Exploratory Factor Analysis

Today . . .

- Oblique rotation
- Cross-validation
- Additional considerations

Factor analysis offers a more realistic model of measurement by admitting the presence of random error and specific, systematic sources of variability.

Another way to make the model more realistic is to relax the restriction that factors be orthogonal. Allowing *oblique factors* has two potential benefits:

- It allows the model to better match the actual data
- It allows the possibility of higher order factors

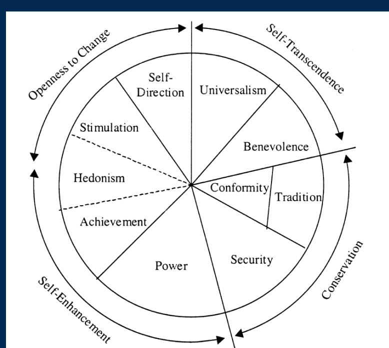
Orthogonal rotation identified four factors that appear to represent "civility," "self-enhancement," "openness," and "conservative."

	PA2
Wealth	0.65
Social_Power	0.62
Successful	0.54
Pleasure	0.53
Public_Image	0.50

	PA3
World_Of_Beauty	0.65
Unity_With_Nature	0.64
Curious	0.61
Protect_Enviroment	0.60
Creativity	0.58
Varied_Life	0.53

	PA1
Honest	0.72
Helpful	0.64
Equality	0.64
Freedom	0.60
Loyal	0.59
Responsible	0.59
Ambitious	0.58
Family_Security	0.56
Broad_Minded	0.55
Social_Justice	0.55
Politeness	0.52

	PA4
Respect_For_Tradition	0.66
Respect_Elders	0.56
Devout	0.55
Obedient	0.55



The underlying model suggests that groups of values are correlated. An oblique rotation might better represent the value ratings.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (pp. 1-65). San Diego, CA: Academic Press.

By relaxing the requirement that factors be independent, we now require an additional matrix—the *factor pattern matrix*.

One way to appreciate why this matrix is necessary is to remember that when factors are orthogonal, the variables can be reconstructed from those factors using a very simple linear combination of the standardized factor scores:

$$\hat{X} = Z_f \Lambda_f'$$

$$\hat{X} = Z_f \Lambda_f'$$

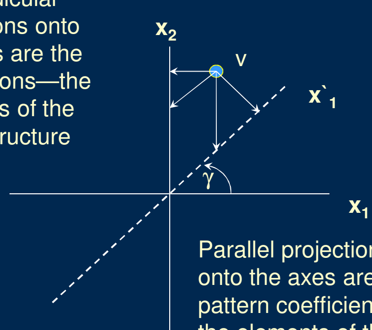
This linear combination resembles a multiple regression equation, but because the “predictors” are independent, the regression coefficients or weights are simply correlations---just the elements of the factor structure matrix.

When the factors are no longer orthogonal, then the correlations are not the appropriate weights and a separate matrix containing those weights is necessary.

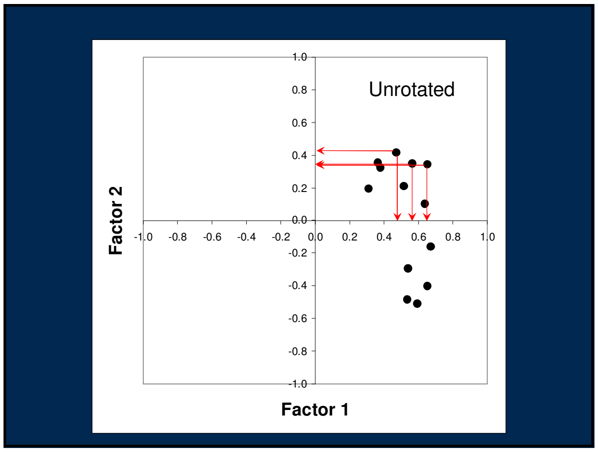
The factor <u>pattern</u> matrix contains the “regression” weights for reconstructing the variable from the factors.

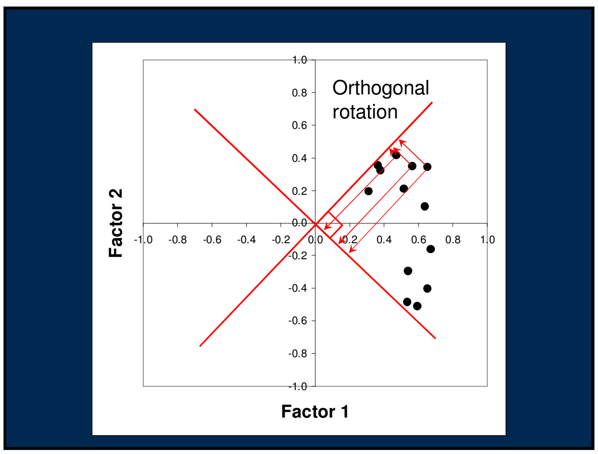
These weights appropriately take into account the correlations among the factors (the “predictors”) and indicate the unique role played by any one of the predictors in reconstructing a given variable.

