

Exploratory Factor Analysis of the Twin Data

```
label
  progmatt = 'Progressive matrices (M)'
  reason   = 'Reasoning Ability (R)'
  verbal   = 'Verbal Ability (V)'
  total    = 'Total Psych test score (T)'
  headlng  = 'Head Length (L)'
  headbrd  = 'Head breadth (B)'
  headcir  = 'Head Circumference (C)'
  cephal   = 'Cephalic index (I): head breadth/length'
  headar   = 'Head area (E): Cephalic index * headcir'
  bizyg    = 'Bizygomatic Breadth (Z): Dist betw eyes?'
  weight   = 'Weight of twin (W)'
  height   = 'Height of twin (S)'
  pondrl   = 'Height/weight**1/3 (P)';

/***** twinfac1.sas *****/

TITLE2 'Exploratory Factor Analysis';

include '/folders/myfolders/431s15/twinread.sas';

proc factor simple corr method=ml nfactor=2 rotate=varimax;
  var progmatt reason verbal /* mental */
      headlng headbrd headcir bizyg height weight; /* physical */
```

Twin Data

1

The FACTOR Procedure

Means and Standard Deviations from 74 Observations

Variable	Mean	Std Dev
progmatt	37.9865	8.642034
reason	53.3514	16.510630
verbal	74.7230	24.217641
headlng	186.3784	7.088874
headbrd	146.8784	6.166531
headcir	543.4865	16.591170
bizyg	130.5541	5.888558
height	1651.1351	83.548989
weight	121.7095	21.808476

This handout was prepared by Jerry Brunner, Department of Statistical Sciences, University of Toronto. It is licensed under a Creative Commons Attribution - ShareAlike 3.0 Unported License. Use any part of it as you like and share the result freely. The OpenOffice.org document is available from the course website:

<http://www.utstat.toronto.edu/~brunner/oldclass/431s15/>

Correlations

	progmatt	reason	verbal	headlng	headbrd
progmatt	1.00000	0.55032	0.61360	0.31380	0.16095
reason	0.55032	1.00000	0.75374	0.15817	0.07281
verbal	0.61360	0.75374	1.00000	0.27770	0.14672
headlng	0.31380	0.15817	0.27770	1.00000	0.30018
headbrd	0.16095	0.07281	0.14672	0.30018	1.00000
headcir	0.33140	0.25887	0.31487	0.83358	0.67752
bizyg	0.18434	0.21728	0.24734	0.45409	0.80457
height	0.28396	0.14031	0.22415	0.59151	0.46087
weight	0.22900	0.14521	0.15569	0.56948	0.49422

Correlations

	headcir	bizyg	height	weight
progmatt	0.33140	0.18434	0.28396	0.22900
reason	0.25887	0.21728	0.14031	0.14521
verbal	0.31487	0.24734	0.22415	0.15569
headlng	0.83358	0.45409	0.59151	0.56948
headbrd	0.67752	0.80457	0.46087	0.49422
headcir	1.00000	0.72474	0.61149	0.69145
bizyg	0.72474	1.00000	0.66308	0.65697
height	0.61149	0.66308	1.00000	0.66769
weight	0.69145	0.65697	0.66769	1.00000

Twin Data

2

The FACTOR Procedure
Initial Factor Method: Maximum Likelihood

Prior Commuality Estimates: SMC (SMC stands for Squared Multiple Correlation)

progmatt	reason	verbal	headlng	headbrd
0.45915496	0.65218538	0.64730716	0.86455862	0.81350040
headcir	bizyg	height	weight	
0.92119668	0.79997089	0.62380170	0.60818565	

Preliminary Eigenvalues: Total = 34.2040852 Average = 3.80045391
(Eigenvalues of the "reduced correlation matrix," which is the correlation matrix with communalities on the main diagonal instead of ones. Recall the trace is the sum of eigenvalues.)

	Eigenvalue	Difference	Proportion	Cumulative
1	25.9921147	21.3392395	0.7599	0.7599
2	4.6528752	0.8099972	0.1360	0.8959
3	3.8428780	2.8054203	0.1124	1.0083
4	1.0374577	0.9141814	0.0303	1.0386
5	0.1232764	0.1192130	0.0036	1.0422
6	0.0040633	0.3647915	0.0001	1.0424
7	-0.3607281	0.0766780	-0.0105	1.0318
8	-0.4374061	0.2130398	-0.0128	1.0190
9	-0.6504459		-0.0190	1.0000

2 factors will be retained by the NFACTOR criterion.

