

BIBLIOTHÈQUE

DEC 21 1983

ÉCOLE POLYTECHNIQUE
MONTREAL



DÉPARTEMENT DE GÉNIE INDUSTRIEL

QUANTITATIVE TECHNOLOGY MODEL:
A STATISTICAL ANALYSIS

by

Laurent Villeneuve, ing., M. Ing.
Bernard Clément, Ph.D.

RAPPORT TECHNIQUE EP83-R-40

Ecole Polytechnique de Montréal

CA2PQ
UP 5
R83-40
(Ang)
ex.2

Campus de l'Université
de Montréal
Case postale 6079
Succursale 'A'
Montréal, Québec
H3C 3A7

BIBLIOTHÈQUE

DEC 21 1983

ÉCOLE POLYTECHNIQUE
MONTREAL

QUANTITATIVE TECHNOLOGY MODEL:

A STATISTICAL ANALYSIS

BY

Laurent Villeneuve, ing., M. Ing.

Bernard Clément, Ph.D

Don

CONTENTS

	page
Abstract	ii
Introduction	1
The Problem	2
The Model	3
Identification of the Relations between variables	7
a) Pearson's correlation	7
b) Contingency tables	8
c) Results	8
Regrouping of similar cases	10
Comparaison of two populations	17
a) T-test for means	17
b) Differences in probabilities for each variable level .	18
c) Discrimination of companies	21
d) Results	22
Conclusion	28
Bibliography	30
Appendix A	32

ABSTRACT

This article presents the methodology used to analyse a mixture of qualitative and quantitative information in the study of an industrial sector. The "model T" gives a global indicator on the technological situation of a sector and of each enterprise in this sector. To this tool some statistical tests were joined to identify the strengths and weaknesses of the production units. This technological model was tested in the wood furniture sector of the province of Quebec on the request of the Trade and Commerce Department.

INTRODUCTION

The problem of analyzing data banks is very well known and has been considered by many researchers. However, when these data are a mixture of qualitative and quantitative information relative to a population, it is obvious that the treatment is very different than when the data deal solely with quantitative information. Furthermore, when the sample is small (40) and each element of this set contains a large number (500) of variables, the situation is not the same as when the sample is large (1 000) and the number of variables (30) is small. The above describes the conditions that are encountered when a study of a specific industrial sector has to be made through a survey of a small number of manufacturing companies.

THE PROBLEM

The Trade and Commerce department of the Province of Quebec requested that strengths and weaknesses of the production units in an industrial sector (wood furniture) be identified. To answer this request, a questionnaire (containing more than 500 questions) was elaborated and utilized to survey the sector. Financial support was received to visit 45 companies. Finally, a similar study with the same sample size was made in France.

THE MODEL

A model was developed [4] [7] defining technology (T) as a multi-dimensional function:

$$T = F (X_1, X_2, X_3, X_4, X_5) \quad (1)$$

where X_i are the component vectors

X_1 : inputs

X_2 : physical processes

X_3 : organizational processes

X_4 : human resources

and X_5 : outputs

each vector being a multi-dimensional function of variables

$$X_1 = f_1 (x_{11}, x_{12}, \dots, x_{1n_1})$$

$$X_2 = f_2 (x_{21}, x_{22}, \dots, x_{2n_2})$$

\vdots

$$X_5 = f_5 (x_{51}, x_{52}, \dots, x_{5n_5})$$

The values of n_1, n_2, n_3, n_4, n_5 were established and reviewed by a panel of 18 professionals from academia, industry and government. After consultation with those experts, the following number of variables* was obtained

$$n_1 = 8 \quad n_2 = 8 \quad n_3 = 16 \quad n_4 = 11 \quad n_5 = 8$$

Since the state of each variable is not binary, it was essential to break down each one by levels of technology availability.

