Factor Analysis

SDS 390 Structural Equation Modeling

Monday March 4, 2019

Emma K. T. Benn

Increasing Inclusion, Promotion and Evidence:

Uniquely Merging my Intersectionality
With My Profession as a
Biostatistician

Tuesday, March 5 • 5 p.m.

Seelye Hall, Room 106

Sponsored by the Statistical and Data Sciences Program and The Smith College Lect

Agenda for today

- Factor Analysis
  - Exploratory Factor Analysis (EFA)
  - Confirmatory Factor Analysis (CFA)

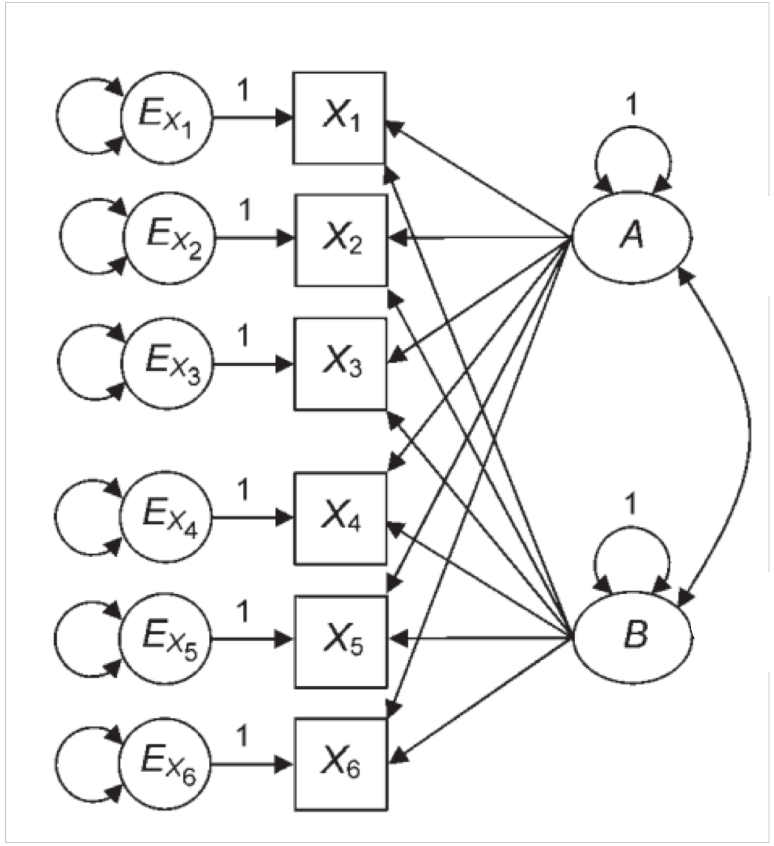
Factor Analysis

- SEM is basically Factor Analysis + Multiple Regression

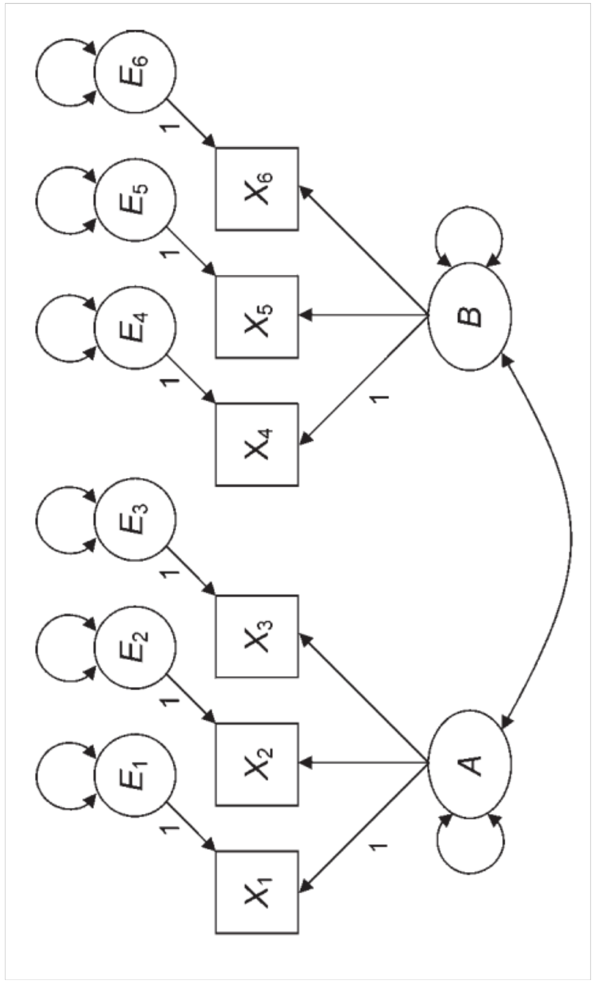
- Factor Analysis
  - Partitioning the variance of indicators into:
    1. Common variance shared among indicators (due to the factor of interest)
    2. Unique variance due to that specific indicator
    3. Measurement error

Exploratory FA vs. Confirmatory FA

EFA



CFA



Exploratory Factor Analysis

Exploratory Factor Analysis (EFA)

- Often we want to be able to describe a relatively large number of **items** by a much fewer number of **factors**.
- In the bfi dataset there are 25 items measuring personality, but are there just a few underlying factors that are responsible for people's scores on those items?
- We might guess what those are (e.g., extroversion, conscientiousness, etc.), but if we didn't know we could use **EFA** to let the data tell us about the underlying dimensions.