Iteration	Criterion	Ridge	Change	Communalities			
1	2.4624138	0.0000	0.3498	0.40014 0.66318 0.55038	0.30234 0.97217	0.41197 0.73835	0.61032 0.49057
2	2.3336477	0.0000	0.1649	0.44002 0.63289 0.56393	0.46060 0.87180	0.51228 0.67471	0.77522 0.45745
3	2.0079966	0.0000	0.2474	0.48736 0.48998 0.56674	0.61763 0.96741	0.75964 0.63163	0.57021 0.50783
4	1.9460303	0.0000	0.1463	0.46456 0.51976 0.55136	0.68353 0.89637	0.82938 0.60128	0.71653 0.44558
5	1.9271433	0.0000	0.1000	0.46505 0.46899 0.55596	0.68397 0.95520	0.83148 0.59638	0.61650 0.48506
6	1.9183528	0.0000	0.0780	0.46435 0.50281 0.54703	0.68339 0.91660	0.83228 0.59385	0.69447 0.44858
7	1.9130477	0.0000	0.0554	0.46473 0.47685 0.55008	0.68403 0.94764	0.83152 0.59255	0.63908 0.47123

Twin Data

3

The FACTOR Procedure
Initial Factor Method: Maximum Likelihood

Iteration	Criterion	Ridge	Change	Communalities			
8	1.9103859	0.0000	0.0421	0.46439 0.49579 0.54580	0.68366 0.92616	0.83196 0.59181	0.68122 0.45190
9	1.9088365	0.0000	0.0306	0.46462 0.48160 0.54782	0.68401 0.94288	0.83154 0.59144	0.65067 0.46479
10	1.9080256	0.0000	0.0230	0.46443 0.49204 0.54568	0.68379 0.93096	0.83181 0.59118	0.67368 0.45443
11	1.9075647	0.0000	0.0168	0.46456 0.48424 0.54690	0.68398 0.94006	0.83158 0.59107	0.65684 0.46165
12	1.9073179	0.0000	0.0126	0.46446 0.48999 0.54579	0.68385 0.93347	0.83174 0.59098	0.66945 0.45603
13	1.9071797	0.0000	0.0093	0.46453 0.48570 0.54649	0.68396 0.93845	0.83161 0.59094	0.66018 0.46004
14	1.9071047	0.0000	0.0069	0.46448 0.48887 0.54589	0.68388 0.93482	0.83170 0.59090	0.66711 0.45697
15	1.9070631	0.0000	0.0051	0.46452 0.48651 0.54629	0.68394 0.93755	0.83163 0.59089	0.66200 0.45919
16	1.9070403	0.0000	0.0038	0.46449 0.48825 0.54597	0.68390 0.93555	0.83168 0.59087	0.66581 0.45751

17	1.9070278	0.0000	0.0028	0.46451 0.48695 0.54618	0.68393 0.93705	0.83165 0.59086	0.66300 0.45873
18	1.9070209	0.0000	0.0021	0.46449 0.48791 0.54601	0.68391 0.93595	0.83167 0.59085	0.66510 0.45780
19	1.9070171	0.0000	0.0015	0.46450 0.48720 0.54613	0.68392 0.93678	0.83165 0.59084	0.66355 0.45848
20	1.9070150	0.0000	0.0012	0.46449 0.48772 0.54603	0.68391 0.93617	0.83167 0.59084	0.66471 0.45797
21	1.9070138	0.0000	0.0008	0.46450 0.48733 0.54610	0.68392 0.93663	0.83166 0.59084	0.66386 0.45834

Convergence criterion satisfied.

The FACTOR Procedure
 Initial Factor Method: Maximum Likelihood
 Significance Tests Based on 74 Observations

Test	DF	Chi-Square	Pr > ChiSq
H0: No common factors HA: At least one common factor	36	490.5715	<.0001
(There are $9(9-1)/2 = 36$ correlations between observed variables.)			
H0: 2 Factors are sufficient HA: More factors are needed	19	129.3591	<.0001
(There are $9(9-1)/2=36$ unique elements in Σ and $9 \times 2=18$ factor loadings. So it seems there should be $36-18=18$ df. But the LR test requires an identifiable parameter, and identifiability can be purchased by setting one factor loading to zero before rotation. So there are 17 factor loadings in the re-parameterized model: $36-17 = 19$ df.)			

