Estimation of textile sales behaviour.

J. J. Denimal (1), W. Diallo (2), F. Boussu (3).

(1) Laboratoire de Statistiques et Probabilités, UFR de Mathématiques Pures et Appliquées
USTL (Université des Sciences et Technologies de Lille)
59655 Villeneuve d’Ascq Cedex
FRANCE (EUROPE).

(2) LASAD (Laboratoire de Statistique et d’Analyse de Données)
Faculté des Sciences de l’Université de Conakry
BP 1147
RÉPUBLIQUE DE GUINÉE (AFRIQUE).
Tel : (224) 464 689 - Fax : (224) 464 808.

(3) Laboratoire GEMTEX (GEnie et Matériaux TEXtiles)
ENSAT (Ecole Nationale Supérieure des Arts et Industries Textiles)
9 rue de l’Ermitage, BP 30329, 59056 Roubaix Cedex 01
FRANCE (EUROPE).

Abstract: - The permanent show of textile items, responding to the customer demand, needs to be prepared by a safe organisation between all the partners of the textile/garment/distribution channel. To cope with this complexity and supply the textile items, the provider can use its own forecasting sales to obtain an estimation of the future needs of raw material. We propose in this paper to complete our approach by using qualitative values in order to estimate a forecasted sales behaviour. Within this framework, different statistical methods will be defined and compared. We put on the fore the interesting results obtained for the textile/garment/distribution manager. Thanks to this method, the estimation of the main sales items behaviour can be known before the sale season and can help the industrial manager to plan its production and raw material needs.

Key-Words: - Data analysis, Discriminant analysis, Regression methods, Textile/Garment/Distribution Channel.

1 Introduction
The permanent show of textile items, responding to the customer demand, needs to be prepared by a safe organisation between all the partners of the textile/garment/distribution channel. The strategy used is to respond as quickly as possible to the consumer demand and change as often as possible the textile purchase offer. But, the production time of textile items, between all the partners of the Textile/Garment/Distribution channel, can spread on four weeks (for a Quick Response application [1]) to four months (for many other organisations). Then, the major difficulty of the order giver is to adjust the short term sales period with the medium and long term time production of the textile items. To solve the problem, the order giver tends to maintain a permanent stock of textile items in its show rooms as small as possible. And, a delivery strategy is engaged with its suppliers to frequently respond to the consumer demand. Thus, the textile items stocks and its problem of management also concern all the partners of the Textile/Garment channel. The other major difficulty of the supplier is to avoid the out of stock of its order giver and minimise the buffer stock size. Moreover, the number of textile references is as much high as the quantity sold are low. To cope with this complexity and supply the textile items, the provider can use its own forecasting sales to obtain an estimation of the future needs of raw material. Two approaches of the forecasting model identification have already been envisaged. The one lies on a first classification stage on the sales profiles of items, and secondly followed by a stage of forecasting models identification applied on the average sales behaviour of the obtained clusters. The other tends to
identify an unique and well fitted forecasting model to the data sales [2]. The first method allows the industrial manager to better understand its sales environment [3]. In addition, the clustering criterion used in this approach is based on the difference of sales values of quantitative type [4][5][6]. This approach tends to be faithful to identify a forecasting model but doesn’t integrate the influence of the items characteristics of qualitative type. Therefore, a forecasting model previously identified can’t be applied to items described by new characteristics.

We propose in this paper to complete the first described approach by using these qualitative values in order to estimate a forecasted sales behaviour.

The first part of our paper presents, step by step, the new proposed analysis method and the principal obtained results. In the second part, all the different steps are described in details and validated at each stage. To conclude, we put on the fore the interesting results obtained for the textile/garment/distribution manager. Thanks to this method, the estimation of the main sales items behaviour can be known before the sales season and can help the industrial manager to plan its production and raw material needs.

2 Problem Formulation

After the clustering of textile items based on their quantitative sales behaviour, a discriminant analysis on the table combining all the items and their qualitative characteristics (colour, kind, tissue type, model, etc.) is engaged. The discriminant linear functions then found are used as explanatory variables within the different kinds of regression methods (classical linear regression [7][8] and nearest neighbour regression [9]). These regression methods allow to explain the different textile sales behaviours.

