015-0596

FORECASTING PRACTICES IN BRAZILIAN FOOD INDUSTRIES

CRISTIANO CECATTO

FEI University Center

Av. Humberto de Alencar Castelo Branco, 3972, Assunção, São Bernardo do Campo - SP - Brasil - Zip code: 09850-901

e-mail: cecatto@qualilog.com

PATRÍCIA BELFIORE

FEI University Center

Av. Humberto de Alencar Castelo Branco, 3972, Assunção, São Bernardo do Campo - SP, Brasil - Zip code: 09850-901

e-mail: prspbelfiore@fei.edu.br

JOÃO CHANG JUNIOR

FEI University Center

Av. Humberto de Alencar Castelo Branco, 3972, Assunção, São Bernardo do Campo - SP, Brasil - Zip code: 09850-901

e-mail: chang.joao@gmail.com

POMS 21st Annual Conference

Vancouver, Canada

May 7 to May 10, 2010

ABSTRACT

This paper studies the forecasting practices that have been used by food industries in Brazil. Based on literature review, a questionnaire was developed and sent to a sample of 450 food companies from ABIA (Brazilian Association of Food Industries) which represents 70% of the universe, and a response rate of 14.4% (65/450) was achieved. The objective is to detect how these companies have been used forecasting methods, what are the main factors that influence their choice and what are the main difficulties in the use of forecasting methods. Data were analyzed by multivariate statistics techniques as discriminant analysis using the SPSS software. The results show that the historical analysis model is the most used. The factors that influence the choice of the models are the type of product and the time spent during the forecasting, and the main difficulties are the availability of software and difficulty in understanding.

1. INTRODUCTION

The interest in forecasting methods in Brazilian food industries have increased in the last years. Companies with advanced forecasting practices have emerged in competitive markets. The limitations of many Brazilian industries about these practices and the adoption for innovations contributed to the development of this work.

Birchall (2004) emphasizes the importance of the food industry. According to the author, the sector was responsible for the first Brazilian industrial impulse in the late 19th century, and at the end of World War I was the second largest activity in Brazil. Nowadays, the sector has a significant role in economic growth, from the increasing employment and exports rates to its significant participation in the Gross Domestic Product (GDP).

According to the Brazilian Association of Food Industries (ABIA), entity responsible by the sector, the food industry represents 8.0% of GDP, with a net profit of US$128 billion and about 1.32 billion of employees in 2007. Besides, the food industry was responsible for 16.6% of total exports in 2007, with revenues of US$ 26.6 billion. On the other hand, Conceição (2007) observe that the rate of innovation practices in this sector is very low and many actions should be adopted to improve its performance.

In this sense, investments in demand planning could bring considerable gains for organizations. The choice of an appropriate forecasting model enables companies to plan their production, inventory and purchasing, reduce waste, supply the costumer demand and define strategic and complex issues, becoming a critical success factor for managing business.

Jain and Malehorn (2006) cite that there isn't a magic and unique forecasting model that could be used in all situations and conclude that the forecasting practice is a mixture of science and art. This paper aims to analyze forecasting practices that have been adopted in the Brazilian food industry.

The organization of this paper is detailed below. Section 2 presents the objectives, research questions and hypotheses adopted, defined from a detailed literature review. The research methodology is discussed in section 3, focusing on the composition of the sample, data collection and description of the statistical methods used. The result analysis is presented in the next section. Finally, in section 5 are the final considerations.

2. Objectives e hypotheses

The main objective of this study is to verify the current state of forecasting practices in Brazilian food industries.

Information about the food sector, using statistical and economic data related to the importance of these industries to the country, were collected from the ABIA which represents 70% of the sector in Brazil. From the 2008 ABIA Yearbook, we selected a total of 450 companies that belong to the sector and a questionnaire was developed and sent to these companies. Data were analyzed using descriptive statistics and multivariate statistical techniques. The paper attempts to answer some questions and research hypotheses defined from a detailed literature review about forecasting practices.

Jain and Malehorn (2006) conducted a survey in U.S. industries from different sectors, to provide which models have been used by these companies. Results showed that the exponential smoothing model is the most used. We check if this model is also the most used by the Brazilian food industries through question 1.

Question 1: Do food companies that compose the database use some forecasting method? If so, what models are being used?

Hypothesis 1: Food companies that compose the database use one of the forecasting models listed in the questionnaire (Simulation, Market research, Delphi, Expert panel, Historical analysis, Moving average, Exponential smoothing, ARIMA, Regression, Econometric, Neural Network).

Hypothesis 2: Most of the companies that compose the database don’t use sophisticated models.

Hypothesis 3: The most used model by food industries in Brazil is the exponential smoothing.

Hypothesis 4: There isn’t consensus between Brazilian food industries in the use of a single forecasting model.

Question 2: How do forecasting models behave in terms of accuracy, time horizon, company size and type of product?

Differently of the most studies in the literature, the “**type of product**” variable was also added to the analysis to evaluate which sectors would require new development and researches.

Hypothesis 5: The type of product influences directly the forecasting models used.

Hypothesis 6: Sophisticated models don’t guarantee, necessarily, better accuracy.

From different works as Winklhofer, Diamantopoulos and Witt (1996), Armstrong and Fildes (2006), Küsters, McCullough and Bell (2006), Wilson and Daubek (1989), Fildes and Hastings (1994) and Naylor (1981), we defined the main variables that could influence the choice of forecasting models. They are: accuracy, ease of understanding, usability, time horizon, cost, time spent, data consistency and availability of software for forecasting. From these variables we define question 3 and its hypotheses.

Question 3: What are the main factors that influence the choice of the model?

Hypothesis 7: Accuracy, ease of understanding, usability, and cost are the main factors that influence the choice of the forecasting model in Brazilian food industries.

From papers as Winklhofer et al. (1996) and Fildes and Hastings (1994), we pointed the main difficulties in the use of forecasting models (difficulty of understanding, difficulty of using, high cost, high time spent, difficulty of data, availability of software, precision), coming to question 4.

Question 4: What are the main difficulties found in the use of forecasting models?

Hypothesis 8: Difficulty of understanding, usability and high cost are the main factors that influence the choice of the model by the companies studied.

Similar to question 2 and hypothesis 5, in addition to the variables found in the literature that support questions 2 and 3, we added the variable **type of product** that can sign deficient areas which require further research.

3 . RESEARCH METHODOLOGY

3.1 Sample composition

From the list of companies which are members of ABIA, we selected all the 450 companies responsible for food production. Contacts were made by telephone and e-mail, from July to November 2008, and the response rate was 14,44%.