Chi-Square without Bartlett's Correction	139.21201
Akaike's Information Criterion	101.21201
Schwarz's Bayesian Criterion	57.43477
Tucker and Lewis's Reliability Coefficient	0.54000

Squared Canonical Correlations

Factor1	Factor2
0.95761280	0.86804173

Eigenvalues of the Weighted Reduced Correlation
 Matrix: Total = 29.1701845 Average = 3.24113161

	Eigenvalue	Difference	Proportion	Cumulative
1	22.5920286	16.0138755	0.7745	0.7745
2	6.5781531	5.3370482	0.2255	1.0000
3	1.2411049	0.7245865	0.0425	1.0425
4	0.5165184	0.4444988	0.0177	1.0603
5	0.0720196	0.0831900	0.0025	1.0627
6	-0.0111704	0.3229722	-0.0004	1.0623
7	-0.3341426	0.3404424	-0.0115	1.0509
8	-0.6745851	0.1351570	-0.0231	1.0278
9	-0.8097420		-0.0278	1.0000

This is the matrix of factor loadings before rotation. For progmatt,
 $> 0.42881^2 + 0.52973^2$
 [1] 0.4644919

	Factor Pattern	
	Factor1	Factor2
progmatt	0.42881	0.52973
reason	0.37446	0.73736
verbal	0.45175	0.79220
headlng	0.80639	-0.11832
headbrd	0.66507	-0.21258
headcir	0.95701	-0.14400
bizyg	0.75642	-0.13628
height	0.66779	-0.10979
weight	0.71347	-0.19215

Twin Data

5

The FACTOR Procedure
 Initial Factor Method: Maximum Likelihood

Variance Explained by Each Factor

Factor	Weighted	Unweighted
Factor1	22.5920286	4.06376784
Factor2	6.5781531	1.59937394

Final Communalities Estimates and Variable Weights
 Total Communality: Weighted = 29.170182 Unweighted = 5.663142

Variable	Communalities	Weight
progmatt	0.46449409	1.8674146
reason	0.68392047	3.1637681
verbal	0.83165645	5.9402206
headlng	0.66427090	2.9749409
headbrd	0.48750360	1.9505690
headcir	0.93660411	15.7799995
bizyg	0.59073481	2.4440025
height	0.45800209	1.8461595
weight	0.54595524	2.2031098

What is the “Weight?” By default, proc factor operates on the correlation matrix. The results can be converted to an analysis of the covariance matrix using weights. The manual says “In ML factor analysis, the weight is the reciprocal of the uniqueness.” So for progmatt, the uniqueness is $1 - 0.46449409 = 0.5355059$, and $1/0.5355059 = 1.867393$.

The FACTOR Procedure
Rotation Method: Varimax

Orthogonal Transformation Matrix

	1	2
1	0.92969	0.36834
2	-0.36834	0.92969

Rotated Factor Pattern

	Factor1	Factor2
progmatt	0.20354	0.65044
reason	0.07653	0.82345
verbal	0.12819	0.90290
headlng	0.79328	0.18703
headbrd	0.69661	0.04734
headcir	0.94277	0.21862
bizyg	0.75343	0.15192
height	0.66128	0.14391
weight	0.73408	0.08415

Variance Explained by Each Factor

Factor	Weighted	Unweighted
Factor1	20.4194016	3.61090172
Factor2	8.7507802	2.05224006

Final Communalities Estimates and Variable Weights

Total Communalities: Weighted = 29.170182 Unweighted = 5.663142

Variable	Communalities	Weight
progmatt	0.46449409	1.8674146
reason	0.68392047	3.1637681
verbal	0.83165645	5.9402206
headlng	0.66427090	2.9749409
headbrd	0.48750360	1.9505690
headcir	0.93660411	15.7799995
bizyg	0.59073481	2.4440025
height	0.45800209	1.8461595
weight	0.54595524	2.2031098

Again for progmatt,
> 0.20354^2 + 0.65044^2
[1] 0.4645007

Exploratory Factor Analysis of Simulated Data

```
/* efac.sas */
options linesize=79 pagesize=200noovp formdlim='_';
title 'Exploratory factor analysis with simulated data';

data latent1; /* True factor structure is like Varimax */
  n = 500;

  /* True factor loadings (All communalities = .25) */
  L11 = .5; L12 = 0;
  L21 = .5; L22 = 0;
  L31 = .5; L32 = 0;
  L41 = .5; L42 = 0;
  L51 = 0; L52 = .5;
  L61 = 0; L62 = .5;
  L71 = 0; L72 = .5;
  L81 = 0; L82 = .5;

  /* Variances of error terms */
  v1 = 1 - L11**2 - L12**2;
  v2 = 1 - L21**2 - L22**2;
  v3 = 1 - L31**2 - L32**2;
  v4 = 1 - L41**2 - L42**2;
  v5 = 1 - L51**2 - L52**2;
  v6 = 1 - L61**2 - L62**2;
  v7 = 1 - L71**2 - L72**2;
  v8 = 1 - L81**2 - L82**2;

do i=1 to n;
  /* Factors are independent standard normal */
  F1 = rannor(0); F2 = rannor(0);
  /* Observed variables all have variance one */
  X1 = L11*F1 + L12*F2 + sqrt(v1)*rannor(0);
  X2 = L21*F1 + L22*F2 + sqrt(v2)*rannor(0);
  X3 = L31*F1 + L32*F2 + sqrt(v3)*rannor(0);
  X4 = L41*F1 + L42*F2 + sqrt(v4)*rannor(0);
  X5 = L51*F1 + L52*F2 + sqrt(v5)*rannor(0);
  X6 = L61*F1 + L62*F2 + sqrt(v6)*rannor(0);
  X7 = L71*F1 + L72*F2 + sqrt(v7)*rannor(0);
  X8 = L81*F1 + L82*F2 + sqrt(v8)*rannor(0);
  output; /* Create a case */
end;

proc factor method=ML rotate=varimax;
  title2 'True factor structure is like Varimax';
  var X1 - X8;
```



```

data latent2; /* Truth is not like Varimax */
  n = 500;

  /* True factor loadings (All communalities = .5^2 + .8^2 = 0.89) */
  L11 = .5; L12 = -.8;
  L21 = .5; L22 = -.8;
  L31 = .5; L32 = -.8;
  L41 = .5; L42 = -.8;
  L51 = .8; L52 = .5;
  L61 = .8; L62 = .5;
  L71 = .8; L72 = .5;
  L81 = .8; L82 = .5;

  /* Variances of error terms */
  v1 = 1 - L11**2 - L12**2;
  v2 = 1 - L21**2 - L22**2;
  v3 = 1 - L31**2 - L32**2;
  v4 = 1 - L41**2 - L42**2;
  v5 = 1 - L51**2 - L52**2;
  v6 = 1 - L61**2 - L62**2;
  v7 = 1 - L71**2 - L72**2;
  v8 = 1 - L81**2 - L82**2;

  do i=1 to n;
    /* Factors are independent standard normal */
    F1 = rannor(0); F2 = rannor(0);
    /* Observed variables all have variance one */
    X1 = L11*F1 + L12*F2 + sqrt(v1)*rannor(0);
    X2 = L21*F1 + L22*F2 + sqrt(v2)*rannor(0);
    X3 = L31*F1 + L32*F2 + sqrt(v3)*rannor(0);
    X4 = L41*F1 + L42*F2 + sqrt(v4)*rannor(0);
    X5 = L51*F1 + L52*F2 + sqrt(v5)*rannor(0);
    X6 = L61*F1 + L62*F2 + sqrt(v6)*rannor(0);
    X7 = L71*F1 + L72*F2 + sqrt(v7)*rannor(0);
    X8 = L81*F1 + L82*F2 + sqrt(v8)*rannor(0);
    output; /* Create a case */
  end;

proc factor method=ML rotate=varimax;
  title2 'Truth is not like Varimax';
  var X1 - X8;

```

Exploratory factor analysis with simulated data
 True factor structure is like Varimax