Exploratory Factor Analysis (EFA)

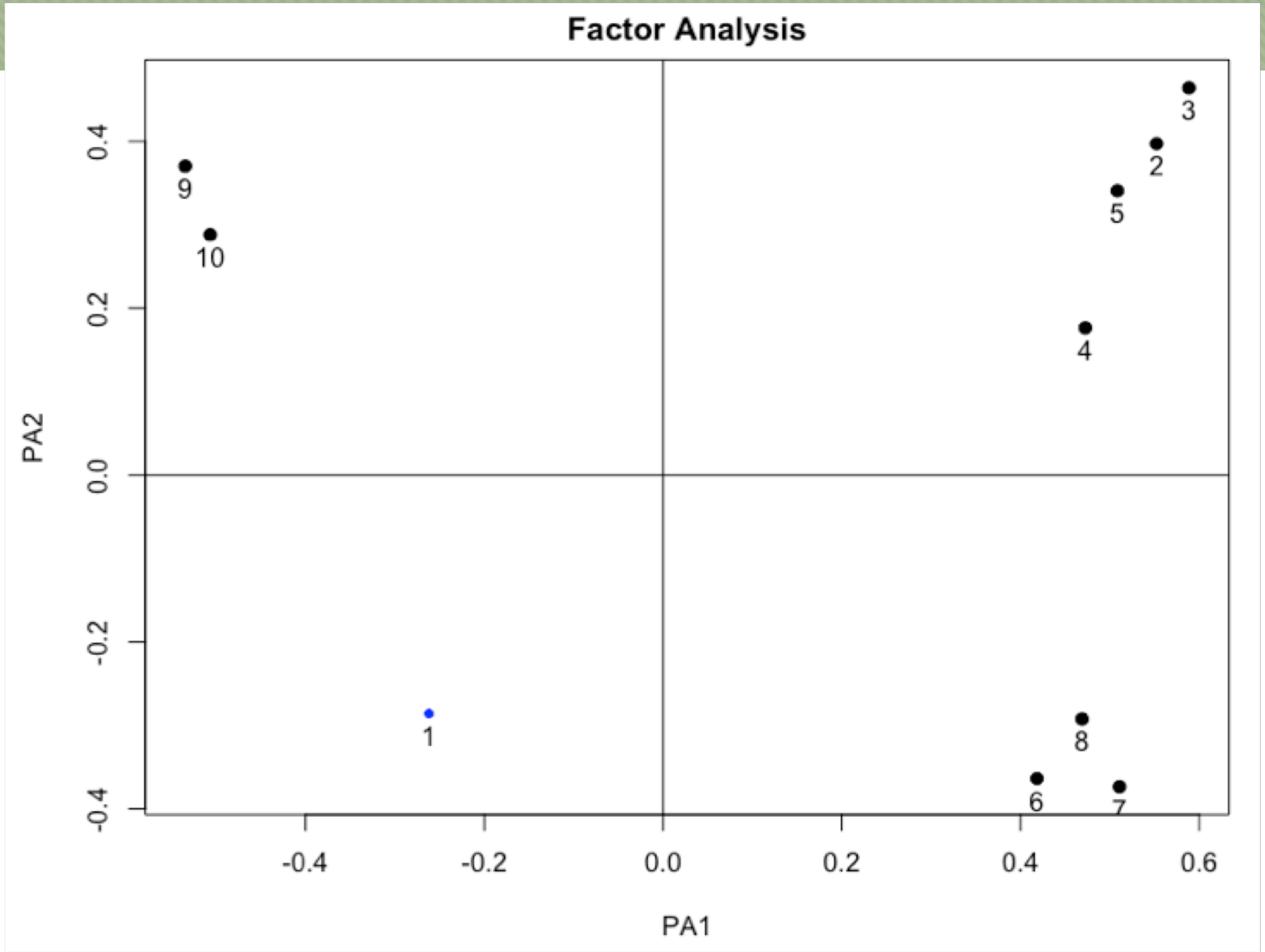
- Exploratory Factor Analysis (EFA) will use inter-correlations among the items to give us a sense of…
  1. how many factors may be present,
  2. which items can be explained by which factors, and
  3. the extent to which these underlying factors are correlated with each other.

- EFA is just that, exploratory
  - It is important to keep in mind that in the end this is a data driven technique. Meaning that peculiarities in the data may lead you to a rather weird solution.
  - It takes some sense finesse, listen to what your data is telling you.

