CHAPTER

EIGHT

FACTOR ANALYSIS OF THE DATA

- THURSTONE'S CENTROID METHOD

- -

Factor Analysis of the Test Data

The shortcomings of the commonly used single score tests have been revealed by the statistical procedure known as 'Factor Analysis'. This procedure was first developed and applied to mental ability tests by Spearman and his co-workers in England. Thurstone has done some pioneering work in this field in United States.

Following the work of Spearman and Thurstone, many other investigators have used factor analysis to study the nature of human abilities. Some of these studies have been used merely to gain a greater understanding of the organisation and components of such abilities. Others have been used as the foundation for the construction of multi-factor tests. Psychologists have contributed much to the understanding of tests and human behaviour.

Factor analysis techniques are now widely applied to varied kinds of instruments such as tests of motor skills, intelligence tests, special aptitude tests and personality inventories. Factor analysis has also been applied in such varying fields as physical measurement, political science, finance and medicine. The method is essentially a statistical tool.

General Nature of Factor Analysis

In the continuing effort of science to establish theoretical systems which will accounts for observed phenomena, identification and measurement of basic variables are of paramount importance. Almost every field of science has set up a number of classificatory categories which can be described and measured. Factor analysis starts with a set of observations obtained from a given sample by means of such a priori measures. It is a method of analysing this set of observations from their inter-correlations to determine whether the variations represented can be accounted for adequately by a number of basic categories smaller than that with which the investigation was started.

Factor analysis as applied to human behaviour is somewhat analogous to quantitative analysis in chemistry where a compound is analysed to discover the nature and amounts of its basic elements.

Factor analysis provides a mathematical model which can be used to describe certain areas of nature.

" In factor analysis " according to B.Fruchter[1] " a series of test scores or other measures are inter-correlated to determine the number of dimensions the test space occupies and to identify these dimensions in terms of traits or other general concepts."

---

[1]Fruchter,B.Introduction to Factor Analysis.(New Jersey:D.Van Nostrand Company,Inc.,1954,p.2

Again in the words of Holzinger and Harman[1] " Factor analysis is a branch of statistical theory concerned with the resolution of a set of descriptive variables in terms of a smaller number of categories or factors. This resolution is accomplished by the analysis of the inter-correlations of the variables."

Thurstone's Centroid Method of Factoring

Since Spearman proposed his criterion of the tetrad difference, a number of procedures for factor analysis have been proposed. All of these start with a correlation matrix. In the present test, the 'Centroid Method' of Thurstone is followed.

The Thurstone's theory and methods have been developed on the basis of matrix algebra.

The term 'Centroid' as used in factor analysis is closely allied with its mechanical concept. In mechanics, the centroid is a point in a mass where the centre of gravity is located. In factor analysis, the centroid of the end points of the test factors might be considered the location of the centre of gravity of equal weights at the points. A centroid is then a centre of gravity. Statistically regarded, it is a mean.

Although the method does not give a mathematically unique, least squares solution, as does the Principal axes method, it is based on summations and is computationally less laborious.

[1]Holzinger,K.J. and Harman,H.H.,Factor Analysis,(Chicago: University of Chicago Press,1941),p.3.

The purpose of factoring a correlation is to account for the inter-correlations with fewer factors than there are tests. Thus factoring should be done so as to minimise the residuals after each factor has been determined. The main centroid axis is regarded as an approximation to the major principal axis of the factor configuration. The main centroid axis is so placed that it has zero projections on all the remaining cooordinate axes. This fact leads to the theorem that " the sum of the coefficients in the correlation matrix is equal to the square of the sum of the first centroid factor loadings, "[1] permitting factoring through simple summational procedure after appropriate reflections. By 'reflection' it is meant that each test factor retains its same length but it extends in the opposite direction. The general policy is to reflect one test vector at a time and note the results, then reflect a second one and so on.

The extraction of each factor loading reduces the residuals in the correlation matrix. The factoring process is ordinarily stopped when the standard deviation of the residuals is less than the standard error of a zero correlation.

Analysis

As stated earlier, Thurstone's Centroid Method is applied to the present test data for the extraction factors.

The details of the steps with regard to the process of factoring suggested by Benjamin Fruchter[2] is followed by the

[1]Thurstone, L.L. Multiple Factor Analysis. (Chicago: University of Chicago Press, 1947), p.152.

[2]Fruchter, Benjamin. Introduction to Factor Analysis. (New Jersey: D.Van Nostrand Company, Inc., 1954), pp.59-85.

investigator.

The coefficients of inter-correlations between the seven sub-tests of the test battery are calculated using the product moment method taking the whole sample N = 2000. The correlation coefficients of the inter-correlations between the sub-tests are given in the correlation matrix in Table 46.

Table :46: Correlation Matrix

No.	Test	1	2	3	4	5	6	7
1.	Numerical ability	-	0.47	0.38	0.37	0.16	0.23	0.32
2.	Spatial ability	0.47	-	0.55	0.43	0.27	0.23	0.15
3.	Reasoning ability	0.38	0.55	-	0.66	0.26	0.25	0.11
4.	Ability to inter-relate	0.37	0.43	0.66	-	0.30	0.30	0.27
5.	Mechanical ability	0.16	0.27	0.26	0.30	-	0.34	0.30
6.	Ability to relate cause and effect	0.23	0.23	0.25	0.30	0.34	-	0.11
7.	Ability to infer from an experimental data	0.32	0.15	0.11	0.27	0.30	0.11	-

The correlation matrix and computations for the first centroid loadings are given in Table 47 on the next page.

Table :47: Correlation Matrix and Computations for the First Centroid Loadings

Test	1	2	3	4	5	6	7	Σ_{j1}
1	(.47)	.47	.38	.37	.16	.23	.32	1.93
2	.47	(.55)	.85	.43	.27	.23	.15	2.10
3	.38	.55	(.66)	.66	.26	.25	.11	2.21
4	.37	.43	.66	(.66)	.30	.30	.27	2.33
5	.16	.27	.26	.30	(.34)	.34	.30	1.63
6	.23	.23	.25	.30	.34	(.34)	.11	1.46
7	.32	.15	.11	.27	.30	.11	(.32)	1.26
Σ_{j1}	1.93	2.10	2.21	2.33	1.63	1.46	1.26	$\Sigma\Sigma j1$ = 12.92
t_{j1}	2.40	2.65	2.87	2.99	1.97	1.80	1.58	T_1 = 16.26
a_{j1}	.3958	.6579	.7125	.7423	.4891	.4469	.3922	$\sqrt{T_1}$ = 4.032
								$\frac{1}{\sqrt{T_1}}$ = .2482
								Σa_{j1} = 3.837 (check)

Note: t_{j1} values are obtained by adding the estimate of the communality of each variable to the column total of the variable i.e. Σ_{j1} values.

Table :48: First Residual-Correlation Matrix

a_{j1}	Test	1	2	3	4	5	6	7	Σ_o
		.396	.658	.713	.742	*.489	*.447	*.392	
.396	1	.2095 (.3132)	.2095	.0977	.0761	±.0336	.0530	.1648	.8807
.658	2	.2095	.2095 (.1181)	.0809	-.0582	±.0518	±.0641	±.1077	.1267
.713	3	.0977	.0809	.1696 (.1516)	.1310	±.0886	±.0687	±.1696	.1343
.742	4	.0761	-.0582	.1310	.1310 (.1095)	±.0628	±.0316	±.0209	.1431
*.489	5	±.0336	±.0518	±.0886	±.0628	.2083 (1009)	.1214	.2083	.1938
*.447	6	.0530	±.0641	±.0687	±.0316	.1214	.1214 (.1402)	±.0652	.0850
*.392	7	.1648	±.1077	±.1696	±.0209	.2083	±.0652	.2083 (.1663)	.1760
	Σ_o	.8807	.1267	.1343	.1431	.1938	.0850	.1760	(check)
	Σ_{j2}	.5675	.0086	-.0173	.0336	.0929	-.0552	.0097	$\Sigma\Sigma_{j2}$ = .6398
Column	6	.4615	.1368	.1201	.0968	-.1499	.0552	.1401	.9015
	5	.5287	.2404	.2973	.2224	.1499	.2980	-.2765	1.4602
	7	.1991	.4558	.6365	.2642	.5665	.4284	.2765	2.8270
	t_{j2}	.4086	.6653	.8061	.3952	.7748	.5498	.4848	T_2 = 4.0846
	a_{j2}	.2023	.3295	.3992	.1957	.3836	.2723	.2401	$\sqrt{T_2}$ = 2.021

* asterisk mark indicates the column reflected.

$1/\sqrt{T_2} = .4952$

$\Sigma a_{j2} = 2.0227$

Table :49: Second Residual-Correlation Matrix

a_{j2}		* .202	* .330	.399	.196	.384	* .272	.240	
	Test	1	2	3	4	5	6	7	Σ_o
*.202	1	.1428 (.1687)	.1428	.0171	.0365	±.0440	±.002	.1163	.4354
*.330	2	.1428	.1428 (.1006)	±.0508	±.1229	±.0749	±.0257	.0348	.0039
.399	3	.0171	±.0508	.0738 (.0104)	.0528	-.0646	±.0398	.0738	-.0011
.196	4	.0365	±.1229	.0528	.1229 (.0926)	-.0125	±.0217	-.0261	-.0013
.384	5	±.044	±.0749	-.0646	-.0125	.1161 (.0609)	.017	.1161	-.0020
*.272	6	±.002	±.0257	±.0398	±.0217	.017	.0398 (.0474)	±.0002	∓.0250
.240	7	.1163	.0348	.0738	-.0261	.1161	±.0002	.1163 (.1507)	.4654
	Σ_o	.4354	.0039	-.0011	-.0013	-.0020	-.0250	.4654	(check)
	Σ_{j3}	.2667	-.0967	-.0115	-.0939	-.0629	-.0724	.3147	
Column	2	-.0189	.0967	.0901	.1519	.0869	-.0210	.2451	
	6	-.0149	.1481	.1697	.1953	.0529	.0210	.2455	
	1	.0149	.4337	.1355	.1223	.1409	.0250	.0129	
	t_{j3}	.1577	.5765	.2093	.2452	.2570	.0648	.1292	T_3 = 1.6397
	a_{j3}	.1233	.4508	.1636	.1918	.2009	.0507	.1013	$\sqrt{T_3}$ = 1.281
									$1/\sqrt{T_3}$ = .7820
									Σa_{j3} = 1.2824

Table :50: Third Residual-Correlation Matrix

a_{j3}	Test	.123 1	.451 2	.164 3	.192 4	* .201 5	.051 6	.101 7	Σo
.123	1	.1039 (.1277)	.0873	-.0031	.0129	.0193	-.0043	.1039	.3437
.451	2	.0873	.0873 (-.0606)	-.0232	.0364	±.0158	.0027	-.0108	.0160
.164	3	-.0031	-.0232	.0976 (.0469)	.0213	±.0976	.0314	.0572	.0329
.192	4	.0129	.0364	.0213	.0511 (.0860)	±.0511	.0119	-.0455	.0719
*.201	5	.0193	±.0158	±.0976	±.0511	.0976 (.0757)	.0067	.0958	.0330
.051	6	-.0043	.0027	.0314	.0119	.0067	.0314 (.0372)	-.0050	.0806
.101	7	.1039	-.0108	.0572	-.0455	.0958	-.0050	.1039 (.1061)	.3017
	Σ_o	.3437	.0160	.0329	.0719	.0330	.0806	.3017	(check)
	Σ_{j4}	.2160	.0766	-.0140	-.0141	-.0427	.0434	.1956	
Column 5		.1774	.1082	.1812	.0881	.0427	.0300	.0040	
	t_{j4}	.2813	.1955	.2788	.1392	.1403	.0614	.1079	$T_4 = 1.2044$
	a_{j4}	.2599	.1808	.2575	.1286	.1296	.0567	.0997	$\sqrt{T_4} = 1.097$

$1 / \sqrt{T_4} = .9238$

$\Sigma a_{j4} = 1.128$

Table :51: Fourth Residual-Correlation Matrix

a_{j4}		* .260	* .181	.258	* .129	.130	.057	.100	
	Test	1	2	3	4	5	6	7	Σ_o
*.260	1	.0779 (.0363)	.0402	±.0702	±.0207	±.0145	±.0191	.0779	.0299
*.181	2	.0402	.0699 (.0545)	±.0699	.0130	±.0077	±.0076	±.0289	-.0064
.258	3	±.0702	±.0699	.0702 (.0310)	-.0120	.0640	.0167	.0314	-.0090
*.129	4	±.0207	.0130	-.0120	.0584 (.0345)	.0343	.0045	-.0584	-.0048
.130	5	±.0145	±.0077	.0640	.0343	.0828 (.0807)	-.0007	.0828	.2389
.057	6	±.0191	±.0076	.0167	.0045	-.0007	.0191 (.0281)	-.0107	.0112
.100	7	.0779	±.0289	.0314	-.0584	.0828	-.0107	.0828 (.0939)	.1880
	Σ_o	.0299	-.0064	-.0090	-.0048	.2389	.0112	.1880	(check)
	Σ_{j5}	-.0064	-.0609	-.0400	-.0393	.1582	-.0169	.0941	
Column	2	-.0868	.0609	.0998	-.0653	.1736	-.0017	.1519	
	1	.0868	.1413	.2402	-.0239	.2026	.0365	-.0039	
	4	.1262	.1153	.2642	.0239	.1340	.0275	.1129	
	t_{j5}	.2041	.1852	.3344	.0823	.2168	.0466	.1957	$T_5 = 1.2651$
	a_{j5}	.1828	.1659	.2995	.0737	.1943	.0418	.1753	$\sqrt{T_5} = 1.124$
									$1/\sqrt{T_5} = .8960$
									$\Sigma a_{j5} = 1.1333$

Table :52: Fifth Residual-Correlation Matrix

a_{j5}		.183	.166	.300	.074*	.194*	.042	.175	
	Test	1	2	3	4	5	6	7	Σo
.183	1	.0459 (.0440)	.0098	.0153	.0072	±.0210	.0114	.0459	.1130
.166	2	.0098	.0245 (.0423)	.0201	.0007	±.0245	.0006	-.0002	.0488
.300	3	.0153	.0201	.0342 (-.0198)	±.0342	.0058	.0041	-.0211	-.0298
* .074	4	.0072	.0007	±.0342	.0342 (.0529)	.0199	.0014	±.0714	-.0235
* .194	5	±.0210	±.0245	.0058	.0199	.0489 (.0482)	±.0089	.0489	.0684
.042	6	.0114	.0006	.0041	.0014	±.0089	.0181 (.0173)	-.0181	.0078
.175	7	.0459	-.0002	-.0211	±.0714	.0489	-.0181	.0714 (.0522)	.0362
	Σ_o	.1130	.0488	-.0298	-.0235	.0684	.0078	.0362	(check)
	Σ_{j6}	.0686	.0065	.0100	-.0764	.0202	-.0095	-.0160	
Column	4	.0542	.0051	.0784	.0764	-.0196	-.0123	.1268	
	5	.0962	.0541	.0668	.1162	.0196	.0055	.0190	
	t_{j6}	.1421	.0786	.1010	.1504	.0685	.0236	.0904	T_6 = .6546
	a_{j6}	.1756	.0971	.1248	.1858	.0847	.0292	.1117	$\sqrt{T_6}$ = .8091

$1/\sqrt{T_6} = 1.236$

$\Sigma a_{j6} = .8089$

Table :53: Centroid Factor Matrix

Sr.No.	Test	Factor					h^2	No.of Reflections
		I	II	III	IV	V		
1.	Numerical ability	.40	.20	-.12	-.26	.18	.31	2
2.	Spatial ability	.66	.33	-.45	-.18	.17	.81	2
3.	Reasoning ability	.71	.40	.16	.26	.30	.85	0
4.	Ability to inter-relate	.74	.20	.19	.13	-.07	.65	1
5.	Mechanical ability	.49	-.38	-.20	+.13	+.19	.48	2
6.	Ability to relate cause and effect	.45	-.27	.05	.06	.04	.28	2
7.	Ability to infer from an experimental data	.39	-.24	-.10	-.10	-.18	.26	1

'h^2'= represents the communality of the test indicating the proportion of the total variance of the test which is held in common with the general factor.

The t_{j1} values shown in the correlation matrix given in Table 47 are obtained by adding the estimate of the communality of each variable to the column total of the variable that is Σ_{j1}.

The residual correlation matrices Tables 48 to 52 indicate the Σ_o values, which are nothing but the algebraic sums of the column totals inclusive of the residual diagonal values while the $\Sigma_{j1...}$ values shown in the Tables are exclusive of the residual diagonal values.

Criteria for Sufficient Factors

There are no exact criteria for stopping extraction of factors. A number of empirical criteria however, have been developed but in the present study Humphrey's rule[1] is applied. This criterion takes into account N, the size of the sample and is dependent on the loadings of only 2 variables (which should be sufficient to establish a factor) rather that on the entire matrix. As per the rule ;

i) The product of the two highest loadings in a column of the Centroid factor matrix is to be calculated.

In the present test the highest loadings in the Fifth residual correlation matrix (Table 53) are 0.1756 and 0.1858. The product is given by

$$0.1756 \times 0.1858 = 0.03262$$

---

[1]Fruchter, B. Introduction to Factor Analysis. (New Jersey: D.Van Nostrand Company, Inc., 1954), p.79-80.

ii) The standard error of a correlation coefficient of zero for the type of correlation and size of the sample being used, is to be found out. It is $1/\sqrt{N}$ for the Pearson Product Moment r in the present test:

$$\frac{1}{\sqrt{N}} = \frac{1}{\sqrt{2000}} = 0.0224$$

(iii) If the product found in step (i) (i.e. 0.03262) above does not exceed twice the standard error found in step (ii) (i.e. 0.0448 = 2 x 0.0224), the factor is probably not significant.

Here the product of the two highest loadings is less than twice the standard error of a correlation coefficient of 0 and hence the factor concerned is not significant. As such the further extraction of factors is stopped at the fifth residual correlation matrix Table 52 and the analysis revealed the presence of five factors in the test.

Benjamin Fruchter[1] gives the following formula to find out the maximum number of factors which can be uniquely determined by 'n' variables.

$$r = \frac{2n + 1 - \sqrt{8n - 1}}{2}$$

where n = the No.of variables and r = the No.of factors present

$$= \frac{2 \times 7 + 1 - \sqrt{8 \times 7 - 1}}{2}$$

$$= 3.792$$

$$= 4$$

[1] Ibid., pp.68-69.

Thurstone's Centroid Method of factor analysis in the present test revealed the presence of 5 factors (the details of the loadings of the factors arrived at are given in Table 53.) while the application of the formula ... given by Benjamin Fruchter showed the presence of four factors only. A further discussion of the factors obtained is given at the proper place in the next chapter.