1

The FACTOR Procedure
 Initial Factor Method: Maximum Likelihood

Prior Communalities Estimates: SMC

X1	X2	X3	X4
0.15896611	0.19179222	0.18988932	0.15876910
X5	X6	X7	X8
0.13435261	0.15335863	0.14831620	0.15189696

Preliminary Eigenvalues: Total = 1.53904092 Average = 0.19238011

	Eigenvalue	Difference	Proportion	Cumulative
1	1.34909241	0.31098963	0.8766	0.8766
2	1.03810279	1.02830964	0.6745	1.5511
3	0.00979314	0.04366770	0.0064	1.5575
4	-.03387455	0.10575369	-0.0220	1.5354
5	-.13962824	0.05421296	-0.0907	1.4447
6	-.19384120	0.01716791	-0.1259	1.3188
7	-.21100911	0.06858522	-0.1371	1.1817
8	-.27959432		-0.1817	1.0000

2 factors will be retained by the PROPORTION criterion.

Iteration	Criterion	Ridge	Change	Communalities			
1	0.0388682	0.0000	0.1529	0.25244	0.34472	0.33210	0.25354
				0.23131	0.27923	0.28150	0.26357
2	0.0388635	0.0000	0.0016	0.25286	0.34506	0.33189	0.25505
				0.22973	0.27799	0.28253	0.26450
3	0.0388634	0.0000	0.0002	0.25278	0.34516	0.33186	0.25506
				0.22957	0.27778	0.28271	0.26470

Convergence criterion satisfied.

Significance Tests Based on 500 Observations

Test	DF	Chi-Square	Pr > ChiSq
H0: No common factors	28	393.9502	<.0001
HA: At least one common factor			
H0: 2 Factors are sufficient	13	19.2050	0.1169
HA: More factors are needed			

Chi-Square without Bartlett's Correction	19.392832
Akaike's Information Criterion	-6.607168
Schwarz's Bayesian Criterion	-61.397074
Tucker and Lewis's Reliability Coefficient	0.963480

Squared Canonical Correlations

Factor1	Factor2
0.63895620	0.57831146

Eigenvalues of the Weighted Reduced Correlation Matrix: Total = 3.14116548 Average = 0.39264569

	Eigenvalue	Difference	Proportion	Cumulative
1	1.76974707	0.39832858	0.5634	0.5634
2	1.37141849	1.20393915	0.4366	1.0000
3	0.16747934	0.05932794	0.0533	1.0533
4	0.10815140	0.09241129	0.0344	1.0877
5	0.01574011	0.05452570	0.0050	1.0928
6	-.03878559	0.04434192	-0.0123	1.0804
7	-.08312751	0.08633032	-0.0265	1.0539
8	-.16945783		-0.0539	1.0000

Factor Pattern

	Factor1	Factor2
X1	0.47672	0.15972
X2	0.54544	0.21829
X3	0.54343	0.19117
X4	0.47671	0.16676
X5	-0.13685	0.45916
X6	-0.24082	0.46879
X7	-0.24621	0.47128
X8	-0.23501	0.45770

Variance Explained by Each Factor

Factor	Weighted	Unweighted
Factor1	1.76974707	1.23990696
Factor2	1.37141849	0.99970068

Final Communalities Estimates and Variable Weights
 Total Communality: Weighted = 3.141166 Unweighted = 2.239608

Variable	Communality	Weight
X1	0.25277411	1.33828993
X2	0.34515767	1.52707894
X3	0.33186198	1.49669969
X4	0.25505982	1.34238297
X5	0.22955587	1.29798207
X6	0.27776158	1.38461238
X7	0.28272060	1.39413158
X8	0.26471602	1.35998793

Exploratory factor analysis with simulated data
 Truth is not like Varimax

3

The FACTOR Procedure
 Initial Factor Method: Maximum Likelihood

Prior Communality Estimates: SMC

X1	X2	X3	X4
0.84388116	0.84717600	0.84959658	0.84791201
X5	X6	X7	X8
0.85689298	0.86937749	0.87227084	0.84695200

Preliminary Eigenvalues: Total = 47.1791673 Average = 5.89739591

	Eigenvalue	Difference	Proportion	Cumulative
1	27.0704456	5.5407046	0.5738	0.5738
2	21.5297410	21.6806815	0.4563	1.0301
3	-0.1509405	0.0229821	-0.0032	1.0269
4	-0.1739226	0.0458219	-0.0037	1.0232
5	-0.2197445	0.0247911	-0.0047	1.0186
6	-0.2445357	0.0508435	-0.0052	1.0134
7	-0.2953792	0.0411176	-0.0063	1.0071
8	-0.3364968		-0.0071	1.0000

2 factors will be retained by the PROPORTION criterion.