Factor Rotation

- Unrotated solution

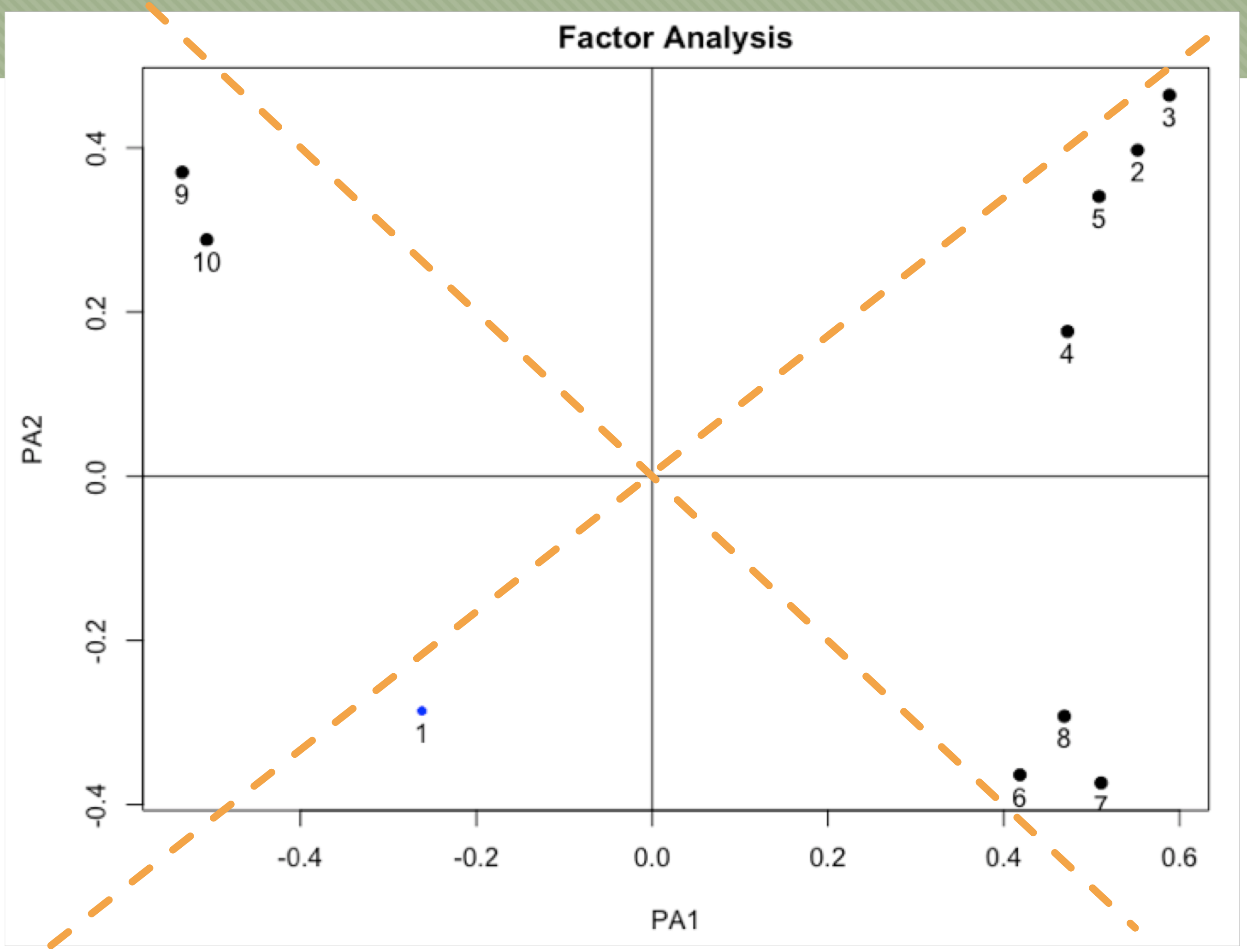
	PA1	PA2
A1	-0.26	-0.29
A2	0.55	0.40
A3	0.59	0.46
A4	0.47	0.18
A5	0.51	0.34
C1	0.42	-0.36
C2	0.51	-0.37
C3	0.47	-0.29
C4	-0.53	0.37
C5	-0.51	0.29



Factor Rotation

- Unrotated solution

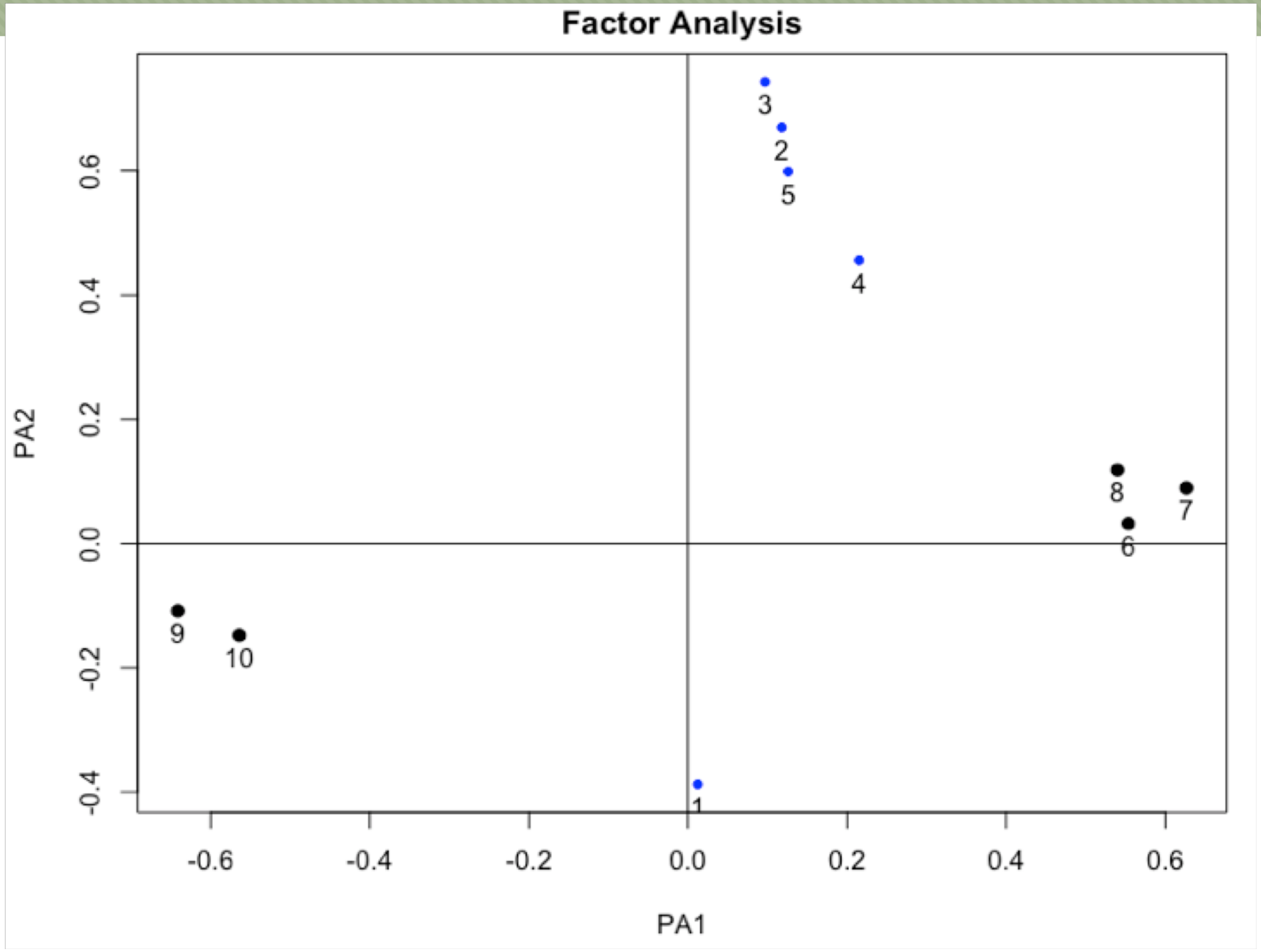
	PA1	PA2
A1	-0.26	-0.29
A2	0.55	0.40
A3	0.59	0.46
A4	0.47	0.18
A5	0.51	0.34
C1	0.42	-0.36
C2	0.51	-0.37
C3	0.47	-0.29
C4	-0.53	0.37
C5	-0.51	0.29



