Suppose we are measuring statistics anxiety with the *SAQ-8* – an 8-item "statistics anxiety questionnaire". Each item is Likert scaled with 1 = strongly disagree and 5 = strongly agree.
Items:

1. Statistics makes me cry
2. My friends will think I'm stupid for not being able to use statistical software
3. Standard deviations excite me
4. I dream that Pearson is attacking me with correlation coefficients
5. I don't understand statistics
6. I have little experience with computers
7. All computers hate me
8. I have never been good at mathematics

On Canvas, you can download a file called `SAQ8.csv`. Let's open that file in JASP and look at the inter-item correlations

Recall: the goal of *factor analysis* is to uncover clusters of items that are related to each other

- exploratory factor analysis conceptualizes these clusters as reflective of *latent* constructs.
- measurements vary along some number of *dimensions* or *factors*
- examples:
  - two dimensions of intelligence (fluid / crystallized)
  - five dimensions of personality (OCEAN)

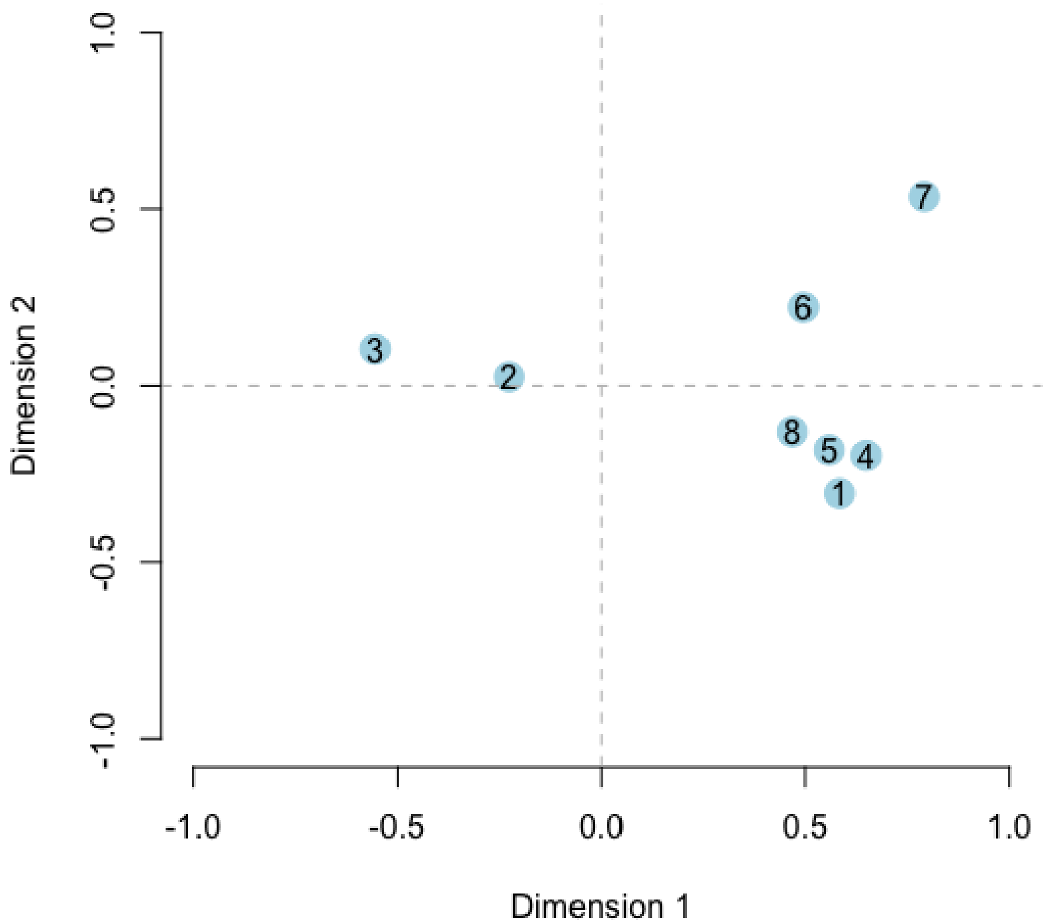
Let's do an exploratory factor analysis in JASP with the `SAQ8` dataset.

