

Advanced issues in Factor Analysis

Telemetrics lab

Department of Psychology
Northwestern University
Evanston, Illinois USA

May, 2013



NORTHWESTERN
UNIVERSITY

Outline

- 1 Simplex structures
 - simulating a simplex
 - factoring
 - ICLUST of a simplex
- 2 The problems of dichotomous items
- 3 Polytomous items



A simplex

- In developmental, or any time process, nearby items are more correlated
 - An underlying growth process
 - localized errors
- grades in progressive quarters
- reaction times during a long session



Simulate a simplex

```
> set.seed(42) # for reproducible results
> s9 <- sim.simplex(9,n=1000)
> str(s9) #show the structure
List of 4
 $ model      : num [1:9, 1:9] 1 0.8 0.64 0.512 0.41 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:9] "V1" "V2" "V3" "V4" ...
  .. ..$ : chr [1:9] "V1" "V2" "V3" "V4" ...
 $ r          : num [1:9, 1:9] 1 0.789 0.625 0.492 0.42 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:9] "V1" "V2" "V3" "V4" ...
  .. ..$ : chr [1:9] "V1" "V2" "V3" "V4" ...
 $ observed: num [1:1000, 1:9] -0.659 -0.858 0.241 0.714 1.268 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : NULL
  .. ..$ : chr [1:9] "V1" "V2" "V3" "V4" ...
 $ Call      : language sim.simplex(nvar = 9, n = 1000)
- attr(*, "class")= chr [1:2] "psych" "sim"
```



A simplex correlation matrix

```
> lowerCor(v9$items)
```

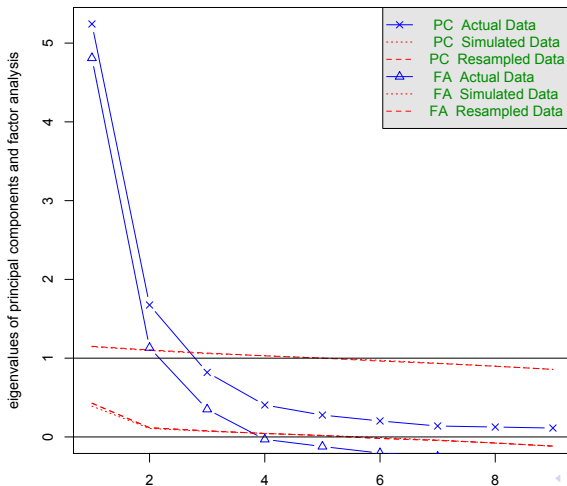
	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	1.00	0.80	0.64	0.51	0.41	0.33	0.26	0.21	0.17
V2	0.80	1.00	0.80	0.64	0.51	0.41	0.33	0.26	0.21
V3	0.64	0.80	1.00	0.80	0.64	0.51	0.41	0.33	0.26
V4	0.51	0.64	0.80	1.00	0.80	0.64	0.51	0.41	0.33
V5	0.41	0.51	0.64	0.80	1.00	0.80	0.64	0.51	0.41
V6	0.33	0.41	0.51	0.64	0.80	1.00	0.80	0.64	0.51
V7	0.26	0.33	0.41	0.51	0.64	0.80	1.00	0.80	0.64
V8	0.21	0.26	0.33	0.41	0.51	0.64	0.80	1.00	0.80
V9	0.17	0.21	0.26	0.33	0.41	0.51	0.64	0.80	1.00



How many factors?

```
fa.parallel(s9$observed)
```

Parallel Analysis Scree Plots



Factor a simplex

```
> f2 <- fa(s9$observed,2)
> f2
Factor Analysis using method = minres
Call: fa(r = s9$observed, nfactors = 2)
Standardized loadings based upon correlation matrix
```

	MR1	MR2	h2	u2
V1	-0.07	0.76	0.53	0.47
V2	-0.06	0.89	0.75	0.25
V3	-0.01	0.92	0.83	0.17
V4	0.21	0.72	0.71	0.29
V5	0.48	0.49	0.69	0.31
V6	0.72	0.23	0.74	0.26
V7	0.90	0.04	0.85	0.15
V8	0.91	-0.10	0.76	0.24
V9	0.79	-0.09	0.57	0.43

