Managerial Factor Analysis (MFA)

Bhavya Gulati
Department of Information Technology
Maharaja Agrasen Institute of Technology,
Affiliated to GGSIP University, Delhi, INDIA

ABSTRACT

Factor analysis is a method which is used to reduce number of observed variables into less number of sub groups known as factors which act as a describing identity for the overall outlook of the data. It is a very simple statistical tool available used as data reduction technique which not only an exploratory data analysis technique but also have different subjective approach which can be used as basic functionary ground for various other analysis. The problem arises on how to apply it for further analysis to make productive decision from the data. This paper focuses on managerial aspects of factor analysis, its application classification in segmentation and regression modeling, proposed changes, code and data with result explanation.

Keywords

Exploratory Factor Analysis, PCA, Eigenvalue, Analysis, Managerial Factor Analysis, MFA, SAS.

1. INTRODUCTION

Factor analysis in a way is simple and short analysis, however, there are two things in which analyst struggle the most. It is not about how to perform the analysis but what is the objective of doing this after doing this factor analysis how to apply it further. Doing this is an easier task but how to use it or what is the objective of doing this after doing this factor analysis how to apply it further. There should have absolute clarity why it is needed secondly when it is done how it is used further.

2. APPLICATION

It qualifies as a data reduction technique. In sample data there may be 'n' number of variables but when there is task about multivariate analysis like

(i) Segmentation (cluster analysis): it is recommended to have variables less than 20 as more then that it become complex algorithm and also become difficult to interpret and take some business actionable insight from it so for practical and balance it should be less than 20 (this is guideline it may be possible to use more than that but it make more complex and difficult to use it efficiently)

(ii) Regression modelling (linear & logistic): It is highly subjective. In this model handling depends on case to case basis anything can't be taken as thumb rule in that but the crust for this is that at the end result should be actionable and mutually independent or exclusive variable (non-overlapping) Focus on a variety of variable and reducing overlap must be done.

3. LITERATURE

3.1 EFA

In Multivariate statistics, EFA i.e. Exploratory Factor Analysis is a statistical tool which is use to explore the underlying structure of relatively large set of variables. EFA within factor analysis is aimed to explore all set of data or factors through which it is to identify relationship between them.

3.2 PCA

PCA i.e. Principle Component Analysis a statistical procedure that uses and orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called the principle components. The number of principle components is less than or equal to the number of original variables called factors

4. TECHNIQUE AND METHOD USED

For the proposed analysis process, the method getting used is Exploratory Factor Analysis because here data is present there is no insight about type of grouping going to get, actually way is to put the data into the algorithm and let the algorithm to be supportive to give the structure that is we are basically exploring data with algorithm and getting to know about what are the variables given and which among them are going to be capable of getting combined together to avoid over lapping with much more clarity of data available to us.

It is nothing but Principle Component Analysis getting used because process is way to apply mathematical transformation of data for better result.

Factor analysis is not standalone itself used, it's a step before other multivariate analysis process which aim to prepare data for the analysis, to manage variables in the data and understanding it's texture of working and their co-relation which is pre requisite of analysis process to get a clear picture of data at first then what aim have to be achieved must be clear and what type of variables or parameters required for the process with consideration of results.

Factor analysis help us to group these variables into less number of possible sub groups where all subgroups are of similar nature and are talking about similar thing.

5. RESOURCES INSTIGATION

Let us consider a sample data for reference (Table 1.) in appendix. This is a data table which is used in SAS as reference data over which proceeded functioning is performed. In this table data is in creation order of variables, showing their type, length and label which is nothing but explanation of type of variable it is referring to.

Code 1. (Ref. appendix) is a SAS code representation which directly used in SAS with little modification for performing Factor Analysis at (Table 1.) data. On running this code make encounter with some output file containing various tables, charts and output. Out of these a few of them are shown and their usage is mentioned in this content.

Among those output chart or tables there is a table called Eigen Table (Table 2.) having Eigen value, difference, proportion and Cumulative values.