Main output = "factor loading matrix"

- idea: split the total observed variance for each item into **two** components:
  - *common variance* – proportion of variance that is due to variation in latent constructs
  - *unique variance* – proportion of variance that is unique to that particular item (i.e., not due to latent constructs)
- *factor loadings* – correlation between item and given factor
  - sum of squared loadings for an item = common variance
  - 1 - common variance = unique variance

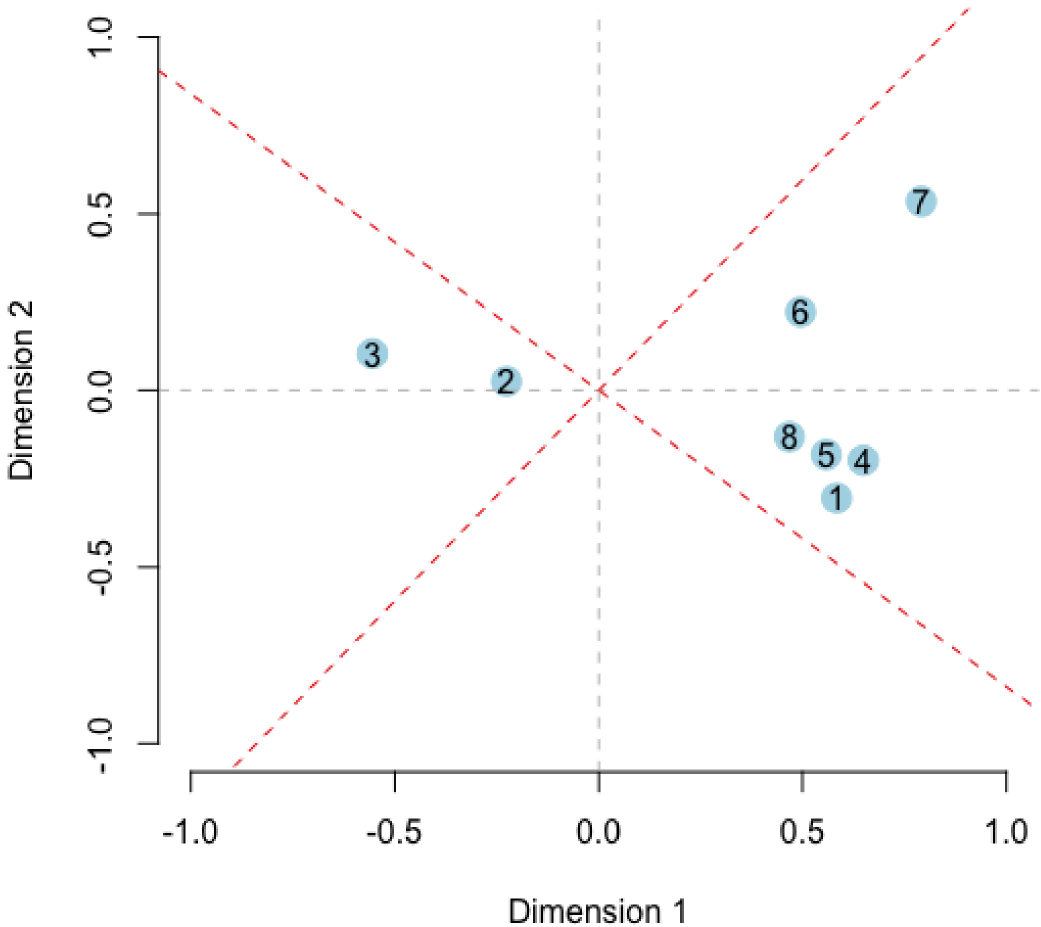
Let's "see" the factors. First, let's make a plot, where

- x-axis = loadings for each item on factor 1
- y-axis = loadings for each item on factor 2



Notice:

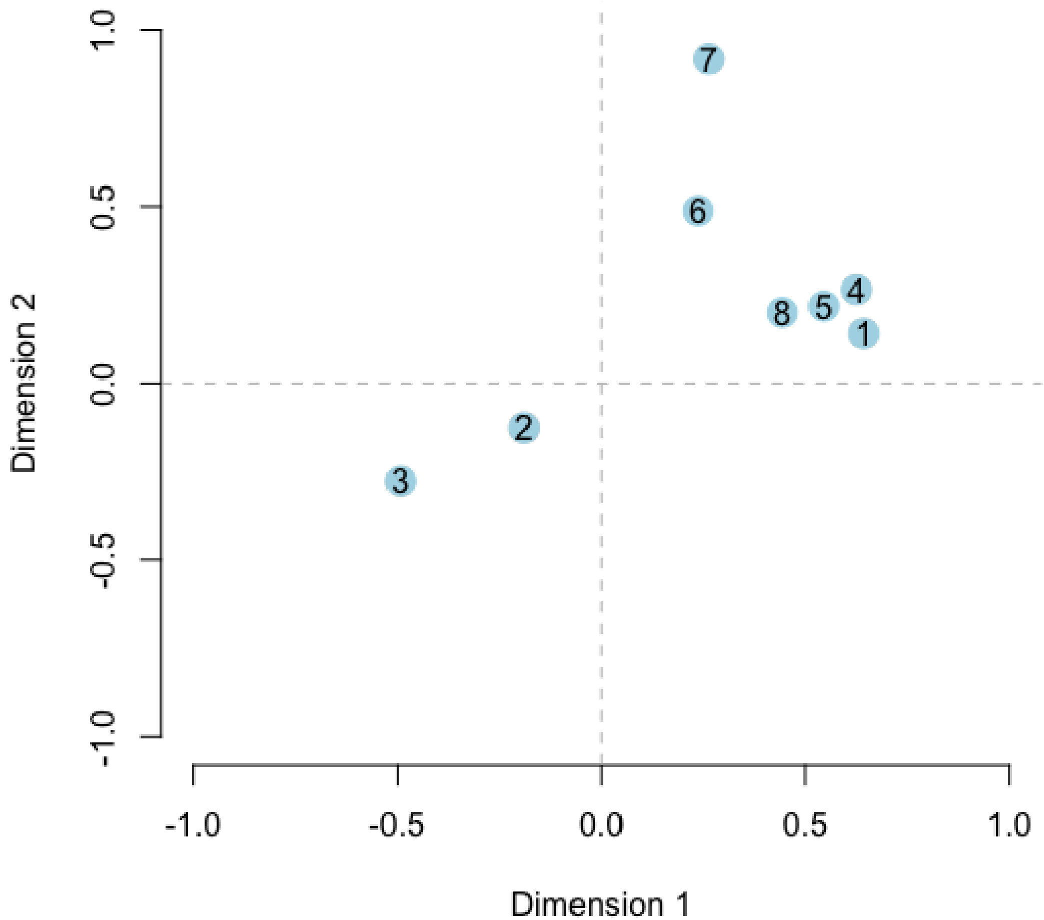
- each item is a combination of measurements on each dimension
- simpler structure if we <u>rotate</u> axes by 40° clockwise



We can do this in JASP by performing a "varimax rotation"

- notice that the factor loadings change
- let's plot the new factor loadings

Varimax rotation:



Note:

- the variance decomposition (common + unique) remains the same. The rotation only improves interpretability of the factors
- Items 6,7 vary along a different dimension than the rest
- perhaps items 6,7 reflect something different than "statistics anxiety"

Think about scale development:

- our goal is to see that the scale measures what we think it does
- in JASP, choose "highlight = 0.4" – see what happens:

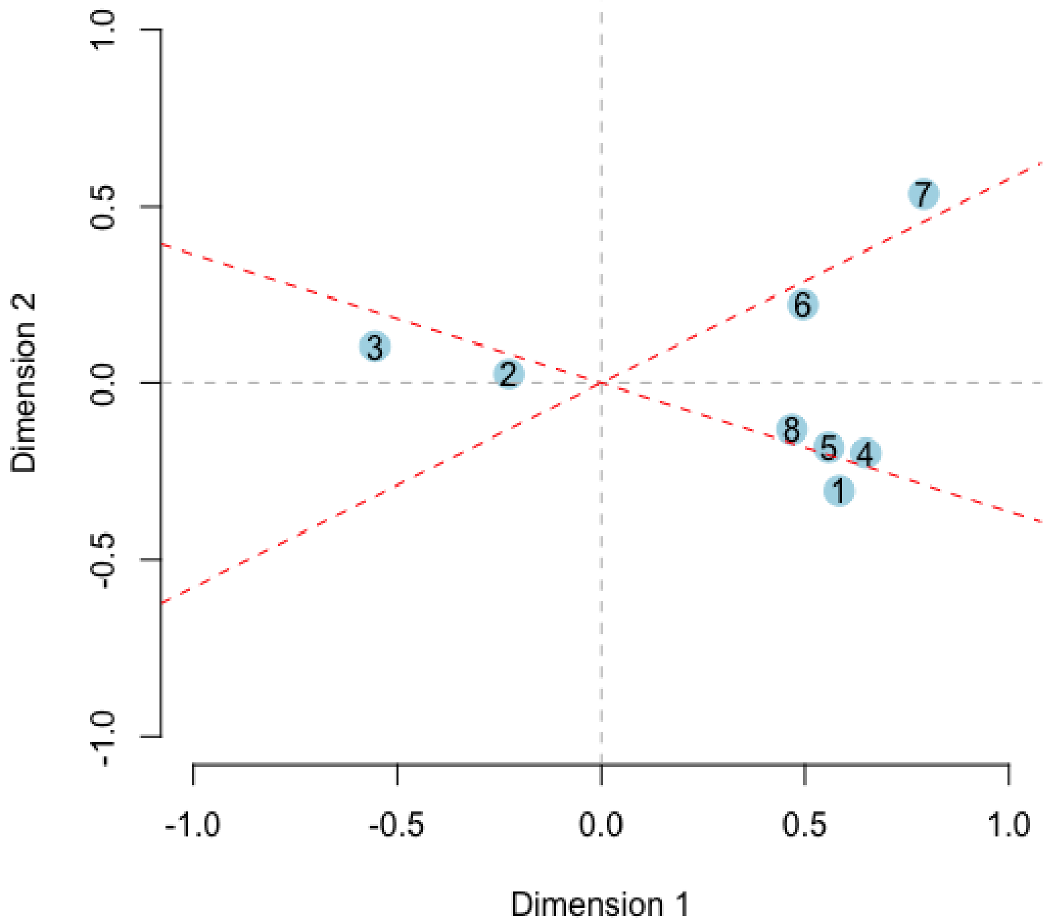
Outcomes:

- item 2 doesn't load heavily on either factor
- items 1,4,5,8 load onto factor 1 – "statistics anxiety"
- items 6,7 load onto factor 2 – "computer self concept"
- item 3 has a negative loading on factor 1 – negatively worded item