	MR1	MR2
SS loadings	3.22	3.19
Proportion Var	0.36	0.35
Cumulative Var	0.36	0.71

With factor correlations of

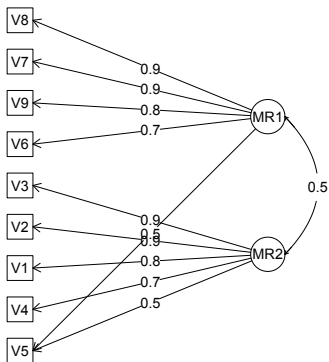
	MR1	MR2
MR1	1.00	0.47
MR2	0.47	1.00



factor diagram

> fa.diagram(f2,simple=FALSE) #show large cross loadings

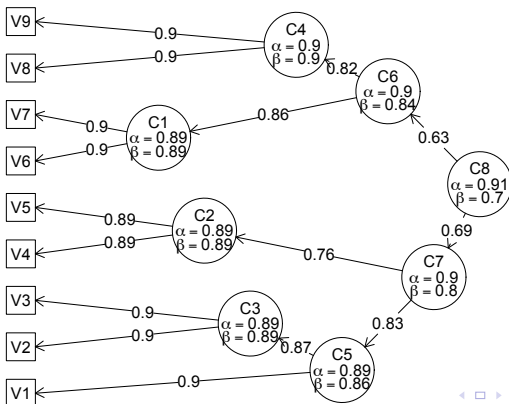
Factor Analysis



ICLUST of a simplex

> iclust(s9\$observed) #cluster analyze the data

iclust



Item difficulty leads to a simplex structure

- Dichotomous items (e.g., ability items) differ in difficulty
 - Easy items have high endorsement rates
 - Hard items have low endorsement rates
- Φ coefficient is sensitive to differences in response
- Items with similar difficulties will correlate more highly



How many factors

```
> set.seed(42)
> v9 <- sim.rasch(9)
> lowerCor(v9$items)
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	1.00								
V2	0.11	1.00							
V3	0.01	0.16	1.00						
V4	0.12	0.23	0.14	1.00					
V5	0.06	0.09	0.07	0.23	1.00				
V6	0.09	0.08	0.17	0.23	0.21	1.00			
V7	0.11	0.09	0.10	0.13	0.07	0.21	1.00		
V8	0.03	0.10	0.08	0.12	0.11	0.05	0.12	1.00	
V9	0.06	0.09	0.07	0.12	0.06	0.16	0.09	0.02	1.00

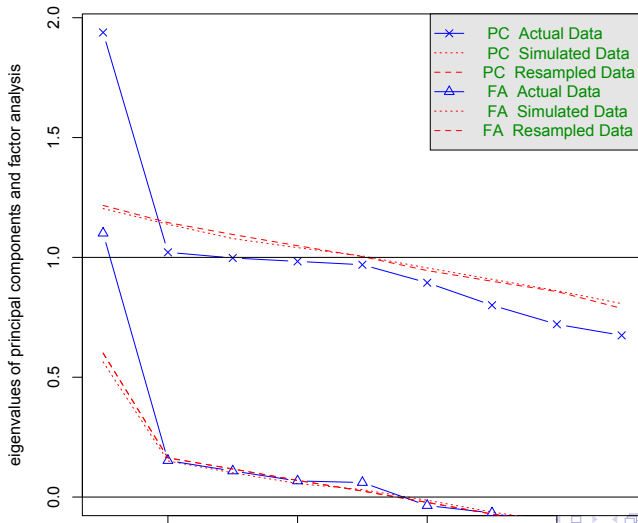
```
> fa.parallel(v9$items)
```

Parallel analysis suggests that the number of factors = 5



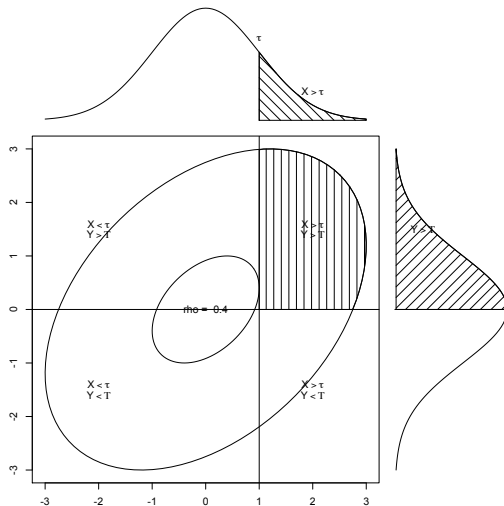
Parallel analysis of dichotomous items

Parallel Analysis Scree Plots



Find the tetrachoric correlations

> draw.tetra(.4,1,0) #rho, cut 1, cut 2



The tetrachoric correlation matrix

