Package ‘psychTools’

June 18, 2023

Version 2.3.6
Date 2023-06-17
Title Tools to Accompany the ‘psych’ Package for Psychological Research
Description Support functions, data sets, and vignettes for the ‘psych’ package. Contains several of the biggest data sets for the ‘psych’ package as well as four vignettes. A few helper functions for file manipulation are included as well. For more information, see the <https://personality-project.org/r/> web page.
License GPL (>= 2)
Imports foreign,psych
Suggests parallel, GPArotation, lavaan, knitr
Depends R(>= 2.10)
LazyData yes
ByteCompile TRUE
VignetteBuilder knitr
URL https://personality-project.org/r/psych/
https://personality-project.org/r/psych-manual.pdf
NeedsCompilation no
Author William Revelle [aut, cre] (<https://orcid.org/0000-0003-4880-9610>)
Maintainer William Revelle <revelle@northwestern.edu>
Repository CRAN
Date/Publication 2023-06-18 00:20:02 UTC

R topics documented:

ability ............................................. 3
affect ............................................. 5
Athenstaedt .................................... 6
bfi .................................................. 10
BFI.adjectives.dictionary ......................... 12
R topics documented:

big5.100.adjectives ........................................... 14
blant ......................................................... 15
blot ......................................................... 16
burt ......................................................... 17
cities ......................................................... 19
colom ......................................................... 20
cubits ......................................................... 21
cushny ......................................................... 22
df2latex ....................................................... 24
dfOrder ....................................................... 26
eminence ....................................................... 28
epi ............................................................. 29
epi.bfi ........................................................ 32
galton ......................................................... 33
GERAS ........................................................ 34
globalWarm .................................................. 37
heights ......................................................... 39
holzinger.swineford ......................................... 40
income ......................................................... 44
iqitems ....................................................... 45
msq ........................................................... 47
msqR .......................................................... 53
neo ............................................................ 61
peas ........................................................... 63
Pollack ......................................................... 64
psychTools ................................................... 65
read.file ...................................................... 67
recode ......................................................... 72
sai ............................................................. 73
sat.act ......................................................... 76
Schutz ......................................................... 78
selectBy ..................................................... 79
Spengler ..................................................... 80
spi ............................................................. 81
usaf ........................................................... 83
Utility ......................................................... 86
vegetables .................................................... 87
vJoin .......................................................... 89
zola ............................................................ 90

Index ......................................................... 92
ability

16 ability items scored as correct or incorrect.

Description

16 multiple choice ability items 1525 subjects taken from the Synthetic Aperture Personality Assessment (SAPA) web based personality assessment project are saved as iqitems. Those data are shown as examples of how to score multiple choice tests and analyses of response alternatives. When scored correct or incorrect, the data are useful for demonstrations of tetrachoric based factor analysis irt.fa and finding tetrachoric correlations.

Usage

data(iqitems)

Format

A data frame with 1525 observations on the following 16 variables. The number following the name is the item number from SAPA.

reason.4 Basic reasoning questions
reason.16 Basic reasoning question
reason.17 Basic reasoning question
reason.19 Basic reasoning question
letter.7 In the following alphanumeric series, what letter comes next?
letter.33 In the following alphanumeric series, what letter comes next?
letter.34 In the following alphanumeric series, what letter comes next
letter.58 In the following alphanumeric series, what letter comes next?
matrix.45 A matrix reasoning task
matrix.46 A matrix reasoning task
matrix.47 A matrix reasoning task
matrix.55 A matrix reasoning task
rotate.3 Spatial Rotation of type 1.2
rotate.4 Spatial Rotation of type 1.2
rotate.6 Spatial Rotation of type 1.1
rotate.8 Spatial Rotation of type 2.3
Details

16 items were sampled from 80 items given as part of the SAPA (https://www.sapa-project.org/) project (Revelle, Wilt and Rosenthal, 2009; Condon and Revelle, 2014) to develop online measures of ability. These 16 items reflect four lower order factors (verbal reasoning, letter series, matrix reasoning, and spatial rotations. These lower level factors all share a higher level factor (‘g’).

This data set may be used to demonstrate item response functions, tetrachoric correlations, or irt.fa as well as omega estimates of of reliability and hierarchical structure.

In addition, the data set is a good example of doing item analysis to examine the empirical response probabilities of each item alternative as a function of the underlying latent trait. When doing this, it appears that two of the matrix reasoning problems do not have monotonically increasing trace lines for the probability correct. At moderately high ability (theta = 1) there is a decrease in the probability correct from theta = 0 and theta = 2.

Source

The example data set is taken from the Synthetic Aperture Personality Assessment personality and ability test at https://www.sapa-project.org/. The data were collected with David Condon from 8/08/12 to 8/31/12.

Similar data are available from the International Cognitive Ability Resource at https://www.icar-project.org/.

References


Examples

data(ability)
cs<- psych::cs
keys <- list(ICAR16=colnames(ability),reasoning = cs(reason.4,reason.16,reason.17,reason.19),
             letters=cs(letter.7, letter.33,letter.34,letter.58),
             matrix=cs(matrix.45,matrix.46,matrix.47,matrix.55),
             rotate=cs(rotate.3,rotate.4,rotate.6,rotate.8))
psych:::scoreOverlap(keys,ability)
#this next step takes a few seconds to run and demonstrates IRT approaches
ability.irt <- psych:::irt.fa(ability)
ability.scores <- psych:::scoreIrt(ability.irt,ability)
affect

ability.sub.scores <- psych::scoreIrt.2pl(keys,ability) #demonstrate irt scoring

#It is sometimes asked how to handle missing data when finding scores
#this next example compares 3 ways of scoring ability items from icar
#Just sum the items
#Sum the means for the items
#IRT score the items

total <- rowSums(ability, na.rm=TRUE)
means <- rowMeans(ability, na.rm=TRUE)
irt <- psych::scoreIrt(items=ability)[1]
df <- data.frame(total, means,irt)
psych:: pairs.panels(df)

affect

Two data sets of affect and arousal scores as a function of personality and movie conditions

Description

A recurring question in the study of affect is the proper dimensionality and the relationship to various personality dimensions. Here is a data set taken from two studies of mood and arousal using movies to induce affective states.

Usage

data(affect)

Details

These are data from two studies conducted in the Personality, Motivation and Cognition Laboratory at Northwestern University. Both studies used a similar methodology:

Collection of pretest data using 5 scales from the Eysenck Personality Inventory and items taken from the Motivational State Questionnaire (see msq. In addition, state and trait anxiety measures were given. In the “maps” study, the Beck Depression Inventory was given also.

Then subjects were randomly assigned to one of four movie conditions: 1: Frontline. A documentary about the liberation of the Bergen-Belsen concentration camp. 2: Halloween. A horror film. 3: National Geographic, a nature film about the Serengeti plain. 4: Parenthood. A comedy. Each film clip was shown for 9 minutes. Following this the MSQ was given again.

Data from the MSQ were scored for Energetic and Tense Arousal (EA and TA) as well as Positive and Negative Affect (PA and NA).

Study flat had 170 participants, study maps had 160.

These studies are described in more detail in various publications from the PMC lab. In particular, Revelle and Anderson, 1997 and Rafaeli and Revelle (2006). An analysis of these data has also appeared in Smillie et al. (2012).
For a much more complete data set involving film, caffeine, and time of day manipulations, see the 
`msqR` data set.

**Source**

Data collected at the Personality, Motivation, and Cognition Laboratory, Northwestern University.

**References**


**Examples**

```r
data(affect)
paste::describeBy(affect[-1], group="Film")
paste::pairs.panels(affect[14:17], bg=c("red","black","white","blue") [affect$Film], pch=21,
    main="Affect varies by movies ")
paste::errorCircles("EA2","TA2", data=affect, group="Film", labels=c("Sad","Fear","Neutral","Humor")
    , main="Energetic and Tense Arousal by Movie condition")
paste::errorCircles(x="PA2",y="NA2", data=affect, group="Film", labels=c("Sad","Fear","Neutral","Humor")
    , main="Positive and Negative Affect by Movie condition")
```

**Athenstaedt**

*Gender Role Self Concept data from Athenstaedt (2003)*

**Description**

Athenstaedt (2003) examined Gender Role Self-Concept. She reports two independent dimensions of Male and Female behaviors. While there are large gender/sex differences on both of these dimensions, the two represent independent factorsl Eagly and Revelle (2022) have used these data to explore the power of aggregation when examining sex differences. This data set is also useful to show various graphical display procedures.

**Usage**

```r
data("Athenstaedt")
```
Format

A data frame with 576 observations on the following 117 variables.

STUDIE  a numeric vector

gender  Male =1, Female= 2

V1 - V74  self report items (see Athenstaedt.dictionary)

V1  Gender (Male = 1, Female =2)

V2  To pay attention to ones appearance in the office

V3  Offer fire to somebody

V4  Paint an Apartment

V5  Mow the Lawn

V6  Make the Bed

V7  Hold the Door Open for your Partner

V8  Do the Dishes

V9  Do Extreme Sports

V10  Tinker with the Car

V11  Talk about Sports

V12  Assemble Prefabricated Furniture

V13  Drive a Car in a Risky Way

V14  Listen Attentively to Others

V15  Tell your Partner about Problems at Work

V16  Play on a Computer

V17  Set the Table

V18  Watch ones Weight

V19  Care for a Partner if he/she is Ill

V20  Play Chess

V21  Meet with friends at a Regulars Table

V22  Watch Soap Operas

V23  Take a Friends Arm

V24  Wrap Presents Beautifully

V25  In case of Vacation with Partner Packing the Luggage for Both

V26  To admit own Occupational Weakness

V27  Work Overtime

V28  Openly Show Vulnerability

V29  Babysit

V30  Change Fuses

V31  Clean a Drain

V32  Take Care of Somebody
V33 Do Repair Work
V34 Change Light Bulbs
V35 Wash the Car
V36 Ride a Motorcycle
V37 Cook Meat on the Grill
V38 Thump Carpets
V39 Dust the Furniture
V40 Buy Electric Appliances
V41 Go Dancing
V42 Go for a Walk through Town
V43 Go to the Ballet
V44 Hug a Friend
V45 Do Handiwork (e.g. Knitting)
V46 Change Bed Sheets
V47 Sew on a Button
V48 Do Aerobics
V49 Watch Sports on Television
V50 Talk about Problems
V51 Play Parlor Games
V52 Talk about Politics
V53 Take Care of Flowers
V54 Make Coffee in the Office
V55 Shovel Snow
V56 Read non-Fiction Books
V57 Organize Company Parties
V58 Do Home Improvement Jobs
V59 Plead for the Socially Disadvantaged
V60 Buy a Present for a Colleague
V61 To Talk with Colleagues about Family Matters
V62 Make Jam
V63 Frequently Ask Colleagues Questions
V64 Decorate the Office with Flowers
V65 Pick up the Dinner Bill
V66 Shop for the Family
V67 Have Problem using Technical Devices
V68 Care for Family Besides a Job
V69 Watch Action Movies
V70 Cook
V71 Help your Partner Put on His or Her Coat
V72 Wash Windows
V73 Do the Ironing
V74 Do the Laundry
V75 Put on Make-up
V76 Femininity Scale
V77 Masculinity Scale
V78 Femininity Scale
V79 Masculinity Scale
V80 Pooled Scale

MMINUS1 - MPLUS see the original Athenstaedt paper

FBEHAV a numeric vector
MBEHAV a numeric vector
Femininity a numeric vector
Masculinity a numeric vector
MF a numeric vector

Details

Ursala Athenstaedt (2003) reported several analyses of items and scales measuring Gender Role Self-Concept. Eagly and Revelle (2022) have used these data in an analysis of the power of aggregation. Here are the original items as well as the three scales Eagly and Revelle (2022). The accompanying Athenstaedt.dictionary may be used to see the items.

See the GERAS data set for a related example.

Source

Ursala Athenstaedt, personal communication, 2022, provided a SPSS sav file with the original data from which the complete cases in this set were selected.

References


Examples

```r
data(Athenstaedt)
psych::scatterHist(Femininity ~ Masculinity + gender, data = Athenstaedt,
cex.point=.4, smooth=FALSE, correl=FALSE, d.arrow=TRUE, col=c("red","blue"),
lwd=4, cex.main=1.5, main="Scatter Plot and Density", cex.axis=2)
psych::cohen.d(Athenstaedt[2:76], group="gender", dictionary=Athenstaedt.dictionary)
#show the top 5 items for each scale
select <- c(psych::selectFromKeys(Athenstaedt.keys$MF10), "gender")
psych::corPlot(Athenstaedt[, select], main="F and M items from Athenstaedt")
```

---

**bfi**

### 25 Personality items representing 5 factors

**Description**

25 personality self report items taken from the International Personality Item Pool (ipip.ori.org) were included as part of the Synthetic Aperture Personality Assessment (SAPA) web based personality assessment project. The data from 2800 subjects are included here as a demonstration set for scale construction, factor analysis, and Item Response Theory analysis. Three additional demographic variables (sex, education, and age) are also included.

**Usage**

```r
data(bfi)
data(bfi.dictionary)
```

**Format**

A data frame with 2800 observations on the following 28 variables. (The q numbers are the SAPA item numbers).

A1 Am indifferent to the feelings of others. (q_146)
A2 Inquire about others’ well-being. (q_1162)
A3 Know how to comfort others. (q_1206)
A4 Love children. (q_1364)
A5 Make people feel at ease. (q_1419)
C1 Am exacting in my work. (q_124)
C2 Continue until everything is perfect. (q_530)
C3 Do things according to a plan. (q_619)
C4 Do things in a half-way manner. (q_626)
C5 Waste my time. (q_1949)
E1 Don’t talk a lot. (q_712)
The first 25 items are organized by five putative factors: Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness. The scoring key is created using `make.keys`, the scores are found using `score.items`.

These five factors are a useful example of using `irt.fa` to do Item Response Theory based latent factor analysis of the polychoric correlation matrix. The endorsement plots for each item, as well as the item information functions reveal that the items differ in their quality.

The item data were collected using a 6 point response scale: 1 Very Inaccurate 2 Moderately Inaccurate 3 Slightly Inaccurate 4 Slightly Accurate 5 Moderately Accurate 6 Very Accurate as part of the Synthetic Aperture Personality Assessment (SAPA [https://www.sapa-project.org/]) project. To see an example of the data collection technique, visit [https://www.SAPA-project.org/] or the International Cognitive Ability Resource at [https://icar-project.org]. The items given were sampled from the International Personality Item Pool of Lewis Goldberg using the sampling technique of SAPA. This is a sample data set taken from the much larger SAPA data bank.

**Note**


**Source**

The items are from the ipip (Goldberg, 1999). The data are from the SAPA project (Revelle, Wilt and Rosenthal, 2010) , collected Spring, 2010 ([https://www.sapa-project.org/]).
BFI.adjectives.dictionary

References


See Also

bi.bars to show the data by age and gender, irt.fa for item factor analysis applying the irt model.

Examples

data(bfi)
psych::describe(bfi)
# create the bfi.keys (actually already saved in the data file)
bfi.keys <-
  list(agree=c("-A1","A2","A3","A4","A5"), conscientious=c("C1","C2","C3","-C4","-C5"),
    extraversion=c("-E1","-E2","E3","E4","E5"), neuroticism=c("N1","N2","N3","N4","N5"),
    openness = c("O1","-O2","O3","O4","-O5"))

scores <- psych::scoreItems(bfi.keys,bfi,min=1,max=6) #specify the minimum and maximum values
scores
#show the use of the keys.lookup with a dictionary
psych::keys.lookup(bfi.keys,bfi.dictionary[,1:4])

BFI.adjectives.dictionary

Dictionary for the 100 Big Five Adjectives

Description

Lew Goldberg organized 100 adjectives to measure 5 factors of personality (The Big5). 500 hundred participants were given these adjectives along with other personality measures. This dictionary allows for easy item labeling of the results. ~

Usage

data("BFI.adjectives.dictionary")
Format

A data frame with 100 observations on the following 2 variables.

- `numer` a character vector of the item label
- `Item` a character vector of the actual adjectives

Details

Keying information for the 100 adjectives:

Source

Data collected at the Personality, Motivation, and Cognition Laboratory, Northwestern University.

References


See Also

- `big5.100.adjectives` for examples of the data.
- `msqR` for 3896 participants with scores on five scales of the EPI.
- `affect` for an example of the use of some of these adjectives in a mood manipulation study.

Examples

```r
data(BFI.adjectives.dictionary) # this includes the bfi.adjectives.keys
bfi.adjectives.keys <- list(
  Agreeableness = psych::cs(V2, -V11, V14, V15, -V19, -V21, V29, -V31, V32, V48, V55, -V61, -V63, V69, V76, -V78, -V79, -V90, -V94, V99),
  Conscientiousness = psych::cs(V9, V10, V13, -V20, V22, -V30, V37, -V38, -V39, V50, -V51, V53, V56, V57, -V67, V68, V70, V73, -V82, -V95),
  Extraversion = psych::cs(V1, V5, -V6, V7, V17, V24, V26, -V40, -V45, -V58, -V60, -V65, V71, -V74, -V77, V92, -V96, V97, V98, -V100),
  Openness = psych::cs(V4, V8, V12, V16, V18, V34, -V35, V41, V43, V44, V54, -V64, -V66, -V80, -V86, -V87, -V88, -V89, -V91, -V93)
)
psych::lookupFromKeys(bfi.adjectives.keys, bfi.adjectives.dictionary, 20)
```
Description

Lew Goldberg organized 100 adjectives to measure 5 factors of personality (The Big 5). 500 hundred participants were given these adjectives along with other personality measures in the Personality, Motivation and Cognition (PMC) lab. This data set is for demonstrations of factor and cluster analysis.

Usage

data("big5.100.adjectives")

Format

A data frame with 554 observations on the following 102 variables.

