

Idiom Analysis

Your Name Here

3 May 2020

Due by 23:59, Sunday, 3 May 2020

Load the clean data

The idiom ratings are described in the paper by Geeraert, Newman, & Baayen (2017) The pdf file of this article is in the Idiom Documentation folder. 1. Subject 2. Age: Age of subject 3. Gender (two levels): female and male 4. Hand: (two levels): handedness of subject: left and right 5. NativeLang: (14 levels): subject's native language 2. Idiom: The idiom 1. Condition: Type of idiom or phrase + Canonical + Conceptu + Blend + Lexical + Literal + Partial 2. AcceptRating.RESP: acceptability rating of phrase's ability to be used literally (which corresponds in some degree to transparency) 3. KnowIdiom.RESP: (two levels): yes, no 3. HowOftenUse.RESP: how often the subject uses the idiom 4. HowOftenOthersUse.RESP: how often other people use the idiom 5. LikeUsingIdioms.RESP: how much the subject likes using the idiom

The idioms were rated on a 0-100 scale except KnowIdiom.RESP which has two levels (yes, no). The data set included here has 17 variables: Subject, Condition, Age, Gender and Handness. The item ratings names all end in .RESP and have corresponding reaction time measure that end in .RT. There are 7887 subjects.

```
# read in the clean data
```

```
df <- read.csv("Idiom_Data.csv")
```

```
# here is how to select just the numeric rating columns needed for many of  
# the analyses and put them in a data frame called items
```

```
# select just the numeric test items
```

```
items <- df %>%  
  select(ends_with("RESP"))
```

Task

Use an appropriate subset of the techniques presented by Patrick Mair in Chapters 2, 3, 4 and 8 (perhaps) to find structure and latent variables in the data set. Are the latent variables related to those described in the journal articles? Can you find any relationship of the demographic variables (e.g, age, gender, etc.) to the latent variables you have discovered?

There is no single approach that is correct. The tools for dealing with these data are scattered throughout the above-cited chapters:

- Optimal Scaling
- Tetrachoric and polychoric correlations
- Exploratory Factor Analysis
- Confirmatory Factor Analysis
- Item Response Theory (IRT) Analysis

Follow some examples in the book and adapt them to this particular set of data. Above all ask questions and seek my advice and help.

References

Geeraert, K., Newman, J., & Baayen, R. H. (2017). Idiom Variation: Experimental Data and a Blueprint of a Computational Model. *Topics in Cognitive Science*, 9(3), 653-669. [doi:10.1111/tops.12263](https://doi.org/10.1111/tops.12263)