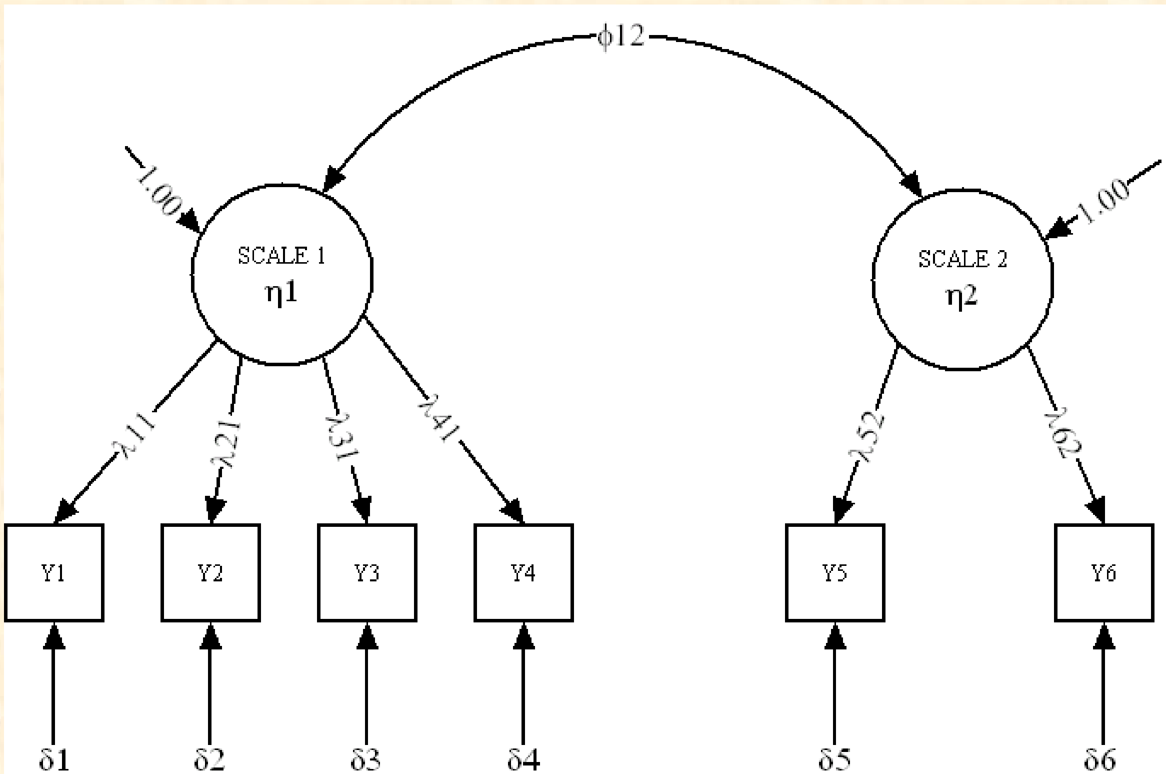
QUESTION: Is it a problem if I use Factor Analysis on dichotomous data in scale development?

By R. L. Brown, Ph.D.
Medical Research Consulting



φ12
1.00
1.00
SCALE 1
η1
SCALE 2
η2
λ11
λ21
λ31
λ41
λ52
λ62
Y1
Y2
Y3
Y4
Y5
Y6
δ1
δ2
δ3
δ4
δ5
δ6

The Model.

