data
●○○○○○

More analysis
○○○○○○○

Advanced analysis
○○○○○○
○

Create the instructions (keys) for scoring the data

```
┌─ R code ─┐
my.keys <-  list(
EA= c("active","alert","aroused", "-sleepy","-tired", "-drowsy"),
 TA = c("anxious","jittery","nervous","-calm", "-relaxed",  "-at.ease"),
                         EAp = c("active","alert","aroused"),
                         EAn = c("sleepy","tired", "drowsy"),
                         TAp = c("anxious","jittery","nervous"),
                         TAn = c("calm", "relaxed", "at.ease")
                         )

 another.keys.list <- list(EA=c(1:3,-4,-5,-6),TA=c(7:9,-10,-11,-12),
                    EAp =1:3,EAn=4:6,TAp =7:9,TAn=10:12)

```

Score the items

```
Call: scoreItems(keys = my.keys, items = my.data)

(Unstandardized) Alpha:
        EA   TA  EAp  EAn  TAp  TAn
alpha 0.85 0.83 0.76 0.89 0.78 0.83

Standard errors of unstandardized Alpha:
        EA    TA   EAp   EAn   TAp  TAn
ASE   0.033 0.037 0.067 0.053 0.065 0.06

Average item correlation:
            EA   TA  EAp  EAn  TAp  TAn
average.r 0.49 0.44 0.51 0.74 0.55 0.62

Median item correlation:
  EA   TA  EAp  EAn  TAp  TAn
0.48 0.40 0.54 0.71 0.57 0.63

 Guttman 6* reliability:
           EA   TA  EAp  EAn  TAp  TAn
Lambda.6 0.89 0.85 0.75 0.87 0.75 0.79

Signal/Noise based upon av.r :
              EA  TA EAp EAn TAp TAn
Signal/Noise 5.9 4.7 3.2 8.5 3.6 4.8

Scale intercorrelations corrected for attenuation
 raw correlations below the diagonal, alpha on the diagonal
 corrected correlations above the diagonal:
        EA    TA   EAp    EAn   TAp    TAn
EA   0.854  0.25  1.025 -1.051  0.40 -0.063
TA   0.212  0.83  0.252 -0.207  1.04 -1.063
EAp  0.826  0.20  0.760 -0.651  0.51  0.032
```

Scores output (continued)

```
Scale intercorrelations corrected for attenuation
 raw correlations below the diagonal, alpha on the diagonal
 corrected correlations above the diagonal:
        EA     TA     EAp    EAn    TAp    TAn
EA    0.854  0.25   1.025 -1.051  0.40  -0.063
TA    0.212  0.83   0.252 -0.207  1.04  -1.063
EAp   0.826  0.20   0.760 -0.651  0.51   0.032
EAn  -0.919 -0.18  -0.537  0.894 -0.26   0.112
TAp   0.330  0.84   0.396 -0.217  0.78  -0.593
TAn  -0.053 -0.88   0.025  0.097 -0.48   0.829

 Average adjusted correlations within and between scales (MIMS)
     EA     TA     EAp    EAn    TAp    TAn
EA    0.49
TA    0.10   0.44
EAp   0.34   0.08   0.51
EAn  -0.58  -0.11  -0.35   0.74
TAp   0.16   0.31   0.18  -0.14   0.55
TAn  -0.03  -0.37   0.01   0.07  -0.26   0.62

 Average adjusted item x scale correlations within and between scales (MIMT)
     EA     TA     EAp    EAn    TAp    TAn
EA    0.76
TA    0.16   0.73
EAp   0.68   0.17   0.82
EAn  -0.84  -0.16  -0.49   0.91
TAp   0.27   0.70   0.33  -0.18   0.84
TAn  -0.05  -0.76   0.02   0.08  -0.41   0.86

 In order to see the item by scale loadings and frequency counts of the data
 print with the short option = FALSE
```

Objects contain objects

```
────────────────── R code ──────────────────
names(my.scales)
my.scores <- my.scales$scores      #the actual scores are saved in the scores ob
dim(my.scores)
```

```
names(my.scales)
 [1] "scores"         "missing"        "alpha"        "av.r"          "sn"
 [6] "n.items"        "item.cor"       "cor"          "corrected"     "G6"
[11] "item.corrected" "response.freq"  "raw"          "ase"           "med.r"
[16] "keys"           "MIMS"           "MIMT"         "Call"
> my.scores <- my.scales$scores      #the actual scores are saved in the scores object
> dim(my.scores)
[1] 200    6
```
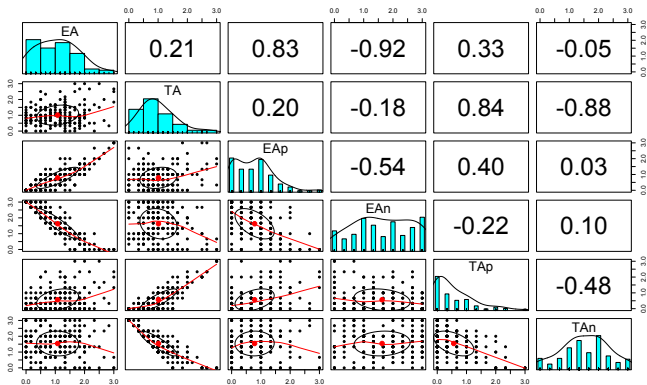
Basic descriptives

```
                                  ┌─── R code ───┐
describe(my.scores)
pairs.panels(my.scores)
```

```
 describe(my.scores)
      vars   n mean   sd median trimmed  mad min max range  skew kurtosis   se
EA       1 200 1.08 0.72   1.08    1.05 0.86   0   3     3  0.36    -0.43 0.05
TA       2 200 1.01 0.63   1.00    0.96 0.49   0   3     3  0.88     0.81 0.04
EAp      3 200 0.78 0.67   0.67    0.71 0.49   0   3     3  0.87     0.75 0.05
EAn      4 200 1.63 0.96   1.67    1.66 0.99   0   3     3 -0.09    -1.17 0.07
TAp      5 200 0.57 0.69   0.33    0.44 0.49   0   3     3  1.42     1.64 0.05
TAn      6 200 1.54 0.78   1.67    1.55 0.99   0   3     3 -0.12    -0.54 0.06
```

Overview
○○○

Using Qualtrics
○○

Analyzing your data
○○○○○●

More analysis
○○○○○○○

Advanced analysis
○○○○○○○
○

Show the SPLOM

The effect of drug and gender

1. The original data set (my.data) included the drug condition (1=placebo, 2 = caffeine) and the gender (1= male, 2 = female) of the subjects.
2. Combine these codes with the scale scores

Data manipulation

```
────────────────── R code ──────────────────
big.data <- cbind(my.scores, my.data[cs(drug,gender)])
describe(big.data)
```

```
> describe(big.data)
        vars   n mean   sd median trimmed  mad min max range  skew kurtosis   se
EA         1 200 1.08 0.72   1.08    1.05 0.86   0   3     3  0.36    -0.43 0.05
TA         2 200 1.01 0.63   1.00    0.96 0.49   0   3     3  0.88     0.81 0.04
EAp        3 200 0.78 0.67   0.67    0.71 0.49   0   3     3  0.87     0.75 0.05
EAn        4 200 1.63 0.96   1.67    1.66 0.99   0   3     3 -0.09    -1.17 0.07
TAp        5 200 0.57 0.69   0.33    0.44 0.49   0   3     3  1.42     1.64 0.05
TAn        6 200 1.54 0.78   1.67    1.55 0.99   0   3     3 -0.12    -0.54 0.06
drug       7 136 1.50 0.50   1.50    1.50 0.74   1   2     1  0.00    -2.01 0.04
gender     8 136 1.54 0.50   2.00    1.55 0.00   1   2     1 -0.18    -1.98 0.04
```

Break it down by group

R code

```
describeBy(big.data ~ drug)
error.bars.by(big.data~drug,by.var=FALSE
     ylab="score" ,main="Effect of caffeine on arousal")
```

```
 Descriptive statistics by group
drug: 1
       vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
EA        1 68 0.82 0.64   0.83    0.78 0.74   0 2.5   2.5  0.35    -0.80 0.08
TA        2 68 0.96 0.68   0.83    0.88 0.49   0 3.0   3.0  1.28     1.76 0.08
EAp       3 68 0.57 0.59   0.33    0.50 0.49   0 2.0   2.0  0.78    -0.43 0.07
EAn       4 68 1.94 0.89   2.00    1.98 1.48   0 3.0   3.0 -0.21    -1.27 0.11
TAp       5 68 0.47 0.74   0.00    0.31 0.00   0 3.0   3.0  2.08     4.04 0.09
TAn       6 68 1.54 0.82   1.67    1.55 0.99   0 3.0   3.0 -0.15    -0.77 0.10
drug      7 68 1.00 0.00   1.00    1.00 0.00   1 1.0   0.0   NaN      NaN 0.00
gender    8 68 1.56 0.50   2.00    1.57 0.00   1 2.0   1.0 -0.23    -1.97 0.06
-------------------------------------------------------------------
drug: 2
       vars  n mean   sd median trimmed  mad  min  max range  skew kurtosis   se
EA        1 68 1.40 0.74   1.33    1.38 0.74 0.00 3.00  3.00  0.22    -0.40 0.09
TA        2 68 1.31 0.63   1.33    1.27 0.74 0.33 2.67  2.33  0.41    -0.59 0.08
EAp       3 68 1.01 0.75   1.00    0.94 0.49 0.00 3.00  3.00  0.82     0.54 0.09
EAn       4 68 1.21 0.90   1.17    1.17 1.24 0.00 3.00  3.00  0.23    -1.11 0.11
TAp       5 68 0.86 0.70   0.67    0.82 0.49 0.00 2.33  2.33  0.53    -0.79 0.08
TAn       6 68 1.24 0.73   1.33    1.25 0.99 0.00 2.67  2.67 -0.16    -1.03 0.09
drug      7 68 2.00 0.00   2.00    2.00 0.00 2.00 2.00  0.00   NaN      NaN 0.00
gender    8 68 1.53 0.50   2.00    1.54 0.00 1.00 2.00  1.00 -0.12    -2.02 0.06
```
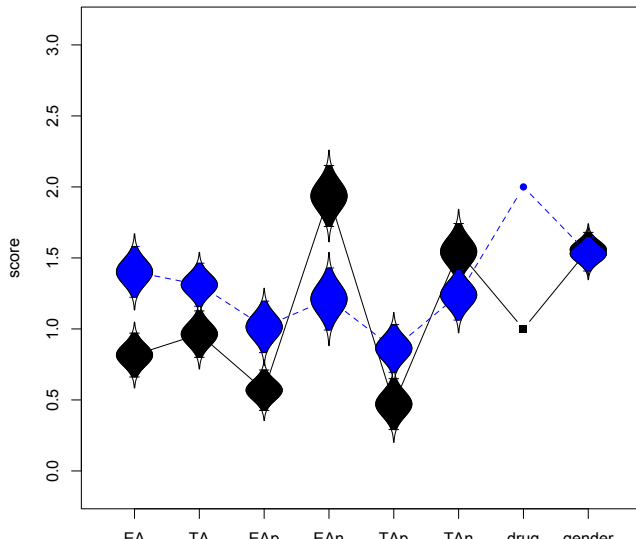
The effect of caffeine on multiple affect measures

Effect of caffeine on arousal

Inferential statistics

```
R code
```
```
setCor(EA+ TA ~ drug*gender, data=big.data)
```

Call: setCor(y = EA + TA ~ drug * gender, data = big.data)

Multiple Regression from raw data

 DV = EA
             slope   se   t   p lower.ci upper.ci VIF Vy.x
(Intercept)   0.00 0.06 0.00 1.0e+00   -0.12   0.12   1 0.00
drug   0.40 0.06 6.40 1.1e-09   0.28   0.53   1 0.17
gender   -0.22 0.06 -3.54 4.9e-04   -0.35   -0.10   1 0.05
drug*gender -0.08 0.06 -1.22 2.2e-01   -0.20   0.05   1 0.01

Residual Standard Error = 0.89 with 196 degrees of freedom

 Multiple Regression
       R   R2 Ruw R2uw Shrunken R2 SE of R2 overall F df1 df2   p
EA 0.47 0.22 0.41 0.17   0.21   0.05   18.81   3 196 9.18e-11

 DV = TA
             slope   se   t   p lower.ci upper.ci VIF Vy.x
(Intercept)   0.00 0.07 0.00 1.00000   -0.14   0.14   1 0.00
drug   0.28 0.07 4.03 0.00008   0.14   0.41   1 0.08
gender   0.03 0.07 0.41 0.68000   -0.11   0.16   1 0.00
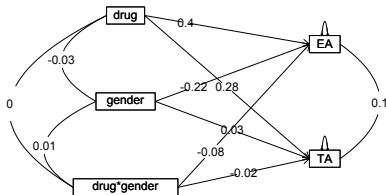drug*gender -0.02 0.07 -0.32 0.75000   -0.16   0.11   1 0.00

