EXPLORATORY DATA ANALYSIS IN COTTON QUALITY MANAGEMENT

Maria Elisabete Cabeço Silva and Antonio Alberto Cabeço Silva
School of Engineering
University of Minho
Guimarães, Portugal

Abstract

Computational exploratory data analysis methods include simple basic statistics and more advanced, multivariate exploratory techniques designed to identify patterns in multivariate data sets that include different methods such as: Factor analysis, Cluster analysis, Discriminant Function Analysis, etc.

This paper provides preliminary results on exploratory data analysis in cotton quality management with the application of a factor analysis approach to a spinning data base.

In this study we present the work that we have developed in a Portuguese spinning textile enterprise.

The performance, a factor analysis is used to detect the relationships between processing spinning variables, is demonstrated.

Introduction

Exploratory Data Analysis (EDA) is a data analysis approach. What other data analysis approaches exist and how does EDA differ from these other approaches? Three popular data analysis approaches are: Classical , Exploratory (EDA), Bayesian.

These three approaches are similar in that they all start with a general science/engineering problem and all yield a science/engineering conclusions. The difference is the sequence and focus of the intermediate steps.

For classical analysis, the sequence is:
Problem => Data => Model => Analysis => Conclusions

For EDA, the sequence is:
Problem => Data => Analysis => Model => Conclusions

For Bayesian, the sequence is:
Problem => Data => Model => Prior Distribution => Analysis => Conclusions

Thus for classical analysis, the data collection is followed by the imposition of a model (normality, linearity, etc.) and the analysis, estimation, and testing that follows is focused on the parameters of that model. For EDA, the data collection is not followed by a model imposition; rather it is followed immediately by analysis whose goal is to infer what model would be appropriate.

Finally for a Bayesian analysis, the analyst attempts to incorporate scientific/engineering knowledge/expertise into the analysis by imposing a data-independent distribution on the parameters of the also-imposed model. The analysis thus consists of formally combining both the prior distribution on the parameters and the collected data to jointly make inferences and/or do test assumptions about the model parameters.

Reprinted from the *Proceedings of the Beltwide Cotton Conference*
Volume 2:1271-1273 (2001)
National Cotton Council, Memphis TN

In the real world, data analysts freely mix elements of all of the above 3 (and other approaches) analysis.

Focusing on EDA versus classical analysis, these two approaches differ as follows: Models, Focus, Techniques, Rigor, Data Treatment, and Assumptions.
There exist several methods in EDA for quickly producing and summarize the data sets.

Multivariate statistics help the researcher to summarize data and reduce the number of variables necessary to describe it.

Most commonly multivariate statistics are employed:

- for developing taxonomies or systems of classification;
- to investigate useful ways to conceptualize our group items;
- to generate hypotheses;
- to test hypotheses.

In multiple regression and analysis of variance, several variables are used, however one - a dependent variable - is generally predicted or explained by means of others - independent variables and covariates. These are called dependence methods.

Factor Analysis

Factor analysis approach can be used to analyze interrelationships among a large number of variables and to explain these variables in terms of common underlying dimensions (factors). The statistical approach involves finding a way of condensing the information contained in a number of original variables into a smaller set of dimension (factors) with a minimum loss of information.

Factor analysis used an estimate of common variance among the original variables to generate the factor solution. So, the number of factors will always be less then the number of original variables.

There are four basic factor analysis steps:

- data collection and generation of correlation matrix;
- extraction of initial factor solution;
- rotation;
- interpretation.

The output of a factor analysis will provide several things. First, it shows how output helps to determine the number of factors to be retained for further analysis, One good ruler of thumb for determining the number of factors, is the eigenvalue greater than 1 criteria. This criteria allows to be fairly sure that any factors will account for at least the variance of one of variables used in the analysis. There are other criteria for selecting the number of factors to keep, but this is the easiest to apply, since it is the default of most statistical computer programs.

After fixing the number of factors it is calculated the unrotated factor matrix. The loading listed under the Factor headings represent a correlation between that item and the overall factor. Like correlation's, the ranges from -1 to 1.

This matrix shows the difficulty of understanding an unrotated factor solution.

One way to obtain more understanding results is to rotate this solution. They are more techniques, but the more usual is the orthogonal varimax rotation. Notice that the loadings are distributed between the factors, and that results are easier to interpret.

Now we have a highly interpretable solution of the data. The next step is to name the factors. There are a few rules suggested by methodologists.

Factor names should:

- be brief,
- communicate the nature of underlying construction.

Textile Approach

In this case study we present the information's of the raw materials, the processing and the final product (yarn) that are organized in knowledge bases, distributed by data bases.

The database is composed by the parameters showed in the Table 1 that have been evaluated by the HVI systems. The variables are the parameters of the fibers properties evaluate by the HVI 900 system and the yarns properties parameters evaluate by TENSORAPID 3 and USTER TESTER IV - SX systems.

The most important information to this study has been selected from a production of 24 months of the historic knowledge bases of the Portuguese enterprise. The raw materials are African cotton blends; namely Chade, Camarões, Mali and 6 combed yarns families produced with these blends.

Results and Discussion

It is important to know if the factor analysis technique is applied to this study. So, we have used the KMO (Kaiser Mayer Oklin) method. The KMO is equal to 0.82 so we can continue this study.

Table 2 shows the factors and the eigenvalues associated with each factor, the percentage of the variance associated to each factor and the percentage of variance cumulated in each factor.

By application of the Kaiser Criteria (to fix the number of factors which eigenvalues are greater than 1), we find 3 factors representing the data.

Since the first three factors are the only ones with eigenvalues greater than 1, the final factors solution will represent 84,74 % of the variance in the data.

Table 3 shows the final communalities of each variable to know the proportion of the variable explained by the commons factors in this variable. In this table it is possible to verify that the communalities in general, present greater values.

Table 4 shows the unrotated factor matrix solution. This represents the correlations between the factors and the variables. We present only the values greater than 0,5 to visualize better the higher correlation's, for a better interpretation of the results.

The analysis of these correlations permits to conclude that the first factor is attached with the fibers properties, except the elongation and the micronaire index, the processing parameters (count and twist) and the tenacity and CV Uster of the yarn.

The second factor is correlated with the unevenness properties, the yarn's energy of rupture and the fiber's elongation.

The third factor is only correlated with the yarn's elongation and hairiness and with the fibers micronaire index

To make a better interpretation of this analysis it is recommended to rotate the factorial matrix in order to the axes are closer the variables where they are saturated.

The factorial matrix after rotation (table 5) is a linear combination of the first one and so, explains the same value of the initial variance, where the communalities are not modified but the variance explained by each factor is modified.

In this case we prefer to apply the orthogonal Varimax method. This method maximizes the number of variables that have high loading in the factor.
The analysis of this table shows that:

> The first factor is correlated with the fibers properties and the yarn twist;
>
> The second factor is related with the yarn properties, except the yarn elongation and hairiness and the third factor includes the fiber micronaire and the yarn elongation and hairiness.

Conclusions

In this case study we can conclude that the most important information concerns:

- The parameters that characterize the fibers properties,
- The spinning process variables in their relationship with the yarns properties parameters,
- The property parameters less important are the elongation and the yarn hairiness and the fiber micronaire.

With this work it was possible to put in evidence patterns of textile information and improving the knowledge that must be part of the technology culture of a spinning textile enterprise.

References

Cabeço Silva, M. E., Cabeço Silva, A. A. - 'Factor Analysis in Textile Quality Assurance', IMTEX'95, p. 227 - 231, Lodz, Poland, 1995, May.

Cabeço Silva, M. E., Cabeço Silva, A. A. - 'Análise de Factores em Garantia da Qualidade', 5ªs Jornadas Nacionais de Projecto, Planeamento e Produção Assistidos Por Computador, Ordem dos Engenheiros, p. 353 - 358, Universidade do Minho, 1995, Maio-Junho.

Cabeço Silva, M. E., Cabeço Silva, A. A. - 'Multivariate Analysis in Quality Design of Cotton Blends', 1996 Beltwide Cotton Conferences, Vol 2, p. 1467 - 1471, Nashville, USA, 1996, January.

Araújo Marques, M. J., Cabeço Silva, M. E., Cabeço Silva, A. A. - 'Multivariate Analysis in the Optimisation and Modelling of the Quality Properties of Surgical Clothing', CESA 98, CD Edition, p 254 - 258, Nabeul-Hammamet, Tunisia, 1998, April.

Cabeço Silva, M. E., Marques, M. J. A., Cabeço Silva, A. A. - 'Involving Statistical Factor Analysis for Disposable Surgical Clothing Properties Prediction', EPMESC-Computational Methods in Engineering and Science, ELSEVIER SCIENCE Ltd, Edited by J. Bento, E. Arantes e Oliveira, E. Pereira,, Vol 2, p. 1085-1093, ISBN 008043570 X HC, 1999, Macau, 1999.

Cabeço Silva, M. E. - 'Análise Exploratória de Dados em Sistemas HVI de Metrologia Têxtil', X Congreso Latino-Iberoamericano de Investigación de Operaciones y Sistemas - X CLAIO, México, D. F., Edição em CD, México, 2000, Septiembre.

Tukey, J. - 'Exploratory Data Analysis', New York, Macmillan, 1973.

Brandt S. - 'Statistical and Computation Methods in Data Analysi', Northholland, Second Revised Edition, 1989.

Karson, M. J. - 'Multivariate Statistical Methods' The Iowa State University Press, 1982.

Kshirasagar, A. M. - 'Multivariate Analysis', Edited by D. B. Owen, 1978.

Table 1. Fibers and Yarns Properties.

PROPERTY/VARIABLE	UNITY
Micronaire Index	mg / "
Length	mm
Uniformity Ratio	%
Tenacity (f)	cN/tex
Elongation (f)	%
Count	10 < tex < 40
Twist	2750 < α tex < 4300
CV Uster	%
Thins Places	/ 1000 m
Thick Places	/ 1000 m
Neps	/ 1000 m
Hairiness	/ 100 m
Tenacity (Y)	cN/tex
Elongation (Y)	%
Energy of Rupture	cN/tex

Table 2. Factors Extraction - Combed.

PROPERTY	Factor	Eigenvalues	% Var.	% Cum.
Micronaire	1	5.642	37.613	37.613
Length	2	3.861	25.740	63.353
Uni.Ratio	3	3.208	21.387	84.740
Tenacity (f)	4	0.948	6.320	91.060
Elongation (f)	5	0.430	2.867	93.927
Count	6	0.378	2.520	96.447
Twist	7	0.206	1.373	97.820
CV Uster	8	0.103	0.687	98.507
Thin Places	9	0.083	0.553	99.060
Thick Places	10	0.058	0.387	99.447
Neps	11	0.034	0.226	99.673
Hairiness	12	0.026	0.173	99.846
Tenacity (Y	13	0.019	0.127	99.973
Elongation (Y)	14	0.003	0.020	99.993
E. of Rupture	15	0.001	0.007	100

Table 3. Communalities - Combed.

PROPERTY	Comm-uality	Factor	Engen-values	%Var.	% Cum.
Micronaire	0.944	1	5.642	37.613	37.6
Length	0.982	2	3.861	25.740	63.3
Uni. Ratio	0.787	3	3.208	21.387	84.7
Tenacity (f)	0.739	4	0.948	6.320	91.0
Elongation (f)	0.757	5	0.430	2.867	93.9
Count	0.756	6	0.378	2.520	96.4
Twist	0.732	7	0.206	1.373	97.8
CV Uster	0.979	8	0.103	0.687	98.5
Thin Places	0.985	9	0.083	0.553	99.1
Thick Places	0.647	10	0.058	0.387	99.4
Neps	0.871	11	0.034	0.226	99.6
Hairiness	0.893	12	0.026	0.173	99.8
Tenacity (Y)	0.737	13	0.019	0.127	99.9
Elongation (Y)	0.953	14	0.003	0.020	99.9
E. of Rupture	0.733	15	0.001	0.007	100

Table 4. Factorial Matrix - Combed.

PROPERTY	Factor 1	Factor 2	Factor 3
Micronaire			0.845
Hairiness			0.802
Elongation (Y)			0.745
CV Uster	0.742		
Elongation (f)		0.555	
Count	- 0.671		
Tenacity (Y)	0.784		
Thin Places		- 0.534	
Thick Places		- 0.664	
Neps		- 0.732	
Uni. Ratio	0.719		
Tenacity (f)	0.654		
Twist	0.622		
Length	0.686		
E. of Rupture		0.601	

Table 5. Factorial Matrix after Rotation - Combed.

PROPERTY	Factor 1	Factor 2	Factor 3
Uni. Ratio	0.966		
Length	0.962		
Tenacity (f)	0.884		
Twist	0.836		
Elongation (f)	0.706		
Thick Places		0.925	
CV Uster		0.922	
Neps		0.921	
Thin Places		0.732	
E. of Rupture		-0.674	
Tenacity (Y)		0.620	
Count		-0.518	
Hairiness			0.926
Elongation (Y)			0.924
Micronaire			0.851