Iteration	Criterion	Ridge	Change	Communalities			
1	0.0203369	0.0000	0.0369	0.87871	0.88407	0.88651	0.88447
				0.88840	0.90368	0.90615	0.87752
2	0.0203349	0.0000	0.0002	0.87872	0.88398	0.88646	0.88459
				0.88829	0.90367	0.90636	0.87742

Convergence criterion satisfied.

Significance Tests Based on 500 Observations

Test	DF	Chi-Square	Pr > ChiSq
H0: No common factors	28	5263.1031	<.0001
HA: At least one common factor			
H0: 2 Factors are sufficient	13	10.0488	0.6899
HA: More factors are needed			

Chi-Square without Bartlett's Correction	10.147125
Akaike's Information Criterion	-15.852875
Schwarz's Bayesian Criterion	-70.642781
Tucker and Lewis's Reliability Coefficient	1.001214

Squared Canonical Correlations

Factor1	Factor2
0.97291189	0.96620418

Eigenvalues of the Weighted Reduced Correlation
 Matrix: Total = 64.50603 Average = 8.06325375

	Eigenvalue	Difference	Proportion	Cumulative
1	35.9165695	7.3271091	0.5568	0.5568
2	28.5894605	28.4773099	0.4432	1.0000
3	0.1121506	0.0337536	0.0017	1.0017
4	0.0783971	0.0576221	0.0012	1.0030
5	0.0207749	0.0342166	0.0003	1.0033
6	-0.0134417	0.0605386	-0.0002	1.0031
7	-0.0739803	0.0499203	-0.0011	1.0019
8	-0.1239006		-0.0019	1.0000

Factor Pattern

	Factor1	Factor2
X1	0.51001	0.78652
X2	0.48985	0.80251
X3	0.49799	0.79904
X4	0.49959	0.79686
X5	0.84107	-0.42532
X6	0.85006	-0.42552
X7	0.85035	-0.42810
X8	0.84329	-0.40776

Variance Explained by Each Factor

Factor	Weighted	Unweighted
Factor1	35.9165695	3.86187517
Factor2	28.5894605	3.24760431

Final Communalities Estimates and Variable Weights

Total Communality: Weighted = 64.506030 Unweighted = 7.109479

Variable	Communality	Weight
X1	0.87872066	8.2457066
X2	0.88398014	8.6190458
X3	0.88646106	8.8073129
X4	0.88458628	8.6646020
X5	0.88828997	8.9514229
X6	0.90366870	10.3805801
X7	0.90636477	10.6797209
X8	0.87740788	8.1576387

Exploratory factor analysis with simulated data
Truth is not like Varimax

4

The FACTOR Procedure
Rotation Method: Varimax

Orthogonal Transformation Matrix

	1	2
1	0.87215	0.48923
2	-0.48923	0.87215

Rotated Factor Pattern

	Factor1	Factor2
X1	0.06001	0.93548
X2	0.03461	0.93956
X3	0.04341	0.94052
X4	0.04587	0.93941
X5	0.94162	0.04053
X6	0.94956	0.04476
X7	0.95108	0.04264
X8	0.93497	0.05694

Repeat Rotated factor loadings. Compare them to the truth.

Rotated Factor Pattern

	Factor1	Factor2
X1	0.06001	0.93548
X2	0.03461	0.93956
X3	0.04341	0.94052
X4	0.04587	0.93941
X5	0.94162	0.04053
X6	0.94956	0.04476
X7	0.95108	0.04264
X8	0.93497	0.05694

The Truth

```
/* True factor loadings */  
  
L11 = .5; L12 = -.8;  
L21 = .5; L22 = -.8;  
L31 = .5; L32 = -.8;  
L41 = .5; L42 = -.8;  
L51 = .8; L52 = .5;  
L61 = .8; L62 = .5;  
L71 = .8; L72 = .5;  
L81 = .8; L82 = .5;
```