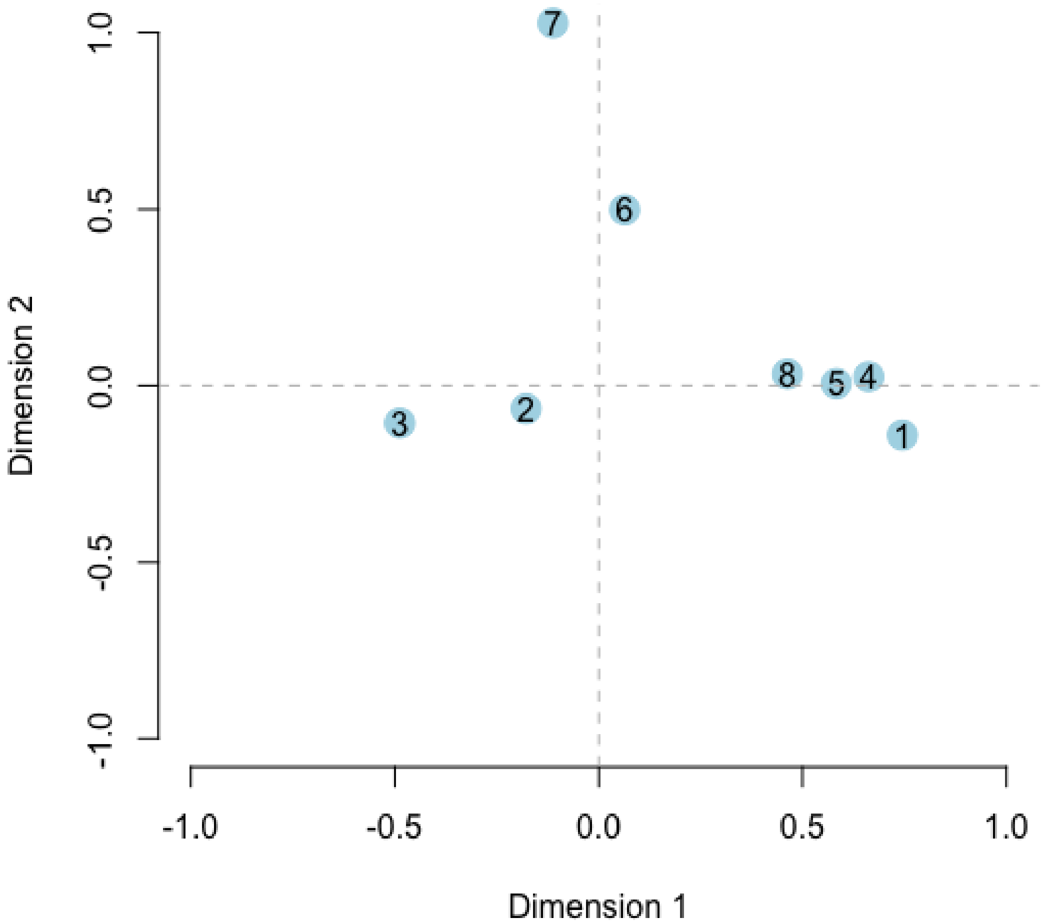
Goal in exploratory factor analysis – achive *simple structure*

- each item loads highly onto <u>one factor only</u>
- each factor has high loading for only <u>some</u> of the items
- this may require an *oblique rotation*
  - i.e., allow the factors to correlate

Notice that a better factor structure might be achieved if we allow the axes to cross *non-orthogonally* (i.e., not 90 degrees)



If we choose "oblique - promax" in JASP and plot the resulting factor loadings, we get the following:



Note: if you check "Factor correlations", you can see the following correlations:

- Orthogonal rotation: 0.144
- Oblique rotation: 0.748

Last thing – how many factors should I choose?

- there are a lot of "rules of thumb", but I prefer using "model fit" to inform number of factors
- factor analysis *simplifies* the data structure by grouping items into a smaller number of factors
- as such, the "recovered" structure is only approximate.

To see this, let's talk about how to "recover" the correlation between two items: In general (assuming *orthogonal* factors),

$$\begin{aligned}\rho_{ij} \approx & \text{ loading of item } i \text{ on factor } 1 \times \text{ loading of item } j \text{ on factor } 1 \\ & + \text{ loading of item } i \text{ on factor } 2 \times \text{ loading of item } j \text{ on factor } 2\end{aligned}$$

So, let's estimate the correlation between items 1 and 3:

$$\begin{aligned}\rho_{13} \approx & (0.584)(-0.555) \\ & + (-0.305)(0.104) \\ & = -0.355\end{aligned}$$

Compared to the observed correlation of -0.337, this has a little bit of error. How much is acceptable?

- use RMSEA = "root mean squared error of approximation"
- acceptable fit = $\mathbf{RMSEA} < \mathbf{0.08}$ (Browne & Cudeck, 1993)
- in JASP, check RMSEA for 1 factor, then 2 factors, etc. Stop extracting factors once RMSEA goes below 0.08