Cumulative is cumulative value. It is the value percentage of data covered upto particular step or value, in other words it represent the share of data covered by the particular step as share of whole data.

Eigen Value tell about orthogonality.

Reordered factor table: In this table all factors are reordered such that factors in groups are arranged so that the most important factor is at the top of table which is arranged on basis of relation value between factor and its group known as factor loading which is importance of a factor in a factor group.

6. PROPOSED STRUCTURE

Eigen value which is about orthogonality. At standard Eigen value is taken to be 1 or near to 1, to consider the result which means all the factors till then are completely independent of each other. In the above example if the Eigenvalue is near to 1 then its wrap upto 5 groups which means that those 5 groups are completely independent of each other there will be no overlap of the information.

But practically overlapping of groups are not only be the consideration but gathering decent amount of data is also important to be considered.

In many books it is mentioned that take solution where Eigenvalue is 1 or just near to 1 also in some B School that is what is taught. *But the problem is the decisions to be taken are not just limits to statistical decision practically managerial decisions are to be made. So orthogonality is to be balanced out but also capturing enough amount of information in the data so variance is also important to for consideration..*

What is discussed is true while working for segmentation analysis if performing factor analysis for segmentation analysis then capturing enough amount of variance get much importance and correlation between variables and overlap between variables are not that detrimental it impact but not so bad that can't be handled.

So for segmentation triad is to be settle down where Eigenvalue and correlation are together close to 70% because while doing cluster analysis variety in data is needed so that good clusters can be built also in cluster analysis multi collinearity is not a major problem that can't be handled in an easy way because doing this is a working systematic part of the process.

However while doing factor analysis for regression modeling then in this case multi collinearity is a major problem. Yet to some extent what text books say have to agreed that to consider groups with Eigenvalue near to 1(It must be noticed that very rarely perfect 1 is achieved because there will be some noise in the data). Still many people are blindly following this with Eigenvalue just near to 1

Both actually goes with a flip together it's up to situation when we are using them just to keep the standard of keeping the factors (i.e. reduced sub groups) non overlapping and aid to get better productive result.

To decide group is all subjective that is not mathematics. At times have to check +- 1 of number of groups so as to check, analyze and decide to get a better insight of data.

Factor analysis is a subjective technique some of the variables which are believed can be separate out as mentioned if the number of factors are increased they might separate out if

reduces them some might get reshuffled not too much but to some extend but the main thing which should not miss is the decisions to be made should not be pure statistical they must be more managerial. So factor analysis is just a guide for us on which judgement is laid and common sense how to end up using is to know about data & variables which can be broadly divided into these groups and our objective is to reduce the number of variables. All these variables are continuous variables at nominal so factor analysis can't be performed easily.

7. VALIDATION OF RESULT

There is no proven and accepted test to verify that whether produced factor solution is valid or not. However by using a number of random test that can be determined by hit and trial sometimes i.e. performing a similar test with lower of ideal taught values and analyze the result then again do the same with higher values on analyzing all the factors and result with above proposed way will result for a better solution.

But with working and dealing with data and gathering some experience about the structural working of factor analysis and managerial aspects required for the working that can be determined in a quick way.

Factor analysis is a step before analysis so results after analysis will prove better when deal with data in both ways i.e. with proposed way or standard ways.

8. CONCLUSION

In regards to above stated proposed system and guidelines for managerial factor analysis and with study of example from its explanations it can be concluded that the stated proposal is working when deal according to the stated aspects of analysis. Validation of result state that there is no specific way to prove the result and its working with specific measures.

It is very well understandable that factor analysis is subjective and work specific technique which is deal and used in required way. For the managerial aspects this proposed structure work very well with stated guidelines

9. REFERENCES

[1] Rachel J. Goldberg SAS Proceedings paper 268, PROC FACTOR How to interpret the Output of Real World Example.

[2] SAS Documentation 63347, Factor Analysis (9.22 User's Guide).

[3] Multivariate Data Analysis by Joseph F. Hair, William C. Black, Barry J. Babin & Rolph E. Anderson 7th Edition.

[4] INFOGIX blog Feb 04, 2016, Differentiating Between Cluster Analysis and Factor analysis.

[5] Statistics in Plain English by Timothy C. Urdan 3rd Edition.

[6] Data Mining Cookbook, Modeling Data for Marketing, Risk and Customer Relationship Management by Olivia Parr Rud, Wiley Computer Publishing.

[7] Data Analytics – A Practical Approach in ISACA 2016 Volume1.

[8] Predictions for Big Data Analytics in 2016 by Hawlett Packard Enterpise.

[9] Application of Factor Analysis to K- Means Clustering Algorithm on Transportaation Data by Sesham Anand

and P. Padmanabham and A. Govardhan IJCA -2014 Volume 95.

[10] Operational excellence through business process orientation: an intra and inter – organizational analysis by Bahar Movahedi, Kayvan miri Lavassani, Uma Kumar, The TQM Journal, Vol. 28 Iss: 3, pp.

[11] Eigenvalue Statistics of Finite Dimensional Random Matrices for MIMO Wireless Communications by Giuseppa Alfano, Antonia M. Tulino, Angel Lozano, Sergio Verdu.

[12] Sample EigenValue based detection Of High-Dimensional Signals in White noise using relatively few samples by Raj Rao Nadakuditi and Alan Edelman.

[13] Practical Predictive Analytics for Healthcare 101 by Steven S. Eisenberg.

[14] Optimization Solution an overview of Honeywell's Layered Optimization Solution

[15] Rapid Predictive Modeling for Customer Intelligence by Wayne Thompson and David Duling SAS Institure Inc. Paper 113-2010.

[16] The Rise and Value of Predictive Analytics in Enterprise Decision Making by CGI.

[17] The Business Value of Predictive Analytics by Dan Vesset and Henry D. Morrls in IDC June2011.

10. APPENDIX

CODE:1 To be implemented in SAS

ods html file='F:\analytics\Business Analytics\Factor Analysis\Factor.xls';

PROC FACTOR DATA= s.SAQ

METHOD = PRINCIPAL SCREE MINEIGEN=0 NFACTOR = 5
// principle show its PCA, SCREE draw and plot graph MINEIGEN = minimum Eigen, NFACTOR is how many group of factors u want to create

ROTATE = VARIMAX REORDER OUT= Factor;
//Rotate will rotate values so as to get maximum variance in the result, REORDER will arrange such that in each group variable with maximum importance should be at the top, out will give output in factor named file

Varq01 q02
q03 q04
q05 q06
q07 q08
q09 q10
q11 q12
q13 q14
q15 q16
q17 q18
q19 q20
q21 q22
q23;

run;

ods html close;

Table 1. Sample Data used

				Variables in Creation Order
#	Variable	Type	Len	Label
1	q01	Num	8	Statistics makes me cry
2	q02	Num	8	My friends will think I'm stupid for not being able to cope with SAS
3	q03	Num	8	Standard deviations excite me
4	q04	Num	8	I dream that Pearson is attacking me with correlation coefficients
5	q05	Num	8	I don't understand statistics

6	q06	Num	8	I have little experience of computers
7	q07	Num	8	All computers hate me
8	q08	Num	8	I have never been good at mathematics
9	q09	Num	8	My friends are better at statistics than me
10	q10	Num	8	Computers are useful only for playing games
11	q11	Num	8	I did badly at mathematics at school
12	q12	Num	8	People try to tell you that SAS makes statistics easier to understand but it doesn't
13	q13	Num	8	I worry that I will cause irreparable damage because of my incompetance with computers
14	q14	Num	8	Computers have minds of their own and deliberately go wrong whenever I use them
15	q15	Num	8	Computers are out to get me
16	q16	Num	8	I weep openly at the mention of central tendency
17	q17	Num	8	I slip into a coma whenever I see an equation
18	q18	Num	8	SAS always crashes when I try to use it
19	q19	Num	8	Everybody looks at me when I use SAS
20	q20	Num	8	I can't sleep for thoughts of Eigen vectors
21	q21	Num	8	I wake up under my duvet thinking that I am trapped under a normal distribution
22	q22	Num	8	My friends are better at SAS than I am
23	q23	Num	8	If I'm good at statistics my friends will think I'm a nerd

Table 2. Eigen Value (Output of code)

The SAS System The FACTOR Procedure *Initial Factor Method: Principal Components* *Prior Communality Estimates: ONE*			
Eigenvalues of the Correlation Matrix: Total= 23 Average = 1			
	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1	7.2900471	5.55121832	0.317	32%
2	1.7388288	0.42207722	0.0756	39%
3	1.3167515	0.08955337	0.0573	45%
4	1.2271982	0.23932026	0.0534	50%
5	0.9878779	0.09254749	0.043	55%
6	0.8953304	0.08977002	0.0389	59%
7	0.8055604	0.02274045	0.035	62%
8	0.7828199	0.03184875	0.034	65%
9	0.7509712	0.03401347	0.0327	69%

|				
10	0.7169577	0.03336999	0.0312	72%
11	0.6835877	0.01408618	0.0297	75%
12	0.6695016	0.05750398	0.0291	78%
13	0.6119976	0.03425984	0.0266	80%
14	0.5777377	0.0285502	0.0251	83%
15	0.5491875	0.02603716	0.0239	85%
16	0.5231504	0.01475421	0.0227	88%
17	0.5083962	0.05245632	0.0221	90%
18	0.4559399	0.03213628	0.0198	92%
19	0.4238036	0.01601264	0.0184	94%
20	0.4077909	0.02831107	0.0177	95%
21	0.3794799	0.0154576	0.0165	97%
22	0.3640223	0.03096046	0.0158	99%
23	0.3330618		0.0145	100%
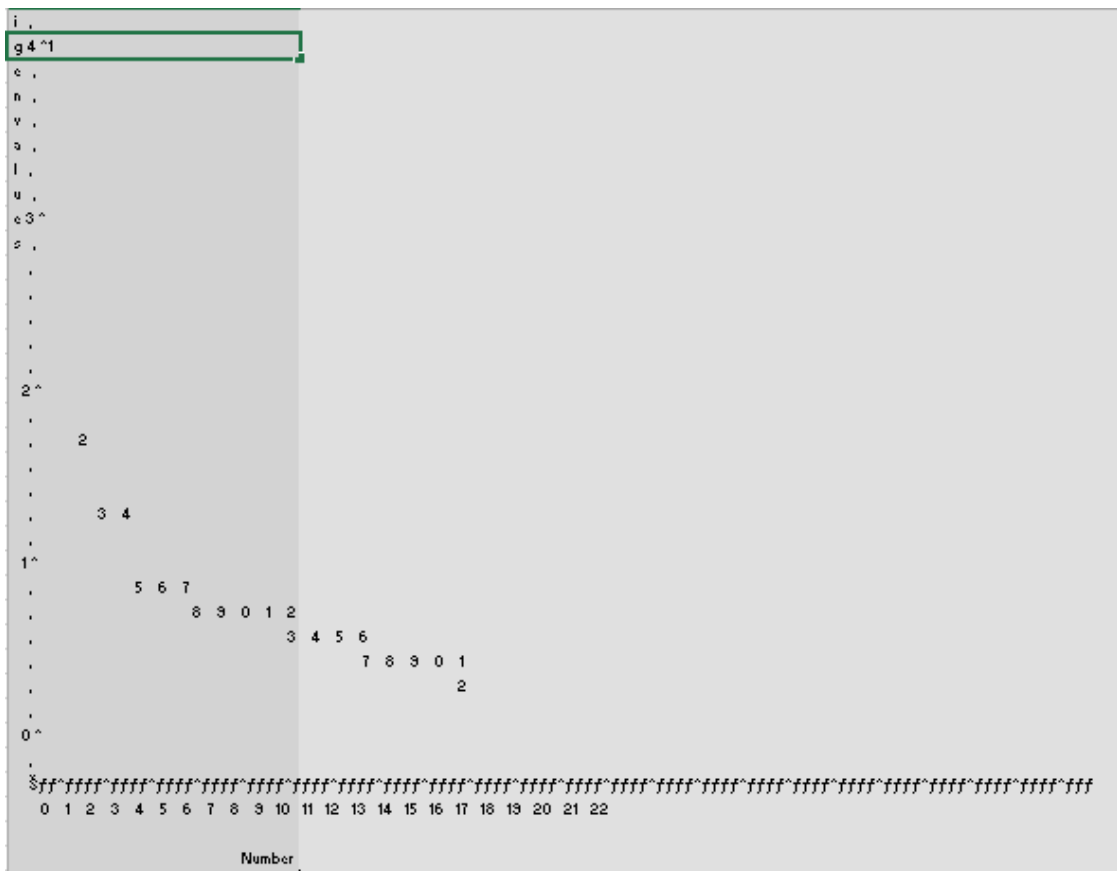
Figure 1. Scree Plot of Eigen Value