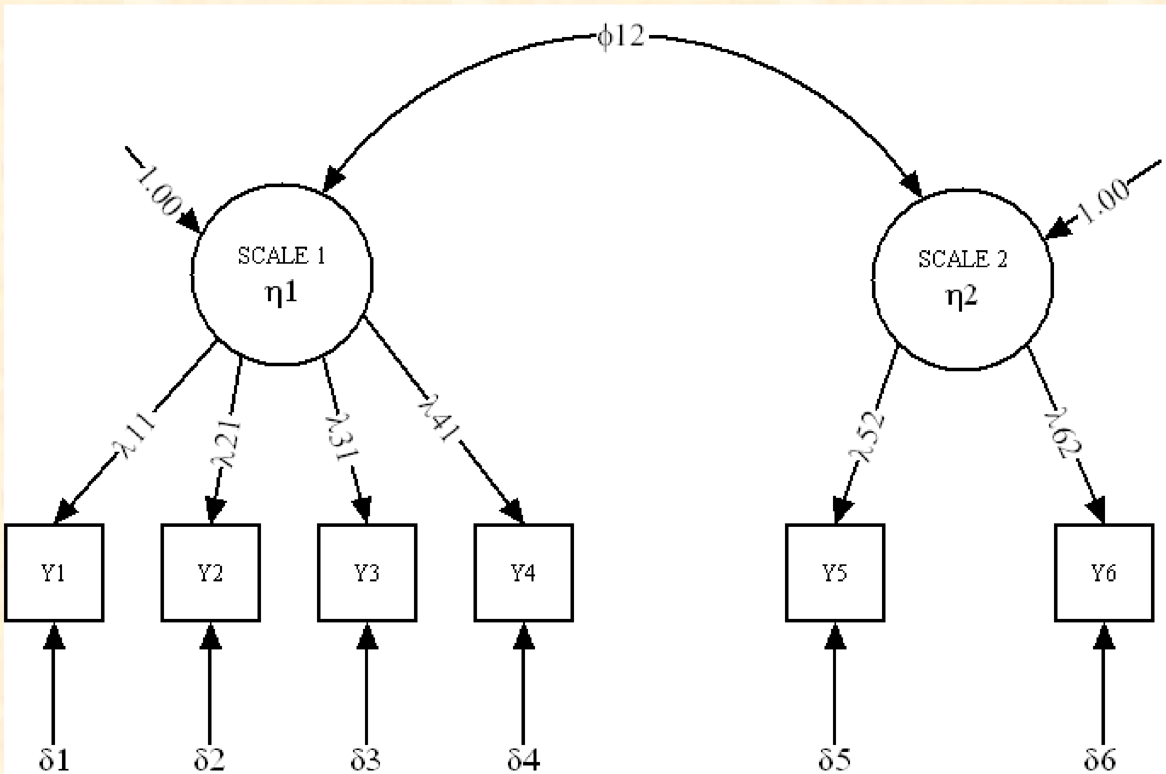
The measurement model used for scale development in structural covariance analyses consists of a set of (p) regression equations linking the (p) measured (Y) variables with (m) scales or factors (latent variables η). The system of equations may be written as the following, using standard terminology;

$$y_1 = \lambda_{11}\eta_1 + \lambda_{12}\eta_2 + \ldots + \lambda_{1m}\eta_m + \varepsilon_1$$
$$y_2 = \lambda_{21}\eta_1 + \lambda_{22}\eta_2 + \ldots + \lambda_{2m}\eta_m + \varepsilon_2$$
$$\vdots$$
$$y_p = \lambda_{p1}\eta_1 + \lambda_{p2}\eta_2 + \ldots + \lambda_{pm}\eta_m + \varepsilon_p$$

Or in matrix notation as;

$$y = \Lambda\eta + \varepsilon$$

$$\Sigma = \Lambda \Phi \Lambda' + \Psi$$



Assumption: Unlimited normally distributed continuous variables

Estimate Pearson/Phi Correlations

$$F = (1/2)\ \mathrm{tr}[(S - \Sigma)\ W\ ]^2$$

Weight matrix $\mathbf{W} = \boldsymbol{\Sigma}^{-1}$ for ML

Ho: $S = \Sigma$

What if the variables are not unlimited continuous, but dichotomous, and possibly asymmetrical?

