Factor analysis: Some simple examples: Homework for 405: Introduction to psychometric theory

William Revelle
Northwestern University

Three problems

You can do all of these problems using the *psych* package (Revelle, 2022) and some of its functions. For more information about doing factor analysis, see the *How To* at http://personality-project.org/r/psych/HowTo/factor.pdf.

The ability data set

16 multiple choice ability items were given to 1525 subjects using the principals of Synthetic Aperture Personality Assessment (SAPA) http://sapa-project.org web based personality assessment project (Revelle et al., 2016) and are saved as the `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 and are saved as the `ablity` data set as well as the file on the class server at http://personality -project.org/revelle/syllabi/405/ability.txt. (Note, these data are also available as the `ability` data set. The option of reading a remote file is given to be more realistic.)

1. Get the data set from the remote server and `describe` it.
2. Show the correlations using `corPlot`
3. Find the scree plot of the data using `scree`
4. Estimate the number of factors using `fa.parallel`
5. Compare that to what you get when you use `nfactors`
6. Compare that result when you specify the correlation = "tet" for tetrachoric option.
7. Extract 1, 2, 3, and 4 factors

---

8. Examine the `fa.diagram` of your solutions

9. Examine all of these results and try to interpret them

The Big 5 data set

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 http://sapa-project.org. 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. The data are at http://personality-project .org/revelle/syllabi/405/bfi.txt. (Note, these data are also available as the `bfi` data set. The option of reading a remote file is given to be more realistic.)

1. Get the data set from the remote server and `describe` it.

2. Find the scree plot of the data using `scree`

3. Estimate the number of factors using `fa.parallel`

4. Compare that to what you get when you use `nfactors`

5. Compare that result when you specify the correlation = "poly" for polychoric

6. Extract 4, 5 and 6 factors

7. Label the factor analysis output using the `fa.lookup` function with the `bfi.dictionary`. You only need columns 1 and 2 of the bfi.dictionary.

8. Examine the `fa.diagram` of your solutions

9. Examine all of these results and try to interpret them

The Motivational State Qustionnaire

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. The data were collected as pretest information during multiple studies in the Personality-Motivation-Cognition lab at Northwestern.

he Motivational States Questionnaire (MSQ) is composed of 72 items, which represent the full affective range (Revelle & Anderson, 1997). 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 et al., 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 tfour point rating scale.

The original version of the MSQ included 72 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 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)

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.

1. Get the data set `msq` and `describe` it.
2. Use just the first 72 variables.
3. Find the scree plot of the data using `scree`

4. Estimate the number of factors using `fa.parallel`

5. Compare that to what you get when you use `nfactors`

6. Compare that result when you specify the correlation = "poly" for polychoric

7. Extract 2 factors

8. Examine the `fa.diagram` of your solutions

9. Better yet, use the fa.plot function to show the data (use labels = colnames(msq)[1:72], show.points=FALSE)

10. Examine all of these results and try to interpret them

References

Revelle, W. (2022). psych: Procedures for psychological, psychometric, and personality research (2.2.3 ed.) [Computer software manual]. psych. (R package version 2.2.3)

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

Revelle, W., Condon, D. M., Wilt, J., French, J. A., Brown, A., & Elleman, L. G. (2016). Web and phone based data collection using planned missing designs. In N. G. Fielding, R. M. Lee, & G. Blank (Eds.), *Sage handbook of online research methods* (2nd ed., p. 578-595). Sage Publications, Inc.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063-1070. doi: 10.1037/0022-3514.54.6.1063