```
> rtet <- tetrachoric(v9$items)
Loading required package: mvtnorm
> rtet
Call: tetrachoric(x = v9$items)
tetrachoric correlation
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	1.000	0.27	0.024	0.27	0.15	0.24	0.33	0.115	0.253
V2	0.268	1.00	0.299	0.41	0.18	0.15	0.21	0.281	0.322
V3	0.024	0.30	1.000	0.24	0.11	0.32	0.20	0.195	0.182
V4	0.273	0.41	0.242	1.00	0.36	0.38	0.23	0.253	0.288
V5	0.147	0.18	0.114	0.36	1.00	0.33	0.12	0.221	0.130
V6	0.239	0.15	0.316	0.38	0.33	1.00	0.35	0.111	0.335
V7	0.330	0.21	0.195	0.23	0.12	0.35	1.00	0.247	0.212
V8	0.115	0.28	0.195	0.25	0.22	0.11	0.25	1.000	0.048
V9	0.253	0.32	0.182	0.29	0.13	0.34	0.21	0.048	1.000

with tau of

	V1	V2	V3	V4	V5	V6	V7	V8	V9
	-1.46	-1.00	-0.72	-0.32	0.00	0.39	0.69	1.16	1.38



Compare the Pearson and the Tetrachoric

Use the helper function `lowerUpper`

```
both <- lowerUpper(rtet$rho, lowerCor(v9$items))  
round(both, 2)
```

	V1	V2	V3	V4	V5	V6	V7	V8	V9
V1	NA	0.11	0.01	0.12	0.06	0.09	0.11	0.03	0.06
V2	0.27	NA	0.16	0.23	0.09	0.08	0.09	0.10	0.09
V3	0.02	0.30	NA	0.14	0.07	0.17	0.10	0.08	0.07
V4	0.27	0.41	0.24	NA	0.23	0.23	0.13	0.12	0.12
V5	0.15	0.18	0.11	0.36	NA	0.21	0.07	0.11	0.06
V6	0.24	0.15	0.32	0.38	0.33	NA	0.21	0.05	0.16
V7	0.33	0.21	0.20	0.23	0.12	0.35	NA	0.12	0.09
V8	0.12	0.28	0.19	0.25	0.22	0.11	0.25	NA	0.02
V9	0.25	0.32	0.18	0.29	0.13	0.34	0.21	0.05	NA



Factor the raw items

```
> fa(v9$items)
```

```
Factor Analysis using method = minres
```

```
Call: fa(r = v9$items)
```

```
Standardized loadings (pattern matrix) based upon correlation matrix
```

	MR1	h2	u2
V1	0.21	0.046	0.95
V2	0.34	0.118	0.88
V3	0.30	0.090	0.91
V4	0.53	0.282	0.72
V5	0.36	0.132	0.87
V6	0.47	0.219	0.78
V7	0.33	0.106	0.89
V8	0.22	0.049	0.95
V9	0.24	0.060	0.94

	MR1
SS loadings	1.10
Proportion Var	0.12

```
Test of the hypothesis that 1 factor is sufficient.
```

```
The root mean square of the residuals (RMSR) is 0.04
```

```
The df corrected root mean square of the residuals is 0.06
```

```
The harmonic number of observations is 500 with the empirical chi square = 45.5
```

```
The total number of observations was 500 with MLE Chi Square = 30.89 with p
```



Factor the tetrachorics, report as normal factor analysis output

```
> fa(rtet$rho)
```

```
Factor Analysis using method = minres
```

```
Call: fa(r = rtet$rho)
```

```
Standardized loadings (pattern matrix) based upon correlation matrix
```