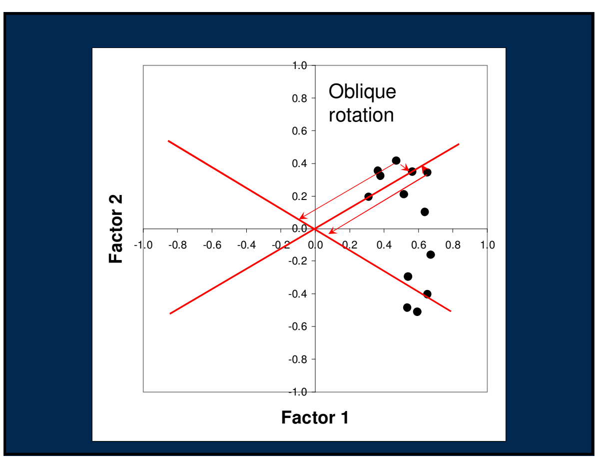
Perpendicular projections onto the axes are the correlations—the elements of the factor structure matrix.



Parallel projections onto the axes are the pattern coefficients—the elements of the factor pattern matrix.


Unrotated
Factor 2
Factor 1
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0


Orthogonal
rotation
Factor 2
Factor 1
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0


Oblique
rotation
Factor 2
Factor 1
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0 0.2 0.4 0.6 0.8 1.0

Because the factor pattern matrix represents the unique contribution of each factor to the reconstruction of any variable in X, it provides a better basis for judging simple structure.

When we have both a structure matrix and a pattern matrix, the convention is to refer to the pattern matrix as the loading matrix and to search for “simple structure” there.

One common oblique rotation method (there are many) is known as oblimin. Oblimin seeks to minimize the covariance among squared loadings of separate columns.

This will also induce correlation among factors while producing simple structure in the pattern matrix (which represents the unique contribution of factors to variables).

	MR1	MR2	MR3	MR4
Honest	0.72	0.00	0.06	0.18
Helpful	0.64	-0.19	0.24	0.20
Equality	0.64	-0.15	0.34	-0.03
Freedom	0.60	0.25	0.29	0.00
Loyal	0.59	0.10	0.01	0.21
Responsible	0.59	0.20	0.03	0.30
Ambitious	0.58	0.32	0.06	0.15
Family_Security	0.56	0.09	0.01	0.24
Broad_Minded	0.55	-0.05	0.35	-0.12
Social_Justice	0.55	-0.17	0.33	0.09
Politeness	0.52	0.14	0.00	0.43

Varimax

	MR1	MR2	MR4	MR3
Honest	0.75	-0.06	-0.06	0.07
Helpful	0.64	-0.25	0.15	0.10
Equality	0.64	-0.22	0.23	-0.14
Loyal	0.62	0.05	-0.09	0.12
Responsible	0.59	0.15	-0.06	0.20
Ambitious	0.59	0.28	-0.05	0.04
Family_Security	0.58	0.05	-0.08	0.15
Freedom	0.58	0.19	0.18	-0.13
Broad_Minded	0.55	-0.11	0.24	-0.23
Social_Justice	0.53	-0.23	0.25	-0.01
Politeness	0.51	0.10	-0.06	0.35

Oblimin

World_Of_Beauty	0.15	0.01	0.65	0.14
Unity_With_Nature	0.01	-0.11	0.64	0.31
Curious	0.19	0.18	0.61	0.04
Protect_Enviroment	0.12	-0.10	0.60	0.13
Creativity	0.23	0.16	0.58	0.03
Varied_Life	0.22	0.31	0.53	-0.08
Daring	0.05	0.29	0.49	0.06
Exciting_Life	0.29	0.44	0.46	-0.10

Varimax

Unity_With_Nature	-0.13	-0.15	0.71	0.26
World_Of_Beauty	0.03	-0.03	0.66	0.06
Protect_Enviroment	0.01	-0.14	0.63	0.07
Curious	0.08	0.14	0.60	-0.06
Creativity	0.13	0.12	0.57	-0.06
Varied_Life	0.12	0.28	0.50	-0.18
Daring	-0.06	0.26	0.50	-0.02
Exciting_Life	0.21	0.40	0.41	-0.22

Oblimin

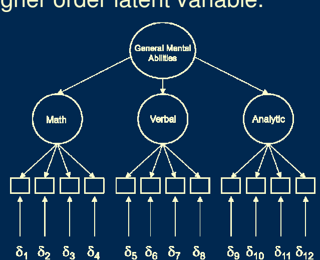
The minor changes in the loadings from varimax compared to oblimin suggest that the orthogonal factors imposed by varimax rotation actually represent the data fairly well. This is verified by the correlations among the factors from oblique rotation.