Factor Rotation

- Orthogonal rotation

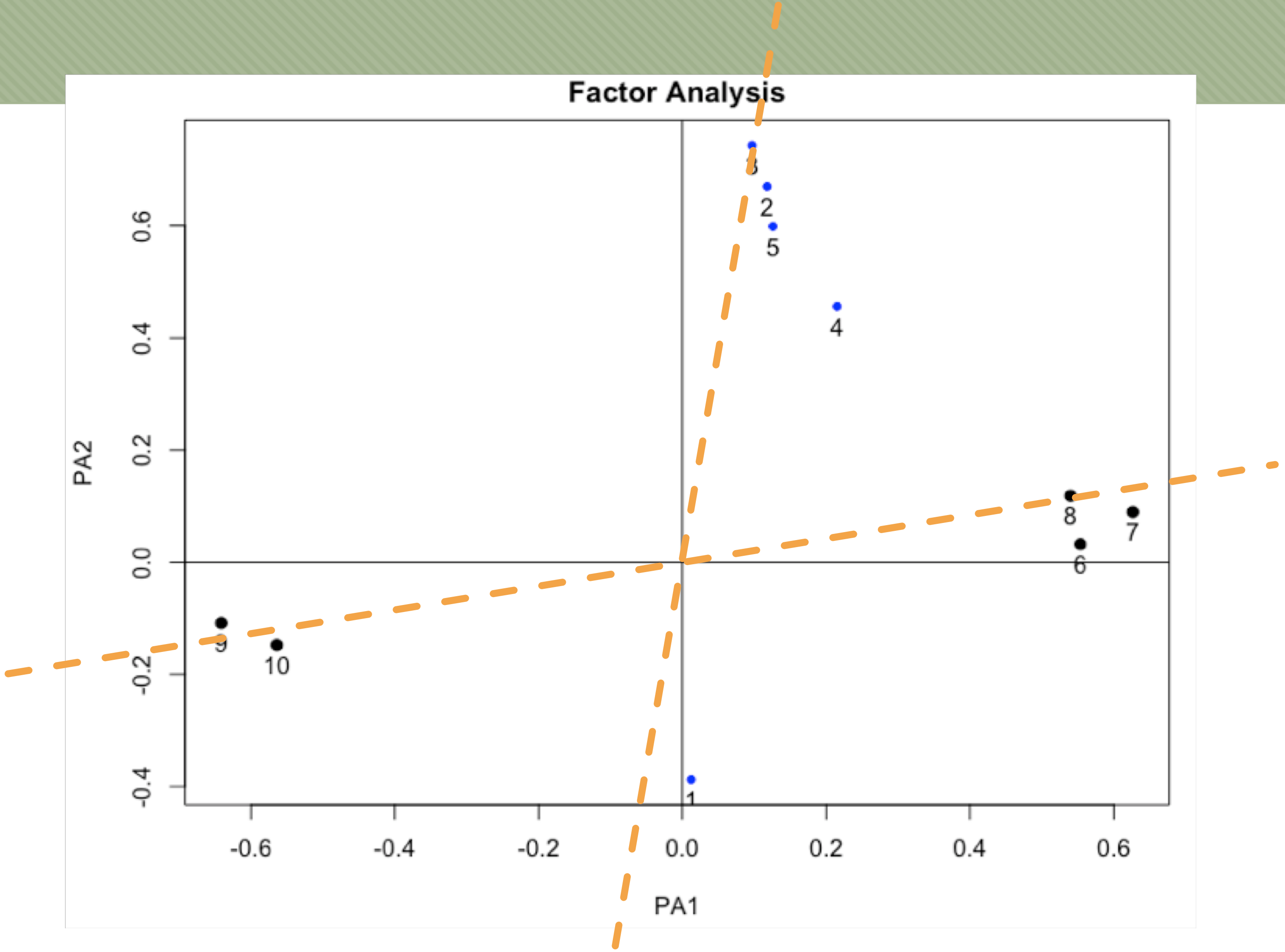
	PA1	PA2
A1	0.01	-0.39
A2	0.12	0.67
A3	0.10	0.74
A4	0.21	0.46
A5	0.13	0.60
C1	0.55	0.03
C2	0.63	0.09
C3	0.54	0.12
C4	-0.64	-0.11
C5	-0.56	-0.15



Factor Rotation

- Orthogonal rotation

	PA1	PA2
A1	0.01	-0.39
A2	0.12	0.67
A3	0.10	0.74
A4	0.21	0.46
A5	0.13	0.60
C1	0.55	0.03
C2	0.63	0.09
C3	0.54	0.12
C4	-0.64	-0.11
C5	-0.56	-0.15



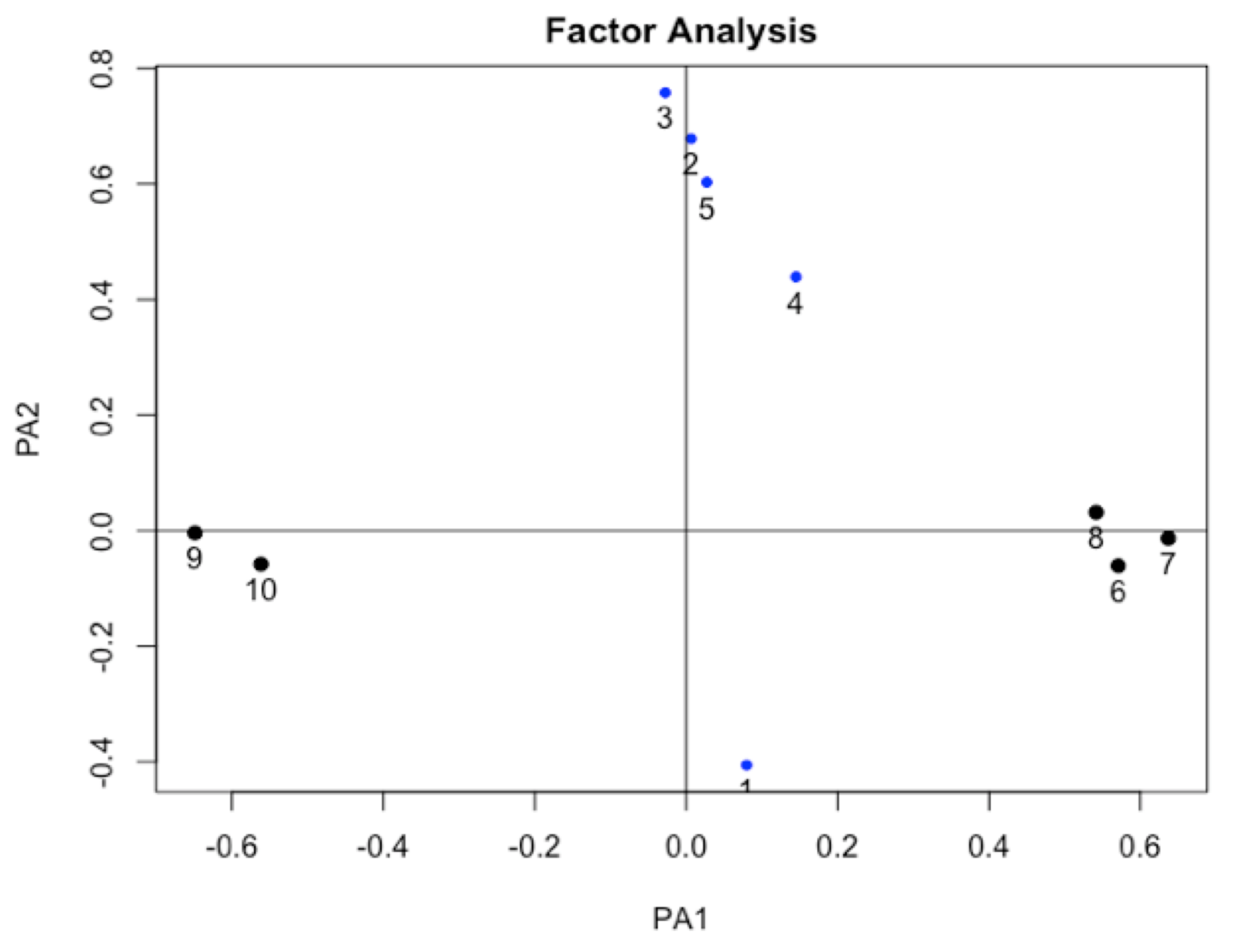
Exploratory Factor Analysis (EFA)

- Oblique factor rotation

	PA1	PA2
A1	0.08	-0.41
A2	0.01	0.68
A3	-0.03	0.76
A4	0.14	0.44
A5	0.03	0.60
C1	0.57	-0.06
C2	0.64	-0.01
C3	0.54	0.03
C4	-0.65	0.00
C5	-0.56	-0.06

With factor correlations of

	PA1	PA2
PA1	1.00	0.32
PA2	0.32	1.00



Exploratory Factor Analysis (EFA)

- We will use the **psych** package

Inference Test	R function
Factor Analysis	fa()
Principal Component Analysis	principal()

R Practice

Confirmatory Factor Analysis

Confirmatory Factor Analysis

- Kenny's (1979) rule of thumb about the number of indicators is apropos: "Two *might* be fine, three is better, four is best, and anything more is gravy" (p. 143; emphasis in original.)

- If the researcher's model is reasonably correct, then one should see the following pattern of results:
  1. (1) all indicators specified to measure a common factor have relatively high standardized factor loadings on that factor (e.g., > .70); and
  2. (2) estimated correlations between the factors are not excessively high (e.g., < .90 in absolute value).
- The first result indicates convergent validity; the second, discriminant validity.

Confirmatory Factor Analysis (CFA)