	MR1	h2	u2
V1	0.43	0.18	0.82
V2	0.54	0.30	0.70
V3	0.41	0.17	0.83
V4	0.66	0.44	0.56
V5	0.43	0.18	0.82
V6	0.58	0.33	0.67
V7	0.47	0.22	0.78
V8	0.36	0.13	0.87
V9	0.47	0.22	0.78

	MR1
SS loadings	2.18
Proportion Var	0.24



Factor analyze the items using tetrachorics

Report in terms of IRT parameters

```
> f.irt <- irt.fa(v9$items)
```

```
> f.irt
```

Item Response Analysis using Factor Analysis =

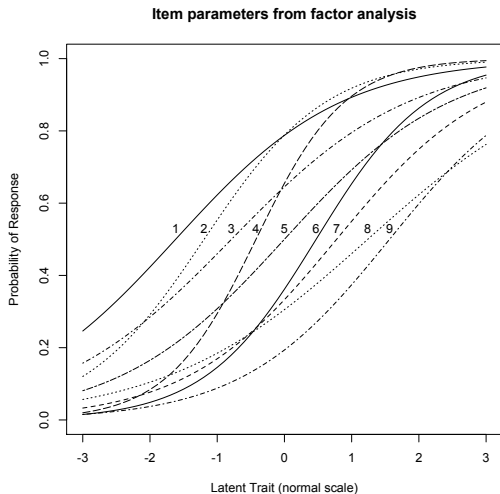
Call: irt.fa(x = v9\$items)

	Location	Discrimination	tau	Loading
V1	-1.62	0.48	-1.46	0.43
V2	-1.19	0.65	-1.00	0.54
V3	-0.79	0.45	-0.72	0.41
V4	-0.42	0.89	-0.32	0.66
V5	0.00	0.48	0.00	0.43
V6	0.47	0.71	0.39	0.58
V7	0.78	0.53	0.69	0.47
V8	1.24	0.39	1.16	0.36
V9	1.56	0.54	1.38	0.47

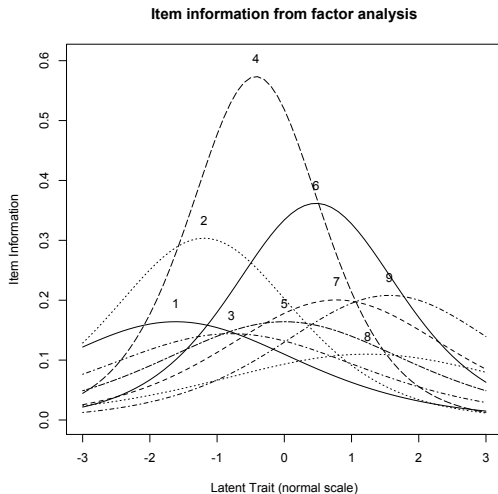


Show the items

```
> plot(f.irt,type="ICC")
```



Show the item information functions



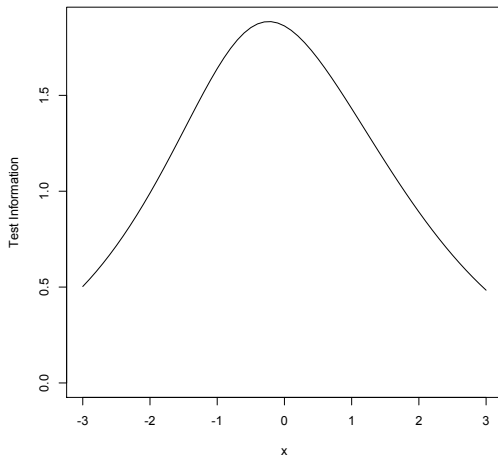
```
> plot(f.irt)
```



Show the Test information function

```
> plot(f.irt,type="test")
```

Test information -- item parameters from factor analysis



Polytomous items

- Most personality items have 3-6 alternatives
 - The fewer the alternatives, the more the correlation is restricted
 - For 6 choice items this is not too serious, but for 4, it probably is
- Find the polychoric correlation (What would be the Pearson if the data were bivariate normal?)
- polychoric function in R



Find polychoric correlations for Big 5 items

```
data(bfi)
rbfi <-polychoric(bfi[1:25]) #this takes awhile
```