To synthesize this structured information in order to obtain the picture of the industrial sector and to compare each company to its sector, it was decided to write the technology function as a linear combination of the component vectors, X_i , as follows:

$$T = \sum_{i=1}^5 W_i X_i \quad (2)$$

* See appendix A

where i designates the component vector $i = 1, \dots, 5$

W_i : weight of the vector

X_i : component vector

Also, each component vector was expressed as

$$X_i = \sum_{j=1}^{n_i} w_{ij} x_{ij} \quad (3)$$

where j designates the variable $j = 1, \dots, n_i$

w_{ij} represents the weight of the variable x_{ij}

therefore, the model T is

$$T = \sum_{i=1}^5 W_i \sum_{j=1}^{n_i} w_{ij} x_{ij}. \quad (4)$$

Also, the weights W_i and w_{ij} must satisfy the constraints

$$\sum_{i=1}^5 W_i = 1 \quad (5)$$

and

$$\sum_{j=1}^{n_i} w_{ij} = 1 \quad i = 1, \dots, 5 \quad (6)$$

To compute the value of T , w_i and w_{ij} had to be defined.

a) w_i and w_{ij}

The same previously mentioned panel of experts, used Delphi's method to establish each weight. The ranking of the weights was verified using the Kendall [2] coefficient of concordance (W) i.e.

$$W = \frac{S}{\frac{k^2}{12} (N^3 - N)} \quad (7)$$

where W : degree of agreement among k judges

$$S = \sum (R_j - \frac{\sum R_j}{N})^2$$

k : number of ranking judges

N : number of individual variables ranked

R_j : sum of the ranks assigned by the experts for each variable.

$\frac{k^2}{12} (N^3 - N)$: maximum possible sum of the squared deviation i.e. the sum S which would occur with perfect agreement among k rankings.

Two rounds were necessary to arrive at a satisfactory coefficient of concordance.

b) x_{ij}

Since in any given sector of technology, the variables are numerous and can take on different units and that, inevitably, there would be different types of variables, it was decided to use an ordinal level for all the variables. The values of x_{ij} were restricted between 0 and 1, taking from 3 to 8 discrete equidistant levels.

This model T gives a score for each of the visited companies T_i and a score for the industrial sector \bar{T} . These results allow a classification of these enterprises and the identification of the features i.e. the strengths and weaknesses of both the visited companies and the sector as a whole.

Now that this data bank is organized, it is possible to utilize conventional mathematical tools to pursue a detailed analysis, for example:

- a) to identify the relations between variables
- b) to regroup similar cases
- c) to compare two populations.

IDENTIFICATION OF THE RELATIONS BETWEEN VARIABLES

After a first compilation or aggregation of the collected data, the next step is the analysis of all the answers received. This is necessary in order to comprehend the state of technology in the sector (and firms), to verify some hypotheses and finally, to identify the relations that exist between some variables. This can be accomplished by using two available techniques, namely the Pearson correlation coefficient (r) and the contingency tables.

a) Pearson's correlation

r is used to measure the strength of the relationship between two interval-level variables and is computed using the following formula:

$$r = \frac{\sum_{k=1}^c (x_k - \bar{x})(y_k - \bar{y})}{s_x s_y} \quad (8)$$

where x_k : value of variable x on company k

y_k : value of variable y on company k

c : number of companies

s_x : standard deviation of variable x

s_y : standard deviation of variable y

b) Contingency tables

The contingency tables which are more appropriate for ordinal variables give the relationship between variables. Different statistical treatments can be applied namely: Chi-Square test of statistical significance, measures of association such as Cramer's V, the contingency coefficient, lambda, etc...

c) Results

The following two examples are the results that were obtained with the above mentioned techniques. Pearson's correlation was used to obtain the relationship between two quantitative variables, while the contingency tables were employed for two qualitative variables.

Example 1

Question: Is there a correlation between the variables within each vector?

Results: Tables 1, 2, 3, 4, 5 give the Pearson's correlation coefficients.

Comments: There is no correlation between the variables within each vector. Out of 259 coefficients, only 40 have an absolute value greater than 0,40 and only 1 (one) is greater than 0,75.

Therefore, one can conclude that each variable is independent and that each set of variables within each vector gives a good coverage of the theme.

Example 2

Question: Is there a relation between the remuneration mode (piece rate or hourly rate) of the direct labor workers in this sector and the fact that those workers belong to a Union? (2 qualitative variables).

Results: (see TABLE 6)

χ^2 : 2,86

P: 0,091

C: 0,29

where χ^2 : Chi-square statistical test

P: significance level associated with testing zero correlation

C: contingency coefficient

Comments: Yes, there seems to be a relation between the remuneration mode of the direct labor workers and the fact those workers belong to a Union. One can observe that 18 out of 21 non-unionized companies pay their employees on an hourly rate basis and that 10 out of 13 piece rate remuneration mode companies are unionized.

REGROUPING OF SIMILAR CASES

Usually companies are grouped on the basis of size and/or the type of products that comes out of the company. However, when one starts to add on more points of comparison, it becomes more difficult eventually to group similar cases.

Cluster analysis was the tool utilized to arrive at a grouping of companies where no a priori classification was made. This technique usually groups variables according to some measure of relationship between them, such as a correlation coefficient or a distance coefficient. The criterion for clustering usually involves minimizing the within-group variance of the variables and maximizing the variance between groups [6].

The grouping of homogeneous enterprises [3] is done sequentially with the transpose of the matrix shown in table 7 serving as input [1].

TABLE 3

PEARSON'S CORRELATION: ORGANIZATIONAL PROCESSES VS ORGANIZATIONAL PROCESSES

	3.1	3.2	3.3	3.4	3.5	3.6	3.7	3.8	3.9	3.10	3.11	3.12	3.13	3.14	3.15	3.16
3.1	1,0	0,43	0,52	0,47	0,18	0,43	0,33	0,28	0,38	0,17	0,18	0,12	0,39	0,15	0,05	0,19
3.2		1,0	0,38	0,43	0,38	0,49	0,13	0,21	-0,02	0,36	0,12	0,27	0,08	0,05	-0,10	-0,03
3.3			1,0	0,31	0,30	0,36	-0,04	0,37	0,46	0,12	0,50	0,16	0,30	0,17	-0,09	0,18
3.4				1,0	0,22	0,30	0,26	0,27	0,11	0,24	0,04	0,01	0,07	0,28	-0,01	-0,01
3.5					1,0	0,42	0,22	0,42	-0,04	0,35	0,16	0,31	0,30	0,04	-0,07	0,05
3.6						1,0	0,14	0,45	0,05	0,04	0,16	0,43	0,29	0,18	0,05	0,20
3.7							1,0	0,23	0,11	0,27	-0,17	0,01	0,15	0,18	0,24	-0,15
3.8								1,0	-0,02	0,07	0,18	0,11	0,31	0,12	0,08	-0,04
3.9									1,0	0,01	0,21	0,04	0,20	0,16	0,16	0,09
3.10										1,0	0,03	-0,01	0,09	0,09	-0,02	0,01
3.11											1,0	-0,04	-0,01	-0,51	0,25	0,23
3.12												1,0	0,03	-0,13	0,01	0,13
3.13													1,0	0,03	0,09	-0,13
3.14														1,0	-0,27	0,30
3.15															1,0	-0,33
3.16																1,0

TABLE 6

REMUNERATION MODE PER ADHERENCE TO A UNION

COUNT COL PCT TOT PCT	UNION		ROW TOTAL
	YES	NO	
HOURLY RATE	14 58.3 31.1	18 35.7 40.0	32 71.1
PIECE RATE	10 41.7 22.2	3 14.3 6.7	13 28.9
COLUMN TOTAL	24 53.3	21 46.7	45 100.0

TABLE 7

DATA MATRIX

COMPANY	VARIABLE					
	x_1	x_2	x_k	x_{n-2}	x_{n-1}	x_n
c_1						
c_2						
• • c_j • •			c_{ik}			
c_{m-2}						
c_{m-1}						
c_m						c_{mn}

The problem is: given a set of companies (c_1, c_2, \dots, c_m , $c_i \in S$) each defined by a set of variables (x_1, x_2, \dots, x_n), to group S into subgroups such that the subsets are as internally homogeneous and as mutually dissimilar as possible. The definition of the dissimilarity between any two firms c_i and c_j is:

$$d_{ij}^2 = \{c_i - c_j\}^2 = \sum_{k=1}^n (c_{ik} - c_{jk})^2 \quad (9)$$

i.e. squared Euclidean distance.