Estimate tetrachoric correlations not Pearson/Phi

$$F(\theta) = (s - \sigma)' W^{-1} (s - \sigma)$$

Weight matrix **W** = ∗∗∗ for WLS

Simulation Study

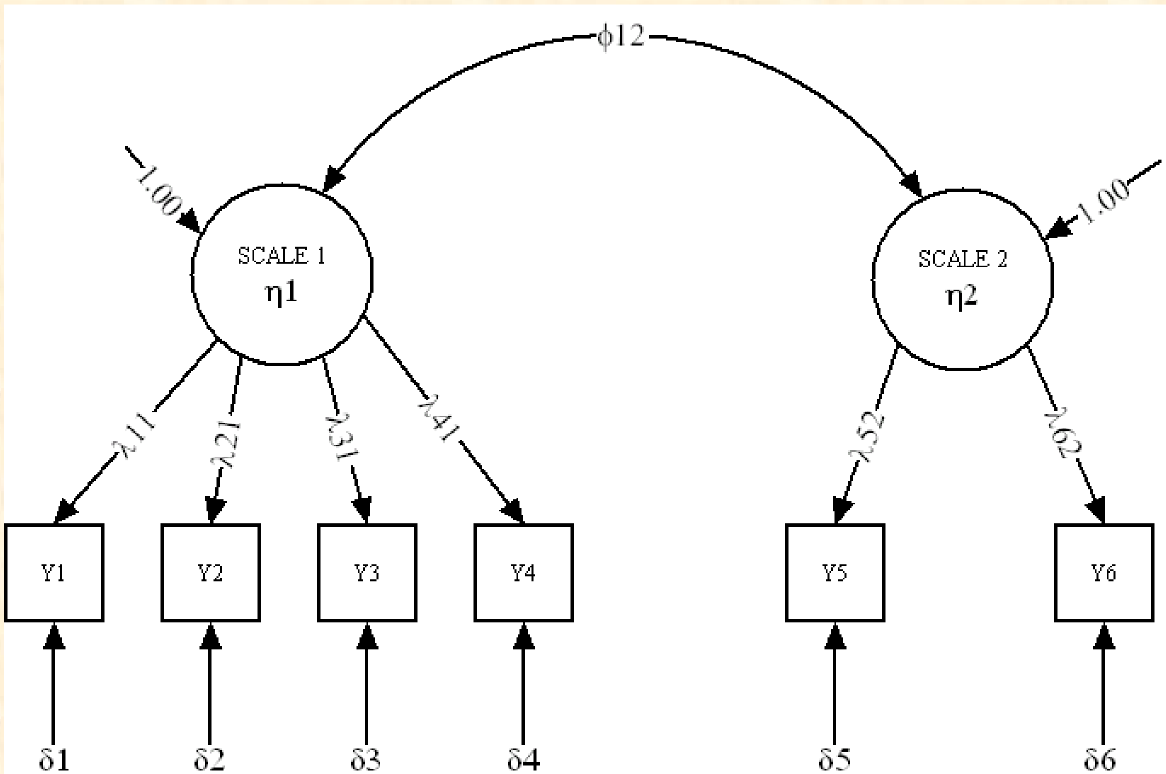

φ12
1.00
1.00
SCALE 1
η1
SCALE 2
η2
λ11
λ21
λ31
λ41
λ52
λ62
Y1
Y2
Y3
Y4
Y5
Y6
δ1
δ2
δ3
δ4
δ5
δ6

Table 1. Population Estimates for the two scale model.

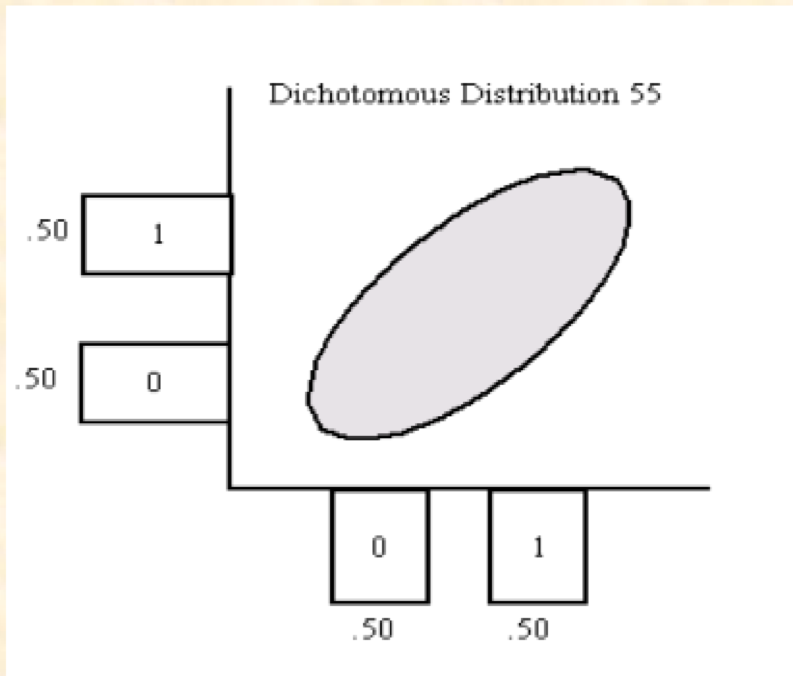
	SCALE 1 λ's	SCALE 2 λ's	ERROR δ's
Y1	0.87	0.00	0.24
Y2	0.83	0.00	0.32
Y3	0.77	0.00	0.41
Y4	0.66	0.00	0.56
Y5	0.00	0.74	0.46
Y6	0.00	0.93	0.14