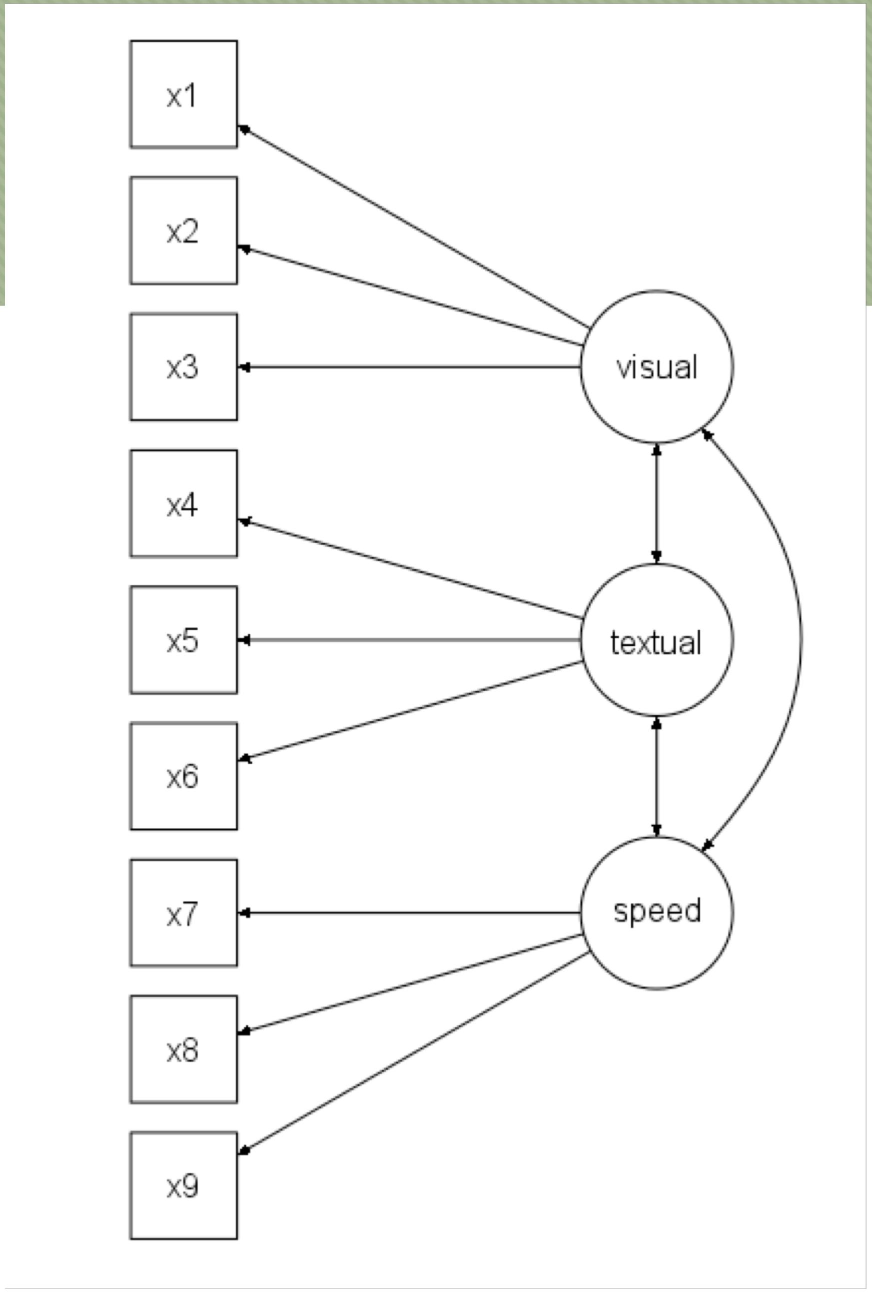
```{r}
library(lavaan)
data(HolzingerSwineford1939)
```

- Mental ability test score from 7th and 8th grade children from two schools
  - A *visual* factor measured by 3 variables: x1, x2 and x3
  - A *textual* factor measured by 3 variables: x4, x5 and x6
  - A *speed* factor measured by 3 variables: x7, x8 and x9
- We want to test if indeed these measures fall on these three scales as we hypothesize.
- We are *confirming* a hypothesized factor structure instead of exploring.

Visual factor:
x1, x2 and x3

Textual factor:
x4, x5 and x6

Speed factor:
x7, x8 and x9



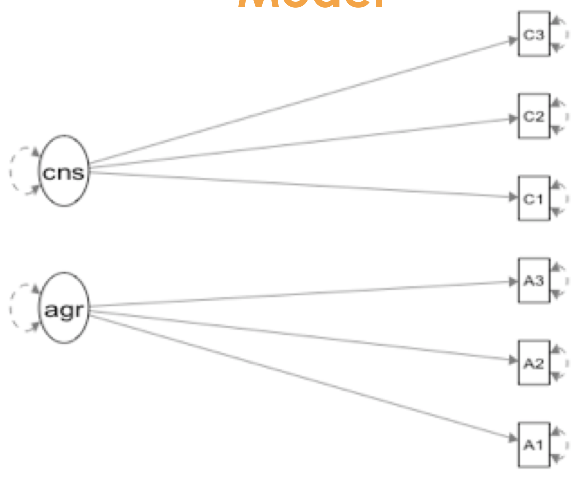
Confirmatory Factor Analysis (CFA)

- Does the model we have in our heads actually fit the data?

Data Cor matrix

	A1	A2	A3	C1	C2	C3
A1	1.000	-0.340	-0.265	0.028	0.016	-0.019
A2	-0.340	1.000	0.485	0.092	0.136	0.192
A3	-0.265	0.485	1.000	0.097	0.141	0.132
C1	0.028	0.092	0.097	1.000	0.428	0.308
C2	0.016	0.136	0.141	0.428	1.000	0.356
C3	-0.019	0.192	0.132	0.308	0.356	1.000

Model



Model implied Cor matrix

	A1	A2	A3	C1	C2	C3
A1	1.000					
A2	-0.337	1.000				
A3	-0.256	0.492	1.000			
C1	-0.063	0.122	0.093	1.000		
C2	-0.074	0.143	0.109	0.418	1.000	
C3	-0.056	0.108	0.082	0.316	0.370	1.000

Fit?

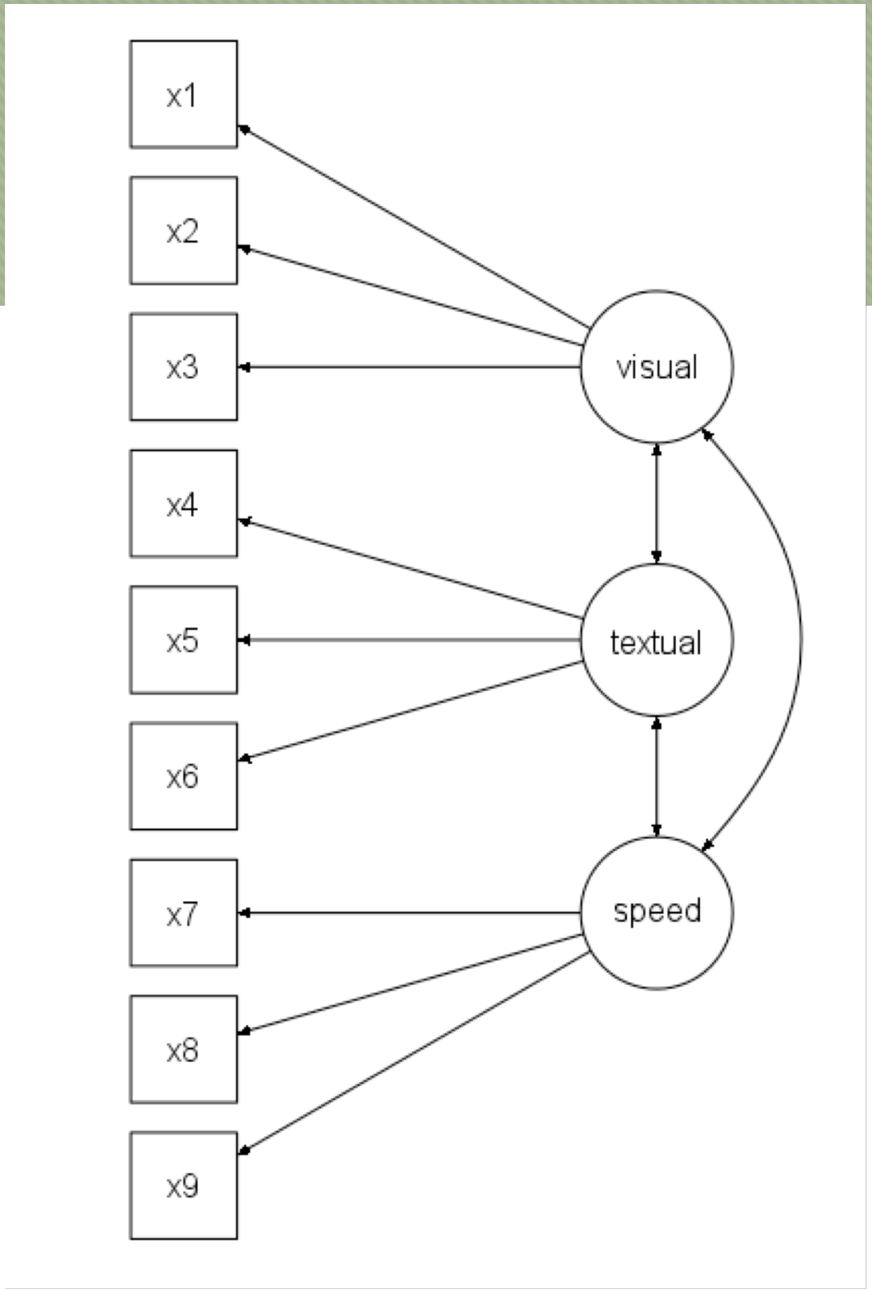
Confirmatory Factor Analysis (CFA)

- We will use the R package lavaan to fit CFAs

- lavaan steps:
  - Step 1: Specify the model
  - Step 2: Fit the model
  - Step 3: Ask for the output you want

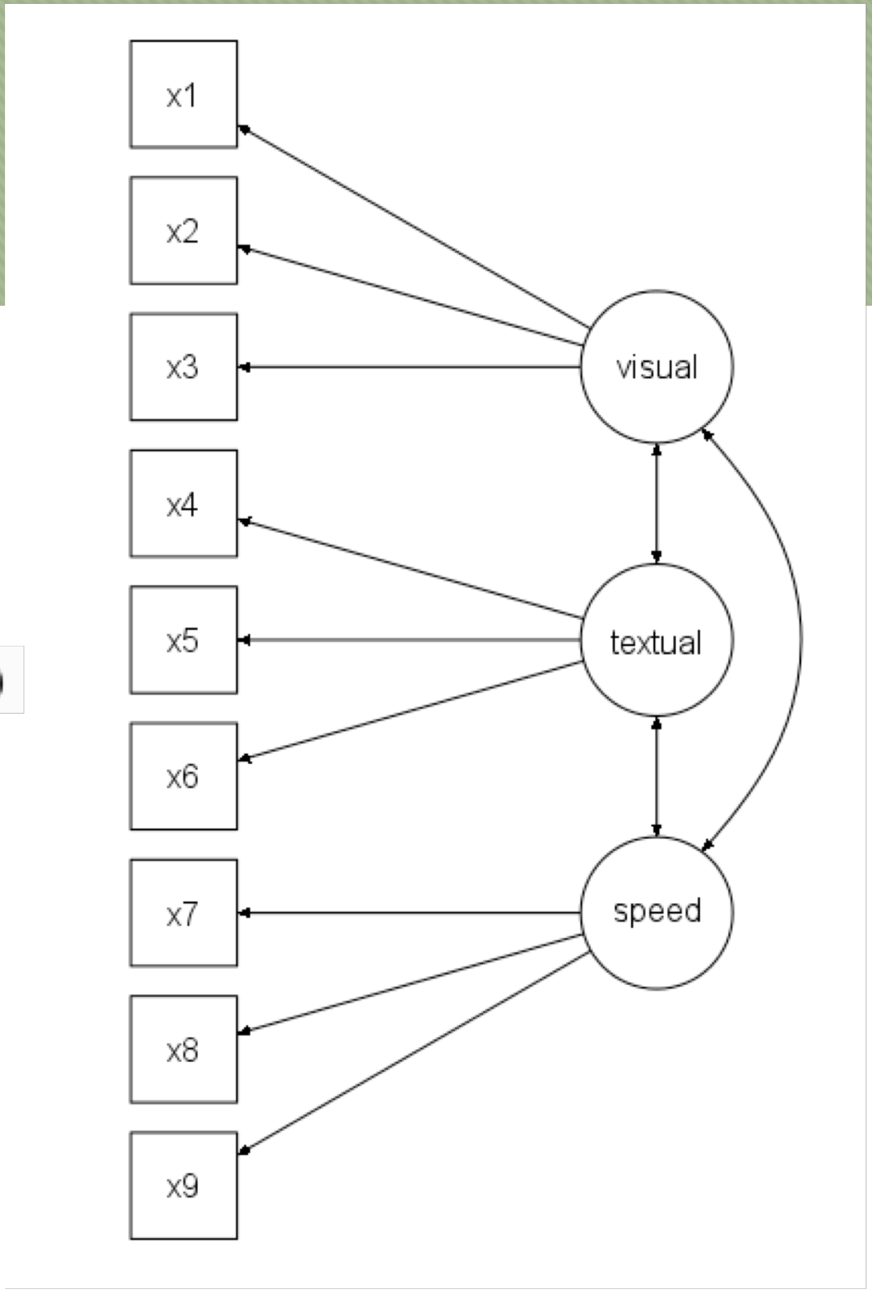
Step 1:
Specify the Model

```
HS.model <- ' visual  =~ x1 + x2 + x3
              textual =~ x4 + x5 + x6
              speed   =~ x7 + x8 + x9 '
```