Example 3

Question: Can a stratified configuration of the companies be established?

Results: (see TABLE 8)

Comments: The clustering has been executed with 99 variables (51 from the model and 48 complementary data on product and financial information). Since this technique does not identify the specific variables that have been employed to define each cluster, a brief investigation was performed to discover those variables that were probably used for the grouping. The analysis of the 3 clusters reveals that cluster 1 contains companies (less 2) that produce furniture in large quantity, cluster 2

groups firms (less 4) that are in the high -rice bracket and cluster 3 associates the enterprises (less 2) that are in the medium-price bracket.

COMPARISONS OF TWO POPULATIONS

Having gathered data on two sets of companies in two countries, it was then necessary to confront the two groups to highlight the similarities and the disparities. The t-test for means, Chi-square test for differences in probabilities, and the discriminant analysis were utilized to achieve this comparison.

a) T-test for means

In this case, one wants to verify that the means of the two populations are equal ($H_0: \mu_1 = \mu_2$) and establish a confidence interval.

if $t > t_{\alpha/2; n_1 + n_2 - 2}$ H_0 is rejected

where

$$t = \frac{(\bar{X}_1 - \bar{X}_2)}{\sqrt{\frac{(n_1-1) s_1^2 + (n_2-1) s_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (10)$$

\bar{X}_1 : means of Quebec firms

\bar{X}_2 : means of France firms

s_1 : standard deviation of Quebec firms

s_2 : standard deviation of France firms

n_1 : number of Quebec firms

n_2 : number of France firms

α : level of significance

$t_{\alpha/2; n_1 + n_2 - 2}$: upper percentage point of a Student variable with $n_1 + n_2 - 2$ degrees of freedom.

b) Differences in probabilities for each variable level

"The Chi-square test for differences in probabilities" has been used in this case. Thus for a question Q, one can represent the results as follows.

		ANSWERS			
		1	2	c	TOTAL
QUEBEC	0_{11}	0_{12}	-----	0_{1c}	n_1
FRANCE	0_{21}	0_{22}	-----	0_{2c}	n_2
		s_1	s_2	s_c	

where $0_{i\ell}$: number of observations by country (i) and by variable level (ℓ)

$i = 1 = \text{Quebec}$

$i = 2 = \text{France}$

$\ell = 1, \dots, c$

s_ℓ : total by column

letting $n_i = \sum_{\ell=1}^c 0_{i\ell}$ where $i = 1, 2$

in this case

$$n_1 = n_2 = 41$$

$$S_\ell = O_{1\ell} + O_{2\ell}$$

Letting $p_{i\ell}$ represent the probability that a firm of country i answers ℓ to a question, the hypothesis to be tested is: all the probabilities in the same column are equal to each other, i.e.

$$H_0: p_{1\ell} = p_{2\ell} \quad \text{when } \ell = 1, \dots, c$$

$$H_1: p_{1\ell} \neq p_{2\ell}, \quad \text{for at least one } \ell'$$

The statistical test T is given by

$$T = \frac{1}{\frac{n_1}{n_1 + n_2} \cdot \frac{n_2}{n_1 + n_2}} \sum_{\ell=1}^c \frac{(O_{i\ell} - \frac{n_1 S_\ell}{n_1 + n_2})^2}{S_\ell} \quad (11)$$

when $n_1 = n_2$, T becomes

$$T = 4 \sum_{\ell=1}^c \frac{(O_{1\ell} - \frac{S_\ell}{2})^2}{S_\ell}$$

and if $\hat{p}_{1\ell} = \frac{O_{1\ell}}{n_1}$, $\hat{p}_{2\ell} = \frac{O_{2\ell}}{n_2}$

T can be written

$$T = n_1 \sum_{\ell=1}^c \frac{(\hat{p}_{1\ell} - \hat{p}_{2\ell})^2}{\hat{p}_{1\ell} + \hat{p}_{2\ell}} \quad (12)$$

The exact sample distribution of T when H_0 is true is complicated, but the asymptotic distribution ($n_1, n_2 \rightarrow \infty$) of T can be approximated by χ^2 with $c-1$ degrees of freedom.

c) Discrimination of companies

When one possesses data on companies and when those firms are grouped on a basis (e.g. regional) different than the one on which questions are asked, it is interesting to verify their belonging to a population. In this case two questions must be answered.