	SCALE 1 ϕ's	SCALE 2 ϕ's
SCALE 1	1.00	
SCALE 2	0.58	1.00

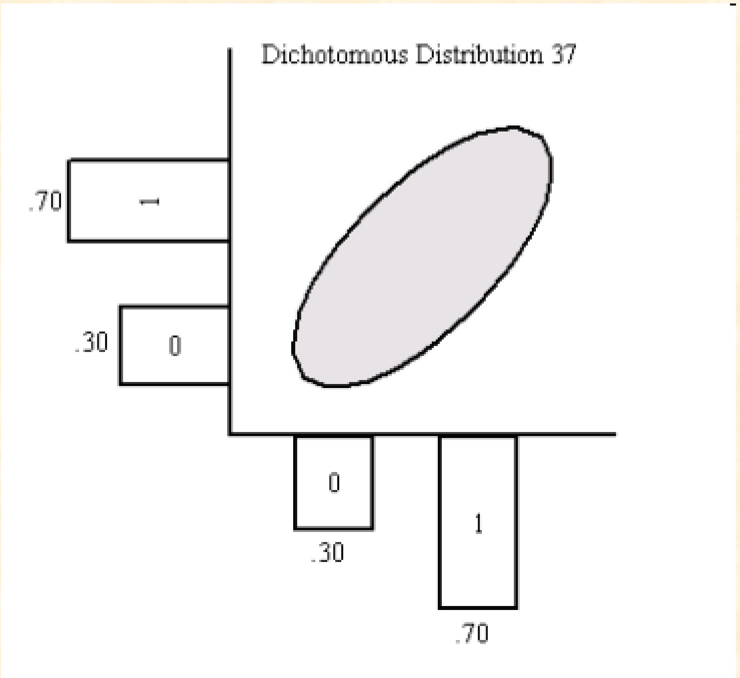
Table 2. Population Correlation Matrix based on theoretical two dimensional model.

	Y1	Y2	Y3	Y4	Y5	Y6
Y1	1.00					
Y2	0.72	1.00				
Y3	0.68	0.63	1.00			
Y4	0.56	0.56	0.52	1.00		
Y5	0.39	0.34	0.32	0.30	1.00	
Y6	0.49	0.44	0.38	0.37	0.68	1.00

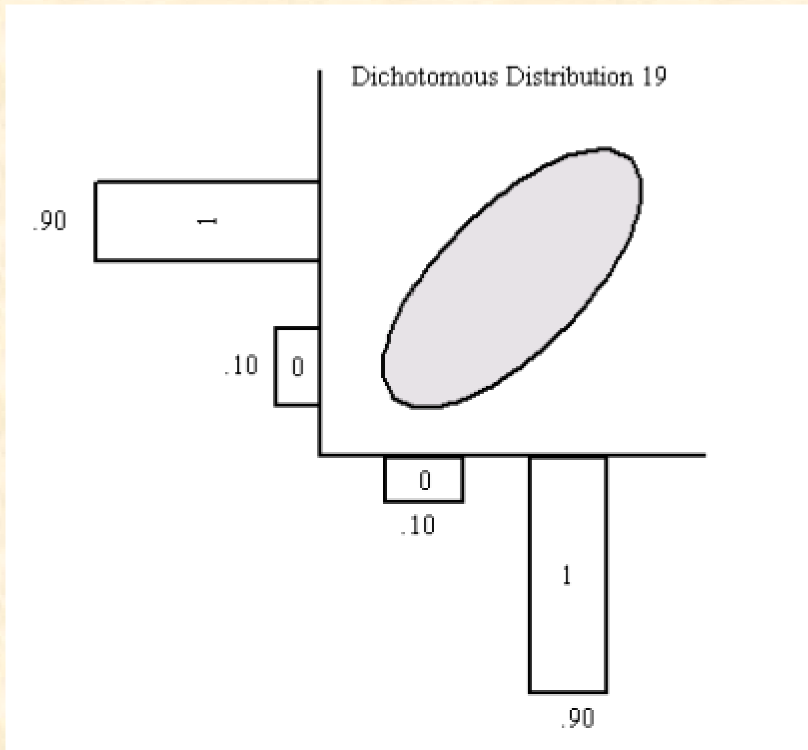
Simulation of asymmetry in dichotomous variables



Dichotomous Distribution 55
.50
1
.50
0
0
1
.50
.50



Dichotomous Distribution 37
.70
1
.30
0
0
.30
1
.70



Dichotomous Distribution 19
.90
1
.10
0
0
.10
1
.90

Assess Issue of

- Different correlational estimates (Pearson vs tetrachoric)
- Different estimation procedures (ML vs WLS)
- Levels of asymmetry

Table 2. Population Correlation Matrix based on theoretical two dimensional model.

	Y1	Y2	Y3	Y4	Y5	Y6
Y1	1.00					
Y2	0.72	1.00				
Y3	0.68	0.63	1.00			
Y4	0.56	0.56	0.52	1.00		
Y5	0.39	0.34	0.32	0.30	1.00	
Y6	0.49	0.44	0.38	0.37	0.68	1.00

100 correlation matrices were simulated for each asymmetrical condition

Percent bias in the correlation =

$((r - \rho)/\rho) * 100$

Distribution 55

Pearson (Phi) Correlations

	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6
Var 1	1.00	18.61	27.35	31.60	30.76	35.30
Var 2	.586	1.00	28.09	28.57	23.82	33.40
Var 3	.494	.453	1.00	30.57	55.00	39.47
Var 4	.383	.400	.361	1.00	28.66	35.40
Var 5	.270	.259	.144	.214	1.00	39.41
Var 6	.317	.293	.230	.239	.412	1.00

ML

Tetrachoric Correlations

	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6
Var 1	1.00	13.05	6.32	5.71	12.82	3.06
Var 2	.814	1.00	6.66	8.03	22.94	5.68
Var 3	.723	.672	1.00	7.50	23.75	1.05
Var 4	.592	.605	.559	1.00	17.00	4.86
Var 5	.440	.418	.244	.351	1.00	7.94
Var 6	.505	.465	.376	.388	.626	1.00

WLS

Distribution 37

Pearson (Phi) Correlations

	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6
Var 1	1.00	38.05	34.26	41.60	39.23	42.44
Var 2	.446	1.00	20.95	28.57	52.94	41.81
Var 3	.447	.498	1.00	44.03	57.81	53.42
Var 4	.327	.400	.291	1.00	79.33	45.94
Var 5	.237	.160	.135	.062	1.00	42.50
Var 6	.282	.256	.177	.200	.391	1.00

Tetrachoric Correlations

	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6
Var 1	1.00	4.16	2.94	2.50	7.94	1.42
Var 2	.690	1.00	19.52	13.75	12.64	0.68
Var 3	.700	.753	1.00	2.88	17.81	12.89
Var 4	.546	.637	.505	1.00	59.00	2.97
Var 5	.421	.297	.263	.123	1.00	7.50
Var 6	.483	.443	.331	.359	.629	1.00

Distribution 19

Pearson (Phi) Correlations

	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6
Var 1	1.00	33.88	61.61	53.92	35.12	25.10
Var 2	.476	1.00	59.04	31.78	26.76	58.40
Var 3	.261	.258	1.00	13.84	33.75	47.36
Var 4	.258	.382	.448	1.00	16.33	42.16
Var 5	.253	.249	.212	.251	1.00	39.26
Var 6	.367	.183	.200	.214	.413	1.00

Tetrachoric Correlations

	Var 1	Var 2	Var 3	Var 4	Var 5	Var 6
Var 1	1.00	1.94	3.08	0.53	1.53	5.51
Var 2	.706	1.00	3.01	5.17	10.88	8.63
Var 3	.659	.611	1.00	0.19	2.81	8.68
Var 4	.563	.589	.521	1.00	15.66	20.00
Var 5	.384	.377	.329	.347	1.00	0.44
Var 6	.517	.478	.413	.444	.683	1.00

Percent bias in the parameters =

$$((\theta^{\varepsilon} - \theta)/\theta) * 100$$

Table 4c. Scale estimates and percent of bias in the estimates for the DI55 distribution.

	Pearson (Phi) Correlation Method		Tetrachoric Correlation Method	
	Parameter Estimate	Percent Bias	Parameter Estimate	Percent Bias
λ_{11}	0.754	13.333	0.8894	2.2299
λ_{21}	0.700	15.663	0.8310	0.1205
λ_{31}	0.630	18.182	0.7596	1.3506
λ_{41}	0.506	23.333	0.6438	2.4545
λ_{52}	0.640	13.514	0.7672	3.6757
λ_{62}	0.770	17.204	0.9114	2.0000
ϕ_{12}	0.532	8.276	0.5718	1.4138
δ_1	0.430	79.167	0.2072	13.6667
δ_2	0.508	58.750	0.3078	3.8125
δ_3	0.604	47.317	0.4214	2.7805
δ_4	0.742	32.500	0.5834	4.1786
δ_5	0.590	28.261	0.4108	10.6957
δ_6	0.408	191.429	0.1682	20.1429
$\chi2$	12.686	45.315	9.4274	7.9885

Table4b. Scale estimates and percent of bias in the estimates for the DI37 distribution.

	Pearson (Phi) Correlation Method		Tetrachoric Correlation Method	
	Parameter Estimate	Percent Bias	Parameter Estimate	Percent Bias
λ_{11}	0.748	14.023	0.8922	2.5517
λ_{21}	0.696	16.145	0.8316	0.1928
λ_{31}	0.616	20.000	0.7576	1.6104
λ_{41}	0.536	18.788	0.6898	4.5152
λ_{52}	0.626	15.405	0.7618	2.9459
λ_{62}	0.756	18.710	0.9048	2.7097
ϕ_{12}	0.522	10.000	0.5726	1.2759
δ_1	0.442	84.167	0.2022	15.7500
δ_2	0.518	61.875	0.3066	4.1875
δ_3	0.622	51.707	0.4254	3.7561
δ_4	0.708	26.429	0.5226	6.6786
δ_5	0.608	32.174	0.4168	9.3913
δ_6	0.428	205.714	0.1794	28.1429
$\chi 2$	13.602	55.808	8.5392	2.1856

Table 4a. Scale estimates and percent of bias in the estimates for the DI19 distribution.

	Pearson (Phi) Correlation Method		Tetrachoric Correlation Method	
	Parameter Estimate	Percent Bias	Parameter Estimate	Percent Bias
λ_{11}	0.632	27.356	0.83360	4.1839
λ_{21}	0.666	19.759	0.85660	3.2048
λ_{31}	0.572	25.714	0.77698	0.9065
λ_{41}	0.470	28.788	0.67260	1.9091
λ_{52}	0.512	30.811	0.70180	5.1622
λ_{62}	0.760	18.280	0.96460	3.7204
ϕ_{12}	0.476	17.931	0.53260	8.1724
δ_1	0.590	145.833	0.30020	25.0833
δ_2	0.554	73.125	0.26480	17.2500
δ_3	0.670	63.415	0.39100	4.6341
δ_4	0.778	38.929	0.54560	2.5714
δ_5	0.734	59.565	0.50160	9.0435
δ_6	0.412	194.286	0.18200	30.0000
χ^2	15.860	81.672	7.43640	14.8179