Residual Standard Error = 0.97 with 196 degrees of freedom

 Multiple Regression
       R   R2 Ruw R2uw Shrunken R2 SE of R2 overall F df1 df2   p
TA 0.22 0.02 0.12 0.02   0.06   0.04   5.42   3 196 0.00134

Show the graphic

Regression Models

Overview
000

Using Qualtrics
00

Analyzing your data
000000

More analysis
0000000●

Advanced analysis
000000
0

More detailed analysis

```
┌─────── R code ───────┐
setCor(EAp + EAn + TAp + TAn ~drug +gender,data=big.data)
```

```
Call: setCor(y = EAp + EAn + TAp + TAn ~ drug + gender, data = big.data)
Multiple Regression from raw data
 DV = EAp
            slope   se     t       p lower.ci upper.ci VIF Vy.x
(Intercept)  0.00 0.06  0.00 1.0e+00    -0.13     0.13   1 0.00
drug         0.33 0.06  5.09 8.5e-07     0.20     0.45   1 0.11
gender      -0.29 0.06 -4.54 9.6e-06    -0.42    -0.16   1 0.09
Residual Standard Error = 0.9  with  197  degrees of freedom
 Multiple Regression
       R   R2  Ruw R2uw Shrunken R2 SE of R2 overall F df1 df2        p
EAp 0.44 0.2 0.44  0.2        0.19     0.05      23.97   2 197 4.83e-10

 DV = EAn
            slope   se     t       p lower.ci upper.ci VIF Vy.x
(Intercept)  0.00 0.07  0.00 1.0e+00    -0.13     0.13   1 0.00
drug        -0.38 0.07 -5.75 3.0e-08    -0.50    -0.25   1 0.14
gender       0.13 0.07  2.01 4.5e-02     0.00     0.26   1 0.02
Residual Standard Error = 0.92  with  197  degrees of freedom
 Multiple Regression
      R   R2  Ruw R2uw Shrunken R2 SE of R2 overall F df1 df2        p
EAn 0.4 0.16 0.36 0.13        0.15     0.05      18.94   2 197 2.99e-08

 DV = TAp
            slope   se    t       p lower.ci upper.ci VIF Vy.x
(Intercept)  0.00 0.07 0.00 1.0e+00    -0.13     0.13   1 0.00
drug         0.29 0.07 4.22 3.7e-05     0.15     0.42   1 0.08
gender       0.05 0.07 0.76 4.5e-01    -0.08     0.19   1 0.00
Residual Standard Error = 0.96  with  197  degrees of freedom
 Multiple Regression
```

Dimension reduction through Factor Analysis

R code

```
f2 <- fa(my.data[1:12],2)
```

```
Factor Analysis using method =  minres
Call: fa(r = my.data[1:12], nfactors = 2)
Standardized loadings (pattern matrix) based upon correlation matrix
          MR1   MR2   h2   u2 com
active   0.65  0.07 0.41 0.59 1.0
alert    0.73  0.00 0.53 0.47 1.0
aroused  0.52 -0.13 0.30 0.70 1.1
sleepy  -0.77  0.06 0.60 0.40 1.0
tired   -0.79 -0.05 0.61 0.39 1.0
drowsy  -0.76 -0.03 0.58 0.42 1.0
anxious  0.25 -0.54 0.39 0.61 1.4
jittery  0.34 -0.54 0.46 0.54 1.7
nervous  0.13 -0.62 0.43 0.57 1.1
calm     0.01  0.72 0.52 0.48 1.0
relaxed  0.12  0.75 0.55 0.45 1.0
at.ease  0.13  0.76 0.56 0.44 1.1

                        MR1  MR2
SS loadings            3.25 2.70
Proportion Var         0.27 0.22
Cumulative Var         0.27 0.50
Proportion Explained   0.55 0.45
Cumulative Proportion  0.55 1.00

 With factor correlations of
      MR1   MR2
MR1  1.00 -0.15
MR2 -0.15  1.00
```