		Tech challenge compute	Maths Disliking	Stats paranoid	Stats Averse	Socially conciou s (–)	Computers unfriendl y	Socially conciou s (+)
		Rotated Factor Pattern						
		Factor1	**Factor2**	**Factor3**	**4**	**Factor5**	**Factor6**	**7**
q06	I have little experience of computers	0.775	0.114	-0.019	-0.050	0.001	0.246	-0.077
q18	SAS always crashes when I try to use it	0.730	0.144	0.223	0.162	-0.072	0.089	-0.043
q13	incompetenece with computers	0.693	0.268	0.128	0.063	-0.141	0.099	-0.012
q07	All computers hate me	0.644	0.157	0.246	0.212	-0.020	0.149	-0.060
q14	wrong whenever I use them	0.593	0.149	0.238	0.206	-0.079	0.146	0.002
q12	understand but it doesn't	0.518	0.105	0.433	0.244	-0.165	0.010	0.022
q08	I have never been good at mathematics	0.134	0.851	0.138	0.092	0.043	0.050	-0.018
q11	I did badly at mathematics at school	0.247	0.783	0.124	0.117	-0.119	0.149	-0.037
q17	I slip into a coma whenever I see an equation	0.257	0.748	0.227	0.096	0.007	0.096	-0.063
q05	I don't understand statistics	0.184	0.121	0.734	-0.007	-0.042	0.180	-0.068
q04	coefficients	0.269	0.250	0.641	0.132	-0.040	0.078	0.010
q16	I weep openly at the mention of central tendency	0.187	0.239	0.497	0.290	-0.090	0.407	-0.029
q20	I can't sleep for thoughts of eigen vectors	0.081	0.123	0.192	0.738	-0.104	-0.065	0.081
q21	a normal distribtion	0.376	0.172	0.397	0.548	-0.037	-0.049	0.003
q03	Standard deviations excite me	-0.218	-0.187	-0.386	-0.389	0.369	-0.081	0.115
q19	Everybody looks at me when I use SAS	-0.178	-0.052	0.122	-0.662	0.159	-0.252	0.213
q02	with SAS	-0.090	-0.060	-0.111	-0.089	0.823	0.097	-0.047
q09	My friends are better at statistics than me	-0.024	0.077	-0.025	-0.168	0.665	-0.285	0.165
q10	Computers are useful only for playing games	0.246	0.059	0.168	-0.056	-0.029	0.735	-0.032
q15	Computers are out to get me	0.269	0.295	0.155	0.179	-0.138	0.560	0.040
q23	If I'm good at statistics my friends will think I'm a nerd	0.010	-0.036	-0.087	-0.054	0.013	0.018	0.904
q22	My friends are better at SAS than I am	-0.197	-0.083	0.039	-0.009	0.443	-0.074	0.509

Figure 2. Rotated Factor pattern