According to ABIA, companies vary in size, origin and location, representing different sectors of food production, including: meat products; processing of coffee, tea and cereals; oils and fats; dairy products; wheat products; sugar; fruits and vegetables; chocolate, cocoa and candy; dried and frozen; and canned fish.

3.2 Data collection

The questionnaire attached was applied to a sample of 65 companies and the data were collected and answered by e-mail, by the professional responsible for forecasting.

3.2.1 Questionnaire structure

The questionnaire is divided into two parts. The first part aims to identify the characteristics of the companies, while the second aims to characterize the forecasting process.

The main characteristics of the companies collected in the first part of the questionnaire were: location, size, type of product and origin (national or multinational).

The main characteristics to be identified in the forecasting process by the companies are: sector responsible, model used (if yes), accuracy of the model, time horizon, forecasting interval, factors that influence the choice of the model, and difficulties in the use of these models.

The accuracy of the forecasting model corresponds to the percentage of correctness of the predicted demand to its actual value. According to Armstrong (2001), the accuracy is a relevant criterion to identify the maturity stage of the forecasting process. The time horizon is defined by Pellegrini (2000) as the number of future periods at which the forecasting process will be evaluated. According to the same author, the forecasting interval corresponds to the frequency or period of time at which new forecasts are calculated (weekly, monthly, quarterly). Using these measures, it is possible to analyze if new methodologies that have been implemented in the forecasting systems satisfy the expected goal.

3.3 Multivariate data analysis

From the questionnaires answered by the companies, we aim to answer the research questions and hypotheses specified at section 2, through multivariate statistical methods as correspondence and discriminant analysis solved by SPSS software (*Statistic Package for Social Study)*. A brief description of each method will be described below. More details can be found at Fávero et al. (2009).

3.3.1. Correspondence analysis

Different from dependent techniques designed to identify the relations between variables, correspondence analysis (CA) is classified as an exploratory/descriptive data technique designed to analyze associations between the elements of two sets. Correspondence analysis can also be defined as a statistical visualization method for displaying the associations between the levels (rows and columns) of a two-way contingency table. CA is also a generalization of principal correspondence analysis (PCA) at which the variables are categorical instead of quantitative.

Multiple correspondence analysis (MCA) is an extension of correspondence analysis (CA) and designed to study associations between the levels of more than two sets in a multi-way contingency table.

Associations between variables are calculated by the chi-square test. Thus, CA and MCA allow researches to visualize these associations between different categories of nominal variables in a two or multi-dimensional graph, respectively.

A more detailed description about the method may be found in Fávero et al. (2009).

3.3.2. Discriminant Analysis

According to Fávero et al. (2009), discriminant analysis (DA) is a multivariate and dependence technique used when the dependent variable is qualitative, categorical or non-metric and independent variables are quantitative or metric.

The idea of the discriminant analysis is to determine whether groups differ from the mean of a variable, and then predict in which groups new cases will belong.

Maroco (2003) presents the main objectives of DA:

a) identify which variables discriminate two or more groups;

b) use these variables to create a discriminant function that represents the differences between groups.

c) use this discriminant function to classify new cases in the groups.

Hair et al. (1998) list the hypotheses that must be checked for application of the DA: multivariate normality of independent variables, homogeneity of the matrices in the classification process, absence of multicollinearity and no linearity of the variables.

More details about the technique can be found in Fávero et al. (2009).

4. RESULTS ANALYSIS

The results obtained from the questionnaires answered by the companies are presented in this section.

4.1 Characterization of the companies

Considering the companies surveyed, 67,5% were small, 17,5% medium and 15% large. Most of them (89,3%) are national and the rest (10,7%) are multinational. Considering the location, the Southeast region is more representative (73,85%), followed by the Southern region (8,46%), the Northeast region with 6,15% and finally the Midwest with 1,54%.

The main production of the participating companies is shown in order of representativeness: oils and fat (25,5%), dairy products (15,48%), canned fish (12,68%), meat products (11,27%), processing of coffee, tea and cereals (8,45%), fruits and vegetables (7,04%), sugar (2,82%), dried and frozen (2,82%), wheat products (1,41%), chocolate, cocoa and candy (1,41%) and others (11,27%).

We can conclude that most of the companies studied are located in Southeast region, are national and produce oils and fat.

4.2 Characterization of the forecasting process

The statistical methods proposed will be applied to analyze the data collected from the questionnaires in order to answer the research questions and test the hypotheses described in section 2.

4.2.1 Descriptive statistics to analyze the forecasting models that have been used by the companies

Through descriptive statistics, we will answer the first research question and confirm or not the hypotheses 1, 2, 3 and 4.

The results of descriptive statistics comparing the forecasting models used by the companies are presented below. The frequency distribution is illustrated in table 1.

Table 1 - Frequency distribution

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Simulation	1	1.5	1.5	1.5
	Market_research	13	20.0	20.0	21.5
	Delphi	1	1.5	1.5	23.1
	Historical_analysis	35	53.8	53.8	76.9
	Moving_average	13	20.0	20.0	96.9
	Exponential_smoothing	1	1.5	1.5	98.5
	Arima	1	1.5	1.5	100.0
	Total	65	100.0	100.0	

Table generated by SPSS program

Table 1 shows that **historical analysis** model is the most used (35 companies), followed by **market research** (13 companies) and **moving average** model (13 companies).

Thus, we conclude that **hypothesis 1** is true, i.e., the companies surveyed use various forecasting models among those listed in the questionnaire.

From the results presented, we also conclude that **hypothesis 2** is true, i.e., most of the companies studied do not use sophisticated forecasting models, that is, 93.8% of the companies use **historical analysis**, **moving average** and **market research**. You can also conclude that, differently from the study done by Jain and Malehorn (2006), the **exponential smoothing** is not the most used, rejecting **hypothesis 3**.

The results also show that there is no a single forecasting model for using by all the companies, confirming that **hypothesis 4** elaborated from Jain and Malehorn (2006) and Zhou (1999) is true.

4.2.2 Correspondence analysis and discriminant analysis to answer the second research question and its respective hypotheses

In this section, firstly, we apply the correspondence analysis technique to analyze associations between the methods used by the companies and their sector of actuation (type of product). The variables **forecasting methods used** and **type of product** are categorical, which justifies the application of the technique.

Discriminant analysis (DA) will be used to evaluate the influence of the variables **precision**, **time horizon** and **company size** in the choice of the **forecasting methods used**. As the independent variables are quantitative and the dependent variable is qualitative (group of methods), DA can be applied to determine which independent variables discriminate the groups. Thus, from the initial data of a particular company about accuracy, time horizon and company size, you can determine the most appropriate forecasting model.

Table 1 shows the name of the variables studied, the name adopted in the routine, the scale and type of variable.

Table 1 – Characteristics of the Variables

Variables	Name of the variable	Scale adopted	Type of variable
Model	Model	0=not use 1=simulation 2=market research 3=Delphi 4=expert panel 5=historical analysis 6=moving average 7=exponential smoothing 8=arima 9=decomposition 10=regression 11=econometric 12=neural networks	Nominal
Accuracy	Accuracy	Numerical	Metric
Time horizon	Time horizon	Numerical	Metric
Type of product	Type of product	1=meat products 2=coffee, tea and cereals 3=fruits and vegetables 4=wheat products 5=dried and frozen 6=oils and fat 7=dairy products 8=sugar 9=canned fish 10=chocolate, cocoa and candy	Nominal
Number of employees	Number of employees	Numerical	Metric

The main results of the correspondence analysis to identify associations between **forecasting models** and **type of product** of the company are listed below. We used the guide proposed by Favero et al (2009) to show the results.

Firstly, to check the independence between groups, or association between the categories of the variables **type of product** and **forecasting model**, we used the chi-square test. The null hypothesis of the test states that there is independence between groups.

Table 2 - Results of the Chi-Square tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	45.962(a)	54	.774
Likelihood Ratio	38.587	54	.944
Linear-by-Linear Association	1.004	1	.316
N of Valid Cases	65		

a 68 cells (97.1%) have expected count less than 5. The minimum expected count is .02.

(Table generated by SPSS program)

Table 2 shows a p-value of 0.774 for the chi-square test, value major than 0.05, indicating no rejection of the null hypothesis of independence between groups, considering a significance level of 5% (Bussab and Morettin, 2006). The results show evidence that there is independence between variables, i.e. there is no association between the variables **type of product** and **forecasting models**, which do not allow the application of the correspondence analysis technique.

So, hypothesis 5 that states that there is association between the variables **type of product** and **forecasting models** is rejected

To analyze the influence of some variables as **accuracy**, **time horizon** and **company size** in the choice of **forecasting models**, we used discriminant analysis (DA), as described above. Initially, all the variables in the model were considered. However, the variable **company size** was not significant to the model, and its permanence would have damage the results of the analysis. Therefore, another DA model was constructed considering as predictor (independent) variables only the **accuracy** and **time horizon**, and as response variable (dependent) the **forecasting model**. The main results are listed below, from the guide proposed by Favero et al. (2009).

The ANOVA test (Analysis of Variance) presents the Wilks' Lambda test that determines the differences between groups' means for each variable analyzed. The test indicates equality between groups when its values is close to 1; on the other hand, values close to zero or lower indicate differences between groups. It was found from table 3 that the values of the variables studied are high, noting differences between groups. ANOVA also analyzes the significance of the variables by F test, based on the null hypothesis that the variable is not significant for discriminating groups or the groups' means are equal (Bussab and Morettin, 2006). Note that the null hypothesis was not rejected for the variable **accuracy**, considering the significance level of 5%, and was rejected for the variable **time horizon** (0.047 <0.05), concluding that the last variable discriminate groups.

Table 3 – Test of Equality of Group Means (ANOVA)

	Wilks' Lambda	F	df1	df2	Sig.
Horizonte_tempo	.816	2.186	6	58	.047
Precisao	.847	1.748	6	58	.126

(Table generated by SPSS program)

Through the covariance matrices and correlation for each group, we verified that there was no correlation between the variables (values close to zero). If there are problems of multicollinearity (which would indicate that two or more variables are directly or inversely correlated, i.e., with correlation values near 1 or -1, respectively), it is recommended the application of factor analysis or removal of one of variables (FÁVERO et al., 2009).

Through the covariance matrix, we can identify the presence of homogeneity of covariance, one of the assumptions of discriminant analysis. However, we used Box's M test for this purpose. The null hypothesis of Box's M test states that there are no significant differences between groups or that the covariance matrices of the groups are homogeneous (FÁVERO et al., 2009). For a significance level of 5%, the null hypothesis is not rejected ($0{,}143 > 0{,}05$), indicating that there is equality in dispersion of groups.

Table 4 presents the eigenvalues for the discriminant functions.

Table 4 – Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	.268(a)	65.1	65.1	.460
2	.144(a)	34.9	100.0	.355

a First 2 canonical discriminant functions were used in the analysis.

(Table generated by SPSS software)

According to Gageiro and Pestana (2000) and Favero et al. (2009), eigenvalues represents the percentage of variance explained between groups. Eigenvalues near 1 represents little variation between groups explained by the discriminant function, and eigenvalues far from 1 represents large variations between groups explained by the discriminant function (FÁVERO et al., 2009). From Table 4, we can verify that function 1 explained 65.1% of the variance between groups while function 2 explains only 34.9%, concluding that function 1 is most representative for discriminating groups.

Table 5 presents the results of Wilks' Lambda and Chi-Square test.

Table 5 – Wilks' Lambda and Chi-square

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.689	22.155	12	.036
2	.874	8.007	5	.156

(Table generated by SPSS software)

In the first row of Table 5 we test the significance of all discriminant functions. The second line tests the second function in isolation. Considering the first line, the p-value is 0.036, a value less than 0.05, rejecting the null hypothesis that states that the groups' means are equal. Therefore, we can affirm that at least the first discriminant function is significant, which does not happen with the second function, since the p-value is 0.156. The Wilks' Lambda also reports that the second function has an insignificant discriminant function compared to first one, since its value is closer to 1.

Table 6 presents the standardized coefficients of discriminant functions that evaluate the importance of each variable to the discriminant function, and can be also called coefficients of discriminant weights (Maroco, 2003).

Table 6 – Standardized coefficients of discriminant functions

	Function	
	1	2
Time_horizon	.840	-.545
Accuracy	.584	.813

From table 6 we verify that the variable **time horizon** has a high positive coefficient for the first discriminant function, while the variable **accuracy** has a high positive coefficient for the second discriminant function.

Table 7 shows the classification coefficients that allow you to create models to classify new observations or new companies.

Table 7 – Classification coefficients of discriminant functions

Classification Function Coefficients

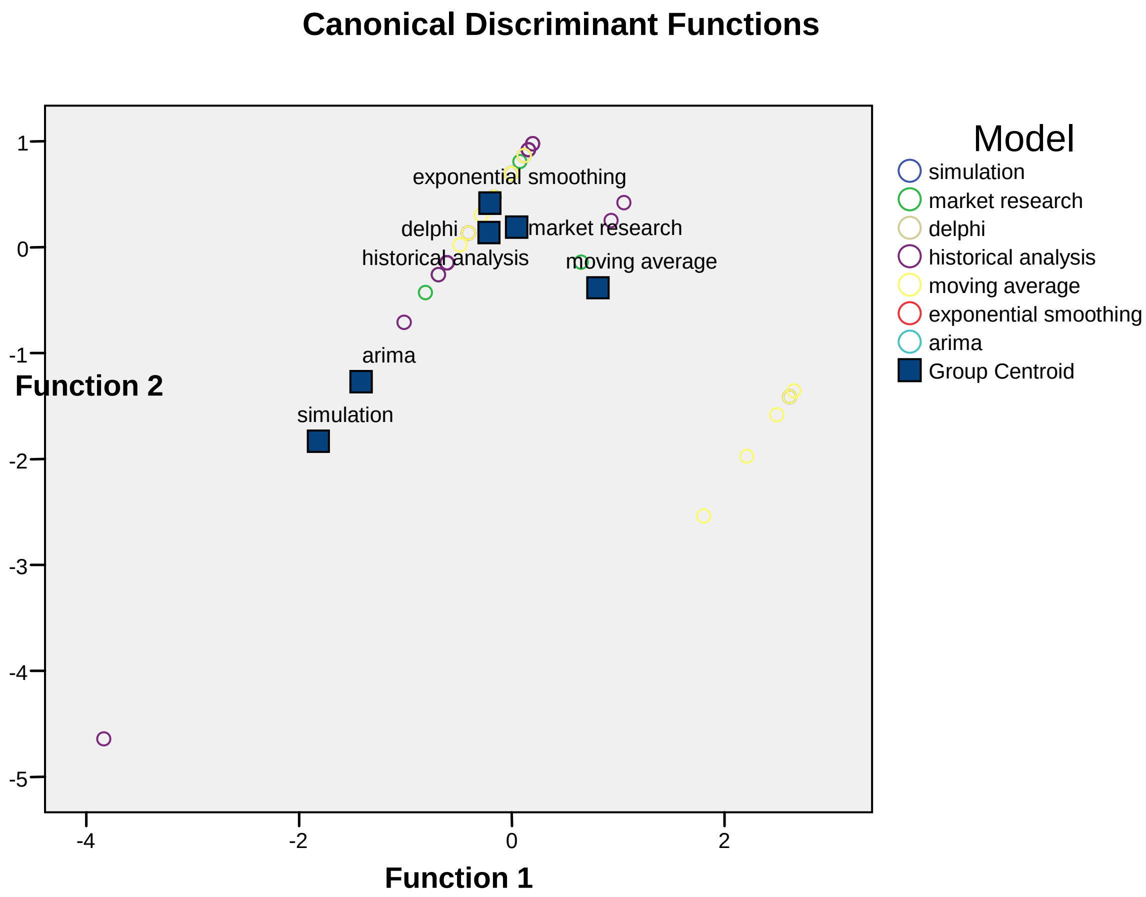
	Model						
	simulation	Market research	delphi	Historical analysis	Moving avarage	Exponential smoothing	arima
Time horizon	36.388	50.669	40.406	44.799	79.874	40.406	37.393
Accuracy	24.250	43.124	43.381	41.792	42.995	43.381	29.033
(Constant)	-10.843	-21.972	-24.369	-19.568	-23.607	-24.369	-13.507

Fisher's linear discriminant functions

(Table generated by SPSS software)

Figure 1 represents graphically the centroids of each group in the discriminant functions.

Figure 1 – Graph of centroids in the discriminant functions



(Figure generated by SPSS software)

This graph includes all possible models, some of them with few observations (**simulation**, **exponential smoothing**, **ARIMA** and **Delphi**), and may generate incorrect classifications.

In summary, it was found that **time horizon** is the variable that most influences the forecasting models used, followed by the variable **accuracy** that is not significant for a significance level of 5%. Note also that the **moving average** model has the most positive discriminant coefficient for the variable **time horizon**, while **simulation** and **ARIMA** models are those with more negative discriminant coefficient. On the other hand, **ARIMA** and **simulation** models have the largest negative discriminant coefficient for the variable **accuracy**, while **exponential smoothing**, **Delphi** and **market research** have the largest positive discriminant coefficient. Therefore, we can conclude that companies using the **moving average** model consider a larger **time horizon** compared to other companies, while companies using **simulation** and **ARIMA** models are considering a lower horizon. We can also conclude that companies that use **ARIMA** and **simulation** models are those with **less accuracy**, while companies that use **exponential smoothing**, **Delphi** and **market research** models are the ones with **better accuracy**.

Thus, we can conclude that the use of sophisticated models do not necessarily guarantee better accuracy, confirming **hypothesis 6**.

4.2.3 Discriminant analysis to evaluate the main factors that influence the choice of the model

Discriminant analysis (DA) will be used to evaluate the influence of several variables in the choice of forecasting models. The independent variables are quantitative (score from 0 to 10 in order of importance) and the dependent variable is qualitative (groups of methods used). Thus, DA can be applied to verify the influence of each variable in the discrimination of groups.

The model considers as dependent variable the **forecasting model** and the predictors or independent variables in the analysis are: **time horizon**, **ease of understanding**, **usability**, **cost**, **time spent**, **accuracy**, **data consistency**, **availability of statistical software** and **type of product**. Table 2 shows all variables studied, the name and scale adopted, and its type.

Table 2 – Name of variables used for question 3.

Variables	Name of the variable	Scale adopted	Type of variable
Model	Model	0=not use 1=simulation 2=market research 3=Delphi 4=expert panel 5=historical analysis 6=moving average 7=exponential smoothing 8=arima 9=decomposition 10=regression 11=econometric 12=neural networks	Nominal
Time horizon	Time horizon	Numerical/integer(0 a 10)	Metric
Ease of understanding	Ease understanding	Numerical/integer(0 a 10)	Metric
Usability	Usability	Numerical/integer(0 a 10)	Metric
Cost	Cost	Numerical/integer(0 a 10)	Metric
Time spent	Time spent	Numerical/integer(0 a 10)	Metric
Accuracy	Accuracy	Numerical/integer(0 a 10)	Metric
Data consistency	Data consistency	Numerical/integer(0 a 10)	Metric
Availability of statistical software	Software availability	Numerical/integer(0 a 10)	Metric
Type of product	Type of product	Numerical/integer(0 a 10)	Metric

However, when all variables listed on table 2 are considered in the process, we identified that the variables **time horizon**, **ease of understanding**, **ease of use**, **cost**, **accuracy**, **data consistency** and **availability of statistical software** were not significant for the statistical model, affecting the results of the analysis. From this analysis, another DA model was created considering as predictor or independent variables **time spent** and **type of product**. The main results are presented below, based on the guide proposed by Favero et al. (2009).

The test of equality of group means is listed on table 8.

Table 8 – Tests of Equality of Group Means (ANOVA)

	Wilks' Lambda	F	df1	df2	Sig.
Time_spent	,787	1,889	8	56	,080
Product_type	,693	3,096	8	56	,006

The Wilks' Lambda test indicates that the variable **time spent** presents the highest value (0.787), indicating a smaller discriminating power compared with the variable **type of product**. The F test indicates that the variable **time spent** is not significant for the statistical model. Moreover, the variable **type of product** indicates its discriminatory power ($0{,}006 < 0{,}05$). Therefore, we can assume that for the predictor variable **type of product** there is at least one group in which the means are different.

Through the covariance matrices and correlation for each group, we verify that there was no correlation between the variables (values close to zero).

For the Box's M test that evaluates the homogeneity of covariance matrices between groups, for a significance level of 5%, the p-value is 0.322, indicating no rejection of the null hypothesis, i.e. there is equality in dispersion of groups.

Table 9 presents the eigenvalues for the discriminant functions.

Tabela 9 – Summary of Canonical Discriminant Functions - Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	,539(a)	72,4	72,4	,592
2	,206(a)	27,6	100,0	,413

a First 2 canonical discriminant functions were used in the analysis.

(Table generated by SPSS program)

We can observe that the function 1 explains 72,4% of the variance between groups, while the function 2 explains only 27,6%.

Table 10 presents the results of Wilks' Lambda and Chi-square test.

Table 10 – Wilks' Lambda and Chi-square test

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	,539	36,197	16	,003
2	,829	10,957	7	,140

(Table generated by SPSS program)

In the first row of Table 10, the p-value is 0.03, value lower than the significance level of 5%, rejecting the null hypothesis that the group means are equal in this function. The same doesn't occur with the second discriminant function. Therefore, we conclude that only the first discriminant function is highly significant. The Wilks' Lambda also reports that the second function has a larger discriminating power compared to the first function, as its value approaches 1.

Table 11 presents the standardized canonical discriminant function coefficients.

Table 11 – Standardized Canonical Discriminant Function Coefficients

	Function	
	1	2
Tempo_despendido	,544	,848
Tipo_produto	,905	-,441

From Table 11, we can verify that the variable **type of product** has a larger positive coefficient for the first discriminant function, while the variable **time spent** has a larger positive coefficient for the second function.

Table 12 presents the classification function coefficients that allow you to create models to classify new observations or companies for the study.

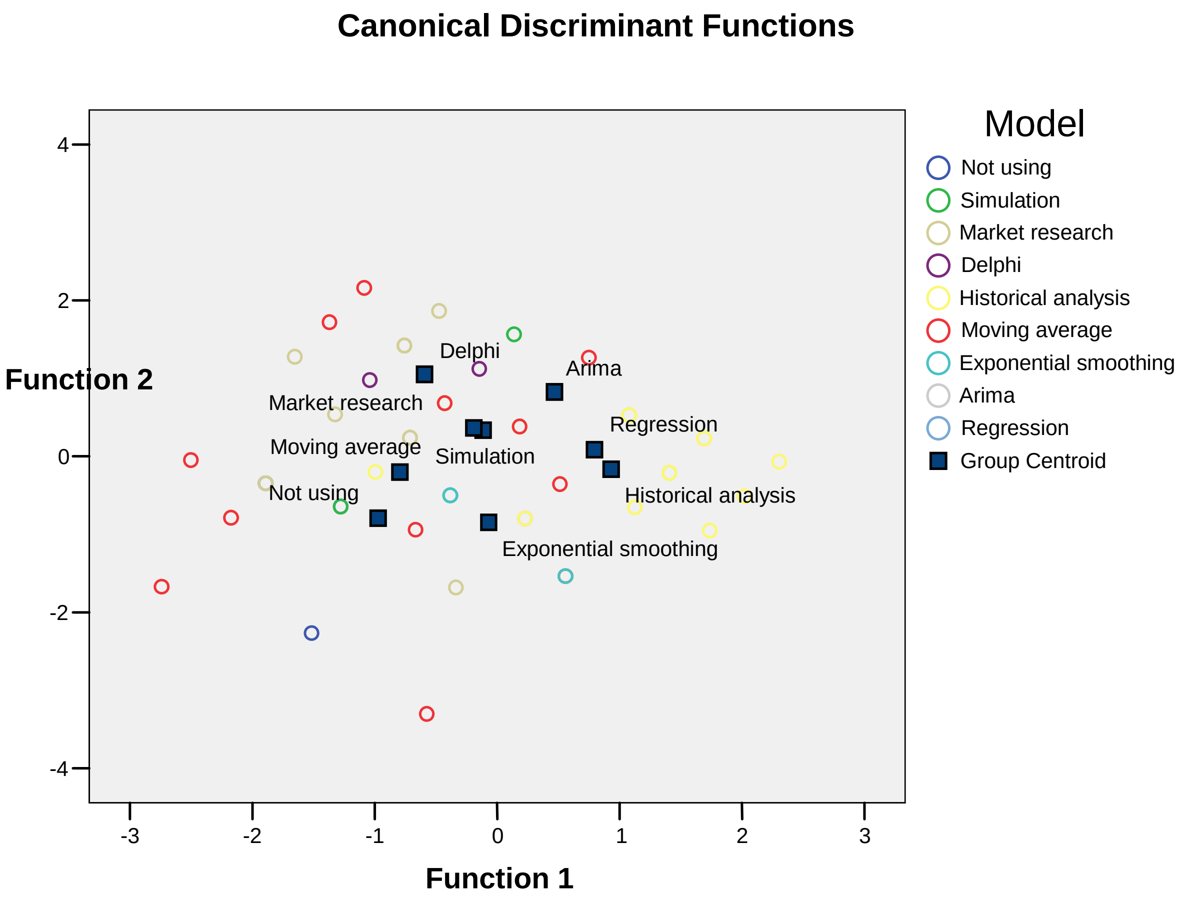
Table 12 – Classification Discriminant Function Coefficients

	Modelo								
	Nao utiliza	Simulacao	Analise de mercado	Delphi	Analise historica	Media Movel	Alisamento exponencial	Arima	Regressao
Tempo_despendido	1,482	2,223	2,214	2,404	2,299	1,793	1,714	2,604	2,371
Tipo_produto	1,368	1,557	1,502	1,051	2,347	1,301	1,937	1,768	2,189
(Constant)	-8,203	-13,674	-11,562	-14,486	-15,190	-8,681	-11,484	-18,545	-18,036

Fisher's linear discriminant functions
(Observation generated by SPSS software)

Figure 2 shows the graph of the canonical discriminant functions where the centroids of each group are listed.