	MR1	MR2	MR4	MR3
MR1	1.00	0.16	0.36	0.21
MR2	0.16	1.00	0.10	0.10
MR4	0.36	0.10	1.00	0.05
MR3	0.21	0.10	0.05	1.00

If factors are correlated, they to can be factor analyzed, perhaps uncovering higher order factors.

Higher order factor analysis makes the same assumption as first-order factor analysis: correlations among lower order factors are accounted for by a higher order latent variable.

Although possible to conduct in a series of exploratory models, higher-order factor analysis is better left for confirmatory methods.



Factor analysis can easily capitalize on chance, making it highly desirable to cross-validate the results. The most convincing cross-validation occurs when the results are replicated in a new sample (rather than simply being repeated in the same sample).

Three ways to determine factor replication are to (a) correlate the factor loadings across samples, (b) calculate Tucker's factor congruence (ϕ) coefficient for corresponding factors, and (c) estimate and correlate factor scores using weights from the other sample.

F_1 and F_2 are the corresponding factor loadings in the two samples.

$$\phi = \frac{\sum_{j=1}^{K}((F_{1j} - a)(F_{2j} - b))}{\sqrt{\sum_{j=1}^{K}(F_{1j} - a)^2 \sum_{j=1}^{K}(F_{2j} - b)^2}}$$

For Tucker's congruence coefficient, a = b =0. For loading correlations, a = mean of F_1 and b = mean of F_2.

Tucker's approach has some distinct advantages: First, the numerator is insensitive to scalar multiplication of the loadings. This means that ϕ measures factor similarity independently of the mean absolute size of the loadings.

This is a desirable feature because factor interpretations should be independent of the amount of variance explained by the factor.

Second, the numerator is sensitive to additive constants.

This is a desirable feature because factor interpretations should also be sensitive to additive constants. For example, the loadings (.6, .5, .1) would be interpreted differently than would the loadings (.1, .0, –.4).

If a sample is large enough, it can be split into random halves to provide cross-validation. In our example, we can use two random subsamples, each containing 269 cases.

Because the analyses with the full sample did not make a particularly compelling case for correlated factors, we'll focus on orthogonal rotation in the cross-validation.

	MR1	MR4	MR3	MR2
MR1	0.83187	0.1964	0.1711	-0.62857
MR2	0.02325	-0.4476	-0.1707	0.81534
MR3	-0.47353	0.8163	-0.4922	0.05344
MR4	-0.25441	-0.4366	0.8529	-0.33315

The corresponding factor loadings are highly correlated (the factors do not necessarily rotate to the same order), although we would like these to be even higher if we are finding the same latent variables.

High correlations for non-corresponding factors are undesirable, but this is partly a function of using correlations rather than congruence coefficients.

Congruence coefficients suggest modest cross-validation of the first three factors.

	MR1	MR4	MR3	MR2
MR1	0.93	0.67	0.62	-0.03
MR2	0.62	0.40	0.46	0.79
MR3	0.36	0.91	0.25	0.33
MR4	0.37	0.25	0.91	0.02

Tucker suggested the following guidelines for judging factor replication:

- .98 to 1.00 = *excellent*
- .92 to .98 = *good*
- .82 to .92 = *borderline*
- .68 to .82 = *poor*
- below .68 = *terrible*

In each sample, actual factor scores are computed (using the weight matrix from that sample). In addition, estimated factor scores are calculated using the weight matrix from the other sample.

If the factors are stable over samples, then the weight matrix from one sample should work well in producing factor scores in the other sample.

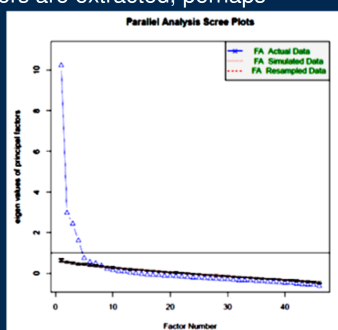
Sample 1

Factor Score Correlations, Sample 1								
	Actual Factor Scores				*Estimated Factor Scores*			
Factor Score	*1*	2	3	4	*1*	2	3	4
Actual 1	1	0	0	0	0.85	0.29	0.19	-0.31
Actual 2	0	1	0	0	0.36	-0.06	0.29	0.91
Actual 3	0	0	1	0	-0.24	0.93	0.08	0.19
Actual 4	0	0	0	1	-0.15	0.01	0.9	0.03
Estimated 1	0.85	0.36	-0.24	-0.15	1	-0.04	0.07	0.01
Estimated 2	0.29	-0.06	0.93	0.01	-0.04	1	0.14	0.02
Estimated 3	0.19	0.29	0.08	0.9	0.07	0.14	1	0.25
Estimated 4	-0.31	0.91	0.19	0.03	0.01	0.02	0.25	1

Sample 2

Factor Score Correlations, Sample 2								
	Actual Factor Scores				*Estimated Factor Scores*			
Factor Score	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*
Actual 1	1	0	0	0	0.8	0.31	-0.13	-0.09
Actual 2	0	1	0	0	0.34	-0.07	0.94	-0.09
Actual 3	0	0	1	0	0.27	0.06	-0.13	0.94
Actual 4	0	0	0	1	-0.37	0.93	0.19	-0.2
Estimated 1	0.8	0.34	0.27	-0.37	1	-0.11	0.09	0.2
Estimated 2	0.31	-0.07	0.06	0.93	-0.11	1	0.04	-0.16
Estimated 3	-0.13	0.94	-0.13	0.19	0.09	0.04	1	-0.22
Estimated 4	-0.09	-0.09	0.94	-0.2	0.2	-0.16	-0.22	1

Cross-validation is especially important when small or weak factors are extracted, perhaps those very close to the rubble in the scree plot or just barely outside the confidence intervals in a parallel analysis. Here we will extract 8 factors.



Tucker's factor congruence indicates that the three factors replicate reasonably well but that five do not. The last two are especially bad.

	MR6	MR1	MR4	MR8	MR2	MR3	MR5	MR7
MR1	0.66	0.61	0.40	0.95	0.35	0.42	0.41	0.12
MR5	0.94	0.45	0.53	0.57	-0.02	0.32	0.46	0.22
MR3	0.50	0.94	0.26	0.44	0.32	0.00	0.48	0.07
MR2	0.09	0.46	0.17	0.54	0.79	0.58	0.55	-0.01
MR8	0.40	0.27	0.75	0.54	0.47	0.68	0.15	0.38
MR4	0.34	0.29	0.82	0.18	0.01	0.10	0.33	-0.08
MR6	0.37	0.05	0.01	0.09	-0.48	0.37	0.06	-0.10
MR7	0.17	0.45	0.26	0.14	0.17	0.06	-0.07	-0.36

The factor score correlations tell a similar story . . .

Sample 1

Factor Score Correlations, Sample 1								
	Actual Factor Scores							
Factor Score	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Estimated 1	0.24	0.13	-0.04	0.93	0.02	0.06	-0.06	-0.13
Estimated 2	0.88	-0.08	0.13	0.08	-0.09	0.01	0.33	0.1
Estimated 3	0.16	0.87	-0.02	-0.12	0.12	-0.12	0.11	0.13
Estimated 4	-0.15	0.15	-0.14	0.1	0.67	0.55	0.54	-0.2
Estimated 5	0.08	0.03	0.52	0.07	0.34	0.51	-0.37	0.38
Estimated 6	-0.04	0.02	0.78	-0.06	-0.11	-0.22	0.32	-0.42
Estimated 7	0.13	0.12	-0.15	0	-0.51	0.5	-0.08	-0.21
Estimated 8	0.03	0.29	0.13	-0.05	0.04	0.06	-0.35	-0.44

Sample 2

Factor Score Correlations, Sample 2								
	Actual Factor Scores							
Factor Score	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Estimated 1	0.09	0.87	0.15	-0.18	0.11	-0.12	0.15	-0.01
Estimated 2	0.22	-0.11	0.86	0.07	-0.09	0.1	0.05	0.28
Estimated 3	-0.08	0.13	-0.04	-0.09	0.6	0.8	-0.03	0.21
Estimated 4	0.9	-0.09	-0.06	0.07	0.03	-0.04	-0.03	0
Estimated 5	0	-0.08	0.1	0.59	0.27	-0.18	-0.63	0.03
Estimated 6	-0.11	0.14	-0.29	0.52	0.51	-0.13	0.56	0.06
Estimated 7	-0.09	0.26	0.15	0.48	-0.28	0.26	0.02	-0.45
Estimated 8	-0.14	0.2	0.16	-0.14	0.35	-0.3	-0.25	-0.51

Other issues in factor analysis:
- How many people?
- The number of variables per factor
- Factoring items versus factoring scales
- Assumptions and distributions
- Reliability for factor score composites
- Confirmatory versus exploratory

Next time . . .
Confirmatory factor analysis