Factor analysis continued

```
Mean item complexity =  1.1
Test of the hypothesis that 2 factors are sufficient.

df null model =  66  with the objective function =  6.17 with Chi Square =  1198.51
df of  the model are 43  and the objective function was  1.65

The root mean square of the residuals (RMSR) is  0.09
The df corrected root mean square of the residuals is  0.11

The harmonic n.obs is  199 with the empirical chi square  223.16  with prob <  3.6e-26
The total n.obs was  200  with Likelihood Chi Square =  319.12  with prob <  7.6e-44

Tucker Lewis Index of factoring reliability =  0.623
RMSEA index =  0.179  and the 90 % confidence intervals are  0.161 0.198
BIC =  91.29
Fit based upon off diagonal values = 0.93
Measures of factor score adequacy
                                                        MR1  MR2
Correlation of (regression) scores with factors         0.94 0.92
Multiple R square of scores with factors                0.87 0.84
Minimum correlation of possible factor scores           0.75 0.69
```
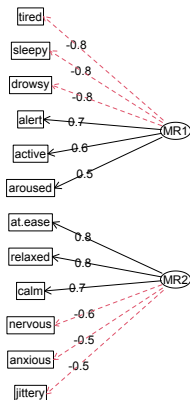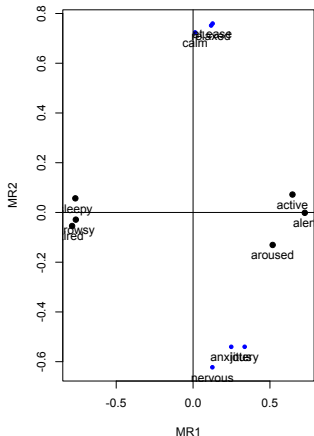
Two ways of showing the results

Factor Analysis

Cluster the items

R code

```
ic <- iclust(my.data[1:12])
```

```
ICLUST (Item Cluster Analysis)
Call: iclust(r.mat = my.data[1:12])

Purified Alpha:
  C9  C10
0.85 0.83
G6* reliability:
  C9  C10
0.71 0.69
Original Beta:
  C9  C10
0.69 0.65
Cluster size:
 C9 C10
  6   6
Item by Cluster Structure matrix:
          O    P    C9    C10
active   C9   C9  -0.65 -0.09
alert    C9   C9  -0.71 -0.19
aroused  C9   C9  -0.55 -0.28
sleepy   C9   C9   0.81  0.24
tired    C9   C9   0.77  0.14
drowsy   C9   C9   0.79  0.15
anxious  C10  C10 -0.31 -0.62
jittery  C10  C10 -0.41 -0.64
nervous  C10  C10 -0.19 -0.68
calm     C10  C10  0.12  0.70
relaxed  C10  C10  0.00  0.70
at.ease  C10  C10  0.01  0.70
```
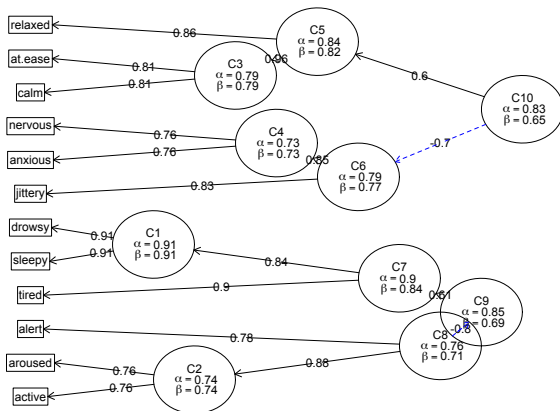
Show the iclust diagram

ICLUST

The two structures are practically identical

```
factor.congruence(ic,f2)
```
R code

```
 factor.congruence(ic,f2)
      MR1  MR2
C9  -0.99 0.20
C10 -0.33 0.96
```

Making an item dictionary to clean up the output

1. When showing the pattern of item responses using factor analysis or scale construction, it is helpful to refer to items not by item number (e.g. Q1, Q2) but rather in terms of the item content.

2. Basic scoring commands (e.g. `makeKeys`) are easier with short names (Q1, Q3) but that results in output which is not as helpful.

3. Thus, we have the ability to create `dictionaries`

4. To do this, first make up a a Excel spreadsheet with at least columns:
   - Column 1: the abbreviation (Q1, Q3, ..) for the item
   - Column 2: The full item (do not use apostrophes)
   - Give column labels (item, content)

5. Copy these two columns to R using `read.clipboard.tab()`
   e.g. `my.dictionary <- read.clipboard.tab()`

6. Give the data the row names of the abbreviations
   `rownames(my.dictionary) <- my.dictionary[,1]`