- study: a character vector
- id: a numeric vector
- V1: numeric vector (see big5.adjectives.dictionary)
- V100: A numeric vector. (see big5.adjectives.dictionary)
- bfi.adjectives.keys: a key list

Details

Procedure. The data were collected over nine years in the Personality, Motivation and Cognition laboratory at Northwestern, as part of a series of studies examining the effects of personality and situational factors on motivational state and subsequent cognitive performance. In each of 38 studies, prior to any manipulation of motivational state, participants signed a consent form and in some studies, consumed 0 or 4mg/kg of caffeine. In caffeine studies, they waited 30 minutes and then filled out the MSQ as well as other personality trait measures (e.g. the Big 5 adjectives).

Source

Data collected at the Personality, Motivation, and Cognition Laboratory, Northwestern University.

References


Examples

data(big5.100.adjectives)
five.scores <- psych::scoreItems(big5.adjectives.keys,big5.100.adjectives)
sweep(five.scores)

blant

A 29 x 29 matrix that produces weird factor analytic results

Description

Normally, min.res factor analysis and maximum likelihood produce very similar results. This data set (from Alexandra Blant) does not. Warnings are given for the min.res solution, the pa solution, but not the old.min nor the mle solution. Included as a test case for the factor analysis function.

Usage

data("blant")

Format

The format is: num [1:29, 1:29] 1 0.77 0.81 0.68 0.717 ... - attr(*, "dimnames")=List of 2 ..$: NULL ..$: chr [1:29] "V1" "V2" "V3" "V4" ...

Details

This data matrix was sent by Alexandra Blant as an example of a problem with the minres solution in the fa function. The default solution, using fm="minres" issues a warning that the solution has improper factor score weights. This is not the case for the fm="old.min" and fm="mle" options, but is for fm="pa", fm="ols".

The residuals are indeed smaller for fm="minres" than for fm="old.min" or fm="mle".

"old.min" attempts to find the minimum residual but uses the gradient for mle. This was the approach until version 1.7.5 but was changed (see the help page for fa) following extensive communication with Hao Wu.

The problem with this matrix is probably that it is almost singular, with some smcs approaching 1 and the smallest three eigenvalues of .006, .004 and .001.

This problem matrix was provided by Alexandra Blant.

Source

Alexandra Blant, personal communication
Examples

data(blant)

# compare
f5 <- psych::fa(blant,5,rotate="none")  # the default minres
f5.old <- psych::fa(blant,5, fm="old.min",rotate="none")  # old version of minres
f5.mle <- psych::fa(blant,5, fm="mle",rotate= "none")  # maximum likelihood

# compare solutions
psych::factor.congruence(list(f5,f5.old,f5.mle))

# compare sums of squared residuals
sum(residuals(f5,diag=FALSE)^2,na.rm=TRUE)  # 1.355489
sum(residuals(f5.old,diag=FALSE)^2,na.rm=TRUE)  # 1.539757
sum(residuals(f5.mle,diag=FALSE)^2,na.rm=TRUE)  # 2.402092

# but, when we divide the squared residuals by the original (squared) correlations, we find
# a different ordering of fit
f5$fit  # 0.9748177
f5.old$fit  # 0.9752774
f5.mle$fit  # 0.9603324

blot

Bond’s Logical Operations Test – BLOT

Description

35 items for 150 subjects from Bond’s Logical Operations Test. A good example of Item Response Theory analysis using the Rasch model. One parameter (Rasch) analysis and two parameter IRT analyses produce somewhat different results.

Usage

data(blot)

Format

A data frame with 150 observations on 35 variables. The BLOT was developed as a paper and pencil test for children to measure Logical Thinking as discussed by Piaget and Inhelder.

Details

Bond and Fox apply Rasch modeling to a variety of data sets. This one, Bond’s Logical Operations Test, is used as an example of Rasch modeling for dichotomous items. In their text (p 56), Bond and Fox report the results using WINSTEPS. Those results are consistent (up to a scaling parameter) with those found by the rasch function in the ltm package. The WINSTEPS seem to produce difficulty estimates with a mean item difficulty of 0, whereas rasch from ltm has a mean difficulty of -1.52. In addition, rasch seems to reverse the signs of the difficulty estimates when reporting the coefficients and is effectively reporting "easiness".

However, when using a two parameter model, one of the items (V12) behaves very differently.

This data set is useful when comparing 1PL, 2PL and 2PN IRT models.
Source

The data are taken (with kind permission from Trevor Bond) from the webpage https://www.winsteps.com/BF3/bondfox3.htm and read using read.fwf.

References


See Also

See also the \texttt{irt.fa} and associated plot functions.

Examples

```r
data(blot)

# ltm is not required by psychTools, but if available, may be run to show a Rasch model

# do the same thing with functions in psych
blot.fa <- psych::irt.fa(blot)  # a 2PN model
plot(blot.fa)
```

11 emotional variables from Burt (1915)

Description

Cyril Burt reported an early factor analysis with a circumplex structure of 11 emotional variables in 1915. 8 of these were subsequently used by Harman in his text on factor analysis. Unfortunately, it seems as if Burt made a mistake for the matrix is not positive definite. With one change from .87 to .81 the matrix is positive definite.

Usage

data(burt)

Format

A correlation matrix based upon 172 "normal school age children aged 9-12".

Sociality  Sociality
Sorrow    Sorrow
The Burt data set is interesting for several reasons. It seems to be an early example of the organization of emotions into an affective circumplex, a subset of it has been used for factor analysis examples (see Harman.Burt, and it is an example of how typos affect data. The original data matrix has one negative eigenvalue. With the replacement of the correlation between Sorrow and Tenderness from .87 to .81, the matrix is positive definite.

Alternatively, using cor.smooth, the matrix can be made positive definite as well, although cor.smooth makes more (but smaller) changes.

Source

(retrieved from the web at https://www.biodiversitylibrary.org/item/95822#790) Following a suggestion by Jan DeLeeuw.

References

Burt, C. General and Specific Factors underlying the Primary Emotions. Reports of the British Association for the Advancement of Science, 85th meeting, held in Manchester, September 7-11, 1915. London, John Murray, 1916, p. 694-696 (retrieved from the web at https://www.biodiversitylibrary.org/item/95822#790)

See Also

Harman.Burt in the Harman dataset and cor.smooth

Examples

data(burt)
eigen(burt)$values #one is negative!
burt.new <- burt
eigen(burt.new)$values #all are positive
bs <- psych::cor.smooth(burt)
round(burt.new - bs,3)
**Description**

Airline distances between 11 US cities may be used as an example for multidimensional scaling or cluster analysis.

**Usage**

```r
data(cities)
```

**Format**

A data frame with 11 observations on the following 11 variables.

- **ATL** Atlanta, Georgia
- **BOS** Boston, Massachusetts
- **ORD** Chicago, Illinois
- **DCA** Washington, District of Columbia
- **DEN** Denver, Colorado
- **LAX** Los Angeles, California
- **MIA** Miami, Florida
- **JFK** New York, New York
- **SEA** Seattle, Washington
- **SFO** San Francisco, California
- **MSY** New Orleans, Louisiana

**Details**

An 11 x 11 matrix of distances between major US airports. This is a useful demonstration of multiple dimensional scaling.

*city.location* is a dataframe of longitude and latitude for those cities.

Note that the 2 dimensional MDS solution does not perfectly capture the data from these city distances. Boston, New York and Washington, D.C. are located slightly too far west, and Seattle and LA are slightly too far south.

**Source**

https://www.timeanddate.com/worldclock/distance.html
**Examples**

```r
data(cities)
city.location[,1] <- -city.location[,1] # included in the cities data set
plot(city.location, xlab="Dimension 1", ylab="Dimension 2",
     main ="Multidimensional scaling of US cities")
# do the mds
city.loc <- cmdscale(cities, k=2) # ask for a 2 dimensional solution
round(city.loc,0)

# flip the axes
city.loc <- -city.loc

# do the mds
city.loc <- psych::rescale(city.loc, apply(city.location,2,mean), apply(city.location,2,sd))
points(city.loc, type="n") # add the date point to the map

# do the mds
text(city.loc, labels=names(cities))

## Not run: # we need the maps package to be available
# an overlay map can be added if the package maps is available
if(require(maps)) {

text(city.loc, labels=names(cities))
}
## End(Not run)
```

---

**Colom**

*Correlations of 14 ability tests from the Spanish version of the WAIS (taken from Colom et al. 2002.)*

---

**Description**

Colom et al. analyze 14 tests from the Spanish version of the WAIS. This is a nice example of a hierarchical structure using the `omega` function. Here are the correlation matrices of the variables (colom), for 4 levels of education.

**Usage**

```r
data("colom")
data("colom.ed0")
data("colom.ed1")
data("colom.ed2")
data("colom.ed3")
```

**Format**

The format is: num [1:14, 1:14] 1 0.755 0.608 0.555 0.715 0.729 0.627 0.616 0.606 0.598 ... - attr(*, "dimnames")=List of 2 ..$ : chr [1:14] "Vocabulary" "Similarities" "Arithmetic" "Digit_span" ...

..$ : chr [1:14] "Vocabulary" "Similarities" "Arithmetic" "Digit_span" ...

```
The Wechsler Adult Intelligence Scale (WAIS) is the "gold standard" measure of intelligence. Here is an example of the correlational structure of 14 tests. It was used by Colom and his colleagues to find correlations of WAIS scores as a function of education. Here we show the complete standardization sample.

The `colom` data set is the complete correlation matrix for all subjects (703 females, 666 males). The four subset data sets for four levels of education. Ns = 301, 432, 525, and 111.

Source

Colom et al, 2002

References


Examples

data(colom)
psych::lowerMat(colom)
psych::omega(colom, 4)  #do the omega analysis

cubits

Galton's example of the relationship between height and 'cubit' or forearm length

Description

Francis Galton introduced the 'co-relation' in 1888 with a paper discussing how to measure the relationship between two variables. His primary example was the relationship between height and forearm length. The data table (cubits) is taken from Galton (1888). Unfortunately, there seem to be some errors in the original data table in that the marginal totals do not match the table.

The data frame, heights, is converted from this table.

Usage

data(cubits)

Format

A data frame with 9 observations on the following 8 variables.

<table>
<thead>
<tr>
<th>Cubit length</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 16.5</td>
<td>16.5</td>
</tr>
<tr>
<td>16.5 &lt;= &lt; 17.0</td>
<td>16.75</td>
</tr>
<tr>
<td>17.0 &lt;= &lt; 17.5</td>
<td>17.25</td>
</tr>
</tbody>
</table>

16.5 Cubit length < 16.5
16.75 16.5 <= Cubit length < 17.0
17.25 17.0 <= Cubit length < 17.5
22

A data set from Cushny and Peebles (1905) on the effect of three drugs on hours of sleep, used by Student (1908)

Description

The classic data set used by Gossett (publishing as Student) for the introduction of the t-test. The design was a within subjects study with hours of sleep in a control condition compared to those in 3 drug conditions. Drug1 was 00mg of L Hscyamine, Drug 2L and Drug2R were said to be .6 mg of Left and Right isomers of Hyoscine. As discussed by Zabell (2008) these were not optical isomers. The detail1, delta2L and delta2R are changes from the baseline control.

Details

Sir Francis Galton (1888) published the first demonstration of the correlation coefficient. The regression (or reversion to mediocrity) of the height to the length of the left forearm (a cubit) was found to .8. There seem to be some errors in the table as published in that the row sums do not agree with the actual row sums. These data are used to create a matrix using `table2matrix` for demonstrations of analysis and displays of the data.

Source

Galton (1888)

References

Galton, Francis (1888) Co-relations and their measurement. Proceedings of the Royal Society. London Series,45,135-145,

See Also

table2matrix, table2df, ellipses, heights, peas, galton

Examples

data(cubits)
cubits
heights <- psych::table2df(cubits,labs = c("height","cubit"))
psych::ellipses(heights,n=1,main="Galton's co-relation data set")
psych::ellipses(jitter(heights$height,3),jitter(heights$cubit,3),pch=".",
main="Galton's co-relation data set",xlab="height",
ylab="Forearm (cubit)") #add in some noise to see the points
psych::pairs.panels(heights,jiggle=TRUE,main="Galton's cubits data set")
cushny

Usage
data(cushny)

Format
A data frame with 10 observations on the following 7 variables.

Control Hours of sleep in a control condition
drug1 Hours of sleep in Drug condition 1
drug2L Hours of sleep in Drug condition 2
drug2R Hours of sleep in Drug condition 3 (an isomer of the drug in condition 2
delta1 Change from control, drug 1
delta2L Change from control, drug 2L
delta2R Change from control, drug 2R

Details
The original analysis by Student is used as an example for the t-test function, both as a paired t-test and a two group t-test. The data are also useful for a repeated measures analysis of variance.

Source
Student (1908) The probable error of the mean. Biometrika, 6 (1), 1-25.

References
See also the data set sleep and the examples for the t.test

Examples
data(cushny)
with(cushny, t.test(drug1,drug2L,paired=TRUE)) #within subjects
psych::error.bars(cushny[1:4],within=TRUE,ylab="Hours of sleep",xlab="Drug condition", main="95% confidence of within subject effects")
df2latex  

Convert a data frame, correlation matrix, or factor analysis output to a LaTeX table

Description

A set of handy helper functions to convert data frames or matrices to LaTeX tables. Although Sweave is the preferred means of converting R output to LaTeX, it is sometimes useful to go directly from a data.frame or matrix to a LaTeX table. cor2latex will find the correlations and then create a lower (or upper) triangular matrix for latex output. fa2latex will create the latex commands for showing the loadings and factor intercorrelations. As the default option, tables are prepared in an approximation of APA format.

Usage

df2latex(x,digits=2,rowlabels=TRUE,apa=TRUE,short.names=TRUE,font.size = "scriptsize", big.mark=NULL,drop.na=TRUE, heading="A table from the psych package in R", caption="df2latex",label="default", char=FALSE, stars=FALSE,silent=FALSE,file=NULL,append=FALSE,cut=0,big=0,abbrev=NULL,drop.na=TRUE,heading="A correlation table from the psych package in R.", caption="cor2latex",label="default",silent=FALSE,file=NULL,append=FALSE,cut=0,big=0)
cor2latex(x,use = "pairwise", method="pearson", adjust="holm",stars=FALSE, digits=2,rowlabels=TRUE,lower=TRUE,apa=TRUE,short.names=TRUE, font.size = "scriptsize", heading="A correlation table from the psych package in R.", caption="cor2latex",label="default",silent=FALSE,file=NULL,append=FALSE,cut=0,big=0)
fa2latex(f,digits=2,rowlabels=TRUE,apa=TRUE,short.names=FALSE,cumvar=FALSE, cut=0,big=.3,alpha=.05,font.size = "scriptsize",long=FALSE, heading="A factor analysis table from the psych package in R", caption="fa2latex",label="default",silent=FALSE,file=NULL,append=FALSE)
omega2latex(f,digits=2,rowlabels=TRUE,apa=TRUE,short.names=FALSE,cumvar=FALSE,cut=.2, big=.3,font.size = "scriptsize", heading="An omega analysis table from the psych package in R", caption="omega2latex",label="default",silent=FALSE,file=NULL,append=FALSE)
irt2latex(f,digits=2,rowlabels=TRUE,apa=TRUE,short.names=FALSE, font.size = "scriptsize", heading="An IRT factor analysis table from R", caption="fa2latex",label="default",silent=FALSE,file=NULL,append=FALSE)
ICC2latex(icc,digits=2,rowlabels=TRUE,apa=TRUE,ci=TRUE, font.size = "scriptsize",big.mark=NULL, drop.na=TRUE, heading="A table from the psych package in R", caption="ICC2latex",label="default",char=FALSE,silent=FALSE,file=NULL,append=FALSE)

Arguments

x A data frame or matrix to convert to LaTeX. If non-square, then correlations will be found prior to printing in cor2latex
digits Round the output to digits of accuracy. NULL for formatting character data
abbrev How many characters should be used in column names –defaults to digits + 3
rowlabels If TRUE, use the row names from the matrix or data.frame
short.names Name the columns with abbreviated rownames to save space
apa If TRUE formats table in APA style
cumvar For factor analyses, should we show the cumulative variance accounted for?
font.size e.g., "scriptsize", "tiny" or any other acceptable LaTeX font size.
heading The label appearing at the top of the table
caption The table caption
lower In cor2latex, just show the lower triangular matrix
f The object returned from a factor analysis using fa or irt.fa.
label The label for the table
big.mark Comma separate numbers large numbers (big.mark=",")
drop.na Do not print NA values
method When finding correlations, which method should be used (pearson)
use use="pairwise" is the default when finding correlations in cor2latex
adjust If showing probabilities, which adjustment should be used (holm)
stars Should probability 'magic astericks' be displayed in cor2latex (FALSE)
char char=TRUE allows printing tables with character information, but does not allow for putting in commas into numbers
cut In omega2latex, df2latex and fa2latex, do not print abs(values) < cut
big In fa2latex and df2latex boldface those abs(values) > big
alpha If fa has returned confidence intervals, then what values of loadings should be boldfaced?
icc Either the output of an ICC, or the data to be analyzed.
ci Should confidence intervals of the ICC be displayed
silent If TRUE, do not print any output, just return silently – useful if using Sweave
file If specified, write the output to this file
append If file is specified, then should we append (append=TRUE) or just write to the file
long if TRUE, then do long tables. (requires the longtables package in latex)

Value

A LaTeX table. Note that if showing "stars" for correlations, then one needs to use the siunitx package in LaTeX. The entire LaTeX output is also returned invisibly. If using Sweave to create tables, then the silent option should be set to TRUE and the returned object saved as a file. See the last example.

Finally, some users have asked for the ability to convert these output tables into HTML. This may be done using the tth package.

Author(s)

William Revelle with suggestions from Jason French and David Condon and Davide Morselli
See Also

The many LaTeX conversion routines in Hmisc.

To convert these LaTeX objects to HTML, you should install the tth package.

Consider the last example

Examples

df2latex(psych::Thurstone,rowlabels=FALSE,apa=FALSE,short.names=FALSE, caption="Thurstone Correlation matrix")
df2latex(psych::Thurstone,heading="Thurstone Correlation matrix in APA style")

df2latex(psych::describe(psych::sat.act)[2:10],short.names=FALSE)
cor2latex(psych::Thurstone)
cor2latex(psych::sat.act,short.names=FALSE)
fa2latex(psych::fa(psych::Thurstone,3),heading="Factor analysis from R in quasi APA style")

#To convert these latex tables to HTML

#f3.lat <- fa2latex(psych::fa(psych::Thurstone,3),
#  heading="Factor analysis from R in quasi APA style")
#library(tth)
#f3.ht <- tth(f3.lat)
#print(as.data.frame(f3.ht),row.names=FALSE)

###

#If using Sweave to create a LateX table as a separate file then set silent=TRUE
#e.g.,
#LaTeX preamble
#....
#<<print=FALSE,echo=FALSE>>=
#f3 <- fa(Thurstone,3)
#fa2latex(f3,silent=TRUE,file='testoutput.tex')
#@
#
\input{testoutput.tex}

dfOrder Sort (order) a dataframe or matrix by multiple columns

Description

Although order will order a vector, and it is possible to order several columns of a data.frame by specifying each column individually in the call to order, dfOrder will order a dataframe or matrix by as many columns as desired. The default is to sort by columns in lexicographic order. If the
object is a correlation matrix, then the selected columns are sorted by the (abs) max value across the columns (similar to fa.lookup in psych). If object is a correlation matrix, rows and columns are sorted.

Usage

dfOrder(object, columns, absolute=FALSE, ascending=TRUE)

Arguments

object The data.frame or matrix to be sorted
columns Column numbers or names to use for sorting. If positive, then they will be sorted in increasing order. If negative, then in decreasing order
absolute If TRUE, then sort the absolute values
ascending By default, order from smallest to largest.

Details

This is just a simple helper function to reorder data.frames and correlation matrices. Originally developed to organize IRT output from the ltm package. It is a basic add on to the order function. (Completely rewritten for version 1.8.1. and then again for 2.2.1 to allow sorting correlation matrices by numeric values.)

Value

The original data frame is now in sorted order. If the input is a correlation matrix, the output is sorted by rows and columns.

Author(s)

William Revelle

See Also

Other useful file manipulation functions include read.file to read in data from a file or read.clipboard from the clipboard, fileScan, filesList, filesInfo, and fileCreate

dfOrder code is used in the test.irt function to combine ltm and sim.irt output.

Examples

# create a data frame and then sort it in lexicographic order
set.seed(42)
x <- matrix(sample(1:4,64,replace=TRUE),ncol=4)
dfOrder(x) # sort by all columns
dfOrder(x,c(1,4)) # sort by the first and 4th column
x.df <- data.frame(x)
dfOrder(x.df,c(1,-2)) # sort by the first in increasing order,
# the second in decreasing order
# now show sorting correlation matrices
r <- cor(sat.act, use="pairwise")
r.ord <- dfOrder(r, columns=c("education", "ACT"), ascending=FALSE)
psych::corPlot(r.ord)

### eminence

**Eminence of 69 American Psychologists**

**Description**

Marco Del Giudice criticized an earlier study by Simonton for using partial regression weights to estimate the importance of various predictors of rated eminence. This is a nice example of the (mis)interpretation of beta weights of highly correlated predictors.

**Usage**

data("eminence")

**Format**

A data frame with 69 observations on the following 9 variables.

- **name**: a character vector
- **reputation**: Log of rated reputation
- **birth.year**: Year of birth
- **first.year**: Year of first cited publication
- **last.year**: Year of last cited publication
- **works**: Log of number of publications
- **citations**: Log of number of citations
- **composite**: A composite index of publications
- **h**: The ’h’ index of citations

**Details**

Simonton (1997, 2014) discusses various estimates of eminence among 69 psychologists born between 1842 and 1912 and reports that the regression weights are small and interprets this as meaning number of publications and citations are not very important. Del Giudice (2020) points out that citations and the number of publications are highly collinear and thus while their independent contributions are small, their joint effect is quite large (R = .69). These data are given here as an example of multiple correlation and partial correlation.

**Source**

Del Giudice (2020) links to a web page with the data.
References


Examples

data(eminence)
psych::lowerCor(eminence)
cs <- psych::cs
psych::partial.r(eminence, x= cs(reputation, works, citations),y=cs(birth.year))
psych::setCor(reputation ~ works + h + first.year,data=eminence)

Description

The EPI is and has been a very frequently administered personality test with 57 measuring two broad dimensions, Extraversion-Introversion and Stability-Neuroticism, with an additional Lie scale. Developed by Eysenck and Eysenck, 1964. Eventually replaced with the EPQ which measures three broad dimensions. This data set represents 3570 observations collected in the early 1990s at the Personality, Motivation and Cognition lab at Northwestern. An additional data set (epiR) has test and retest information for 474 participants. The data are included here as demonstration of scale construction and test-retest reliability.

Usage

data(epi)
data(epi.dictionary)
data(epiR)

Format

A data frame with 3570 observations on the following 57 variables.

id The identification number within the study
time First (group testing) or 2nd time (before a lab experiment) for the epiR data set.
study Four lab based studies and their pretest data
V1 a numeric vector
V2 a numeric vector
V3 a numeric vector
V4 a numeric vector  
V5 a numeric vector  
V6 a numeric vector  
V7 a numeric vector  
V8 a numeric vector  
V9 a numeric vector  
V10 a numeric vector  
V11 a numeric vector  
V12 a numeric vector  
V13 a numeric vector  
V14 a numeric vector  
V15 a numeric vector  
V16 a numeric vector  
V17 a numeric vector  
V18 a numeric vector  
V19 a numeric vector  
V20 a numeric vector  
V21 a numeric vector  
V22 a numeric vector  
V23 a numeric vector  
V24 a numeric vector  
V25 a numeric vector  
V26 a numeric vector  
V27 a numeric vector  
V28 a numeric vector  
V29 a numeric vector  
V30 a numeric vector  
V31 a numeric vector  
V32 a numeric vector  
V33 a numeric vector  
V34 a numeric vector  
V35 a numeric vector  
V36 a numeric vector  
V37 a numeric vector  
V38 a numeric vector  
V39 a numeric vector  
V40 a numeric vector
V41 a numeric vector
V42 a numeric vector
V43 a numeric vector
V44 a numeric vector
V45 a numeric vector
V46 a numeric vector
V47 a numeric vector
V48 a numeric vector
V49 a numeric vector
V50 a numeric vector
V51 a numeric vector
V52 a numeric vector
V53 a numeric vector
V54 a numeric vector
V55 a numeric vector
V56 a numeric vector
V57 a numeric vector

Details
The original data were collected in a group testing framework for screening participants for subsequent studies. The participants were enrolled in an introductory psychology class between Fall, 1991 and Spring, 1995.

The actual items may be found in the epi.dictionary.

The structure of the E scale has been shown by Rocklin and Revelle (1981) to have two subcomponents, Impulsivity and Sociability. These were subsequently used by Revelle, Humphreys, Simon and Gilliland (1980) to examine the relationship between personality, caffeine induced arousal, and cognitive performance.

The epiR data include the original group testing data and matched data for 474 participants collected several weeks later. This is useful for showing that internal consistency estimates (e.g. \textit{alpha} or \textit{omega}) can be low even though the test is stable across time. For more demonstrations of the distinction between immediate internal consistency and delayed test-retest reliability see the \textit{msqR} and \textit{sai} data sets and \textit{testRetest}.

Source
Data from the PMC laboratory at Northwestern.

References

Examples

data(epi)
L = c("V6", "V24", "V36", "V12", "V18", "V30", "V42", "V48", "V54"),
Imp = c("V1", "V3", "V8", "V10", "V13", "V22", "V39", "V5", "V41"),
)
scores <- psych::scoreItems(epi.keys, epi)
psych::keys.lookup(epi.keys[1:3], epi.dictionary) #show the items and keying information
# a variety of demonstrations (not run) of test retest reliability versus alpha versus omega
E <- psych::selectFromKeys(epi.keys$E)
# look at the testRetest help file for more examples

epi.bfi

13 personality scales from the Eysenck Personality Inventory and Big 5 inventory

description

A small data set of 5 scales from the Eysenck Personality Inventory, 5 from a Big 5 inventory, a Beck Depression Inventory, and State and Trait Anxiety measures. Used for demonstrations of correlations, regressions, graphic displays.

Usage

data(epi.bfi)

Format

A data frame with 231 observations on the following 13 variables.
epiE  EPI Extraversion
epiS  EPI Sociability (a subset of Extraversion items
epiImp  EPI Impulsivity (a subset of Extraversion items
epilie  EPI Lie scale
epiNeur  EPI neuroticism
bfagree  Big 5 inventory (from the IPIP) measure of Agreeableness
bfcon  Big 5 Conscientiousness
Details

Self report personality scales tend to measure the “Giant 2” of Extraversion and Neuroticism or the “Big 5” of Extraversion, Neuroticism, Agreeableness, Conscientiousness, and Openness. Here is a small data set from Northwestern University undergraduates with scores on the Eysenck Personality Inventory (EPI) and a Big 5 inventory taken from the International Personality Item Pool.

Source

Data were collected at the Personality, Motivation, and Cognition Lab (PMCLab) at Northwestern by William Revelle.

References

https://personality-project.org/pmc.html

Examples

```r
data(epi.bfi)
psych::pairs.panels(epi.bfi[,1:5])
psych::describe(epi.bfi)
```

---

### galton

*Galton’s Mid parent child height data*

**Description**

Two of the earliest examples of the correlation coefficient were Francis Galton’s data sets on the relationship between mid parent and child height and the similarity of parent generation peas with child peas. This is the data set for the Galton height.

**Usage**

```r
data(galton)
```

**Format**

A data frame with 928 observations on the following 2 variables.

- **parent**  Mid Parent heights (in inches)
- **child**   Child Height
Details

Female heights were adjusted by 1.08 to compensate for sex differences. (This was done in the original data set)

Source

This is just the galton data set from UsingR, slightly rearranged.

References


See Also

The other Galton data sets: heights, peas, cubits

Examples

data(galton)
psych::describe(galton)
  #show the scatter plot and the lowess fit
psych::pairs.panels(galton, main="Galton's Parent child heights")
  #but this makes the regression lines look the same
psych::pairs.panels(galton, lm=TRUE, main="Galton's Parent child heights")
  #better is to scale them
psych::pairs.panels(galton, lm=TRUE, xlim=c(62,74), ylim=c(62,74),
       main="Galton's Parent child heights")

GERAS

Data from Gruber et al, 2020, Study 2: Gender Related Attributes Survey

Description

Gruber et al. (2020) report on the psychometric properties of a multifaceted Gender Related Attributes Survey. Here are the data from their 3 domains (Personality, Cognition and Activities and Interests from their study 2. Eagly and Revelle (2022) include these data in their review of the power of aggregation. The data are included here as demonstrations of the cohen.d and scatterHist functions in the psych package and may be used to show the power of aggregation.
Usage

```r
data("GERAS")
# These other objects are included in the file
# data("GERAS.scales")
# data("GERAS.dictionary")
# data("GERAS.items")
# data("GERAS.keys")
```

Format

A data frame with 471 observations on the following 51 variables (selected from the original 93). The code numbers are item numbers from the bigger set.

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>V15</td>
<td>reckless</td>
</tr>
<tr>
<td>V22</td>
<td>willing to take risks</td>
</tr>
<tr>
<td>V11</td>
<td>courageous</td>
</tr>
<tr>
<td>V6</td>
<td>a adventurous</td>
</tr>
<tr>
<td>V19</td>
<td>dominant</td>
</tr>
<tr>
<td>V14</td>
<td>controlling</td>
</tr>
<tr>
<td>V20</td>
<td>boastful</td>
</tr>
<tr>
<td>V21</td>
<td>rational</td>
</tr>
<tr>
<td>V23</td>
<td>analytical</td>
</tr>
<tr>
<td>V9</td>
<td>pragmatic</td>
</tr>
<tr>
<td>V44</td>
<td>to find an address for the first time</td>
</tr>
<tr>
<td>V45</td>
<td>to find a way again</td>
</tr>
<tr>
<td>V46</td>
<td>to understand equations</td>
</tr>
<tr>
<td>V50</td>
<td>to follow directions</td>
</tr>
<tr>
<td>V51</td>
<td>to understand equations</td>
</tr>
<tr>
<td>V53</td>
<td>day-to-day calculations</td>
</tr>
<tr>
<td>V48</td>
<td>to write a computer program</td>
</tr>
<tr>
<td>V69</td>
<td>paintball</td>
</tr>
<tr>
<td>V73</td>
<td>driving go-cart</td>
</tr>
<tr>
<td>V71</td>
<td>drinking beer</td>
</tr>
<tr>
<td>V68</td>
<td>watching action movies</td>
</tr>
<tr>
<td>V75</td>
<td>playing cards (poker)</td>
</tr>
<tr>
<td>V72</td>
<td>watching sports on TV</td>
</tr>
<tr>
<td>V67</td>
<td>doing certain sports (e.g. soccer, ...)</td>
</tr>
<tr>
<td>V74</td>
<td>Gym (weightlifting)</td>
</tr>
<tr>
<td>V27</td>
<td>warm-hearted</td>
</tr>
</tbody>
</table>
GERAS

V28 loving
V29 caring
V26 compassionate
V32 delicate
V30 tender
V24 family-oriented
V40 anxious
V39 thin-skinned
V41 careful
V55 to explain foreign words
V58 to find the right words to express certain content
V59 synonyms for a word in order to avoid repetitions
V60 to phrase a text
V54 remembering events from your own life
V63 to notice small changes
V57 to remember names and faces
V89 shopping
V92 gossiping
V81 watching a romantic movie
V80 talking on the phone with a friend
V90 yoga
V83 rhythmic gymnastics
V84 going for a walk
V86 dancing
gender gender (M=1 F=2)

Details
These 50 items (+ gender) may be formed into scales using the GERAS.keys The first 10 items are Male Personality, the next 10 are Female Personality, then 7 and 7 M and F Cognition, then 8 and 8 M and F Activity items. The Pers, Cog and Act scales are formed from the M-F scales for the three domains. M and F are the composites of the Male and then the Female scales. MF.all is the composite of the M-F scales. See the GERAS.keys object for scoring directions.

"M.pers" "F.pers" "M.cog" "F.cog" "M.act" "F.act" "Pers" "Cog" "Act" "M" "F" "MF.all" "gender"

See the Athenstaedt data set for a related data set.

Source
Study 2 data downloaded from the Open Science Framework https://osf.io/42jhr/ Used by kind permission of Freya M. Gruber, Tullia Ortner, and Belinda A. Pletzer.
References


Examples

data(GERAS)
GERAS.keys #show the keys
#show the items from the dictionary
psych::lookupFromKeys(GERAS.keys, GERAS.dictionary[,4,drop=FALSE])

#now, use the GERAS.scales to show a scatterHist plot showing univariate d and bivariate # Mahalanobis D.

psych::scatterHist(F ~ M + gender, data=GERAS.scales, cex.point=.3,smooth=FALSE,
                   xlab="Masculine Scale",ylab="Feminine Scale",correl=FALSE,
                   d.arrow=TRUE,col=c("red","blue"), bg=c("red","blue"), lwd=4, title="Combined M and F
                   scales",cex.cor=2,cex.arrow=1.25, cex.main=2)

---

globalWarm

7 attitude items about Global Warming policy from Erik Nisbet

Description

Erik Nisbet reported the relationship between emotions, ideology, and party affiliation as predictors of attitudes towards government action on climate change. The data were used by Hayes (2013) in a discussion of regression. They are available as the glbwarm data set in the processR package. They are copied here for examples of mediation.

Usage

data("globalWarm")
Format

A data frame with 815 observations on the following 7 variables.

govact  Support for government action
posemot  Positive emotions about climate change
negemot  Negative emotions about climate change
ideology  Political ideology (Liberal to conservative)
age  age
sex  female =0, male =1
partyid  Democratic =1, Independent =2, Republican =3

Details

This data set is discussed as an example of regression in Hayes (2013) p 24 - 30 and elsewhere. It is a nice example of moderated regression. It was collected by Erik Nisbet (no citation) who studies communication and the media. E. Nisbet is currently on the faculty at Northwestern School of Communication.

Source

The raw data are available from the processR package (Keon-Woong Moon, 2020) as the glbwarm data set as well as from Hayes’ website. The data set is used by Hayes in several examples. Used here by kind permission of Erik Nisbet.

Although the processR package has been removed from CRAN, an earlier version had the data.

References


Examples

data(globalWarm)
psych::lowerCor(globalWarm)
#compare to Hayes p 254-258
psych::lmCor(govact ~ negemot * age + posemot +ideology+sex,data=globalWarm,std=FALSE)
Francis Galton introduced the 'co-relation' in 1888 with a paper discussing how to measure the relationship between two variables. His primary example was the relationship between height and forearm length. The data table (cubits) is taken from Galton (1888). Unfortunately, there seem to be some errors in the original data table in that the marginal totals do not match the table.

The data frame, *heights*, is converted from this table using *table2df*.

### Usage
```r
data(heights)
```

### Format
A data frame with 348 observations on the following 2 variables.

- **height**: Height in inches
- **cubit**: Forearm length in inches

### Details
Sir Francis Galton (1888) published the first demonstration of the correlation coefficient. The regression (or reversion to mediocrity) of the height to the length of the left forearm (a cubit) was found to .8. The original table (cubits) is taken from Galton (1888). There seem to be some errors in the table as published in that the row sums do not agree with the actual row sums. These data are used to create a matrix using *table2matrix* for demonstrations of analysis and displays of the data.

### Source
Galton (1888)

### References
Galton, Francis (1888) Co-relations and their measurement. Proceedings of the Royal Society. London Series,45,135-145,

### See Also
- *table2matrix*, *table2df*, *cubits*, *ellipses*, *galton*

### Examples
```r
data(heights)
psych::ellipses(heights,n=1,main="Galton's co-relation data set")
```
The raw and transformed data from Holzinger and Swineford, 1939

Description
A classic data set in psychometrics is that from Holzinger and Swineford (1939). A 4 and 5 factor solution to 24 of these variables problem is presented by Harman (1976), and 9 of these are used by the lavaan package. The two data sets were supplied by Keith Widaman.

Usage

data(holzinger.swineford)
data(holzinger.raw)
data(holzinger.dictionary)

Format
A data frame with 301 observations on the following 33 variables. Longer descriptions taken from Thompson, (1998).

  case  a numeric vector
  school School Pasteur or Grant-White
  grade  Grade (7 or 8)
  female  male = 1, female = 2
  ageyr  age in years
  mo  months over year
  agemo  Age in months
  t01_visperc  Visual perception test from Spearman VPT Part I
  t02_cubes  Cubes, Simplification of Brighams Spatial Relations Test
  t03_frmboard  Paper formboard-Shapes that can be combined to form a target
  t04_lozenges  Lozenges from Thorndike-Shapes flipped over then identify target
  t05_geninfo  General Information Verbal Test
  t06_paracomp  Paragraph Comprehension Test
  t07_sentcomp  Sentence Completion Test
  t08_wordclas  Word clasification-Which word not belong in set
  t09_wordmean  Word Meaning Test
  t10_addition  Speeded addition test
  t11_code  Speeded codetest-Transform shapes into alpha with code
  t12_countdot  Speeded counting of dots in shap
  t13_sccaps  Speeded discrimination of straight and curved caps
Details

The following commentary was provided by Keith Widaman:

“The Holzinger and Swineford (1939) data have been used as a model data set by many investigators. For example, Harman (1976) used the “24 Psychological Variables” example prominently in his authoritative text on multiple factor analysis, and the data presented under this rubric consisted of 24 of the variables from the Grant-White school (N = 145). Meredith (1964a, 1964b) used several variables from the Holzinger and Swineford study in his work on factorial invariance under selection. Joreskog (1971) based his work on multiple-group confirmatory factor analysis using the Holzinger and Swineford data, subsetting the data into four groups.

Rosseel, who developed the ‘lavaan’ package for R, included 9 of the manifest variables from Holzinger and Swineford (1939) as a “resident” data set when one downloads the ‘lavaan’ package. Several background variables are included in this “resident” data set in addition to 9 of the psychological tests (which are named x1 – x9 in the data set). When analyzing these data, I found the distributions of the variables (means, SDs) did not match the sample statistics from the original article. For example, in the “resident” data set in ‘lavaan’, scores on all manifest variables ranged between 0 and 10, sample means varied between 3 and 6, and sample SDs varied between 1.0 and 1.5. In the original data set, scores ranges were rather different across tests, with some variables having scores that ranged between 0 and 20, but other manifest variables having scores ranging from 50 to over 300 – with obvious attendant differences in sample means and SDs.

After a bit of snooping (i.e., data analysis), I discovered that the 9 variables in the “resident” data set in ‘lavaan’ had been rescored through ratio transformations. The ratio transformations involved dividing the raw score for each person on a given test by a particular constant for that test that transformed scores on the test to have the desired range.

I decided to perform transformations of all 26 variables so that two data sets could be available to interested researchers:"

holzinger.raw are the raws scores on all variables from Holzinger & Swineford (1939)
holzinger.swineford are rescaled scores on all variables from Holzinger & Swineford.
holzinger.dictionary is a list of the variable names in short and long form.
... Widaman continues:

“As several persons have noted, Harman (1976) used data only from the Grant-White school (N = 145) for his 24 Psychological Variables data set. In doing so, Harman replaced t03_frmbord and t04_lozenges with t25_frmbord2 and t26_flags, because the latter two tests were experimental tests that were designed to be more appropriate for this age level. This substitution is fine, as long as one analyzes data from only the Grant-White school. If one wishes to perform multiple-group analyses and uses school as a grouping variable (as Meredith, 1964a, 1964b, and Joreskog, 1971, did), then tests 25 and 26 should not be used.”

“As have others, Gorsuch (1983) mentioned that analyses based on the raw data reported by Holzinger and Swineford (1939) will not produce statistics (means, SDs, correlations) that match precisely the values reported by Holzinger and Swineford or Harman (1976). Following Gorsuch, I have assumed that the raw data are correct. Applying factor analytic techniques to the raw data from the Grant-White school and to the summary data reported by Harman (1976) will produce slightly different results, but results that differ in only minor, unimportant details.”

These data are interesting not just for the historical completeness of having the original data, but also as an example of suppressor variables. Age and grade are positively correlated, and scores are higher in the 8th grade than in the 7th grade. But age (particularly in months) is negatively correlated with many of the cognitive tasks, and when grade and age are both entered into regression, this negative correlation is enhanced. That is, although increasing grade increases cognitive performance, younger children in both grades do better than the older children.

**Note**

As discussed by Widaman, the descriptive values reported in Harman (1967) (p 124) do not quite match the descriptive statistics in holzinger.raw. Further note that the correlation matrix and factor loadings are trivially different from the Harman.24 factor loadings in the GPA rotation package.

The purpose behind presenting both the raw and transformed data is to show that the fit statistics from factor analysis are identical for these two data sets.

The variables v1 ... v9 in the lavaan package correspond to tests 1, 2, 4, 6, 7, 9, 10, 12 and 13.

**Source**


**References**


See Also

psych::Holzinger

Examples

data(holzinger.raw)
psych::describe(holzinger.raw)
data(holzinger.dictionary)
holzinger.dictionary  # to see the longer names for these data (taken from Thompson)

# Compare these to the lavaan correlation matrix
psych::lowerCor(holzinger.swineford[7+ c(1, 2, 4, 6, 7, 9, 10, 12, 13)])

psych::lmCor(t01_visperc + t05_geninfo + t08_wordclas ~ grade + agemo, data = holzinger.raw)
psych::lmCor(t06_paracomp ~ grade + agemo, data = holzinger.swineford)
psych::mediate(t06_paracomp ~ grade + (agemo), data = holzinger.raw, std=TRUE)

# show the omega structure of the 24 variables
om4 <- psych::omega(holzinger.swineford[8:31], 4)
psych::omega.diagram(om4, sl=FALSE, main="26 variables from Holzinger-Swineford")

# these data also show an interesting suppression effect
psych::lowerCor(holzinger.swineford[c(3,7,12:14)])
psych::lmCor(t06_paracomp ~ grade + agemo, data = holzinger.swineford)
# or show as a mediation effect
mod <- psych::mediate(t06_paracomp ~ grade + (agemo), data = holzinger.raw, std=TRUE, n.iter=50)
summary(mod)

# now, show a plot of these effects
plot(t07_sentcomp ~ agemo, col=c("red", "blue")[holzinger.swineford$grade - 6],
     pch=26-holzinger.swineford$grade, data=holzinger.swineford,ylab="Sentence Comprehension", xlab="Age in Months",
     main="Sentence Comprehension varies by age and grade")
# we use lmCor to figure out the lines
# note that we need to not plot the default graph
by(holzinger.swineford, holzinger.swineford$grade - 6, function(x) abline(
     psych::lmCor(t07_sentcomp ~ agemo, data=x, std=FALSE, plot=FALSE),
     lty=c("dashed", "solid")[x$grade-6])

text(190, 3.3, "grade = 8")
text(190, 2, "grade = 7")
US family income from US census 2008

Description

US census data on family income from 2008

Usage

data(income)

Format

A data frame with 44 observations on the following 4 variables.

- value: lower boundary of the income group
- count: Number of families within that income group
- mean: Mean of the category
- prop: proportion of families

Details

The distribution of income is a nice example of a log normal distribution. It is also an interesting example of the power of graphics. It is quite clear when graphing the data that income statistics are bunched to the nearest 5K. That is, there is a clear sawtooth pattern in the data.

The all.income set is interpolates intervening values for 100-150K, 150-200K and 200-250K

Source

US Census: Table HINC-06. Income Distribution to $250,000 or More for Households: 2008
https://www.census.gov/hhes/www/cpstables/032009/hhinc/new06_000.htm

Examples

data(income)
with(income[1:40,], plot(mean,prop, main="US family income for 2008",xlab="income", ylab="Proportion of families",xlim=c(0,100000)))
with (income[1:40,], points(lowess(mean,prop,f=.3),typ="l"))
psych::describe(income)

with(all.income, plot(mean,prop, main="US family income for 2008",xlab="income", ylab="Proportion of families",xlim=c(0,250000)))
with (all.income[1:50,], points(lowess(mean,prop,f=.25),typ="l"))
16 multiple choice IQ items

Description

16 multiple choice ability items taken from the Synthetic Aperture Personality Assessment (SAPA) web based personality assessment project. The data from 1525 subjects are included here as a demonstration set for scoring multiple choice inventories and doing basic item statistics. For more information on the development of an open source measure of cognitive ability, consult the readings available at the https://personality-project.org/.

Usage

data(iqitems)

Format

A data frame with 1525 observations on the following 16 variables. The number following the name is the item number from SAPA.

reason.4 Basic reasoning questions
reason.16 Basic reasoning question
reason.17 Basic reasoning question
reason.19 Basic reasoning question
letter.7 In the following alphanumeric series, what letter comes next?
letter.33 In the following alphanumeric series, what letter comes next?
letter.34 In the following alphanumeric series, what letter comes next
letter.58 In the following alphanumeric series, what letter comes next?
matrix.45 A matrix reasoning task
matrix.46 A matrix reasoning task
matrix.47 A matrix reasoning task
matrix.55 A matrix reasoning task
rotate.3 Spatial Rotation of type 1.2
rotate.4 Spatial Rotation of type 1.2
rotate.6 Spatial Rotation of type 1.1
rotate.8 Spatial Rotation of type 2.3
Details

16 items were sampled from 80 items given as part of the SAPA (https://www.sapa-project.org/) project (Revelle, Wilt and Rosenthal, 2009; Condon and Revelle, 2014) to develop online measures of ability. These 16 items reflect four lower order factors (verbal reasoning, letter series, matrix reasoning, and spatial rotations. These lower level factors all share a higher level factor (‘g’). Similar data are available from the International Cognitive Ability Resource at https://www.icar-project.org/.

This data set and the associated data set (ability based upon scoring these multiple choice items and converting them to correct/incorrect may be used to demonstrate item response functions, tetrachoric correlations, or irt.fa as well as omega estimates of of reliability and hierarchical structure.

In addition, the data set is a good example of doing item analysis to examine the empirical response probabilities of each item alternative as a function of the underlying latent trait. When doing this, it appears that two of the matrix reasoning problems do not have monotonically increasing trace lines for the probability correct. At moderately high ability (theta = 1) there is a decrease in the probability correct from theta = 0 and theta = 2.

Source

The example data set is taken from the Synthetic Aperture Personality Assessment personality and ability test at https://www.sapa-project.org/. The data were collected with David Condon from 8/08/12 to 8/31/12.

References


Examples

data(iqitems)
iq.keys <- c(4,4,4, 6, 6,3,4, 5,2,2,4, 3,2,6,7)
 psych::score.multiple.choice(iq.keys,iqitems)  # this just gives summary statistics
# convert them to true false
iq.scrub <- psych::scrub(iqitems,isvalue=0)  # first get rid of the zero responses
iq.tf <- psych::score.multiple.choice(iq.keys,iq.scrub,score=FALSE)
# convert to wrong (0) and correct (1) for analysis
psych::describe(iq.tf)
# see the ability data set for these analyses
# now, for some item analysis
iq.irt <- psych::irt.fa(iq.tf)  # do a basic irt
iq.sc <- psych::scoreIrt(iq.irt,iq.tf)  # find the scores
op <- par(mfrow=c(4,4))
psych::irt.responses(iq.sc[,1], iq.tf)
op <- par(mfrow=c(1,1))

Description
Emotions may be described either as discrete emotions or in dimensional terms. The Motivational
State Questionnaire (MSQ) was developed to study emotions in laboratory and field settings. The
data can be well described in terms of a two dimensional solution of energy vs tiredness and tension
versus calmness. Additional items include what time of day the data were collected and a few
personality questionnaire scores.

Usage
data(msq)

Format
A data frame with 3896 observations on the following 92 variables.
active a numeric vector
afraid a numeric vector
alert a numeric vector
angry a numeric vector
anxious a numeric vector
aroused a numeric vector
ashamed a numeric vector
astonished a numeric vector
at.ease a numeric vector
at.rest a numeric vector
attentive a numeric vector
blue a numeric vector
bored a numeric vector
calm a numeric vector
cheerful a numeric vector
clutched up a numeric vector
confident a numeric vector
content a numeric vector
delighted a numeric vector
depressed a numeric vector
determined a numeric vector
distressed a numeric vector
drowsy a numeric vector
dull a numeric vector
elated a numeric vector
energetic a numeric vector
enthusiastic a numeric vector
excited a numeric vector
fearful a numeric vector
frustrated a numeric vector
full of pep a numeric vector
gloomy a numeric vector
grouchy a numeric vector
guilty a numeric vector
happy a numeric vector
hostile a numeric vector
idle a numeric vector
inactive a numeric vector
inspired a numeric vector
intense a numeric vector
interested a numeric vector
irritable a numeric vector
jittery a numeric vector
lively a numeric vector
lonely a numeric vector
nervous a numeric vector
placid a numeric vector
pleased a numeric vector
proud a numeric vector
quiescent a numeric vector
quiet a numeric vector
relaxed a numeric vector
sad a numeric vector
satisfied a numeric vector
scared a numeric vector
serene a numeric vector
sleepy a numeric vector
sluggish a numeric vector
sociable a numeric vector
sorry a numeric vector
still a numeric vector
strong a numeric vector
surprised a numeric vector
tense a numeric vector
tired a numeric vector
tranquil a numeric vector
unhappy a numeric vector
upset a numeric vector
vigorous a numeric vector
wakeful a numeric vector
warmhearted a numeric vector
wide.awake a numeric vector
alone a numeric vector
kindly a numeric vector
scornful a numeric vector
EA Thayer's Energetic Arousal Scale
TA Thayer's Tense Arousal Scale
PA Positive Affect scale
NegAff Negative Affect scale
Extraversion Extraversion from the Eysenck Personality Inventory
Neuroticism Neuroticism from the Eysenck Personality Inventory
Lie Lie from the EPI
Sociability The sociability subset of the Extraversion Scale
Impulsivity The impulsivity subset of the Extraversions Scale
The Motivational States Questionnaire (MSQ) is composed of 72 items, which represent the full affective space (Revelle & Anderson, 1998). The MSQ consists of 20 items taken from the Activation-Deactivation Adjective Check List (Thayer, 1986), 18 from the Positive and Negative Affect Schedule (PANAS, Watson, Clark, & Tellegen, 1988) along with the items used by Larsen and Diener (1992). The response format was a four-point scale that corresponds to Russell and Carroll’s (1999) “ambiguous–likely-unipolar format” and that asks the respondents to indicate their current standing (“at this moment”) with the following rating scale:

0 ———— 1 ———— 2 ———— 3

Not at all A little Moderately Very much

The original version of the MSQ included 70 items. Intermediate analyses (done with 1840 subjects) demonstrated a concentration of items in some sections of the two dimensional space, and a paucity of items in others. To begin correcting this, 3 items from redundantly measured sections (alone, kindly, scornful) were removed, and 5 new ones (anxious, cheerful, idle, inactive, and tranquil) were added. Thus, the correlation matrix is missing the correlations between items anxious, cheerful, idle, inactive, and tranquil with alone, kindly, and scornful.

Procedure. The data were collected over nine years, as part of a series of studies examining the effects of personality and situational factors on motivational state and subsequent cognitive performance. In each of 38 studies, prior to any manipulation of motivational state, participants signed a consent form and filled out the MSQ. (The procedures of the individual studies are irrelevant to this data set and could not affect the responses to the MSQ, since this instrument was completed before any further instructions or tasks). Some MSQ post test (after manipulations) is available in affect.

The EA and TA scales are from Thayer, the PA and NA scales are from Watson et al. (1988). Scales and items:

Energetic Arousal: active, energetic, vigorous, wakeful, wide.awake, full.of.pep, lively, -sleepy, -tired, - drowsy (ADACL)

Tense Arousal: Intense, Jittery, fearful, tense, clutched up, -quiet, -still, - placid, - calm, -at rest (ADACL)

Positive Affect: active, alert, attentive, determined, enthusiastic, excited, inspired, interested, proud, strong (PANAS)
Negative Affect: afraid, ashamed, distressed, guilty, hostile, irritable, jittery, nervous, scared, upset (PANAS)

The PA and NA scales can in turn be thought of as having subscales: (See the PANAS-X) Fear: afraid, scared, nervous, jittery (not included frightened, shaky) Hostility: angry, hostile, irritable, (not included: scornful, disgusted, loathing guilt: ashamed, guilty, (not included: blameworthy, angry at self, disgusted with self, dissatisfied with self) sadness: alone, blue, lonely, sad, (not included: downhearted) joviality: cheerful, delighted, energetic, enthusiastic, excited, happy, lively, (not included: joyful) self-assurance: proud, strong, confident, (not included: bold, daring, fearless ) attentiveness: alert, attentive, determined (not included: concentrating)

The next set of circumplex scales were taken (I think) from Larsen and Diener (1992). High activation: active, aroused, surprised, intense, astonished Activated PA: elated, excited, enthusiastic, lively Unactivated NA : calm, serene, relaxed, at rest, content, at ease PA: happy, warmhearted, pleased, cheerful, delighted Low Activation: quiet, inactive, idle, still, tranquil Unactivated PA: dull, bored, sluggish, tired, drowsy NA: sad, blue, unhappy, gloomy, grouchy Activated NA: jittery, anxious, nervous, fearful, distressed.

Keys for these separate scales are shown in the examples.

In addition to the MSQ, there are 5 scales from the Eysenck Personality Inventory (Extraversion, Impulsivity, Sociability, Neuroticism, Lie). The Imp and Soc are subsets of the the total extraversion scale.

Source

Data collected at the Personality, Motivation, and Cognition Laboratory, Northwestern University.

References


See Also

msqR for a larger data set with repeated measures for 3032 participants measured at least once, 2753 measured twice, 446 three times and 181 four times. affect for an example of the use of some of these adjectives in a mood manipulation study.
make.keys, scoreItems and scoreOverlap for instructions on how to score multiple scales with and without item overlap. Also see fa and fa.extension for instructions on how to do factor analyses or factor extension.

Examples

```r
data(msq)
# in in the interests of time
# basic descriptive statistics
psych::describe(msq)

# score them for 20 short scales -- note that these have item overlap
# The first 2 are from Thayer
# The next 2 are classic positive and negative affect
# The next 9 are circumplex scales
# the last 7 are msq estimates of PANASX scales (missing some items)
keys.list <- list(
    EA = c("active", "energetic", "vigorous", "wide.awake", "full.of.pep", "lively", "sleepy", "tired", "drowsy"),
    TA = c("intense", "jittery", "fearful", "tense", "clutched.up", "quiet", "still", "placid", "calm", "at.rest"),
    PA = c("active", "excited", "strong", "inspired", "determined", "attentive", "interested", "enthusiastic", "proud", "alert"),
    NAF = c("jittery", "nervous", "scared", "afraid", "guilty", "ashamed", "distressed", "upset", "hostile", "irritable"),
    HAct = c("active", "aroused", "surprised", "intense", "astonished"),
    aPA = c("elated", "excited", "enthusiastic", "lively"),
    uNA = c("calm", "serene", "relaxed", "at.rest", "content", "at.ease"),
    pa = c("happy", "warmhearted", "pleased", "cheerful", "delighted"),
    LAct = c("quiet", "inactive", "idle", "still", "tranquil"),
    uPA = c("dull", "bored", "sluggish", "tired", "drowsy"),
    naf = c("sad", "blue", "unhappy", "gloomy", "grouchy"),
    aNA = c("jittery", "anxious", "nervous", "fearful", "distressed"),
    Fear = c("afraid", "scared", "nervous", "jittery"),
    Hostility = c("angry", "hostile", "irritable", "scornful"),
    Guilt = c("guilty", "ashamed"),
    Sadness = c("sad", "blue", "lonely", "alone"),
    Joviality = c("happy", "delighted", "cheerful", "excited", "enthusiastic", "lively", "energetic"),
    Self.Assurance = c("proud", "strong", "confident", "fearful"),
    Attentiveness = c("alert", "determined", "attentive"),
    #, acquiscence = c("sleepy", "wakeful", "relaxed", "tense")
    # dropped because it has a negative alpha and throws warnings
)

msq.scores <- psych::scoreItems(keys.list, msq)

# show a circumplex structure for the non-overlapping items
fcirc <- psych::fa(msq.scores$scores[,5:12],2)
psych::fa.plot(fcirc, labels=colnames(msq.scores$scores)[5:12])

# now, find the correlations corrected for item overlap
msq.overlap <- psych::scoreOverlap(keys.list, msq)
# a warning is thrown by smc because of some NAs in the matrix
```
f2 <- psych::fa(msq.overlap$cor,2)
psych::fa.plot(f2,labels=colnames(msq.overlap$cor),
    title="2 dimensions of affect, corrected for overlap")

#extend this solution to EA/TA PA/NA space
fe <- psych::fa.extension(cor(msq.scores$scores[,5:12],msq.scores$scores[,1:4]),fcirc)
psych::fa.diagram(fcirc,fe=fe,
    main="Extending the circumplex structure to EA/TA and PA/NA ")

#show the 2 dimensional structure
f2 <- psych::fa(msq[1:72],2)
psych::fa.plot(f2,labels=colnames(msq)[1:72],
    title="2 dimensions of affect at the item level",cex=.5)

#sort them by polar coordinates
round(psych::polar(f2),2)

---

data("msqR")

## msqR

### 75 mood items from the Motivational State Questionnaire for 3032 unique participants

---

**Description**

Emotions may be described either as discrete emotions or in dimensional terms. The Motivational State Questionnaire (MSQ) was developed to study emotions in laboratory and field settings. The data can be well described in terms of a two dimensional solution of energy vs tiredness and tension versus calmness. Alternatively, this space can be organized by the two dimensions of Positive Affect and Negative Affect. Additional items include what time of day the data were collected and a few personality questionnaire scores. 3032 unique participants took the MSQ at least once, 2753 at least twice, 446 three times, and 181 four times. The 3032 participants also took the sai state anxiety inventory at the same time. Some studies manipulated arousal by caffeine, others manipulations included affect inducing movies.

**Usage**

```r
data("msqR")
```

**Format**

A data frame with 6411 observations on the following 88 variables.

- `active` a numeric vector
- `afraid` a numeric vector
- `alert` a numeric vector
alone  a numeric vector
angry   a numeric vector
aroused a numeric vector
ashamed a numeric vector
astonished a numeric vector
at.ease a numeric vector
at.rest a numeric vector
attentive a numeric vector
blue    a numeric vector
bored   a numeric vector
calm    a numeric vector
clutched.up a numeric vector
confident a numeric vector
content a numeric vector
delighted a numeric vector
depressed a numeric vector
determined a numeric vector
distressed a numeric vector
drowsy  a numeric vector
dull    a numeric vector
elated  a numeric vector
energetic a numeric vector
enthusiastic a numeric vector
excited a numeric vector
fearful  a numeric vector
frustrated a numeric vector
full.of.pep a numeric vector
gloomy  a numeric vector
grouchy a numeric vector
guilty  a numeric vector
happy   a numeric vector
hostile a numeric vector
inspired a numeric vector
intense a numeric vector
interested a numeric vector
irritable a numeric vector
jittery a numeric vector
lively a numeric vector
lonely a numeric vector
nervous a numeric vector
placid a numeric vector
pleased a numeric vector
proud a numeric vector
quiescent a numeric vector
quiet a numeric vector
relaxed a numeric vector
sad a numeric vector
satisfied a numeric vector
scared a numeric vector
serene a numeric vector
sleepy a numeric vector
sluggish a numeric vector
sociable a numeric vector
sorry a numeric vector
still a numeric vector
strong a numeric vector
surprised a numeric vector
tense a numeric vector
tired a numeric vector
unhappy a numeric vector
upset a numeric vector
vigorous a numeric vector
wakeful a numeric vector
warmhearted a numeric vector
wide.awake a numeric vector
anxious a numeric vector
cheerful a numeric vector
idle a numeric vector
inactive a numeric vector
tranquil a numeric vector
kindly a numeric vector
scornful a numeric vector
Extraversion Extraversion from the EPI
Neuroticism Neuroticism from the EPI
Lie from the EPI
Sociability from the EPI
Impulsivity from the EPI
gender 1 = male, 2 = female (coded on presumed x chromosome). Slowly being added to the data set.
TOD Time of day that the study was run
drug 1 if given placebo, 2 if given caffeine
film 1-4 if given a film: 1=Frontline, 2= Halloween, 3=Serengeti, 4 = Parenthood
time Measurement occasion (1 and 2 are same session, 3 and 4 are the same, but a later session)
id a numeric vector
form msq versus msqR
study a character vector of the experiment name

Details

The Motivational States Questionnaire (MSQ) is composed of 75 items, which represent the full affective space (Revelle & Anderson, 1998). The MSQ consists of 20 items taken from the Activation-Deactivation Adjective Check List (Thayer, 1986), 18 from the Positive and Negative Affect Schedule (PANAS, Watson, Clark, & Tellegen, 1988) along with the affective circumplex items used by Larsen and Diener (1992). The response format was a four-point scale that corresponds to Russell and Carroll's (1999) "ambiguous–likely-unipolar format" and that asks the respondents to indicate their current standing ("at this moment") with the following rating scale:

\[
\begin{array}{c}
0 & 1 & 2 & 3 \\
Not at all & A little & Moderately & Very much
\end{array}
\]

The original version of the MSQ included 70 items. Intermediate analyses (done with 1840 subjects) demonstrated a concentration of items in some sections of the two dimensional space, and a paucity of items in others. To begin correcting this, 3 items from redundantly measured sections (alone, kindly, scornful) were removed, and 5 new ones (anxious, cheerful, idle, inactive, and tranquil) were added. Thus, the correlation matrix is missing the correlations between items anxious, cheerful, idle, inactive, and tranquil with alone, kindly, and scornful.

2605 individuals took Form 1 version, 3806 the Form 2 version. 3032 people (1218 form 1, 1814 form 2) took the MSQ at least once. 2086 at least twice, 1112 three times, and 181 four times.

To see the relative frequencies by time and form, see the first example.

Procedure. The data were collected over nine years in the Personality, Motivation and Cognition laboratory at Northwestern, as part of a series of studies examining the effects of personality and situational factors on motivational state and subsequent cognitive performance. In each of 38 studies, prior to any manipulation of motivational state, participants signed a consent form and in some studies, consumed 0 or 4mg/kg of caffeine. In caffeine studies, they waited 30 minutes and then filled out the MSQ. (Normally, the procedures of the individual studies are irrelevant to this data set and could not affect the responses to the MSQ at time 1, since this instrument was completed before any further instructions or tasks. However, caffeine does have an effect.) The MSQ post test following a movie manipulation) is available in affect as well as here.

The XRAY study crossed four movie conditions with caffeine. The first MSQ measures are showing the effects of the movies and caffeine, but after an additional 30 minutes, the second MSQ seems
to mainly show the caffeine effects. The movies were 9 minute clips from 1) a BBC documentary on British troops arriving at the Bergen-Belsen concentration camp (sad); 2) an early scene from Halloween in which the heroine runs around shutting doors and windows (terror); 3) a documentary about lions on the Serengeti plain, and 4) the "birthday party" scene from Parenthood.

The FLAT study measured affect before, immediately after, and then after 30 minutes following a movie manipulation. See the affect data set.

To see which studies used which conditions, see the second and third examples.

The EA and TA scales are from Thayer, the PA and NA scales are from Watson et al. (1988). Scales and items:

Energetic Arousal: active, energetic, vigorous, wakeful, wide.awake, full.of.pep, lively, -sleepy, -tired, - drowsy (ADACL)

Tense Arousal: Intense, Jittery, fearful, tense, clutched up, -quiet, -still, - placid, - calm, -at rest (ADACL)

Positive Affect: active, alert, attentive, determined, enthusiastic, excited, inspired, interested, proud, strong (PANAS)

Negative Affect: afraid, ashamed, distressed, guilty, hostile, irritable, jittery, nervous, scared, upset (PANAS)

The PA and NA scales can in turn can be thought of as having subscales: (See the PANAS-X) Fear: afraid, scared, nervous, jittery (not included frightened, shaky) Hostility: angry, hostile, irritable, (not included: scornful, disgusted, loathing guilt: ashamed, guilty, (not included: blameworthy, angry at self, disgusted with self, dissatisfied with self) sadness: alone, blue, lonely, sad, (not included: downhearted) joviality: cheerful, delighted, energetic, enthusiastic, excited, happy, lively, (not included: joyful) self-assurance: proud, strong, confident, (not included: bold, daring, fearless ) attentiveness: alert, attentive, determined (not included: concentrating)


Keys for these separate scales are shown in the examples.

In addition to the MSQ, there are 5 scales from the Eysenck Personality Inventory (Extraversion, Impulsivity, Sociability, Neuroticism, Lie). The Imp and Soc are subsets of the the total extraversion scale based upon a reanalysis of the EPI by Rocklin and Revelle (1983). This information is in the msq data set as well.

**Note**

In December, 2018 the caffeine, film and personality conditions were added. In the process of doing so, it was discovered that the EMIT data had been incorrectly entered. This has been fixed.

**Source**

Data collected at the Personality, Motivation, and Cognition Laboratory, Northwestern University.
References


See Also

msq for 3896 participants with scores on five scales of the EPI. affect for an example of the use of some of these adjectives in a mood manipulation study.

make.keys, scoreItems and scoreOverlap for instructions on how to score multiple scales with and without item overlap. Also see fa and fa.extension for instructions on how to do factor analyses or factor extension.

Given the temporal ordering of the sai data and the msqR data, these data are useful for demonstrations of testRetest reliability. See the examples in testRetest for how to combine the sai tai and msqR datasets.

Examples

data(msqR)
table(msqR$form,msqR$time) #which forms?
table(msqR$study,msqR$drug) #Drug studies
table(msqR$study,msqR$film) #Film studies
table(msqR$study,msqR$TOD) #To examine time of day

#score them for 20 short scales -- note that these have item overlap
#The first 2 are from Thayer
#The next 2 are classic positive and negative affect
#The next 9 are circumplex scales
#the last 7 are msq estimates of PANASX scales (missing some items)
keys.list <- list(
  EA = c("active", "energetic", "vigorous", "wakeful", "wide.awake", "full.of.pep", "lively", "-sleepy", "-tired", "-drowsy"),
  TA = c("intense", "jittery", "fearful", "tense", "clutched.up", "-quiet", "-still", "-placid", "-calm", "-at.rest"),

  "active", "energetic", "vigorous", "wakeful", "wide.awake", "full.of.pep", "lively", "-sleepy", "-tired", "-drowsy"),
  TA = c("intense", "jittery", "fearful", "tense", "clutched.up", "-quiet", "-still", "-placid", "-calm", "-at.rest"),
PA = c("active", "excited", "strong", "inspired", "determined", "attentive", "interested", "enthusiastic", "proud", "alert"), 
Naf = c("jittery", "nervous", "scared", "afraid", "guilty", "ashamed", "distressed", 
"upset", "hostile", "irritable"), 
HAct = c("active", "aroused", "surprised", "intense", "astonished"), 
aPA = c("elated", "excited", "enthusiastic", "lively"), 
UAct = c("calm", "serene", "relaxed", "at.rest", "content", "at.ease"), 
pa = c("happy", "warmhearted", "pleased", "cheerful", "delighted"), 
LAct = c("quiet", "inactive", "idle", "still", "tranquil"), 
uPA = c("dull", "bored", "sluggish", "tired", "drowsy"), 
Naf = c("sad", "blue", "unhappy", "gloomy", "grouchy"), 
aNA = c("jittery", "anxious", "nervous", "feared", "distressed"), 
Fear = c("afraid", "scared", "nervous", "jittery"), 
Hostility = c("angry", "hostile", "irritable", "scornful"), 
Guilt = c("guilty", "ashamed"), 
Sadness = c("sad", "blue", "lonely", "alone"), 
Joviality = c("happy", "delighted", "cheerful", "excited", "enthusiastic", "lively", 
"energetic"), 
Self.Assurance = c("proud", "strong", "confident", "-fearful"), 
Attentiveness = c("alert", "determined", "attentive"))

#Yik Russell and Steiger list the following items 
Yik.keys <- list( 
pleasure = psych::cs(happy, content, satisfied, pleased), 
act.pleasure = psych::cs(proud, enthusiastic, euphoric), 
pleasant.activation = psych::cs(energetic, full.of.pep, excited, wakeful, attentive, 
wide.awake, active, alert, vigorous), 
activation = psych::cs(aroused, hyperactivated, intense), 
unpleasant.act = psych::cs(anxious, frenzied, jittery, nervous), 
activated.displeasure = psych::cs(scared, upset, shaky, fearful, clutched.up, tense, 
ashamed, guilty, agitated, hostile), 
displeasure = psych::cs(troubled, miserable, unhappy, dissatisfied), 
Ueactivated.Displeasure = psych::cs(sad, down, gloomy, blue, melancholy), 
Unpleasant.Deactivation = psych::cs(droopy, drowsy, dull, bored, sluggish, tired), 
Deactivation = psych::cs(quiet, still), 
plesant.deactivation = psych::cs(placid, relaxed, tranquil, at.rest, calm), 
deactivated.pleasure = psych::cs(serene, soothed, peaceful, at.ease, secure) 
)

#of these 60 items, 46 appear in the msqR 
Yik.msq.keys <- list( 
Pleasure = psych::cs(happy, content, satisfied, pleased), 
Activated.Pleasure = psych::cs(proud, enthusiastic), 
Pleasant.Activation = psych::cs(energetic, full.of.pep, excited, wakeful, attentive, 
wide.awake, active, alert, vigorous), 
Activation = psych::cs(aroused, intense), 
Unpleasant.Activation = psych::cs(anxious, jittery, nervous), 
Activated.Displeasure = psych::cs(scared, upset, fearful, clutched.up, tense, 
ashamed, guilty, agitated, hostile), 
Displeasure = psych::cs(troubled, miserable, unhappy, dissatisfied), 
Ueactivated.Displeasure = psych::cs(sad, down, gloomy, blue, melancholy), 
Unpleasant.Deactivation = psych::cs(droopy, drowsy, dull, bored, sluggish, tired), 
Deactivation = psych::cs(quiet, still), 
plesant.deactivation = psych::cs(placid, relaxed, tranquil, at.rest, calm), 
deactivated.pleasure = psych::cs(serene, soothed, peaceful, at.ease, secure) 
)
msqR

Pleasant.Deactivation = psych::cs(placid,relaxed,tranquil,at.rest,calm),
Deactivated.Pleasure = psych::cs(serene,at.ease)
)
yik.scores <- psych::scoreItems(Yik.msq.keys,msqR)
yik <- yik.scores$scores
f2.yik <- psych::fa(yik,2) #factor the yik scores
psych::fa.plot(f2.yik,labels=colnames(yik),title="Yik-Russell-Steiger circumplex",cex=.8,
   pos=(c(1,1,2,1,1,1,3,1,4,1,2,4))

msq.scores <- psych::scoreItems(keys.list,msqR)
#show a circumplex structure for the non-overlapping items
fcirc <- psych::fa(msq.scores$scores[,5:12],2)
psych::fa.plot(fcirc,labels=colnames(msq.scores$scores)[5:12])

#now, find the correlations corrected for item overlap
msq.overlap <- psych::scoreOverlap(keys.list,msqR)
f2 <- psych::fa(msq.overlap$cor,2)
psych::fa.plot(f2,labels=colnames(msq.overlap$cor),
   title="2 dimensions of affect, corrected for overlap")

#extend this solution to EA/TA NA/PA space
fe <- psych::fa.extension(cor(msq.scores$scores[,5:12],msq.scores$scores[,1:4]),fcirc)
psych::fa.diagram(fcirc,fe=fe,main="Extending the circumplex structure to EA/TA and PA/NA ")

#show the 2 dimensional structure
f2 <- psych::fa(msqR[1:72],2)
psych::fa.plot(f2,labels=colnames(msqR)[1:72],title="2 dimensions of affect at the item level")

#sort them by polar coordinates
round(psych::polar(f2),2)

#the msqR and sai data sets have 10 overlapping items which can be used for
#testRetest analysis. We need to specify the keys, and then choose the appropriate
#data sets
sai.msq.keys <- list(pos =c( "at.ease" , "calm" , "confident", "content","relaxed"),
   neg = c("anxious", "jittery", "nervous", "tense" , "upset"),
   anx = c("anxious", "jittery", "nervous", "tense", "upset","-at.ease" , 
   "-calm", 
   "-confident", "-content","-relaxed"))
select <- psych::selectFromKeys(sai.msq.keys$anx)
The following is useful for examining test retest reliabilities
msq.control <- subset(msqR,is.element( msqR$study , c("Cart", "Fast", "SHED", "SHOP")))
msq.film <- subset(msqR,(is.element( msqR$study , c("FIAT", "FILM","FLAT","MIXX","XRAY"))
   & (msqR$time < 3 )))
msq.film[((msq.film$study == "FLAT") & (msq.film$time ==3) ),] <- NA
msq.drug <- subset(msqR,(is.element( msqR$study , c("AGES","SALT", "VALE", "XRAY"))
   & (msqR$time < 3 ))
msq.day <- subset(msqR,is.element( msqR$study , c("SAM", "RIM")))
Description

The NEO.PI.R is a widely used personality test to assess 5 broad factors (Neuroticism, Extraversion, Openness, Agreeableness and Conscientiousness) with six facet scales for each factor. The correlation matrix of the facets is reported in the NEO.PI.R manual for 1000 subjects.

Usage

data(neo)

Format

A data frame of a 30 x 30 correlation matrix with the following 30 variables.

N1 Anxiety
N2 AngryHostility
N3 Depression
N4 Self-Consciousness
N5 Impulsiveness
N6 Vulnerability
E1 Warmth
E2 Gregariousness
E3 Assertiveness
E4 Activity
E5 Excitement-Seeking
E6 PositiveEmotions
O1 Fantasy
O2 Aesthetics
O3 Feelings
O4 Ideas
O5 Actions
O6 Values
A1 Trust
A2 Straightforwardness
A3 Altruism
Details

The past thirty years of personality research has led to a general consensus on the identification of major dimensions of personality. Variously known as the “Big 5” or the “Five Factor Model”, the general solution represents 5 broad domains of personal and interpersonal experience. Neuroticism and Extraversion are thought to reflect sensitivity to negative and positive cues from the environment and the tendency to withdraw or approach. Openness is sometimes labeled as Intellect and reflects an interest in new ideas and experiences. Agreeableness and Conscientiousness reflect tendencies to get along with others and to want to get ahead.

The factor structure of the NEO suggests five correlated factors as well as two higher level factors. The NEO was constructed with 6 “facets” for each of the five broad factors.

For a contrasting structure, examine the items of the link{spi} data set (Condon, 2017).

Source


References


Galton’s Peas

Description

Francis Galton introduced the correlation coefficient with an analysis of the similarities of the parent and child generation of 700 sweet peas.

Usage

data(peas)

Format

A data frame with 700 observations on the following 2 variables.

parent  The mean diameter of the mother pea for 700 peas
child   The mean diameter of the daughter pea for 700 sweet peas

Details

Galton’s introduction of the correlation coefficient was perhaps the most important contribution to the study of individual differences. This data set allows a graphical analysis of the data set. There are two different graphic examples. One shows the regression lines for both relationships, the other finds the correlation as well.

Source


The data were generated from this table.

References

Galton, Francis (1877) Typical laws of heredity. paper presented to the weekly evening meeting of the Royal Institution, London. Volume VIII (66) is the first reference to this data set. The data appear in

See Also

The other Galton data sets: heights, galton.cubits

Examples

data(peas)
psych::pairs.panels(peas,lm=TRUE,xlim=c(14,22),ylim=c(14,22),main="Galton's Peas")
psych::describe(peas)
psych::pairs.panels(peas,main="Galton's Peas")

Pollack et al (2012) correlation matrix for mediation example

Description

A correlation matrix taken from Pollack (2012) with 9 variables. Primarily used as an example for setCor and mediation.

Usage

data("Pollack")

Format

A correlation matrix based upon 262 participants.

- sex  Male = 1, Female = 0, 62% male
- age  mean =33
- tenure length of employent, mean = 5.9 years
- self.efficacy self ratings
- competence self rating of competence
- social.ties Contact with business-related social ties
- economic.stress mean of two items on economic stress
- depression 6 items from MAACL measuring depression
- withdrawal Withdrawal intentions in domain of entrepreneurship

Details

This is the correlation matrix from Pollack et al. (2012) p 797. The raw data are available from the processR package (Keon-Woong Moon, 2020). The data set is used by Hayes in example p 179 in example 3.

Source

Pollack et al. 2012
References


Examples

```r
psych::lowerMat(Pollack)
```

Description

PsychTools includes the larger data sets used by the psych package and also includes a few general utility functions such as the `read.file` and `read.clipboard` functions. The data sets are made available for demonstrations of a variety of psychometric functions.

Details

See the various helpfiles listed in the index or as links from here. Also see the main functions in the psych package `?psych-package`.

Data sets from the SAPA/ICAR project:

- `ability`: 16 ICAR ability items scored as correct or incorrect for 1525 participants.
- `iqitems`: multiple choice IQ items (raw responses).
- `affect`: Two data sets of affect and arousal scores as a function of personality and movie conditions.
- `bfi`: 25 Personality items representing 5 factors from the SAPA project for 2800 participants.
- `bfi.dictionary`: Dictionary of the bfi.
- `epi`: Eysenck Personality Inventory (EPI) data for 3570 participants.
- `epi.dictionary`: The items for the epi.
- `epi_bfi`: 13 personality scales from the Eysenck Personality Inventory and Big 5 inventory.
- `epiR`: 474 participants took the epi twice.
- `msq`: 75 mood items from the Motivational State Questionnaire for 3896 participants.
- `msqR`: 75 mood items from the Motivational State Questionnaire for 3032 unique participants.
- `tai`: Trait Anxiety data from the PMC lab matching the sai sample. 3032 unique subjects.
- `sai`: State Anxiety data from the PMC lab over multiple occasions. 3032 unique subjects.
- `sai.dictionary`: items used in the sai.
- `spi`: 4000 cases from the SAPA Personality Inventory including an item dictionary and scoring keys.
- `spi.dictionary`: The items for the spi.
- `spi.keys`: Scoring keys for the spi.
Historically interesting data sets

**burt**
11 emotional variables from Burt (1915)

**galton**
Galton's Mid parent child height data

**heights**
A data.frame of the Galton (1888) height and cubit data set

**cubits**
Galton's example of the relationship between height and cubit or forearm length

**peas**
Galton's Peas

**cushny**
The data set from Cushny and Peebles (1905) on the effect of three drugs on hours of sleep, used by Student (1908)

**holzinger.swineford**
26 cognitive variables + 7 demographic variables for 301 cases from Holzinger and Swineford.

Miscellaneous example data sets

**blant**
A 29 x 29 matrix that produces weird factor analytic results

**blot**
Bonds Logical Operations Test - BLOT

**cities**
Distances between 11 US cities and their geographical location

**city.location**

**income**
US family income from US census 2008

**all.income**
US family income from US census 2008

**neo**
NEO correlation matrix from the NEO_PI_R manual

**Schutz**
The Schutz correlation matrix example from Shapiro and ten Berge

**Spengler**
The Spengler and Damian correlation matrix example from Spengler, Damian and Roberts (2018)

**Damian**
Another correlation matrix from Spengler, Damian and Roberts (2018)

**usaf**
A correlation of 17 body size (anthropometric) measures from the US Air Force. Adapted from the Anthropometric package.

**veg**
Paired comparison of preferences for 9 vegetables (scaling example)

Functions to convert various objects to latex

**fa2latex**
Convert a data frame, correlation matrix, or factor analysis output to a LaTeX table

**df2latex**
Convert a data frame, correlation matrix, or factor analysis output to a LaTeX table

**ICC2latex**
Convert an ICC analysis output to a LaTeX table

**irt2latex**
Convert an IRT analysis output to a LaTeX table

**cor2latex**
Convert a correlation matrix output to a LaTeX table

**omega2latex**
Convert a data frame, correlation matrix, or factor analysis output to a LaTeX table

File manipulation functions

- **fileCreate**
  Create a file

- **fileScan**
  Show the first few lines of multiple files

- **filesInfo**
  Show the information for all files in a directory

- **filesList**
  Show the names of all files in a directory

**dfOrder**
Sorts a data frame

File input/output functions

- **read.clipboard**
  Shortcuts for reading from the clipboard or a file
read.file

Example

psych::describe(ability)

read.file

Usage

read.file(file=NULL, header=TRUE, use.value.labels=FALSE, to.data.frame=TRUE, sep="","quote="\"", widths=NULL, f=NULL, filetype=NULL,...)
# for .txt, .text, TXT, .dat, .DATA, .data, .csv, .rda, .xpt, XPT, or .sav (i.e., data from SPSS sav files may be read as can files saved by SAS using the .xpt option). Data exported by JMP or EXCEL in the csv format are also able to be read. Fixed Width Files saved in .txt mode may be read if the widths parameter is specified. Files saved with writeRDS have suffixes of .rds or Rds, and are read using readRDS. Files associated with objects with suffixes .rda and .Rda are loaded (following a security prompt). The default values for read.spss are adjusted for more standard input from SPSS files. Input from the clipboard is easy but a bit obscure, particularly for Mac users. read.clipboard and its variations are just an easier way to do so. Data may be copied to the clipboard from Excel spreadsheets, csv files, or fixed width formatted files and then into a data.frame. Data may also be read from lower (or upper) triangular matrices and filled out to square matrices. Writing text files may be done using write.file which will prompt for a file name (if not given) and then write or save to that file depending upon the suffix (text, txt, or csv will call write.table, R, or r will dput, rda, Rda will save, Rds, rds will saveRDS).

Examples

psych::describe(ability)

read.file

Shortcut for reading from the clipboard or a file

Description

Input from a variety of sources may be read. data.frames may be read from files with suffixes of .txt, .text, TXT, .dat, .DATA, .data, .csv, .rda, .xpt, XPT, or .sav (i.e., data from SPSS sav files may be read as can files saved by SAS using the .xpt option). Data exported by JMP or EXCEL in the csv format are also able to be read. Fixed Width Files saved in .txt mode may be read if the widths parameter is specified. Files saved with writeRDS have suffixes of .rds or Rds, and are read using readRDS. Files associated with objects with suffixes .rda and .Rda are loaded (following a security prompt). The default values for read.spss are adjusted for more standard input from SPSS files. Input from the clipboard is easy but a bit obscure, particularly for Mac users. read.clipboard and its variations are just an easier way to do so. Data may be copied to the clipboard from Excel spreadsheets, csv files, or fixed width formatted files and then into a data.frame. Data may also be read from lower (or upper) triangular matrices and filled out to square matrices. Writing text files may be done using write.file which will prompt for a file name (if not given) and then write or save to that file depending upon the suffix (text, txt, or csv will call write.table, R, or r will dput, rda, Rda will save, Rds, rds will saveRDS).

Usage

read.file(file=NULL, header=TRUE, use.value.labels=FALSE, to.data.frame=TRUE, sep="","quote="\"", widths=NULL, f=NULL, filetype=NULL,...)
# for .txt, .text, TXT, .dat, .DATA, .data, .csv, .xpt, XPT, or .sav (i.e., data from SPSS sav files may be read as can files saved by SAS using the .xpt option). Data exported by JMP or EXCEL in the csv format are also able to be read. Fixed Width Files saved in .txt mode may be read if the widths parameter is specified. Files saved with writeRDS have suffixes of .rds or Rds, and are read using readRDS. Files associated with objects with suffixes .rda and .Rda are loaded (following a security prompt). The default values for read.spss are adjusted for more standard input from SPSS files. Input from the clipboard is easy but a bit obscure, particularly for Mac users. read.clipboard and its variations are just an easier way to do so. Data may be copied to the clipboard from Excel spreadsheets, csv files, or fixed width formatted files and then into a data.frame. Data may also be read from lower (or upper) triangular matrices and filled out to square matrices. Writing text files may be done using write.file which will prompt for a file name (if not given) and then write or save to that file depending upon the suffix (text, txt, or csv will call write.table, R, or r will dput, rda, Rda will save, Rds, rds will saveRDS).

Examples

psych::describe(ability)
read.clipboard.lower(diag=TRUE, names=FALSE,...)
read.clipboard.upper(diag=TRUE, names=FALSE,...)

#read in data using a fixed format width (see read.fwf for instructions)
read.clipboard.fwf(header=FALSE, widths=rep(1,10),...)

read.https(filename, header=TRUE)
read.file.csv(file=NULL, header=TRUE, f=NULL,...)

#For output:
write.file(x, file=NULL, row.names=FALSE, f=NULL,...)
write.file.csv(x, file=NULL, row.names=FALSE, f=NULL,...)

Arguments

header Does the first row have variable labels (generally assumed to be TRUE).
sep What is the designated separator between data fields? For typical csv files, this will be a comma, but if commas designate decimals, then a ; can be used to designate different records.
quote Specified to
diag for upper or lower triangular matrices, is the diagonal specified or not
names for read.clipboard.lower or upper, are colnames in the first column
widths how wide are the columns in fixed width input. The default is to read 10 columns of size 1.
filename Name or address of remote https file to read.
... Other parameters to pass to read
f A file name to read from or write to. If omitted, file.choose is called to dynamically get the file name.
file A file name to read from or write to. (same as f, but perhaps more intuitive) If omitted and f is omitted, then file.choose is called to dynamically get the file name.
x The data frame or matrix to write to f
row.names Should the output file include the rownames? By default, no.
to.data.frame Should the spss input be converted to a data frame?
use.value.labels Should the SPSS input values be converted to numeric?
filetype If specified the reading will use this term rather than the suffix.

Details

A typical session of R might involve data stored in text files, generated online, etc. Although it is easy to just read from a file (particularly if using read.file, an alternative is to use one's local system to copy from the file to the clipboard and then read from the clipboard read.clipboard. This
is very convenient (and somewhat more intuitive to the naive user). This is particularly convenient when copying from a text book or article and just moving a section of text into R.) However, copying from a file and then reading the clipboard is hard to automate in a script. Thus, code

read.file will read from a file.

The `read.file` function combines the `file.choose` and either `read.table`, `read.fwf`, `read.spss` or `read.xport` (from foreign) or `load` or `readRDS` commands. By examining the file suffix, it chooses the appropriate way to read the file. For more complicated file structures, see the foreign package. For even more complicated file structures, see the rio or haven packages.

Note that `read.file` assumes by default that the first row has column labels (header = TRUE). If this is not true, then make sure to specify header = FALSE. If the file is fixed width, the assumption is that it does not have a header field. In the unlikely case that a fwf file does have a header, then you probably should try fn <- file.choose() and then my.data <- read.fwf(fn, header=TRUE, widths=widths)

Further note: If the file is a .Rda, .rda, etc. file, the read.file command will return the name and location of the file. It will prompt the user to load this file. In this case, it is necessary to either assign the output (the file name) to an object that has a different name than any of the objects in the file, or to call read.file() without any specification. Notice that loading an .Rda file can overwrite existing objects. Thus the warning and the need to do the second step.

If the file has no suffix the default action is to quit with a warning. However, if the filetype is specified, it will use that type in the reading (e.g. filetype="txt" will read as text file, even if there is no suffix.)

If the file is specified and has a prefix of http:// https:// it will be downloaded and then read.

Currently supported input formats are

- `.sav` SPSS.sav files
- `.csv` A comma separated file (e.g. from Excel or Qualtrics)
- `.txt` A typical text file
- `.TXT` A typical text file
- `.text` A typical text file
- `.data` A data file
- `.dat` A data file
- `.rds` A R data file
- `.Rds` A R data file (created by a write)
- `.Rda` A R data structure (created using save)
- `.rda` A R data structure (created using save)
- `.RData` A R data structure (created using save)
- `.data` A R data structure (created using save)
- `.R` A R data structure created using dput
- `.r` A R data structure created using dput
- `.xpt` A SAS data file in xport format
- `.XPT` A SAS data file in XPORT format

Some data files have an extra ' in the data (e.g. the NYT covid data base). These files can be read specifying quote ""

The foreign function `read.spss` is used to read SPSS `.sav` files using the most common options.
Just as `read.spss` issues various warnings, so does `read.file`. In general, these can be ignored. For more detailed information about using `read.spss`, see the help pages in the foreign package.

If you have a file written by JMP, you must first export to a csv or text file.

The `write.file` function combines the `file.choose` and either `write.table` or `saveRDS`. By examining the file suffix, it chooses the appropriate way to write. For more complicated file structures, see the foreign package, or the save function in R Base. If no suffix is added, it will write as a .txt file. `write.file.csv` will write in csv format to an arbitrary file name.

Currently supported output formats are

- `.csv` A comma separated file (e.g. for reading into Excel)
- `.txt` A typical text file
- `.text` A typical text file
- `.rds` A R data file
- `.Rds` A R data file (created by a write)
- `.Rda` A R data structure (created using save)
- `.rda` A R data structure (created using save)
- `.R` A R data structure created using `dput`
- `.r` A R data structure created using `dput`

Many Excel based files specify missing values as a blank field. When reading from the clipboard, using `read.clipboard.tab` will change these blank fields to NA.

Sometimes missing values are specified as "." or "999", or some other values. These can be converted by the `read.file` command specifying what values are missing (e.g., `na = "."`). See the example for the reading from the remote mtcars.csv file.

`read.clipboard` was based upon a suggestion by Ken Knoblauch to the R-help listserv.

If the input file that was copied into the clipboard was an Excel file with blanks for missing data, then `read.clipboard.tab()` will correctly replace the blanks with NAs. Similarly for a csv file with blank entries, `read.clipboard.csv` will replace empty fields with NA.

`read.clipboard.lower` and `read.clipboard.upper` are adapted from John Fox’s `read.moments` function in the `sem` package. They will read a lower (or upper) triangular matrix from the clipboard and return a full, symmetric matrix for use by `factanal`, `fa`, `ICLUST`, `pca`, `omega`, etc. If the diagonal is false, it will be replaced by 1.0s. These two function were added to allow easy reading of examples from various texts and manuscripts with just triangular output.

Many articles will report lower triangular matrices with variable labels in the first column. `read.clipboard.lower` will handle this case. Names must be in the first column if `names=TRUE` is specified.

Other articles will report upper triangular matrices with variable labels in the first row. `read.clipboard.upper` will handle this. Note that labels in the first column will not work for `read.clipboard.upper`. The names, if present, must be in the first row.

Consider the following lower triangular matrix. To read it, copy it to the clipboard and `read.clipboard.lower(names=TRUE)`

```
A1 1.00
A2 -0.34 1.00
A3 -0.27 0.49 1.00
A4 -0.15 0.34 0.36 1.00
```
However, if the data are strung out e.g.,

```
-0.34
-0.27
-0.15
-0.18
0.03
0.49
0.34
0.39
0.09
0.36
0.50
0.10
0.31
0.09
0.12
```

Then one needs to read it using the `read.clipboard.upper(names=FALSE,diag=FALSE)` option.

`read.clipboard.fwf` will read fixed format files from the clipboard. It includes a patch to `read.fwf` which will not read from the clipboard or from remote file. See `read.fwf` for documentation of how to specify the widths.

### Value

the contents of the file to be read or of the clipboard.

### Author(s)

William Revelle

### Examples

```
#All of these functions are meant for interactive Input
#Because these are dynamic functions, they need to be run interactively and
# can not be run as examples.
#Thus they are not to be tested by CRAN

if(interactive()) {
  my.data <- read.file() #search the directory for a file and then read it.
  #return the result into an object
  #or, if the file is a rda, etc. file
  my.data <- read.file() #return the path and instructions of how to load
  # without assigning a value.
```
filesList() #search the system for a particular file and then list all the files in that directory
fileCreate() #search for a particular directory and create a file there.
write.file(Thurstone) #open the search window, choose a location and name the output file,
# write the data file (e.g., Thurstone) to the file chosen

#the example data set from read.delim in the readr package to read a remote csv file
my.data <-read.file(
  na=".") #the na option is used for an example, but is not needed for these data

#These functions read from the local clipboard and thus are interactive
my.data <- read.clipboard() #space delimited columns
my.data <- read.clipboard.csv() #, delimited columns
my.data <- read.clipboard.tab() #typical input if copied from a spreadsheet
my.data <- read.clipboard(header=FALSE) #data start on line 1
my.matrix <- read.clipboard.lower()

recode

<table>
<thead>
<tr>
<th>recode</th>
<th>Recode variable values to new values</th>
</tr>
</thead>
</table>

**Description**

Given a set of numeric codes, change their values to different values given a mapping function

**Usage**

recode(x, where, isvalue, newvalue)

**Arguments**

- **x**: A matrix or data frame of numeric values
- **where**: The column numbers to fix
- **isvalue**: A vector of values to change
- **newvalue**: A vector of the new values

**Details**

Sometime, data are entered as levels in an incorrect order. Once converted to numeric values, this can lead to confusion. recoding of the data to the correct order is straightforward, if tedious.

**Value**

The reordered data
Note

Perhaps useful, but definitely ugly code. For smaller data sets, the results from char2numeric back to the original will not work. char2numeric works column wise and orders the data in each column.

Author(s)

William Revelle

Examples

```r
temp <- bfi[1:100,1:5]
isvalue <- 1:6
newvalue <- psych::cs(one,two,three,four,five,six)
newtemp <- recode(temp,1:5,isvalue,newvalue)
newtemp  #characters
temp.num <- psych::char2numeric(newtemp)  #convert to numeric
temp.num  #notice the numerical values have changed
new.temp.num <- recode(temp.num, 1:5, isvalue=c(3,6,5,2,1,4), newvalue=1:6)
#note that because char2numeric works column wise, this will fail for small sets
```

sai  
State Anxiety data from the PMC lab over multiple occasions.

Description

State Anxiety was measured two-three times in 11 studies at the Personality-Motivation-Cognition laboratory. Here are item responses for 11 studies (9 repeated twice, 2 repeated three times). In all studies, the first occasion was before a manipulation. In some studies, caffeine, or movies or incentives were then given to some of the participants before the second and third STAI was given. In addition, Trait measures are available and included in the tai data set (3032 subjects).

Usage

data(sai)
data(tai)
data(sai.dictionary)

Format

A data frame with 3032 unique observations on the following 23 variables.

id  a numeric vector
study  a factor with levels ages cart fast fiat film flat home pat rob salt shedshop xray
time  1=First, 2= Second, 3=third administration
TOD  TOD (time of day 1= 8:50-9:30 am,2= 1=3 pm, 3= 7:-8pm
drug  drug (placebo (0) vs. caffeine (1))
film (1=Frontline (concentration camp), 2 = Halloween 3= National Geographic (control), 4- Parenthood (humor)

anxious anxious
at ease at ease
calm calm
comfortable comfortable
confident confident
content content
high strung high strung
jittery jittery
joyful joyful
nerveous nervous
pleasant pleasant
rattled over excited and rattled
regretful regretful
relaxed relaxed
rested rested
secure secure
tense tense
upset upset
worried worried
worrying worrying

Details

The standard experimental study at the Personality, Motivation and Cognition (PMC) laboratory (Revelle and Anderson, 1997) was to administer a number of personality trait and state measures (e.g. the epi, msq, msqR and sai) to participants before some experimental manipulation of arousal/effort/anxiety. Following the manipulation (with a 30 minute delay if giving caffeine/placebo), some performance task was given, followed once again by measures of state arousal/effort/anxiety.

Here are the item level data on the sai (state anxiety) and the tai (trait anxiety). Scores on these scales may be found using the scoring keys. The affect data set includes pre and post scores for two studies (flat and maps) which manipulated state by using four types of movies.

In addition to being useful for studies of motivational state, these studies provide examples of test-retest and alternate form reliabilities. Given that 10 items overlap with the msqR data, they also allow for a comparison of immediate duplication of items with 30 minute delays.

Studies CART, FAST, SHED, RAFT, and SHOP were either control groups, or did not experimentally vary arousal/effort/anxiety.

AGES, CITY, EMIT, RIM, SALT, and XRAY were caffeine manipulations between time 1 and 2 (RIM and VALE were repeated day 1 and day 2)
FIAT, FLAT, MAPS, MIXX, and THRU were 1 day studies with film manipulation between time 1 and time 2.

SAM1 and SAM2 were the first and second day of a two day study. The STAI was given once per day. MSQ not MSQR was given.

VALE and PAT were two day studies with the STAI given pre and post on both days.

RIM was a two day study with the STAI and MSQ given once per day.

Usually, time of day 1 = 8:50-9am am, and 2 = 7:30 pm, however, in rob, with paid subjects, the times were 0530 and 22:30.

Source

Data collected at the Personality, Motivation, and Cognition Laboratory, Northwestern University, between 1991 and 1999.

References


Examples

data(sai)

table(sai$study,sai$time) #show the counts for repeated measures

#Here are the keys to score the sai total score, positive and negative items
   sai.p = c("calm", "at.ease", "rested", "comfortable", "confident", "secure", "relaxed", "content", "joyful", "pleasant" ),
   sai.n = c("tense", "anxious", "nervous", "jittery", "rattled", "high.strung", "upset", "worrying", "worried", "regretful" )
)

   tai.pos = c("pleasant", "-wish.happy", "rested", "calm", "happy", "secure", "content", "steady" ),
   tai.n = c("tense", "anxious", "nervous", "jittery", "rattled", "high.strung", "upset", "worrying", "worried", "regretful" )
)
tai.neg = c("nervous", "not.satisfied", "failure","difficulties", "worry",  
"disturbing.thoughts","lack.self.confidence","decisive","inadequate" ,  
"thoughts.bother","disappointments","tension" )

#using the is.element function instead of the %in% function  
#just get the control subjects  
control <- subset(sai,is.element(sai$study,c("Cart", "Fast", "SHED", "RAFT", "SHOP")))

#pre and post drug studies  
drug <- subset(sai,is.element(sai$study, c("AGES","CITY","EMIT","SALT","VALE","XRAY")))

#pre and post film studies  
film <- subset(sai,is.element(sai$study, c("FIAT","FLAT", "MAPS", "MIXX") ))

#this next set allows us to score those sai items that overlap with the msq item sets  
msq.items <- c("anxious", "at.ease", "calm", "confident","content", "jittery",  
"nervous", "relaxed", "tense", "upset" ) #these overlap with the msq

sai.msq.keys <- list(pos =c("at.ease", "calm", "confident", "content","relaxed"),  
neg = c("anxious", "jittery", "nervous", "tense", "upset"),  
anx = c("anxious", "jittery", "nervous", "tense", "upset","-at.ease", "-calm",  
"-confident", "-content","-relaxed")

sai.not.msq.keys <- list(pos=c("secure","rested","comfortable","joyful","pleasant" ),  
neg=c("regretful","worrying","high.strung","worried","rattled"),  
anx = c("regretful","worrying","high.strung","worried","rattled", 
"-secure", "-rested", "-comfortable","-joyful","-pleasant")).

sai.alternate.forms <- list( pos1 =c("at.ease","calm","confident","content","relaxed"),  
neg1 = c("anxious", "jittery", "nervous", "tense", "upset"),  
anx1 = c("anxious", "jittery", "nervous", "tense", "upset","-at.ease", "-calm",  
"-confident", "-content","-relaxed"),  
pos2=c("secure","rested","comfortable","joyful","pleasant" ),  
neg2=c("regretful","worrying","high.strung","worried","rattled"),  
anx2 = c("regretful","worrying","high.strung","worried","rattled", 
"-secure", "-rested", "-comfortable","-joyful","-pleasant")).

sai.repeated <- c("AGES","Cart","Fast","FIAT","FILM","FLAT","HOME","PAT","RIM","SALT",  
"SAM","SHED","SHOP","VALE","XRAY")
sai12 <- subset(sai,is.element(sai$study, sai.repeated)) #the subset with repeated measures

#Choose those studies with repeated measures by:

sai.control <- subset(sai,is.element(sai$study, c("Cart", "Fast", "SHED", "SHOP")))  
sai.day <- subset(sai,is.element(sai$study, c("SAM", "RIM")))
Description

Self reported scores on the SAT Verbal, SAT Quantitative and ACT were collected as part of the Synthetic Aperture Personality Assessment (SAPA) web based personality assessment project. Age, gender, and education are also reported. The data from 700 subjects are included here as a demonstration set for correlation and analysis.

Usage

data(sat.act)

Format

A data frame with 700 observations on the following 6 variables.

- gender  males = 1, females = 2
- education self reported education 1 = high school ... 5 = graduate work
- age    age
- ACT    ACT composite scores may range from 1 - 36. National norms have a mean of 20.
- SATV   SAT Verbal scores may range from 200 - 800.
- SATQ   SAT Quantitative scores may range from 200 - 800

Details

These items were collected as part of the SAPA project (https://www.sapa-project.org/) to develop online measures of ability (Revelle, Wilt and Rosenthal, 2009). The score means are higher than national norms suggesting both self selection for people taking on line personality and ability tests and a self reporting bias in scores.

See also the iq.items data set.

Source

https://personality-project.org/

References


Examples

data(sat.act)
psych::describe(sat.act)
psych::pairs.panels(sat.act)
The Schutz correlation matrix example from Shapiro and ten Berge

Description

Shapiro and ten Berge use the Schutz correlation matrix as an example for Minimum Rank Factor Analysis. The Schutz data set is also a nice example of how normal minres or maximum likelihood will lead to a Heywood case, but minrank factoring will not.

Usage

data("Schutz")

Format

The format is: num [1:9, 1:9] 1 0.8 0.28 0.29 0.41 0.38 0.44 0.4 0.41 0.8 ... - attr(*, "dimnames")=List of 2

Details

These are 9 cognitive variables of importance mainly because they are used as an example by Shapiro and ten Berge for their paper on Minimum Rank Factor Analysis.

The solution from the \texttt{fa} function with the \texttt{fm='minrank'} option is very close (but not exactly equal) to their solution.

This example is used to show problems with different methods of factoring. Of the various factoring methods, \texttt{fm = "minres"}, "uls", or "mle" produce a Heywood case. Minrank, alpha, and pa do not.

See the blant data set for another example of differences across methods.

Source


References


Examples

data(Schutz)
psych::corPlot(Schutz,numbers=TRUE,upper=FALSE)

f4min <- psych::fa(Schutz,4,fm="minrank") #for an example of minimum rank factor Analysis
#compare to
f4 <- psych::fa(Schutz,4,fm="mle") #for the maximum likelihood solution which has a Heywood case
**selectBy**

Select a subset of rows (subjects) meeting one or more criteria for columns

**Description**

Select a subset of a data.frame or matrix for columns meeting specific criteria. Can do logical AND (default) or OR of the resulting search. Columns (variables) are specified by name and the conditions to meet include equality, less than, more than or inequality to a specified set of values. SplitBy creates new dichotomous variables based on the splitting criteria.

**Usage**

```r
selectBy(x, by)
splitBy(x, by, new=FALSE)
```

**Arguments**

- `x`: A data frame or matrix
- `by`: A quote delimited string of variables and criteria values. Multiple variables may be separated by commas (default to AND)
- `new`: If true, return a new data frame with just the dichotomous variables otherwise concatenate the new variables to the right margin of `x`

**Details**

Two relatively trivial functions to help those less familiar with the subset function or how to use `[]` to select variables.

**Value**

The subset of the original data.frame with just the cases that meet the criteria (`selectBy`) or new variables, recoded 0,1

selectBy is equivalent to subsetting x by an x value: small <- x[x[by=criterion]] or the subset function small <- subset(x, x$variable == value)

**Author(s)**

William Revelle

**See Also**

vJoin for another data manipulation function.
Examples

```r
  testand <- selectBy(attitude, 'rating < 70 & complaints > 60') # AND
dim(testand)
  testor <- selectBy(attitude, 'rating < 60 | complaints > 60') # OR
dim(testor)
  test <- splitBy(attitude, 'rating > 70 , complaints > 60')
  psych::headTail(test)
```

Description

Project Talent gave 440,000 US high school students a number of personality and ability tests. Of these, the data for 346,000 were available for followup. Subsequent followups were collected 11 and 50 years later. Marion Spengler and her colleagues Rodica Damian, and Brent Roberts reported on the stability and change across 50 years of personality and ability. Here is the correlation matrix of 25 of their variables (Spengler) as well as a slightly different set of 19 variables (Damian). This is a nice example of mediation and regression from a correlation matrix.

Usage

data("Damian")

Format

A 25 x 25 correlation matrix of demographic, personality, and ability variables, based upon 346,660 participants.

- Race/Ethnicity: 1 = other, 2 = white/caucasian
- Sex: 1 = Male, 2 = Female
- Age: Cohort = 9th grade, 10th grade, 11th grade, 12th grade
- Parental: Parental SES based upon 9 questions of home value, family income, etc.
- IQ: Standardized composite of Verbal, Spatial and Mathematical
- Sociability etc.: 10 scales based upon prior work by Damian and Roberts
- Maturity: A higher order factor from the prior 10 scales
- Extraversion: The second higher order factor
- Interest: Self reported interest in school
- Reading: Self report reading skills
- Writing: Self report writing skills
- Responsible: Self reported responsibility scale
- Ed.11: Education level at 11 year followup
- Educ.50: Education level at 50 year followup
Occupational Prestige at 11 year followup
Occupational Prestige at 50 year followup
Income at 11 year followup
Income at 50 year followup

Details
Data from Project Talent was collected in 1960 on a representative sample of American high school students. Subsequent follow up 11 and 50 years later are reported by Spengler et al (2018) and others.

Source
Marion Spengler, supplementary material to Damian et al. and Spengler et al.

References

Marian Spengler and Rodica Ioana Damian and Brent W. Roberts (2018), How you behave in school predicts life success above and beyond family background, broad traits, and cognitive ability Journal of Personality and Social Psychology, 114 (4) 600-636

Examples
data(Damian)
Spengler.stat #show the basic descriptives of the original data set
psych::lowerMat(Spengler[psych::cs(IQ,Parental,Ed.11,OccPres.50),
psych::cs(IQ,Parental,Ed.11,OccPres.50)])
psych::setCor(OccPres.50 ~ IQ + Parental + (Ed.11),data=Spengler)
#we reduce the number of subjects for faster replication in this example
mod <- psych::mediate(OccPres.50 ~ IQ + Parental + (Ed.11),data=Spengler,
n.iter=50,n.obs=1000) #for speed
summary(mod)

spi
A sample from the SAPA Personality Inventory including an item dictionary and scoring keys.

Description
The SPI (SAPA Personality Inventory) is a set of 135 items primarily selected from International Personality Item Pool (ipip.ori.org). This is an example data set collected using SAPA procedures the sapa-project.org web site. This data set includes 10 demographic variables as well. The data set with 4000 observations on 145 variables may be used for examples in scale construction and validation, as well as empirical scale construction to predict multiple criteria.
Usage

data("spi")
data(spi.dictionary)
data(spi.keys)

Format

A data frame with 4000 observations on the following 145 variables. (The q numbers are the SAPA item numbers).

age  Age in years from 11 -90
sex  Reported biological sex (coded by X chromosomes => 1=Male, 2 = Female)
health  Self rated health 1-5: poor, fair, good, very good, excellent
p1edu  Parent 1 education
p2edu  Parent 2 education
education Respondents education: less than 12, HS grad, current univ, some univ, associate degree, college degree, in grad/prof, grad/prof degree
wellness  Self rated "wellness" 1-2
exer  Frequency of exercise: very rarely, < 1/month, < 1/wk, 1 or 2 times/week, 3-5/wk, > 5 times/week
smoke  never, not last year, < 1/month, <1/week, 1-3 days/week, most days, up to 5 x /day, up to 20 x /day, > 20x/day
ER  Emergency room visits none, 1x, 2x, 3 or more times
q_253  see the spi.dictionary for these items (q_253
q_1328  see the dictionary for all items q_1328)

Details

Using the data contributed by about 125,000 visitors to the https://www.SAPA-project.org/ website, David Condon has developed a hierarchical framework for assessing personality at two levels. The higher level has the familiar five factors that have been studied extensively in personality research since the 1980s – Conscientiousness, Agreeableness, Neuroticism, Openness, and Extraversion. The lower level has 27 factors that are considerably more narrow. These were derived based on administrations of about 700 public-domain IPIP items to 3 large samples. Condon describes these scales as being "empirically-derived" because relatively little theory was used to select the number of factors in the hierarchy and the items in the scale for each factor (to be clear, he means relatively little personality theory though he relied on quite a lot of sampling and statistical theory). You can read all about the procedures used to develop this framework in his book/manual. If you would like to reproduce these analyses, you can download the data files from Dataverse (links are also provided in the manual) and compile this script in R (he used knitR). Instructions are provided in the Preface to the manual.

The content of the spi items may be seen by examining the spi.dictionary. Included in the dictionary are the item_id number from the SAPA project, the wording of the item, the source of the item, which Big 5 scale the item marks, and which "Little 27" scale the item marks.

This small subset of the data is provided for demonstration purposes.
Source


References


An analysis using the spi data set and various tools from the psych package may be found at


Examples

data(spi)
data(spi.dictionary)
psych::bestScales(spi, criteria="health",dictionary=spi.dictionary)

sc <- psych::scoreVeryFast(spi.keys,spi) #much faster scoring for just scores
sc <- psych::scoreOverlap(spi.keys,spi) #gives the alpha reliabilities and various stats #these are corrected for overlap
psych::corPlot(sc$corrected,numbers=TRUE,cex=.4,xlas=2,min.length=6,
               main="Structure of SPI (Corrected for overlap) disattenuated r above the diagonal")

---

**usaf**

17 anthropometric measures from the USAF showing a general factor

Description

The correlation matrix of 17 anthropometric measures from the United States Air Force survey of 2420 airmen. The data are taken from the Anthropometry package and included here as a demonstration of a hierarchical factor structure suitable for analysis by the omega or omegaSem.

Usage

data("USAF")

Format

The format is: num [1:17, 1:17] 1 0.1148 -0.0309 -0.028 -0.0908 ... - attr(*, "dimnames")=List of 2
..$ : chr [1:17] "age" "weight" "grip" "height" ...
..$ : chr [1:17] "age" "weight" "grip" "height" ...
Details

The original data were collected by the USAF and reported in Churchill et al, 1977. They are included as a data file of 2420 participants and 202 variables (the first being an id) in the Anthropometry package. The list of variable names may be found in Churchill et al, on pages 96-99.

The three (correlated) factor structure shows a clear height, bulk, and head size structure with an overall general factor (g) which may be interpreted as body size.

The variables included (and their variable numbers in Antropometry) are:
<table>
<thead>
<tr>
<th>Variable</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>V1</td>
</tr>
<tr>
<td>weight</td>
<td>V2</td>
</tr>
<tr>
<td>grip strength</td>
<td>V12</td>
</tr>
<tr>
<td>height (stature)</td>
<td>V13</td>
</tr>
<tr>
<td>leg length</td>
<td>V26</td>
</tr>
<tr>
<td>knee height</td>
<td>V37</td>
</tr>
<tr>
<td>upper arm</td>
<td>V42</td>
</tr>
<tr>
<td>thumb tip reach</td>
<td>V47</td>
</tr>
<tr>
<td>in sleeve</td>
<td>V49</td>
</tr>
<tr>
<td>chest breadth</td>
<td>V52</td>
</tr>
<tr>
<td>hip breadth</td>
<td>V55</td>
</tr>
<tr>
<td>waist circumference</td>
<td>V71</td>
</tr>
<tr>
<td>thigh circumference</td>
<td>V97</td>
</tr>
<tr>
<td>scye circumference</td>
<td>V103</td>
</tr>
<tr>
<td>head circumference</td>
<td>V141</td>
</tr>
<tr>
<td>bitragion coronal</td>
<td>V145</td>
</tr>
<tr>
<td>head length</td>
<td>V150</td>
</tr>
<tr>
<td>glabella to wall</td>
<td>V181</td>
</tr>
<tr>
<td>external canthus to wall</td>
<td>V183</td>
</tr>
</tbody>
</table>

Note that these numbers are equivalent to the numbers in Churchill et al. The numbers in Anthropometry are these + 1.

**Source**


**References**


**Examples**

data(USAF)
psych::corPlot(USAF,xlas=3)
psych::omega(USAF[,4:19],c(4:10,19)) #just the size variables
Utility

Useful utility functions for file/directory exploration and manipulation.

Description

Wrappers for dirname, file.choose, readLines, file.create, file.path to be called directly for listing directories, creating files, showing the files in a directory, and listing the content of files in a directory. fileCreate gives the functionality of file.choose(new=TRUE). filesList combines file.choose, dirname, and list.files to show the files in a directory, fileScan extends this and then returns the first few lines of each readable file.

Usage

fileScan(f = NULL, nlines = 3, max = NULL, from = 1, filter = NULL)
filesList(f=NULL)
filesInfo(f=NULL,max=NULL)
fileCreate(newName="new.file")

Arguments

f File path to use as base path (will use file.choose() if missing. If f is a directory, will list the files in that directory, if f is a file, will find the directory for that file and then list all of those files.)
nlines How many lines to display
max maximum number of files to display
from First file (number) to display
filter Just display files with "filter" in the name
newName The name of the file to be created.

Details

Just a collection of simple wrappers to powerful core R functions. Allows the user more direct control of what directory to list, to create a file, or to display the content of files. The functions called include file.choose, file.path, file.info, file.create, dirname, and dir.exists. All of these are very powerful functions, but not easy to call interactively.

fileCreate will ask to locate a file using file.choose, set the directory to that location, and then prompt to create a file with the new.name. This is a workaround for file.choose(new=TRUE) which only works for Macs not using R.studio.

filesInfo will interactively search for a file and then list the information (size, date, ownership) of all the files in that directory.

filesList will interactively search for a file and then list all the files in same directory.

Note

Work arounds for core-R functions for interactive file manipulation.
Author(s)

William Revelle

See Also

read.file to read in data from a file or read.clipboard from the clipboard. dfOrder to sort data.frames.

Examples

if(interactive()) {
  # all of these require interactive input and thus are not given as examples

  fileCreate("my.new.file.txt")
  filesList()  # show the items in the directory where a file is displayed
  fileScan()  # show the content of the files in a directory
  # or, if you have a file in mind
  f <- file.choose()  # go find it
  filesList(f)
  fileScan(f)
}

| vegetables | Paired comparison of preferences for 9 vegetables |

Description

A classic data set for demonstrating Thurstonian scaling is the preference matrix of 9 vegetables from Guilford (1954). Used by Guilford, Nunnally, and Nunnally and Bernstein, this data set allows for examples of basic scaling techniques.

Usage

data(vegetables)

Format

A data frame with 9 choices on the following 9 vegetables. The values reflect the percentage of times where the column entry was preferred over the row entry.

Turn Turnips  
Cab Cabbage  
Beet Beets  
Asp Asparagus
Louis L. Thurstone was a pioneer in psychometric theory and measurement of attitudes, interests, and abilities. Among his many contributions was a systematic analysis of the process of comparative judgment (Thurstone, 1927). He considered the case of asking subjects to successively compare pairs of objects. If the same subject does this repeatedly, or if subjects act as random replicates of each other, their judgments can be thought of as sampled from a normal distribution of underlying (latent) scale scores for each object. Thurstone proposed that the comparison between the value of two objects could be represented as representing the differences of the average value for each object compared to the standard deviation of the differences between objects. The basic model is that each item has a normal distribution of response strength and that choice represents the stronger of the two response strengths. A justification for the normality assumption is that each decision represents the sum of many independent inputs and thus, through the central limit theorem, is normally distributed.

Thurstone considered five different sets of assumptions about the equality and independence of the variances for each item (Thurston, 1927). Torgerson expanded this analysis slightly by considering three classes of data collection (with individuals, between individuals and mixes of within and between) crossed with three sets of assumptions (equal covariance of decision process, equal correlations and small differences in variance, equal variances).

This vegetable data set is used by Guilford and by Nunnally to demonstrate Thurstonian scaling.

**Source**


**References**


**See Also**

thurstone

**Examples**

data(vegetables)
psych::thurstone(veg)
vJoin

Combine two matrices or data frames into one based upon variable labels

Description
A typical problem in data analysis is to combine two data sets into one. vJoin will combine two matrices or data.frames into one data.frame. Unique column names from set 1 and set 2 are combined as are unique rows. Column names can differ, as can row names. Basically an extension of rbind and cbind without the requirement of matching column and row names. combineMatrices solves a similar problem for correlation matrices.

Usage

vJoin(x, y, rnames = TRUE, cnames=TRUE)
combineMatrices(x,y, r=NULL)

Arguments

x a matrix or data frame with column and row names.
y a matrix or data frame with column and row names
rnames If TRUE, the default, match on row names, extend to new names. If FALSE then add the y data following the x data.
cnames If TRUE colnames are NULL then create unique colnames for x and y
r should we add the diagonal of y?

Details
For an X and Y matrices/data.frames with column and row names, combine the two data sets. Match on column and row names if they exist, extend to unique names if they do not match.
Matrices by default do not have column or rownames. They will be created for x and for y (depending upon the rnames and cnames options).
combineMatrices takes a square matrix (x) and combines with a rectangular matrix y to produce a larger xy matrix.

Value

xy: a data frame

Note
Inspired by the functionality of full_join and the other related dplyr functions.

Author(s)
William Revelle
Examples

X1 <- bfi[1:10,1:5]
Y1 <- bfi[6:15,4:10]
xy <- vJoin(X1,Y1) # match on rownames
xy1 <- vJoin(X1,Y1,rnames=FALSE) # add Y1 items after X1 items

x <- matrix(1:30, ncol=5)
y <- matrix(1:40, ncol=8)
vJoin(x,y)
vJoin(x,y,cnames=FALSE)
vJoin(x,y, rnames= FALSE, cnames=FALSE)

R <- cor(sat.act,use="pairwise")
r1 <- R[1:4,1:4]
r2 <- R[1:4,5:6]
newr <- combineMatrices(r1,r2)

zola  
Correlation matrix of 135 self report and 30 peer report personality items

Description

Zola et al., (2021) reported the validity of self report personality items from the SAPA personality inventory (SPI) (Condon, 2018) in terms of 30 peer reports on 8 dimensions. Here are the polychoric correlations of these items. spi items were collected using SAPA procedures for 158,631 participants (mean n/item = 18,180), 908 of whom received peer ratings.

Usage

data("zola")

Format

The format is: num [1:165, 1:165] 1 -0.242 0.282 0.65 0.223 ... - attr(*, "dimnames")=List of 2 ..$ : chr [1:165] "q_253" "q_4296" "q_1855" "q_90" ... ..$: chr [1:165] "q_253" "q_4296" "q_1855" "q_90" ...

Details

The polychoric correlation matrix of the spi and peer report data. To see the item labels, use the lookupFromKeys .

This data set is a nice example of a multi-trait, multi-method correlation matrix. (see the scoring example). Five dimensions of self report show high correlations with the corresponding peer report scales.
Source

References

Examples
```r
data(zola)
psych::lookupFromKeys(zola.keys, zola.dictionary)
scores <- psych::scoreOverlap(zola.keys[1:5, 33:37], zola) # MTMM of Big 5
scores
```
Index

* IO
  - read.file, 67
  - Utility, 86
  - vJoin, 89
* datasets
  - ability, 3
  - affect, 5
  - Athenstaedt, 6
  - bfi, 10
  - BFI.adjectives.dictionary, 12
  - big5.100.adjectives, 14
  - blot, 15
  - burt, 17
  - cities, 19
  - colom, 20
  - cubits, 21
  - cushny, 22
  - eminence, 28
  - epi, 29
  - epi.bfi, 32
  - galton, 33
  - GERAS, 34
  - globalWarm, 37
  - heights, 39
  - holzinger.swineford, 40
  - income, 44
  - iqitems, 45
  - msq, 47
  - msqR, 53
  - neo, 61
  - peas, 63
  - Pollack, 64
  - psychTools, 65
  - sai, 73
  - sat.act, 76
  - Schutz, 78
  - Spengler, 80
  - spi, 81
  - usafl, 83
  - vegetables, 87
  - zola, 90
* file
  - Utility, 86
* manip
  - dfOrder, 26
* multivariate
  - read.file, 67
  - recode, 72
  - selectBy, 79
  - vJoin, 89
* utilities
  - df2latex, 24
  - dfOrder, 26

ability, 3, 46, 65
affect, 5, 13, 50, 51, 56–58, 65, 74
all.income (income), 44
alpha, 31
Athenstaedt, 6, 36

bfi, 10, 65
bfi.adjectives (big5.100.adjectives), 14
BFI.adjectives.dictionary, 12
bfi.adjectives.dictionary
  - (BFI.adjectives.dictionary), 12
bi.bars, 12
big5.100.adjectives, 13, 14
big5.adjectives.dictionary
  - (BFI.adjectives.dictionary), 12
big5.adjectives.keys
  - (big5.100.adjectives), 14
blant, 15, 66
blot, 16, 66
burt, 17, 66

cities, 19, 66
city.location (cities), 19
cohen.d, 34
INDEX

colom, 20, 21
combineMatrices (vJoin), 89
cor.smooth, 18
cor2latex, 66
cor2latex (df2latex), 24
cubits, 21, 34, 39, 64, 66
cushny, 22, 66

Damian, 66
Damian (Spengler), 80
df2latex, 24, 66
dfOrder, 26, 26, 27, 66, 87
dir.exists, 86
dirname, 86

eclipses, 22, 39
eminence, 28
epi, 29, 65, 74
epi.bfi, 32, 65
epi.dictionary, 31
epiR, 65
epiR (epi), 29
estress (Pollack), 64

fa, 15, 25, 52, 58, 70, 78
fa.extension, 52, 58
fa2latex, 66
fa2latex (df2latex), 24
file.choose, 68–70, 86
file.create, 86
file.info, 86
file.path, 86
fileCreate, 27, 66, 86
fileCreate (Utility), 86
fileScan, 27, 86
fileScan (Utility), 86
filesInfo, 27, 86
filesInfo (Utility), 86
filesList, 27, 86
filesList (Utility), 86
flat (affect), 5
galton, 22, 33, 39, 64, 66
GERAS, 9, 34
glbwarm (globalWarm), 37
globalWarm, 37

Harman, 18
Harman.Burt, 18

heights, 21, 22, 34, 39, 39, 64, 66
holzinger.dictionary
(holzinger.swineford), 40
holzinger.raw, 42
holzinger.raw (holzinger.swineford), 40
holzinger.swineford, 40, 66

ICC2latex, 66
ICC2latex (df2latex), 24
ICLUST, 70
income, 44, 66
iqitems, 3, 45, 65
irt.fa, 3, 4, 11, 12, 17, 25, 46
irt2latex, 66
irt2latex (df2latex), 24

load, 69
lookupFromKeys, 90

make.keys, 11, 52, 58
maps (affect), 5
msq, 5, 47, 57, 58, 65, 74
msg.keys (msqR), 53
msqR, 6, 13, 31, 53, 58, 65, 74

neo, 61, 66

omega, 4, 20, 31, 46, 70, 83
omega2latex (df2latex), 24
omegaSem, 83
order, 26

pca, 70
peas, 22, 34, 63, 66
Pollack, 64
polychoric, 11
psych, 65
psychTools, 65

read.clipboard, 27, 65–68, 70, 87
read.clipboard (read.file), 67
read.clipboard.lower, 70
read.clipboard.tab, 70
read.clipboard.upper, 70
read.file, 27, 65, 67, 67, 68–70, 87
read.fwf, 69
read.https (read.file), 67
read.spss, 69, 70
read.table, 69
read.xport, 69
INDEX

readRDS, 69
recode, 72

sai, 31, 53, 58, 65, 73, 74
sat.act, 76
saveRDS, 70
scatterHist, 34
Schutz, 66, 78
score.items, 11
scoreItems, 52, 58
scoreOverlap, 52, 58
selectBy, 79, 79
sim.irt, 27
Spengler, 66, 80
spengler (Spengler), 80
Spengler.stat (Spengler), 80
spi, 65, 81
splitBy (selectBy), 79

table2df, 22, 39
table2matrix, 22, 39
tai, 58, 65, 74
tai (sai), 73
test.irt, 27
testRetest, 31, 58
tetrachoric, 4, 46
thurstone, 88

USAF (usaf), 83
usaf, 66, 83
Utility, 86

veg (vegetables), 87
vegetables, 87
vJoin, 79, 89

write.file, 67, 70
write.file (read.file), 67
write.file.csv, 70
write.table, 70

zola, 90