Figure 2 – Canonical Discriminant Functions



We can conclude that the variable **type of product** has the most influence in the choice of forecasting models, followed by the variable **time spent** with a level of significance of 8%. Furthermore, the **historical analysis** and **regression** models are those with more discriminating positive coefficients for the variable **type of product**, while companies that **do not use** any forecasting model and using the **moving average** model are those with the highest negative discriminant coefficients for the same variable. Moreover, **Delphi** and **ARIMA** are those with more discriminating positive coefficients for the variable **time spent**, while companies that do **not use** forecasting models and using the **exponential smoothing** model are those with lower positive discriminating coefficients. Therefore, we can conclude that companies using the **historical analysis** and **regression** models consider the variable **type of product** as the main factor in choosing the model, while companies using **moving average** model do not consider the variable **type of product** as the main factor in choosing the model. We can also conclude that companies using **Delphi** and **ARIMA** models consider the variable **time spent** as a major factor in choosing the model, while companies using the exponential smoothing model do not consider the variable **time spent** as the main factor in the choice of the model.

Thus, **hypothesis 7** that states that the variable **accuracy**, **easy to understanding**, **usability** and **cost** are the main factors influencing the choice of the food industry in Brazil is rejected, since they were not significant in the discriminant model, remaining the variable **type of product** as the most important.

4.2.4 Discriminant analysis to evaluate the main difficulties found in the use of forecasting models

Similarly to the previous section, we will use the technique of discriminant analysis (DA) to evaluate the main difficulties (**difficulty of understanding**, **difficulty of using**, **high cost**, **high time spent**, **difficulty of data**, **availability of statistical software**, **type of product**, **lack of skilled labor** and **lack of interest**) in the use of forecasting models. The independent variables (**difficulties found**) are quantitative (score from 0 to 10 in order of importance) and the dependent variable (**groups of forecasting models**) is qualitative, which justifies the application of DA to determine the influence of each independent variable to discriminate groups.

Table 3 shows the variables studied for this question, the name adopted in the software, the scale adopted and its type.

Table 3 – Variable studied

Variables	Name of the variables	Scale adopted	Type of variable
2 -Model	Model	0=not use 1=simulation 2=market research 3=Delphi 4=expert panel 5=historical analysis 6=moving average 7=exponential smoothing 8=arima 9=decomposition 10=regression 11=econometric 12=neural networks	Nominal
Difficulty of understanding	Difficulty_understanding	Numerical/integer(0 a 10)	Metric
Difficulty of using	Difficulty_using	Numerical/integer(0 a 10)	Metric
High cost	High_cost	Numerical/integer(0 a 10)	Metric
High time spent	High_time_spent	Numerical/integer(0 a 10)	Metric
Difficulty of data	Difficulty_data	Numerical/integer(0 a 10)	Metric

Availability of statistical software	Software_availability	Numerical/integer(0 a 10)	Metric
Type of product	Product_type	Numerical/integer(0 a 10)	Metric
Lack of skilled labor	Lack_ labor	Numerical/integer(0 a 10)	Metric
Lack of interest	Lack_ interest	Numerical/integer(0 a 10)	Metric

Initially, all independent variables were considered in the analysis. However, we found that the variables **difficulty in using**, **high cost**, **high time spent**, **difficulty of data**, **type of product**, **lack of skilled labor** and **lack of interest** were not significant for the statistical model. From this conclusion, another discriminant analysis model was constructed considering as predictor or independent variables **availability of statistical software** and **difficulty in understanding**, and the response or dependent variable is the forecasting model. The results are below.

Tests of equality of group means are listed on table 13.

Tabela 13 – Tests of Equality of Group Means (ANOVA)

	Wilks' Lambda	F	df1	df2	Sig.
Software_availability	,679	3,312	8	56	,004
Difficulty_understanding	,759	2,228	8	56	,039

(Table generated by SPSS software)

For both variables, we can verify that Wilks' Lambda test is close to 1, indicating the absence of differences between the groups. Note that the null hypothesis of F test stating that the groups' means are equal was rejected for the variable **availability of software** at a significance level of 5% $(0{,}004 < 0{,}05)$. The same occurs for the variable **difficulty of understanding**. We can verify that the variable **availability of software** has the most discriminant power, and its Wilks' Lambda value is smaller compared to the variable **difficulty of understanding**. Therefore, for both predictor variables, there is at least one group in which the means are different.

Through the covariance matrices and correlation for each group, we verify that there was no correlation between the variables (values close to zero).

For the equality test of covariance matrices among groups of Box's M, for a significance level of 5%, the p-value was 0.096, indicating no rejection of the null hypothesis, i.e. there is equality in dispersions of the groups.

Table 14 presents the eigenvalues for the discrimination functions.

Tabela 14 – Summary of Canonical Discriminant Functions

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	,650(a)	81,8	81,8	,628
2	,145(a)	18,2	100,0	,355

a First 2 canonical discriminant functions were used in the analysis.
(Table generated by SPSS software)

It can be seen in Table 14 that the function 1 explains 81.8% of the variance between groups, while the function 2 explains only 18.2%.

Table 15 presents the results of Wilks' Lambda and Chi-square tests.

Table 15 – Wilks' Lambda and Chi-square test

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	,530	37,193	16	,002
2	,874	7,896	7	,342

(Table generated by SPSS softaware)

The null hypothesis stating that the group means are equal is rejected for the first line, since the p-value is less than 5%, concluding that at least the first discriminant function is significant. On the other hand, the p-value is more than 5% for the second discriminant function ($0{,}342 > 0{,}05$). The value of Wilks' Lambda also reports that the second function has a larger discriminant power compared to the first function, since its value is closer to 1.