Step 2: Fit the Model

```
fit <- cfa(HS.model, data = HolzingerSwineford1939)
```



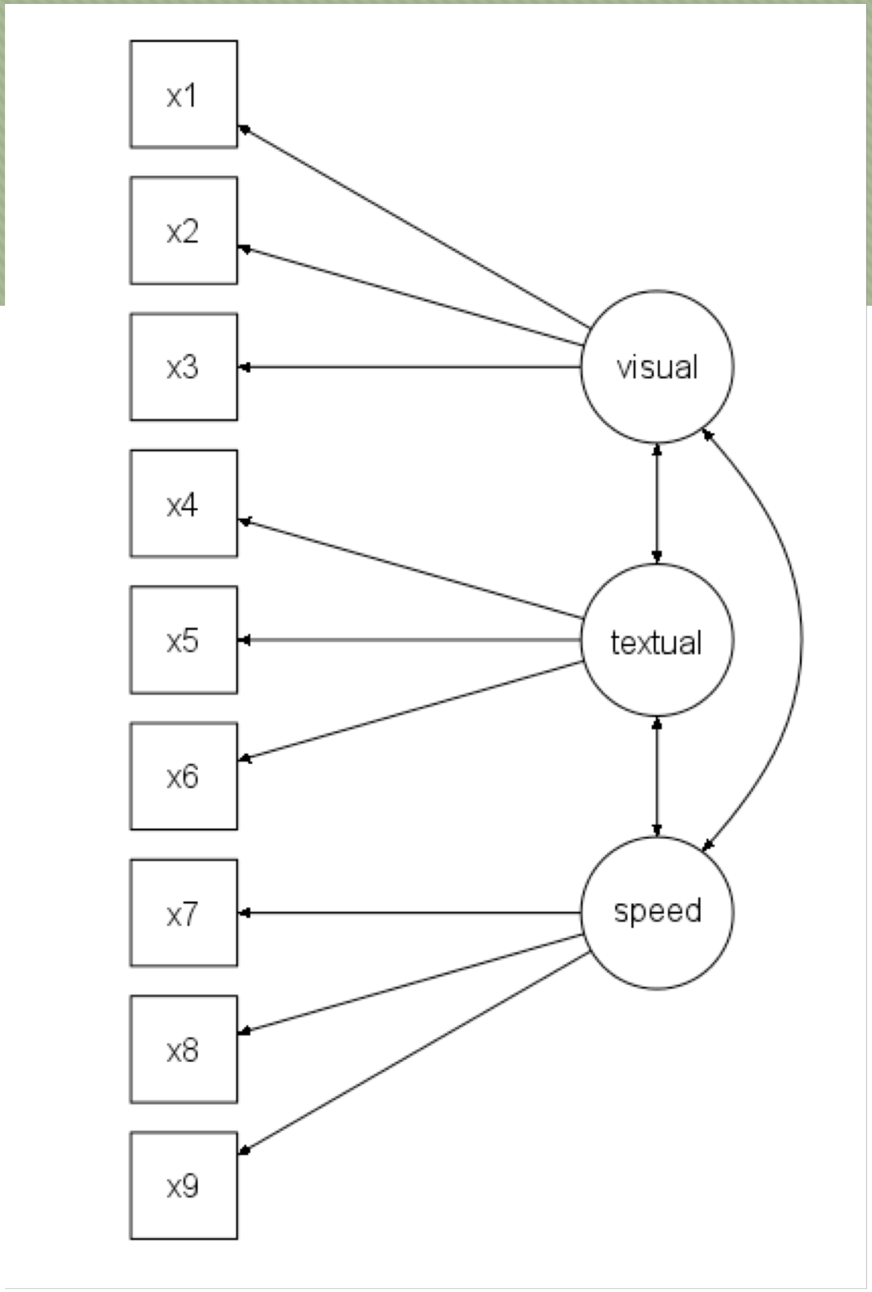
Step 3: Ask for the output you want

```
summary(fit, fit.measures = TRUE)
```

```
parameterEstimates(fit)
```

```
inspect(fit)
```

```
modindices(fit)
```



R Practice

Hierarchical CFA

- Leach et al (2008), Multicomponent Model of group Identification