2.1 Stepwise Description of our approach

Step 1. Clustering of textile items based on their quantitative sales behaviour. Identification with our data sales of a four clusters partition.

Step 2. Random distribution of items (respecting a suitable number of elements) in two sample groups (basis and test).

Step 3. Application of the principal components analysis [10] to the binary table, due to the coding of qualitative description, combining the items and their characteristics. The items of the test group are considered as additional individuals. By using our data sales, a reduction of the 83 qualitative variables to the first 27 factorial axis is obtained (which also corresponds to 96.31% of the conserved total inertia).

Step 4. Selection of the most discriminant axis resulting from the stepwise linear discriminant analysis [11][12] applied on the co-ordinates board combining all the items groups (basis and test) and the 27 factorial axis. Identification of the 7 more discriminant axis corresponding to a maximum percentage of “well clustered” items of the test group (72% with our data application). The percentage of “well clustered” items is calculated in using the discriminant rules to the items clusters of the test group built thanks to the discriminant analysis results applied on the basis group.

Step 5. Identification of 3 linear discriminant functions resulting from the discriminant analysis applied to the table which is reduced to the 7 previous factorial axis (the discriminant values obtained with our data are all upper to 0.81).

Step 6. At each sale period, a multiple linear regression [7], explaining the sold quantity of the basis group items from the 3 linear discriminant functions, is applied. Estimation of the sold quantities of the test group items and comparison with the real sold quantities.

Step 7. Back to the step 6 by applying to each sale period the nearest neighbour regression [9]. The number of neighbour can be chosen with respect to the average lower value of the chi-square distances between the estimated and real quantities of the test group items.

Step 8. Comparison of the obtained estimation for the two regression methods thanks to the calculated value of chi-square distances between the estimated and real quantities for each item of the test group, and validation at the 5% level.

3 Problem Solution

A description of the different value of the qualitative characteristics, and their numbers, is given into the Table 1.

<table>
<thead>
<tr>
<th>Qualitative characteristics</th>
<th>Example</th>
<th>Number of elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kind</td>
<td>Swim or Gym</td>
<td>2</td>
</tr>
<tr>
<td>Type</td>
<td>Child, woman or man</td>
<td>3</td>
</tr>
<tr>
<td>Design</td>
<td>Swimsuit, body, etc.</td>
<td>11</td>
</tr>
<tr>
<td>Tissue number</td>
<td>1345, 1226, etc.</td>
<td>16</td>
</tr>
<tr>
<td>Model number</td>
<td>2014, 2045, etc.</td>
<td>23</td>
</tr>
<tr>
<td>Colour</td>
<td>Black, White, Pink, Yellow, etc.</td>
<td>28</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>83</td>
</tr>
</tbody>
</table>

Table 1. Description of Textile items characteristics.
From the quantitative variables of our 58 textile items, a clustering method is applied (step 1) and the number of clusters is found into the resulting partition thanks to the first minimum value of a validity criterion [13]. The four clusters of data points, resulting from the obtained partition, are represented into (Fig. 1) the main plan of the principal components analysis (with a representation rate of 66.83% of the total inertia).

![Fig. 1. Representation of the four clusters of items (step 1).](image1)

The representation of the average sales behaviour of each clusters, on 52 weeks, is given into Fig. 2.

![Fig. 2. Representation of the four clusters average sales (step 1).](image2)

The random distribution of items leads to achievement of the basis and test groups (step 2). The number of elements contained in each clusters of each groups is given in the following table.

<table>
<thead>
<tr>
<th></th>
<th>Basis sample</th>
<th>Test sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>9</td>
<td>5</td>
</tr>
</tbody>
</table>

| **Total** | **36** | **22** |

Table 2. Distribution of items in the basis and test groups (step 2).

From the binary table of the 83 qualitative variables, combining the items and their characteristics, a principal components analysis is done (step 3) and leads to reduce the variables to the first 27 factorial axis (96.317% of the total inertia is conserved). After the 27th factorial axis, all the eigenvalues are under 0.1; which means that no more additional information is contained in the next factorial axis.