Table 16 presents the standardized canonical discriminant function coefficients.

Table 16 – Standardized Canonical Discriminant Function Coefficients

	Function	
	1	2
Software_availability	,810	-,586
Difficulty_understanding	,591	,806

From the discriminant weights, it is observed that the variable **availability of software** has a high positive coefficient for the first discriminant function. On the other hand, the variable **difficulty of understanding** has a high positive coefficient for the second discriminant function.

Table 17 shows the classification function coefficients that allow you to create models to classify new cases.

Table 17 – Classification Function Coefficients

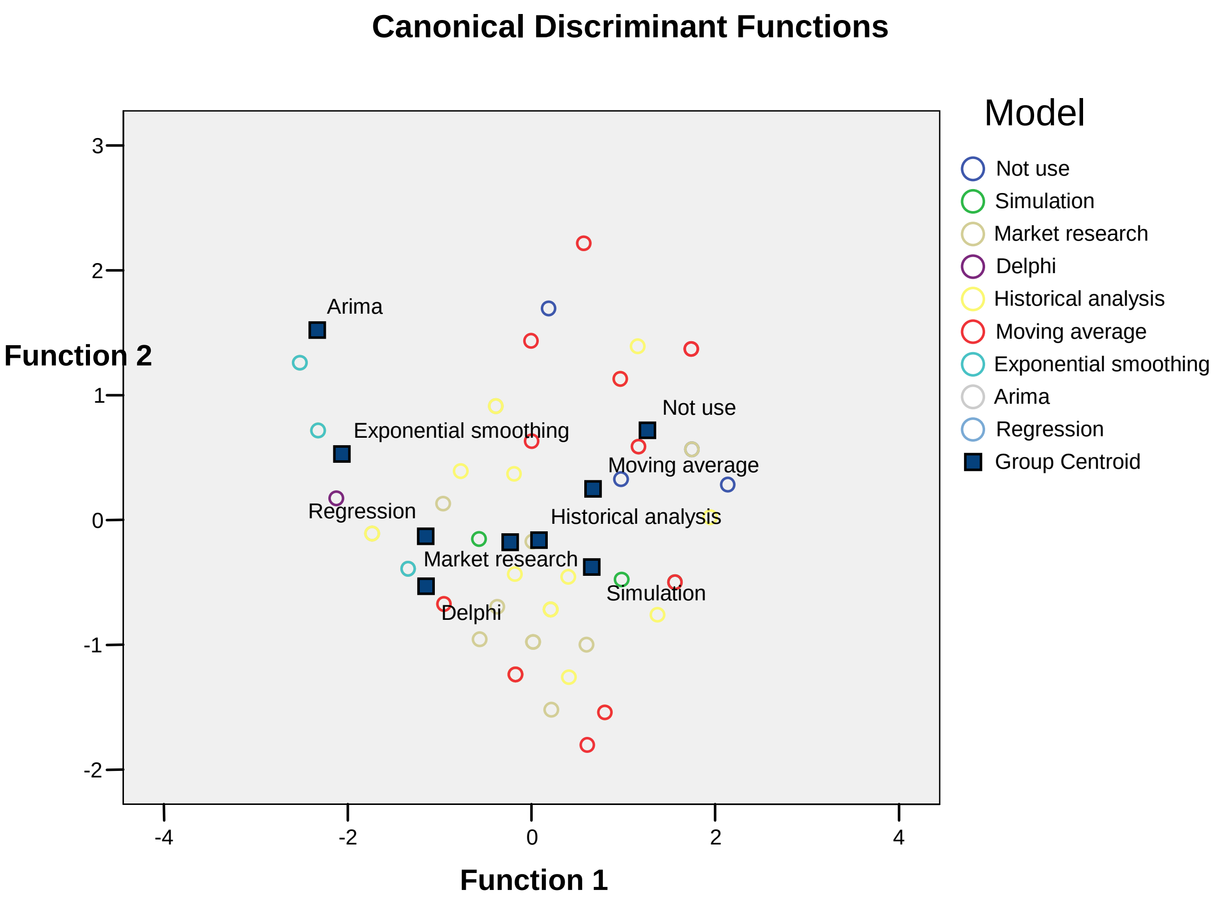
	Model								
	Not use	Simulation	Market research	Delphi	Historical analysis	Moving average	Exponential smoothing	Arima	Regression
Software_availability	1,864	1,937	1,534	1,276	1,652	1,765	,620	,235	1,161
Difficulty_understanding	,793	,392	,273	,006	,338	,557	,107	,315	,110
(Constant)	-13,216	-11,867	-6,590	-6,991	-7,632	-9,843	-3,956	-4,764	-7,132

Fisher's linear discriminant functions

(Observation generated by SPSS program)

Figure 3 shows graphically the centroids in the canonical discriminant functions.

Figure 3 – Canonical Discriminant Functions



We can conclude that the variable **availability of software** is the main difficulty in using forecasting models, followed by variable **difficulty of understanding**. You can also note that companies which do **not use** forecasting models, together with those which using **simulation** models and **moving average** are those with more positive discriminant coefficients for the variable **availability of software**, while companies using **ARIMA** models and **exponential smoothing** are those with more negative discriminant coefficients for the same variable. Moreover, **ARIMA** model has the most positive discriminant coefficient for the variable **difficulty of understanding**. Therefore, we conclude that companies using **simulation** model and **moving average** consider the variable **availability of software** as the main difficulty in using of forecasting models, while the companies that use **ARIMA** and **exponential smoothing** does not consider the variable **availability of software** as the main difficulty. We also conclude that companies using ARIMA model consider the variable **difficulty of understanding** as the main difficulty in using of the model.

Based on the results of discriminant analysis, we can conclude that hypothesis 8 can not be fully confirmed, because the main variable that influences the choice of forecasting models for Brazilian food industries is the **availability of software** that was not mentioned. Furthermore, the variable **difficulty of understanding**, listed on hypothesis 8, also influences directly the choice of forecasting models, however, has a less discriminant power. On the other hand, the variable **high cost** was not significant in the choice of forecasting models for the Brazilian food industries.

5. CONCLUSION

This paper presented an analysis of the characteristics, choices, uses and difficulties of Brazilion food industries surveyed in the use of forecasting models.

According to the results presented, the sample included all the productive sectors of food, being oil and fat the most representative, followed by the dairy industry. The region with the largest number of industry participants was the Southeast, followed by South, while the remaining did not have significant participations. Considering the number of employees declared by each company, we noted a larger participation of small and midsize companies. Furthermore, national companies were more representative than multinational.

Unlike the U.S. survey by Jain and Malehorn (2006) that pointed the exponential smoothing model as the most used by the industrial sector, most of the Brazilian companies have been using the historical analysis model, followed by market research and moving average. Simulation models, Delphi, exponential smoothing and ARIMA are less used.

We can infer from the sample results that type of product is the most influential variable in the choice of forecasting models, followed by the variable time spent. The variable time horizon and accuracy are the ones that support the use of the models chosen.

The major difficulties of the companies interviewed about the use of forecasting models were availability of software and difficulty of understanding them.

It was not possible to conclude which professionals are involved in the planning process, since each industry designated an employee to answer the questionnaire, varying from owner and directors to managers and assistants. It is suggested a deeper research in this aspect.

For future researches, we propose that the same study can be applied to a larger sample, adding companies from different parts of the supply chain in the food sector. This analysis will allow exploring different nuances of the current phase of demand planning in the Brazilian productive sector.

REFERENCES

ARMSTRONG, J. S. Evaluating forecasting methods. In: **Principles of Forecasting**. Norwell, MA: Kluwer Academic Publishers. 2001.

ARMSTRONG, J.S., FIELDS, R. Making progress in forecasting. **International Journal of Forecasting**, USA, v.22, n.3. p. 433– 441, 2006.

ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DA ALIMENTAÇÃO. **Anuário ABIA.** São Paulo: Segmento Mc Editores Ltda., 2008.

BIRCHAL, S.O. **Empresa e Indústria Alimentícia no Brasil.** Minas Gerais, 2004. Available at: <http://www.ceaee.ibmecmg.br/wp/wp17.pdf> access in 01/03/2009.

BUSSAB, W.O., MORETTIN, P.A. **Estatística Básica.** São Paulo: Saraiva, 2006.

CONCEIÇÃO, J.C.P.R. Radiografia da Indústria de Alimentos no Brasil: identificação dos principais fatores referentes à exportação, inovação e ao food safety**.** Em **Texto para discussão 1303.** Brasília 2007. Available at < http://www.ipea.gov.br > acess in 01/03/2009.

FÁVERO, L. P., BELFIORE, P., SILVA, F.L., CHAN, B.L. **Análise de dados:** modelagem multivariada para tomada de decisões. Rio de Janeiro: Campus Elsevier, 2009.

FILDES, R., HASTINGS, R. The organization and improvement of market forecasting **The Journal of the Operational Research Society**, Oxford, v.45, n.1, p. 1 -16, jan 1994.

HAIR JR., J.F., ANDERSON, R.E., TATHAM, R.L., BLACK, W.C. **Multivariate Data Analysis.** 5. ed., New Jersey: Prentice –Hall, Inc. Upper Saddle Riveder, 1998.

JAIN, C. , L., MALEHORN, J. **Benchmarking forecasting practices:** a guide to improving forecasting performance. New York: Graceway Publishing Company, 2006.

KÜSTERS, U, McCULLOUGH, B.C., BELL, M. Forecasting software: Past, present and future. **International Journal of Forecasting**, USA, v. 22, n. 3, p. 599– 615, set. 2006.

MAROCO, João. **Análise estatística com a utilização do SPSS**. Lisboa: Sílabo, 2003.

NAYLOR, T. H.. Experience with Corporate Econometric Models: A Survey. **Business Economics**, Washington, v. 16, n. 1, p.79, jan. 1981.

PELLEGRINI, F.R. **Metodologia para Implantação de Sistemas de Previsão de Demanda**, 2000, 146f - Dissertation (Master in Industrial Engineering), Escola de Engenharia da Universidade Federal do Rio Grande do Sul, Porto Alegre.

WILSON, J. H., DAUBEK, H. G. Marketing Managers Evaluate Forecasting Models. **The Journal of Business Forecasting Methods & Systems**, USA, v. 8, n.1, p. 19-22, mar.1989.

WINKLHOFER, H.; DIAMANTOPOULOS, A.; WITT, S. F. Forecasting practice: a review of the empirical literature and an agenda for future research. **International Journal of Forecasting**, USA, v. 12, n. 2, p. 193-221, jun.1996.

ZHOU, W. Integration of Different Forecasting Models **The Journal of Business Forecasting Methods & Systems**, USA, v.18, n.3, p. 26-31, set. 1999.

Appendix - Questionnaire

Characterization of the company

Name of the company:	
Name of interviewed:	
Time in the company:	
Post:	
Telephone:	
E-mail:	
Location of the company	
Number of employees	

Forecasting planning

1. What is the main production of the company?

	Derivados de carne		Óleos e gorduras
	Beneficiamento de café, chá e cereais		Laticínios
	Derivados de frutas e vegetais		Acúcares
	Derivados de trigo		Conservas de pescado
	Desidratados e supergelados		Chocolate, cacau e balas

2. What is the origin of the company?

	Nacional		Multinacional

3. How many facilities does the company have?

____

Characterization of the forecasting process

4. Which sectors of the company contribute in the forecasting process?

____

5. In the forecasting process, does the company use some method?

	Sim		Não

5.1. If No, how is the forecasting process? Describe.

____

5.2 . If Yes, what is the model used by the company? You can mark one or more option.

	Simulation	Judgement method
	Market research	
	Delphi	
	Panel experts	
	Historical analysis	
	Moving average	Time series method
	Exponential smoothing	
	ARIMA	
	Regression	Cause and effect method

	Econometric	
	Neural networks	
	Other: (Specify):	

6. What is the accuracy of the model used (percentage of correctness of the predicted demand to its actual value)?

____

7. What is the time horizon considered in the model (number of future periods covered by the model)?

____

8. What is the frequency in which new forecasts are prepared?

____

9. What are the main factors influencing the choice of the model used? For each factor, give a score from 0 to 10 in order of importance.

FACTOR	SCORE
Time horizon	
Easy of understanding	
Easy of using	
Cost	
Time spent	
Accuracy	
Data Consistency	
Availability of statistical software	
Type of product	
Other. Specify:	

10. What are the main difficulties encountered of using forecasting models? For each factor, give a score from 0 to 10 in order of importance.

FACTOR	SCORE
Difficulty of understanding	
Difficulty of using	
High cost	
High time spent	
Difficulty of data	
Availability of statistical software	
Type of product	
Lack of skilled labor	
Lack of interest	
Other. Specify:	