- Can the quantitative variables be used to identify or label a company to its set, and if so, what is the percentage of good classification?
- Can the number of quantitative variables be reduced while keeping a high probability of a satisfactory classification?

Discriminant analysis [5] appears to be the tool that can be used to answer those two questions by forming one or more linear combinations of the discriminating variables. These "discriminating functions" have the following form:

$$D_i = d_{i1} Z_1 + d_{i2} Z_2 + \dots + d_{ip} Z_p$$

where D_i : score of the discriminant function

d_{ip} : weighting coefficient

Z_p : standardized values of the p discriminant variables.

A company belongs to a group if its $D_i > C$ and to another group if its $D_i < C$ when C represents the decision factor. Therefore the problem is the determination of d_{ij} and C in such a manner that the probabilities of an unacceptable classification is minimized.

d) Results

Examples of the results that were obtained with the above mentioned techniques follow:

Example 4

Question: Are there any similarities between the 2 groups (Quebec and France) of enterprises?

Results: for $\alpha = 0,05$

H_0 is rejected if $t > 1,96$

where

α : level of significance

t: critical ratio

\bar{X}_1 : means of Quebec firms

\bar{X}_2 : means of France firms

Table 9 gives the results.

Comments: The means equality test on the technological scores has been used to evaluate the similarities between the two groups. T, or total score, has a t value that does not reject H_0 ; thus it indicates that there is no significant dissimilarity. The same has been obtained for the organizational processes. It must be underlined that H_0 is rejected for 4 vectors; the physical processes and the outputs are the components that highlight the strengths of the Quebec firms while the inputs and the human resources are the strengths of France firms.

TABLE 9
RESULTS OF VECTORS

VECTORS	\bar{X}_1	\bar{X}_2	t
INPUTS	43,44	46,99	-4,53
PHYSICAL PROCESSES	57,62	49,98	+2,13
ORGANIZATIONAL PROCESSES	47,66	50,05	-0,67
HUMAN RESOURCES	40,42	51,07	-5,08
OUTPUTS	56,57	51,42	+2,11
T	50,43	49,88	+0,22

Example 5

Question: Are there any similarities, on each variable, between the two groups (Quebec and France) of enterprises?

Results: for $\alpha = 0,05$

H_0 is rejected if $T > 0,19$

where

α : level of significance

T: statistical test

Table 10 gives part of all the results

Comments: The Chi-square test for differences in probabilities was applied to all the variables and the T's obtained reject H_0 for 24 of the 51 variables. Therefore, one can easily identify the variables that contribute to the differences noted in example 4.

TABLE 10
EXAMPLES OF DISTRIBUTION IN PERCENTAGE

	QUEBEC	FRANCE	T
1.1 Supplier's dependence			0,58
a)*	10%	61%	
b)	61%	24%	
c)	7%	3%	
d)	22%	12%	
1.2 Quality control on inputs			0,49
a)	22%	58%	
b)	41%	15%	
c)	37%	15%	
d)	0%	12%	
1.3 Average tolerances			1,19
a)	0%	0%	
b)	85%	8%	
c)	15%	92%	

* Levels of technology available for each variable within each X_i .

Example 6

Question: Do the furniture companies of Quebec and France belong to the same population?

Result: All the D_i 's obtained did permit to perfectly discriminate all the cases in their respective groups; namely, all the France firms in one group and all the Quebec firms in the other group.

Comments: The furniture enterprises in Quebec and in France do belong to two different populations. The 51 variables were utilized to perform the discrimination. However, if the variables are reduced to a few, the risk of misclassification is inferior to 2% [5].

CONCLUSION

The use of the technological model with the statistical tests did permit the study of a specific industrial sector.

The model was developed in six stages:

1. Five vectors were established.
2. These vectors were ranked according to importance.
3. Weights were accordingly assigned to the vectors.
4. Each vector was defined by a group of variables.
5. These variables were also assigned weights in order of importance.
6. The assigned weights at the vector level and at the variable level were then used to compute the technological potential of a given company.

This new management tool gives measurable elements to compare the manufacturing potential of different enterprises and to identify

those points on which an action must be undertaken. After a sectorial study, the model can be utilized as a diagnosis instrument for any firm that wants to identify its "technological level" and to direct its efforts towards the improvement of the management. In other terms, this tool is a model that can be used to measure the factors of technological performance and management in enterprises of a same industrial sector. The managers would then be in a better position to identify the strengths and weaknesses of their company.

The sample size is the main difficulty encountered with all the statistical tests that were utilized. Since the financial support was limited, only a portion of the existing firms were visited. However, the sample size (45 firms) represents more than 50% of the population.

BIBLIOGRAPHY

- [1] HOWARD, N., HARRIS, B., "A Hierarchical Grouping, IBM 360/65 Fortran IV Program", University of Pennsylvania Computer Center.

- [2] KENDALL, M.G., "Rank Correlation Methods", Hafner Publishing Company, New York, 1955.

- [3] LAINJU, B., "Analysis of Manufacturing Systems", Master Degree Thesis, École Polytechnique de Montréal, Montréal, 1980, 183 pages.

- [4] LEBLANC, D., "Un modèle d'évaluation technologique pour un système productif", Prospective socio-économique du Québec, Tome 5.2, L'Éditeur officiel du Québec, avril 1977, pp. 58-59.

- [5] RIOPEL, D., "Analyse comparative des facteurs de performance technologique de l'industrie québécoise et française du meuble de maison en bois", Master Degree Thesis, École Polytechnique de Montréal, Montréal, 1981, 158 pages.

- [6] RUMMEL, R.J., "Applied Factor Analysis", North-Western University Press, 1970, 617 pages.
- [7] VILLENEUVE, L., RIOPEL, D., "Modèle de mesure de la technologie: un nouvel outil de gestion", Gestion, revue internationale de gestion, Montréal, février 1979, pp. 75-81.

APPENDIX A

1.0 INPUTS

- 1.1 Supplier's dependence
- 1.2 Quality control on inputs
- 1.3 Average tolerances
- 1.4 Scientific and technical information
- 1.5 Loss of raw material
- 1.6 Fabrication under licence and patent
- 1.7 Subcontracting
- 1.8 Consultation

2.0 PHYSICAL PROCESSES

- 2.1 Degree of automation
- 2.2 Age of the equipment
- 2.3 Handling
- 2.4 Tolerances
- 2.5 Set-up times
- 2.6 Continuity of the process
- 2.7 Technological state perception
- 2.8 Process design

3.0 ORGANIZATIONAL PROCESSES

- 3.1 Standard times
- 3.2 Computer utilization
- 3.3 Production planning
- 3.4 Manufacturing cost
- 3.5 Ergonomic measures
- 3.6 Worker's security

- 3.7 Value analysis
- 3.8 Environment
- 3.9 Sales forecasts
- 3.10 Control
- 3.11 Quality control
- 3.12 Layout
- 3.13 Maintenance
- 3.14 Research and development
- 3.15 Operation research
- 3.16 Controls in manufacturing

4.0 HUMAN RESOURCES

Production staff

- 4.1 Education
- 4.2 Years of experience
- 4.3 Specialization
- 4.4 Turn-over

Administration staff

- 4.5 Education
- 4.6 Years of experience
- 4.7 Turn-over

Management staff

- 4.8 Training improvement
- 4.9 Description of tasks
- 4.10 Staff selection
- 4.11 Staff evaluation

5.0 OUTPUTS

- 5.1 Quality control
- 5.2 Product complexity
- 5.3 Value added
- 5.4 New products
- 5.5 Export
- 5.6 Product life
- 5.7 Innovation
- 5.8 Product transformation degree

CA2PQ
UP 5
R83-40
(Ang)
ex.2

647 130

~~04 JAN. 1985~~

ÉCOLE POLYTECHNIQUE DE MONTRÉAL



3 9334 00289224 6