![Fig. 3. Choice of the factorial axis number (step 3).](image3)

The stepwise linear discriminant analysis (step 4) is applied to the co-ordinates table combining all the items groups (basis and test) and the previous 27 factorial axis. It allows us to extract 7 factorial axis whose the “well clustered” items ratio raises its maximum value of 72.73%. The discriminant rule, built by the discriminant analysis applied on the basis group, allows to distribute the 22 items of the test group in the following table cases:

<table>
<thead>
<tr>
<th>Corresponding group</th>
<th>Original group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C1</td>
</tr>
<tr>
<td>C1</td>
<td>5</td>
</tr>
<tr>
<td>C2</td>
<td>2</td>
</tr>
<tr>
<td>C3</td>
<td>0</td>
</tr>
<tr>
<td>C4</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3. Classification board of the test group (step 4).
The linear discriminant analysis, achieved to the 7 previous factorial axis, gives 3 (inferior to the number of clusters) linear discriminant functions (step 5) whose values are:

- Discriminant value of the first factor: 0.99519
- Discriminant value of the second factor: 0.86440
- Discriminant value of the third factor: 0.81689

The discriminant value of each factor represents a dividing measure between the four clusters. The closer of 1 the discriminant value is, the better separated the clusters are. The graphical representations, of the clusters of items contained into the basis and test groups, in the different discriminant plans allow to appreciate the clusters division.

![Fig. 4](image.png)  
**Fig. 4.** Representation of the four clusters of the basis group with respect to the first and second discriminant axis (step 5).

![Fig. 5](image.png)  
**Fig. 5.** Representation of the four clusters of the basis group with respect to the second and third discriminant axis (step 5).

At each sale period, two type of regressions allow to explain the vector of sold quantities of the basis group items from the 3 linear discriminant functions assumed to be qualitative variables. The obtained model forecasts the sold quantities of the items of the test group which are also compared to the real sold quantities (step 6 and 7). The number of neighbour can be found with respect to the minimum average chi-square value of the nearest neighbour regression method given in Table 4.

![Fig. 6](image.png)  
**Fig. 6.** Representation of the four clusters of the test group with respect to the first and second discriminant axis (step 5).

![Fig. 7](image.png)  
**Fig. 7.** Representation of the four clusters of the test group with respect to the second and third discriminant axis (step 5).
So, 4 neighbours are needed to the nearest neighbour regression method.

Remark: the average chi-square distances calculated for 2, 3 and 4 neighbours are quite the same.

The two regression methods are compared thanks to the chi-square distances between the forecasted and real quantities of each items of the test group. For each item, these previous distances allow to validate the regression method to the 5% level of the chi-square distribution. The minimum value obtained for the same item helps to identify the best of the two regression methods (step 8).

Fig. 8. Validation and comparison of the two regression methods (step 8).

On one hand, the classical linear regression applied to some items gives better results, as the item number 41 of the cluster 4 of the test group (Fig. 9), than the other method. On the other hand, the inverse phenomenon is observed with some other items, as the item number 47 of the cluster 2 of the test group (Fig. 10).

Fig. 9. Representation of the item number 41 with the two regression methods.

Fig. 10. Representation of the item number 47 with the two regression methods.

As regards the obtained results (Fig. 8), the choice of the well fitted regression methods is difficult. Moreover, the two regression methods provide suitable estimations of the sale profiles of many textile items.

4 Conclusion

Our paper allows us to validate the assumption of the influence of qualitative characteristics on sale profiles. Thus, the different sales behaviour can be explained and estimated thanks to a priori known qualitative characteristics. The method yet developed will help the industrial manager to estimate the sales behaviour of new items. However, the only use of these variables seems to be no sufficient to estimate the sale profiles of some items. On one hand, the introduction, in our model, of other exogenous parameters as the climatic, seasonal or others index; to improve the sales behaviour comprehension, will
be envisaged in soon coming study. On the other hand, some improvements on the method are also engaged. As the use of other regression types, in particular the PLS regression, to avoid the different steps like the classification (step 1) and the discriminant analysis (step 4 and 5). Finally, our future works will take into account the auto-correlation between two successive weeks in our model.

References: