

FIRST INSIGHT INTO REVEALED PREFERENCES FOR CURED HAM IN SPAIN

Helena Resano, Ana I. Sanjuán and Luis M. Albisu

Documento de trabajo 04/12

FIRST INSIGHT INTO REVEALED PREFERENCES FOR CURED HAM IN SPAIN

Helena Resano, Ana I. Sanjuán and Luis M. Albisu

Unidad de Economía Agroalimentaria.

Centro de Investigación y Tecnología Agroalimentaria de Aragón. C.I.T.A.

Gobierno de Aragón

Second Draft, December 2004

Typical Food products in Europe: consumer preference and Objective
Assessment

TYPIC

PROJECT N° QLK1-CT-2002-02225

SUMMARY

This study analyses scanner data on cured ham purchases, from April 2003 to March 2004, from two hypermarkets located in Zaragoza, belonging to the same chain. The raw data comprehend actual purchases of 5640 customers, of 91 ham cured products or articles sold in two sections: self service and delicatessen (carved on request). Articles have been classified into 6 different categories, according to the breed (white or Iberian) and the fact that the product carries a collective quality label (Protected Designation of Origin - PDO-; Traditional Speciality Guaranteed - TSG.-) or not. Consumers have been segmented in different clusters based on diverse criteria, such as the number of purchases, the purchase value, or the categories of ham purchased (i.e. the influence of quality labels on the selection).

The first results obtained show that there are not significant differences between the supermarkets under investigation. Also, the analysis by categories demonstrates the predominance of the "Serrano" category, with the greatest number of purchases, customers, and expenditure. Furthermore, when considering the customers who buy products of a certain category solely, it can be said that these customers choose the "Serrano" category as the favourite one. For all categories, the price seems to be the most important factor influencing consumer's choice, well above the quality label or the product origin. It is also important to notice that the purchases of Iberian ham are very low in terms of number of purchases in comparison to white ham or to Teruel ham. Finally, it is important to emphasise the low frequency of consumer's purchases, partially due to out-of-home consumption but also to the fact that some purchases are made at other speciality shops outside the hypermarket.

CONTENTS

Introduction	1
Results.....	4
Final Remarks	30
Appendix 1.....	32
Appendix 2	40

TABLES CONTENTS

Table 1 Purchase value per ham category and hypermarket (€).....	5
Table 2 Relative purchase value per ham category and hypermarket (% over total ham)	5
Table 3 number of purchases per ham category and hypermarket.....	6
Table 4 Relative number of purchases per ham category and hypermarket (% over total number of purchases)	6
Table 5 Number of customers per ham category and hypermarket.....	7
Table 6 Relative number of customers per ham category and hypermarket (% over the total number of customers).....	7
Table 7 Descriptive statistics of purchases value per customer along the period.	8
Table 8 Descriptive statistics of purchases value per customer along the period. Category 1 "Serrano"	8
Table 9 Descriptive statistics of purchases value per customer along the period. Category 2 "Teruel DO"	9
Table 10 Descriptive statistics of purchases value per customer along the period. Category 3 "Teruel no DO"	9
Table 11 Descriptive statistics of purchases value per customer along the period. Category 4 "Iberian no DO"	9
Table 12 Descriptive statistics of purchases value per customer along the period. Category 5 "TSG Serrano"	9
Table 13 Descriptive statistics of purchases value per customer along the period. Category 6 "Iberian Paleta no DO"	9

Table 14 Position measures of purchase value per customer along the period. Total ham (€)	10
Table 15 Position measures of purchase value per customer along the period. Category 1 "Serrano"	10
Table 16 Position measures of purchase value per customer along the period. Category 2 "Teruel DO" (€)	10
Table 17 Position measures of purchase value per customer along the period. Category 3 "Teruel no DO" (€)	10
Table 18 Position measures of purchase value per customer along the period. Category 4 "Iberian ham no DO" (€)	11
Table 19 Position measures of purchase value per customer along the period. Category 5 "TSG Serrano" (€)	11
Table 20 Position measures of purchase value per customer along the period. Category 6 "Iberian paleta no DO" (€)	11
Table 21 Number of purchases per customer along the period. Total ham	13
Table 22 Number of purchases per customer along the period. Category 1 "Serrano" ..	13
Table 23 Number of purchases per customer along the period. Category 2 "Teruel DO" ..	13
Table 24 Number of purchases per customer along the period. Category 3 "Teruel no DO"	13
Table 25 Number of purchases per customer along the period. Category 4 "Iberian no DO"	13
Table 26 Number of purchases per customer along the period. Category 5 "STG Serrano"	13
Table 27 number of purchases per customer along the period. Category 6 "Iberian paleta no DO"	13
Table 28 Number of customers per intervals along the period	15
Table 29 Number of customers per intervals of number of purchases along the period ..	16
Table 30 Number of customers who only buy one category	17
Table 31 Number of customers who only buy one category (% over the number of customers of the category)	17
Table 32 Number of customers who buy "x" categories"	18
Table 33 Percentage of customers who buy "x" categories over the total number of customers	18

Table 34 Number of categories who buy pairs of categories	19
Table 35 Percentage of consumers who buy specific pairs of categories over the number of customers who buy two categories	19
Table 36 Number of customers who only buy "x" articles.....	20
Table 37 Descriptive statistics of price/kg per category.....	25
Table 38 Cross-tabulation between the clusters defined with respect to purchase value and number of purchases (both hypermarkets)	27
Table 39 Cross-tabulation between the clusters defined with respect to purchase value and number of purchases (%).....	27
Table 40 Cross-tabulation between the clusters defined with respect to purchase value and loyalty to a specific category (%)	28
Table 41 Cross-tabulation between the clusters defined with respect to purchase value and loyalty to the quality label (PDO/TSG) (%).....	28
Table 42 Descriptive statistics of monthly purchase value, across customers. Supermarket 4000.....	32
Table 43 Descriptive statistics of monthly purchase value, across customers. Hypermarket 4000.....	32
Table 44 Descriptive statistics of monthly purchase value, across customers. Both hypermarkets	33

GRAPHS CONTENTS

Graph 1 Percentage of customers per expenditure intervals	15
Graph 2 Percentage of customers per number of purchase intervals	16
Graph 3 Distribution of customers according to the number of articles bought (% for total customers)	20
Graph 4 Monthly evolution of purchase value. Total ham and category 1 "Serrano" ...	21
Graph 5 Monthly evolution of purchase value. Category 2 "Teruel DO"	22
Graph 6 Monthly evolution of purchase value. Category 3 "Teruel no DO"	22
Graph 7 Monthly evolution of purchase value. Category 4 "Iberian no DO"	23
Graph 8 Monthly evolution of purchase value. Category 5 "TSG Serrano"	23

Graph 9 Monthly evolution of purchase value. Category 6 "Iberian paleta no DO"	24
Graph 10 Monthly percentage of weight of ham in the total expenditure of ham	24
Graph 11 Cross-tabulation between the clusters defined with respect to purchase value and exclusive purchases of one ham category. Supermarket 850	33
Graph 12 Cross-tabulation between the clusters defined with respect to number of purchases and exclusive purchases of one ham category. Supermarket 850.....	34
Graph 13 Cross-tabulation between the clusters defined with respect to number of purchases and exclusive purchases of one ham category. Supermarket 850.....	34
Graph 14 Cross tabulation between the clusters defined with respect to purchase value and exclusive purchases of quality labels. Supermarket 850	34
Graph 15. Cross-tabulation between the clusters defined with respect to number of purchases and exclusive purchases of quality labels. Hypermarket 850	37
Graph 1 Cross-tabulation between the clusters defined with respect to purchase value and number of purchases. Hypermarket 4000.....	37
Graph 17 Cross-tabulation between the clusters defined with respect to purchase value and exclusive purchases of one category. Supermarket 4000	36
Graph 18 Cross-tabulation between the clusters defined with respect to number of purchases and exclusive purchases of one category. Supermarket 4000	36
Graph 19 Cross-tabulation between the clusters defined with respect to purchase value and exclusive purchases of quality labels. Supermarket 4000	37
Graph 20 Cross-tabulation between the clusters defined with respect to number of purchases and exclusive purchases of quality labels. Supermarket 4000.....	37

INTRODUCTION

In this working paper, a first description of revealed preferences for cured ham in Spain is provided. The focus is mainly to have a rough idea of the relative weight of different categories of ham

Revealed preferences are based on actual purchasing behaviour, which is studied by means of sales records at two hypermarkets of the same chain located in Zaragoza (Spain), along one year period, starting in April 2003 and finishing in March 2004. Only descriptive statistics are applied along the working paper.

The raw data contains 19599 observations. For each of them, there is information on the following variables: month, year, hypermarket (named Actur and Augusta), customer code, ticket code, article reference or code, article designation, purchase date, purchase value (in €) and the number of units purchased.

A total of 5640 customers exist in the data base. Some of them buy ham at both establishments (442), in which case, a third code is added in the hypermarket coded variable. Accordingly, customers are classified in three groups, regarding the hypermarket where they buy: Actur (coded as 850), Augusta (4000) and both hypermarkets (4850).

The hypermarkets distinguish two sections for the sale of ham: delicatessen or carved on request section, and self service. The raw data include 91 references. These, however are not synonymous of brands. The same brand may be sold in different formats (sliced, small pieces, leg, etc.), which in turn can be sold at one specific section or at both. Each of those combinations leads to a different article reference. Articles have been classified into 6 different categories, according to the breed (white or Iberian) and the fact that the product carries a collective quality label (Protected Designation of Origin - PDO- or Traditional Speciality Guaranteed - TSG.-) or not. The objective is to have a clear picture of the relevance of attributes such as the quality label, the region of origin and breed in consumers' choice. The six categories are:

- Category 1: “Serrano”: white ham which neither carries a collective label nor comes from Teruel. In other words, all white cured ham which do not belong to the rest of categories

- Category 2: “Teruel DO”: white cured ham coming from the area of Teruel with Protected Designation of Origin
- Category 3: “Teruel no DO”: white cured ham coming from the area of Teruel but without PDO
- Category 4: “Iberian no DO”: Iberian ham without PDO
- Category 5: “TSG Serrano” (or ETG using the Spanish Acronym for TSG): white ham with the collective label Traditional Speciality Guaranteed
- Category 6: “Iberian Paleta no DO”. *Paleta* is the front leg, while ham is only the rear leg of the animal. Although strictly speaking *paleta* and ham are different products, and also the organoleptic features differ, both may act as close substitutes in consumption, provided that both satisfy a similar need.

Note that there is not a category for PDO Iberian ham. This is because no reference falling in this category was found along the period of analysis.

The raw data provide information for each purchase situation of the purchase value (in €) and the number of units purchased. The conversion of these two figures to a price per kg is not straightforward because the term “units” not always mean kilogram. For instance, when the sale of a leg is recorded, the variable units will refer to the number of legs or the number of kilograms. Therefore, a very careful inspection of data is needed in order to get the actual price of purchase. To do that, several aspects have been taken into account. First, very often, in the own designation of the article there is an explicit reference to the weight of package or to the fact that the whole leg is sold. In the first case, the price per kg is directly calculated. In the second case, it is necessary to apply a specific weight to the leg in order to get the price per kg. Based on a search on several samples, we used an average weight per leg, different between white ham (7 kg), white paleta (5), Iberian ham (6) and Iberian paleta (4 kg). And second, a combination of the information on the format provided by the designation plus the units sold and the price per unit, are jointly observed, using the price per kg of close products as a reference, in order to delimit a logical range of prices.¹

¹ In a posterior working paper (January 2005) “Scanner database: brands, formats and prices”, a more careful inspection of data has been accomplished, and a more sophisticated procedure to get a closer approximation to the measurement of units and accordingly to price has been carried out. Therefore, it is possible some minor inconsistencies between both working papers because the raw data is not exactly the same. In any case, the present report is a good summary of first results.

Customers are classified in different segments, according to different segmentation criteria related to: how much the client spends on ham along the year or purchase value; how many times the client has purchased ham along the year or number of purchases; and how many and which categories of ham the client has bought along the year. The objective is to get an accurate description on the typology of consumers.

In the next section, results are shown, and finally some primary conclusions are drawn. Given the amount of data, we have used the package GAUSS, designing our own programmes. These programmes are included in the Appendix.

RESULTS

Overall view of purchase value, number of purchases and customers

Table 1 shows the values of sales (purchases) for each of the six categories of ham along the period of observation, and across hypermarkets. In Table 2, these values are expressed in percentages in order to get a clearer idea of the weight of each category on the total sales of ham. The codes 850 and 4000 stand for the hypermarkets named Actur and Augusta. The code 4850 stands for the results based on those consumers who purchase ham at both locations. Note that the total figures correspond to the sum of the values for the 850 and 4850 hypermarkets, and not the sum of all the three codes.

The sales of ham amounted to 166489 € in the two hypermarkets in Zaragoza, with very close shares for each hyper: 48% of the sales in Actur and 52% in Augusta. Serrano ham (category 1), is the most sold, accounting for more than 70% of the total sales value (total value of 121046.96 €). The distribution between supermarkets is also quite similar. The second most important category in value terms, is Iberian ham, which accounts for 11.4% of the total value. As we will see later, this is due to the high sale prices and not the quantity sold. The third and fourth most important categories in value are Teruel ham. PDO Teruel ham amounts to 13489€ and without PDO to 9361€, which in percentage terms, account to 8.1 and 5.6%, respectively, of total value of ham purchases. Both Teruel categories together represent 13.7% of the total ham sales. TSG Serrano accounts for a marginal 1.4% of total sales and “Iberian Paleta no DO” has the lowest weight value (0.8% of total ham sales). It is remarkable, that although distribution across hypermarkets is quite similar, the expenditure in Iberian ham and PDO Teruel ham is relatively higher in the hyper Augusta (13.8% and 9.3%, respectively, versus 8.7 and 6.7, respectively in Actur). This might be a result not only of different preferences on attributes searched when purchasing ham, but also on different purchasing power of the customers attending the two hypermarkets.

Table 1 Purchase value per ham category and hypermarket (€)

Hypermarket	Total	Category 1:	Category 2:	Category 3:	Category 4:	Category 5:	Category 6:
		Serrano	Teruel ham with PDO	Teruel ham without PDO	Iberian ham without PDO	TSG	Iberian "paleta" without PDO
850	79680.83	59712.49	5364.29	5588.80	6947.41	1291.44	776.40
4.000	86808.00	61334.47	8125.12	3772.57	12010.20	992.06	573.56
4.850	17951.99	13733.85	1095.30	1090.56	1511.54	323.05	197.69
Total	166488.83	121046.96	13489.41	9361.37	18957.61	2283.50	1349.96

Table 2 Relative purchase value per ham category and hypermarket (% over total ham)

Hypermarket	Total	Category 1:	Category 2:	Category 3:	Category 4:	Category 5:	Category 6:
		Serrano	Teruel ham with PDO	Teruel ham without PDO	Iberian ham without PDO	TSG	Iberian "paleta" without PDO
850	100	74.94	6.73	7.01	8.72	1.62	0.97
4000	100	70.66	9.36	4.35	13.84	1.14	0.66
4850	100	76.50	6.10	6.07	8.42	1.80	1.10
Total	100	72.71	8.10	5.62	11.39	1.37	0.81

The number of times that each category is purchased is displayed in Table 3, and in Table 4, the percentage with respect to the total number of observations. In the database there is a total of 19599 observations, from which 10475 correspond to hyper Augusta and 9124 to Actur. 2617 purchases by consumers who buy in both hypermarkets. Again, Serrano ham is the category bought the maximum number of occasions (16915) which accounts for 86.3% of the total number of observations, while TSG Serrano and Iberian Paleta are chosen only 53 and 59 times respectively, what means around 0.3% of the times. Teruel ham with and without PDO is chosen a similar number of times (1030 and 1078, respectively), what jointly represents 10.7% of the total number of observations. Iberian ham, on the other hand, only accounts for 2.4% of the total number of purchase occasions, although in economic terms it accounts for 11.4% (Table 2). It is noticeable that although the distribution across hypermarkets is quite similar, there is a higher proportion of purchases of Iberian ham in the hyper Augusta (3.4 versus 1.4%). On the other hand, Teruel ham is more purchased in Actur (6.58 versus 4.56% in Augusta).

Table 3 number of purchases per ham category and hypermarket

Hypermarket	Total	Category 1:	Category 2:	Category 3:	Category 4:	Category 5:	Category 6:
		Teruel ham		Teruel ham	Iberian	Iberian	
		Serrano	with PDO	without	ham	TSG	"paleta"
				PDO	without		without
				PDO	PDO		PDO
850	9124	7837	502	600	129	29	27
4000	10475	9078	528	478	335	24	32
4850	2617	2299	135	121	44	8	10
Total	19599	16915	1030	1078	464	53	59

Table 4 Relative number of purchases per ham category and hypermarket (% over total number of purchases)

Hyper	Total	Category 1:	Category 2:	Category 3:	Category 4:	Category 5:	Category 6:
		Teruel ham		Teruel ham	Iberian ham	Iberian	
		Serrano	with PDO	without	without	TSG	"paleta"
				PDO	PDO		without
				PDO	PDO		PDO
850	100	85.89	5.50	6.58	1.41	0.32	0.30
4000	100	86.66	5.04	4.56	3.20	0.23	0.31
4850	100	87.85	5.16	4.62	1.68	0.31	0.38
Total	100	86.31	5.26	5.50	2.37	0.27	0.30

In Table 5, the number of customers per ham category and hypermarket is shown, and in Table 6, these figures are converted to percentages over the total number of customers. There is a total of 5640 customers, from which, 2741 only buy in hyper Augusta; 2457 only buy in hyper Actur; and 422 buy in both. That means, that hyper Augusta receives a total of 3183 customers; and Actur 2899, what makes a total of 6083. From the 6083 total number, 5538 buy Serrano ham, 671 Teruel ham without PDO and 552 PDO Teruel ham. Therefore, 91% of the customers have bought Serrano ham along the year of study (Table 6), 11% Teruel ham without PDO and 9% PDO Teruel ham. Iberian ham has been chosen by 5.6% of the total number of consumers. Finally, only a marginal 0.8% of customers have ever bought TSG Serrano or Iberian paleta (51 and 54 customers, respectively). The distribution across hypermarkets is quite homogeneous, although there is a significant higher proportion of clients in Augusta hypermarket (7.2% versus 3.7%) who have bought Iberian ham in the year and a slightly higher proportion who have bought Teruel ham with PDO (9.3 versus 8.8%).

Table 5 Number of customers per ham category and hypermarket

Hypermarket	Total	Category 1:	Category 2:	Category 3:	Category 4:	Category 5:	Category 6:
		Teruel ham		Teruel ham	Iberian ham		Iberian
		Serrano	with PDO	without	without	TSG	"paleta"
				PDO	PDO		without
				PDO			PDO
850	2899	2644	256	375	108	29	26
4000	3183	2894	296	296	231	22	28
4850	442	434	59	79	33	7	7
Total	6082	5538	552	671	339	51	54

Table 6 Relative number of customers per ham category and hypermarket (% over the total number of customers)

Hypermarket	Total	Category 1:	Category 2:	Category 3:	Category 4:	Category 5:	Category 6:
		Teruel ham		Teruel ham	Iberian ham		Iberian
		Serrano	with PDO	without	without	TSG	"paleta"
				PDO	PDO		without
				PDO			PDO
850	100	91.20	8.83	12.94	3.73	1.00	0.90
4000	100	90.92	9.30	9.30	7.26	0.69	0.88
4850	100	98.19	13.35	17.87	7.47	1.58	1.58
Total	100	91.06	9.08	11.03	5.57	0.84	0.89

Descriptive statistics of purchase value per customer along the period

In the previous tables we have summarized some of the basic characteristics of the database. In the following tables we present some descriptive statistics for each of the variables analysed previously, across customers.

In Table 7, some descriptive statistics for the value of purchase per customer along the year and across hypermarkets are shown. As position measures, the mean, median (or percentile 50), and other set of percentiles are provided; and as dispersion measures, the standard deviation, the coefficient of variation (st.deviation divided by the mean), the minimum and maximum are presented.

The average expenditure per customer is 27.37€, with a standard deviation equal to 44.78€, and a coefficient of variation of 164%. Purchase values range from a minimum of 0.41 to a maximum of 945.48 €. This gives an idea of the high dispersion of the purchase value across clients, which is due to the heterogeneity of the components. This dispersion is better observed disaggregating by ham categories (Tables 8-13).

The median is 10.17€, i.e., half of the customers spends less (and more) than 10.17€; and 75% of the clients spend less than 33.28€ along the year. Therefore, the vast majority of consumers spend little on ham. There is, however, a 2.5% of the clients who have spent more than 146.29€. The largest customer is located in hyper Augusta. Apart from that, the distribution of customers across hyper is quite similar regarding the percentile distribution.

Table 7 Descriptive statistics of purchases value per customer along the period.

Super	Total	Minimum	Maximum	Mean	St. Dev.	Vari. coef	Perc2.5	Perc25	Perc50	Perc75	Perc97.5
850	79680,8	0.79	630.79	27.48	41.91	1.53	1.65	4.46	11.03	34.76	143.29
4000	86808	0.41	945.48	27.27	47.26	1.73	1.63	3.80	9.41	32.05	149.07
4850	17952	3.02	684.69	40.62	53.50	1.32	4.99	11.86	22.43	51.12	153.76
Total	166489	0.41	945.48	27.37	44.78	1.64	1.63	4.10	10.17	3.28	146.29

The descriptive statistics of the variable purchase value disaggregated for categories, are displayed in Tables 8 to 13. The behaviour within each category is quite similar to the one of total data. "Serrano" and "Iberian no DO" are the categories which show the greatest dispersion (coefficient of variation 158% and 137%, respectively), while "TSG Serrano" shows the smallest dispersion (coefficient of variation, 31%). Serrano category displays the global minimum, and the second maximum, after Iberian ham. With respect to average purchase values, these range from 13.95€ in Teruel no DO to 55,92€ in Iberian ham. Again, a maximum of clients are located at low values. STG Serrano shows a more homogeneous structure: the median is 40.5€, quite close to the mean 44.7€.

Table 8 Descriptive statistics of purchases value per customer along the period.
Category 1 "Serrano"

Super	Total	Minimum	Maximum	Mean	St. Dev.	Vari. coef	Per2.5	Per50	Per97.5
850	59712	0.70	374.24	22.58	33.37	1.48	1.63	9.38	118.82
4000	61334	0.41	581.74	21.19	35.65	1.68	1.59	7.52	116.75
4850	13734	1.68	265.35	31.64	38.77	1.23	2.70	17.21	138.23
Total	121047	0.41	581.74	21.86	34.59	1.58	1.63	8.30	117.00

Table 9 Descriptive statistics of purchases value per customer along the period.
Category 2 "Teruel DO"

Super	Total	Minimum	Maximum	Mean	St. Dev.	Vari.coe	Perc2.5	Perc50	Per97.5
850	5364.3	1.52	240.76	20.95	33.83	1.61	2.44	6.7	107.82
4000	8125.1	1.75	350.15	27.45	43.23	1.57	2.27	7.27	120.33
4850	1095.3	1.75	122.99	18.56	26.92	1.45	1.89	7.09	89.13
Total	13489	1.52	350.15	24.44	39.25	1.61	2.31	6.96	114.63

Table 10 Descriptive statistics of purchases value per customer along the period.
Category 3 "Teruel no DO"

Super	Total	Min.	Max.	Mean	St. Dev.	Vari.coe	Perc2.5	Perc50	Per97.5
850	5588.8	1.58	182.69	14.90	19.33	1.3	3.45	9.2	53.97
4000	3772.6	1	247.85	12.75	21.12	1.66	2.81	7.35	52.7
4850	1090.6	1.97	193.87	13.80	22.90	1.66	3.38	7.61	49.62
Total	9361.4	1	247.85	13.95	20.15	1.44	2.99	8.24	53.95

Table 11 Descriptive statistics of purchases value per customer along the period.
Category 4 "Iberian no DO"

Super	Total	Min.	Max.	Mean	St. Dev.	Vari.coe	Perc2.5	Perc50	Per97.5
850	6947.4	0.89	586.16	64.33	85.97	1.34	1.22	16.96	278.79
4000	12010	0.56	512.21	51.99	71.38	1.37	2.33	18.25	230.35
4850	1511.5	1.25	586.16	45.8	103.48	2.26	N.A.	N.A.	N.A.
Total	18958	0.56	586.16	55.92	76.42	1.37	2.99	18.25	257.22

Table 12 Descriptive statistics of purchases value per customer along the period.
Category 5 "TSG Serrano"

Super	Total	Min.	Max.	Mean	St. Dev.	Vari.coe	Perc2.5	Perc50	Per97.5
850	1291.4	10.66	65.53	44.53	12.99	0.29	N.A.	N.A.	N.A.
4000	992.06	15.82	83.52	45.09	15.66	0.35	N.A.	N.A.	N.A.
4850	323.05	10.66	79.54	46.15	21	0.45	N.A.	N.A.	N.A.
Total	2283.5	10.66	83.52	44.77	14.06	0.31	10.73	40.05	75.69

Table 13 Descriptive statistics of purchases value per customer along the period.
Category 6 "Iberian Paleta no DO"

Super	Total	Min.	Max.	Mean	St. Dev.	Vari.coe	Perc2.5	Perc50	Per97.5
850	776.4	4.95	72	29.86	22.16	0.74	N.A.	N.A.	N.A.
4000	573.56	4.77	98.35	20.48	25.15	1.23	N.A.	N.A.	N.A.
4850	197.69	5.3	57.86	28.24	19.35	0.69	N.A.	N.A.	N.A.
Total	1350	4.77	98.35	25	24	0.96	4.83	10.48	77.21

A more detailed analysis about the distribution of the purchases by customer is obtained through the deciles information that appears in Tables 14 to 20. In Table 14, where results refer total ham, we observe that 70% of the customers buy less than the average value. In other words, there is asymmetry in the distribution with a higher proportion of customers located below the mean values. The same pattern applies to each category. Only in the categories "Iberian no DO" and "TSG Serrano", the distribution is more symmetric, with around 50% of clients placed under and over the mean.

Table 14 Position measures of purchase value per customer along the period. Total ham (€)

Hyper	Mean	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9
850	27.48	2.99	3.66	5.5	7.85	11.03	16.5	26.21	41.34	71.32
4000	27.27	2.68	3.42	4.5	6.59	9.41	14.29	24.04	41.27	74.02
4850	40.62	6.94	10.16	13.21	17.19	22.43	30.66	43.8	59.77	95.2
Total	27.37	2.83	3.53	4.95	7.14	10.17	15.33	25.09	41.34	72.32

Table 15 Position measures of purchase value per customer along the period. Category 1 "Serrano"

Hyper	Mean	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9
850	22.58	2.85	3.5	4.5	6.54	9.38	13.32	21.2	36.16	56.9
4000	21.19	2.42	3.2	3.82	5.33	7.52	11	18.81	33.31	53.47
4850	31.64	5.18	7.18	10.03	13.18	17.21	21.79	32.89	47.27	74.7
Total	21.86	2.66	3.33	4.14	5.96	8.3	12.2	20.04	35.09	55.13

Table 16 Position measures of purchase value per customer along the period. Category 2 "Teruel DO" (€)

Hyper	Mean	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9
850	20.95	3.42	4	4.71	5.26	6.7	8.32	12.08	24.6	72.98
4000	27.45	2.81	3.8	4.7	5.47	7.27	10.65	18.88	65.66	76.74
4850	18.56	3.06	4.3	4.98	6	7.09	8.47	10.82	18.17	65.9
Total	24.44	3.08	3.94	4.71	5.34	6.96	9.41	14.6	41.63	75.72

Table 17 Position measures of purchase value per customer along the period. Category 3 "Teruel no Do" (€)

Hyper	Mean	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9
850	14.9	4.96	6.03	7.12	7.98	9.2	10.64	13.66	17.47	27.77
4000	12.75	4.3	5.28	6.04	6.73	7.35	8.19	10.4	14.04	20.9
4850	13.8	4.88	5.82	6.6	7.01	7.61	8.44	11.24	15.46	22.47
Total	13.95	4.61	5.67	6.59	7.3	8.24	9.89	12.37	15.74	25.26

Table 18 Position measures of purchases value per customer along the period. Category 4 "Iberian ham no DO" (€)

Hyper	Mean	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9
850	64.33	4.97	6.99	8.35	11.4	16.96	75.4	82.4	93.2	127.63
4000	51.99	5.97	7.69	8.65	11.94	18.25	39.91	74.98	80	98.67
4850	45.8	6.83	8.76	10.22	12.02	12.87	18.37	72.65	87.71	N.A.
Total	55.92	7.54	8.35	11.49	18.25	62.53	76.66	90	107.54	N.A.

Table 19 Position measures of purchase value per customer along the period. Category 5 "TSG Serrano" (€)

Hyper	Mean	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9
850	44.53	34	39	39	39	40.05	45.26	51.94	56.25	58.46
4000	45.09	20.4	37.95	38.63	39	39	43.95	50.11	52.58	56.07
4850	46.15	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Total	44.77	36.7	38.26	39	39	40.05	45.48	50.7	55.62	57.99

Table 20 Position measures of purchase value per customer along the period. Category 6 "Iberian paleta no DO" (€)

Hyper	Mean	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9
850	29.86	5.21	5.31	5.35	16.1	29.65	36	43.18	51.18	55.21
4000	20.48	5.15	5.35	5.35	5.38	7.69	10.41	15.87	30.03	53.08
4850	28.24	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Total	25	5.19	5.35	5.35	5.81	10.48	24.96	36	49.68	56.09

Descriptive statistics of number of purchases per customer along the period

On Tables 21 to 27 the same sort of descriptive analysis as in the previous section is carried out, but applied to the variable number of purchases per customer. As we saw before, the distribution of customers and sales across hypermarkets is quite similar, with a slight higher number of customers and sales in hyper 4000 (Augusta). On average, each customer buys ham 3.22 times along the period. This average is higher for those clients who buy in both hypermarkets. The maximum number of times that one customer has bought ham is 87. However, what is more frequent, is customers who have bought twice along the period: 60 % of the customers have bought once or twice along the period of analysis; and only 10% of the around 6000 customers, have bought more than 7 times. If the visits to the hypermarket to buy ham were spread homogenously in time, the latter figure would mean that only around 600 customers (10%) would purchase ham more than once every two months.

Per categories, "Serrano" is the one bought more times, 3 times per customer along the year. This mean drops significantly in the rest of categories (1.8 and 1.6 in the

two Teruel ham, with and without DO; 1.3 times in Iberian ham; and around once in “STG Serrano” and “Iberian Paleta no DO”. Regarding to dispersion, this is higher in Serrano (with a maximum of 82 times) followed with a certain distance by "Teruel DO" (44 times), and “Teruel no DO” (23 times). 90% of the customers who buy Serrano ham, have done it less than 7 times along the period; the frequency is still lower in Teruel ham, with 90% of the customers buying it less than 3 times (or once every four months, if purchases were spread homogeneously along time). STG and Paleta are bought very seldom, and customers do not usually repeat, as 90% of the clients who buy these products have done it only once in the period of analysis.

Table 21 Number of purchases per customer along the period. Total ham

Hyper	Total	Min.	Max.	Mean	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8	Decile9
850	9124	1	56	3.15	1	1	1	1	2	2	3	4	7
4000	10475	1	87	3.29	1	1	1	1	2	2	3	4	7
4850	2617	2	27	5.92	2	2	3	4	4	5	7	9	12
Total	19599	1	87	3.22	1	1	1	1	2	2	3	4	7

Table 22 Number of purchases per customer along the period. Category 1 "Serrano"

Hyper	Total	Min.	Max.	Mean	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8	Decile9
850	7837	1	40	2.96	1	1	1	1	2	2	3	4	6
4000	9078	1	82	3.14	1	1	1	1	2	2	3	4	7
4850	2299	1	26	5.3	2	2	2.2	3	4	5	6	8	11
Total	16915	1	82	3.05	1	1	1	1	2	2	3	4	7

Table 23 Number of purchases per customer along the period. Category 2 "Teruel DO"

Hyper	Total	Min.	Max.	Mean	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8	Decile9
850	502	1	23	1.96	1	1	1	1	1	1	1	2	4
4000	528	1	44	1.78	1	1	1	1	1	1	1	2	3
4850	135	1	17	2.29	1	1	1	1	1	1.4	2	2.2	5.1
Total	1030	1	44	1.87	1	1	1	1	1	1	1	2	3

Table 24 Number of purchases per customer along the period. Category 3 "Teruel no DO"

Hyper	Total	Min.	Max.	Mean	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8	Decile9
850	600	1	18	1.6	1	1	1	1	1	1	1	2	3
4000	478	1	23	1.61	1	1	1	1	1	1	1	1.8	3
4850	121	1	10	1.53	1	1	1	1	1	1	1	2	2.1
Total	1078	1	23	1.61	1	1	1	1	1	1	1	2	3

Table 25 Number of purchases per customer along the period. Category 4 "Iberian no DO"

Hyper	Total	Min.	Max.	Mean	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8	Decile9
850	129	1	6	1.19	1	1	1	1	1	1	1	1	2
4000	335	1	13	1.45	1	1	1	1	1	1	1	1	2
4850	44	1	6	1.33	1	1	1	1	1	1	1	1	1.7
Total	464	1	13	1.37	1	1	1	1	1	1	1	1	2

Table 26 Number of purchases per customer along the period. Category 5 "STG Serrano"

Hyper	Total	Min.	Max.	Mean	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8	Decile9
850	29	1	1	1	1	1	1	1	1	1	1	1	1
4000	24	1	2	1.09	1	1	1	1	1	1	1	1	1
4850	8	1	2	1.14	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Total	53	1	2	1.04	1	1	1	1	1	1	1	1	1

Table 27 numberof purchases per customer along the period. Category 6 "Iberian paleta no DO"

Hyper	Total	Min.	Max.	Mean	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8	Decile9
850	27	1	2	1.04	1	1	1	1	1	1	1	1	1
4000	32	1	2	1.14	1	1	1	1	1	1	1	1	2
4850	10	1	4	1.43	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Total	59	1	2	1.09	1	1	1	1	1	1	1	1	1

Distribution of customers across intervals of purchase value and number of purchases along the period

Expenditure and number of purchases have been split into intervals, and the number of customers falling into each interval calculated. This is complementary to the information provided in Tables 14 and 21 by the percentiles. Detailed results for each hypermarket are shown in Table 28 and 29. The percentage of customers in each interval for both hypermarkets, is shown graphically in Graphs 1 and 2. 34% of the total number of clients have spent less than 6€ along the year, and 19% between 6 and 12€. That implies, that more than half of the total customers (54%) have spent less than 12 € along the period of analysis, what means an average of 1€ per month. There is however a 10% of customers who have spent more than 72€ in the period. Across hypermarkets, the distribution is quite similar, always keeping in mind that in Augusta (hyper 4000) there is a slight higher number of customers. Regarding the distribution of customers according to number of purchases intervals, there is also a predominance of low frequency purchases: 46.8% of the total number of customers have bought ham only once along the year; 19.3% twice; and around 10% 3 times; only 0.6% (39 customers) have bought 12 times, what would lead to an average of once every month. Nevertheless, there is around 4% of customers (277) who have purchased ham from the hypermarket more than 12 times in the period.

In summary, clients of the distribution chain considered, are characterized in a big proportion by devoting a low proportion of the food expenditure to cured ham, and concentrate their purchases in a very few occasions.

Table 28 Number of customers per intervals along the period

Expenditure interval (€)			850	4000	4850	Total
	x	<=6	929	1193	26	2122
6<	x	<=12	586	594	86	1180
12<	x	<=18	297	245	74	542
18<	x	<=24	165	194	46	359
24<	x	<=30	125	126	28	251
30<	x	<=36	101	102	26	203
36<	x	<=42	129	110	19	239
42<	x	<=48	67	87	19	154
48<	x	<=54	74	77	13	151
54<	x	<=60	63	43	18	106
60<	x	<=66	44	36	7	80
66<	x	<=72	37	43	12	80
>72			283	333	68	616
			2900	3183	442	6083

Graph 1 Percentage of customers per expenditure intervals

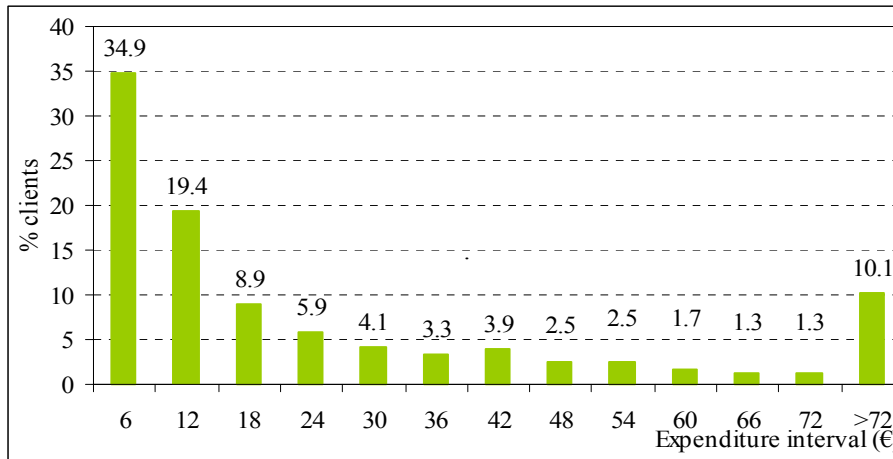
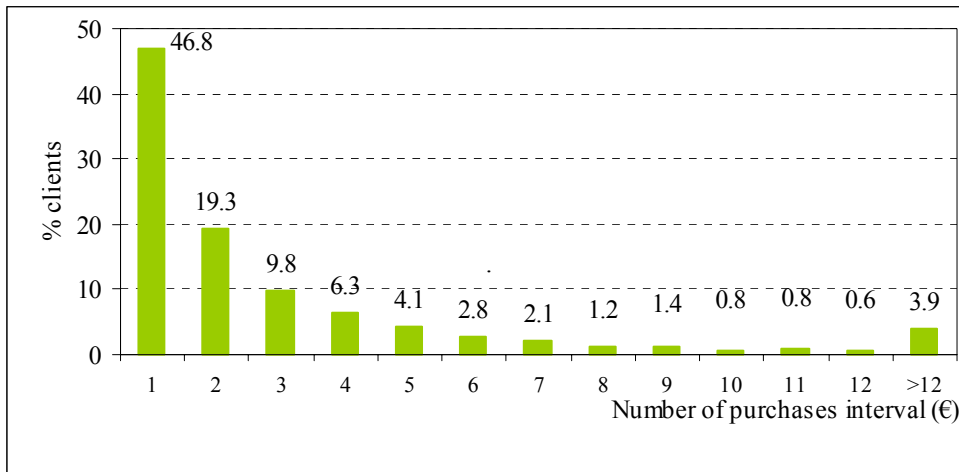


Table 29 Number of customers per intervals of number of purchases along the period

Number of purchases	850	4000	4850	Total
1	1348	1501	0	2849
2	551	622	104	1173
3	298	301	60	599
4	194	189	67	383
5	120	132	40	252
6	94	77	35	171
7	58	70	31	128
8	32	43	13	75
9	40	43	16	83
10	19	27	15	46
11	22	25	13	47
12	19	20	6	39
13	16	13	3	29
14	8	10	8	18
15	11	12	6	23
16	10	11	5	21
17	5	7	7	12
18	11	8	3	19
19	3	4	3	7
20	3	7	1	10
>20	38	61	6	99
Total number of customers	2899	3183	442	6082

Graph 2 Percentage of customers per number of purchase intervals



First insight into purchasing behaviour

From the previous sections we know which categories of ham are purchased the most, in terms of value and frequency of purchase. What we try to do next is to find out if customers are loyal to one category of ham, or on the contrary, they like to try different types. In Table 30 the number of clients who only buy one specific category of ham is shown. In Table 31, these figures are converted into percentages with respect to the number of clients of each category (last row in Table 5).

Of the total number of customers in the data base, 6083, an 84% (5123clients) only buy one category of ham. Therefore, there is a high exclusiveness in terms of types of ham. Within Serrano category, we find the highest level of loyalty, as 83% of the customers who ever buy Serrano, only buy this type of ham. The distribution by hypermarket is quite similar. "Iberian Paleta no DO" and "STG Serrano", also show an important degree of loyalty: 46 and 41% of purchasers of these products, respectively, only buy buy them. Nevertheless, note that the number of clients of these two categories is very low compared to any other type of ham (see Table 5). In these cases the distribution across hyper is slightly different. Interestingly, the clients who buy Teruel ham, with or without PDO, are the less exclusive: only around 28% of the clients who ever buy Teruel ham, buy these products exclusively. The loyalty to Iberian ham, is situated around 37% of the clients of this category, occupying therefore an intermediate position between Serrano and Teruel ham.

Table 30 Number of customers who only buy one category

Number of customers who only buy the following categories:							
	Total	Category 1	Category 2	Category 3	Category 4	Category 5	Category 6
Hypermarket 850	2437	2194	64	112	42	11	14
Hypermarket 4000	2686	2421	92	69	83	10	11
Hypermarket 4850	290	285	0	3	1	0	1
Total	5123	4615	156	181	125	21	25

Table 31 Number of customers who only buy one category (% over the number of customers of the category)

Number of customers who only buy the following categories:							
	Total	Category 1	Category 2	Category 3	Category 4	Category 5	Category 6
Hypermarket 850	84,03	82,98	25,00	29,87	38,89	37,93	53,85
Hypermarket 4000	84,39	83,66	31,08	23,31	35,93	45,45	39,29
Hypermarket 4850	65,61	65,67	0,00	3,80	3,03	0,00	14,29
Total	84,22	83,33	28,26	26,97	36,87	41,18	46,30

In Table 32 the number of customers who buy one or more categories is displayed, and in Table 33 the percentage over the total number of customers. As we have just seen, 84% of the total number of clients only buy one specific category, being Serrano the one that attracts the maximum degree of loyalty. Only two categories are bought by 13.4% of the customers, and in Table 34 which two categories are bought are shown. Of the 815 customers (Table 32) who buy two categories, 780 buy Serrano and something else (Table 34). Teruel ham seems the most preferable ham to combine with Serrano, as 612 customers buy both of them (265 PDO Teruel ham, and 347 Teruel no DO). Finally, only 2% of the consumers buy three categories, while only 13 customers buy up to 4 categories, and nobody buys from the six designed categories.

Table 32 Number of customers who buy "x" categories"

	Number of categories the customer buy					
	1	2	3	4	5	6
Hypermarket 850	2437	392	67	4	0	0
Hypermarket 4000	2686	423	63	9	2	0
Hypermarket 4850	290	129	21	2	0	0
Total	5123	815	130	13	2	0

Table 33 Percentage of customers who buy "x" categories over the total number of customers

	Number of categories the customer buy						
	Total	1	2	3	4	5	6
Hypermarket 850	100	84,03	13,52	2,31	0,14	0,00	0,00
Hypermarket 4000	100	84,39	13,29	1,98	0,28	0,06	0,00
Hypermarket 4850	100	65,61	29,19	4,75	0,45	0,00	0,00
Total	100	84,22	13,40	2,14	0,21	0,03	0,00

Table 34 Number of categories who buy pairs of categories

		Category 1	Category 2	Category 3	Category 4	Category 5	Category 6
Hypermarket 850	Category 1	---	132	190	42	13	3
	Category 2		---	5	2	0	0
	Category 3			---	1	1	1
	Category 4				---	0	2
	Category 5					---	0
	Category 6						---
Hypermarket 4000	Category 1	---	133	157	91	10	9
	Category 2		---	10	6	1	0
	Category 3			---	3	0	1
	Category 4				---	0	2
	Category 5					---	0
	Category 6						---
Hypermarket 4850	Category 1	---	40	62	19	4	1
	Category 2		---	1	0	0	0
	Category 3			---	1	0	1
	Category 4				---	0	0
	Category 5					---	0
	Category 6						---
Total	Category 1	---	265	347	133	23	12
	Category 2		---	15	8	1	0
	Category 3			---	4	1	2
	Category 4				---	0	4
	Category 5					---	0
	Category 6						---

Table 35 Percentage of consumers who buy specific pairs of categories over the number of customers who buy two categories

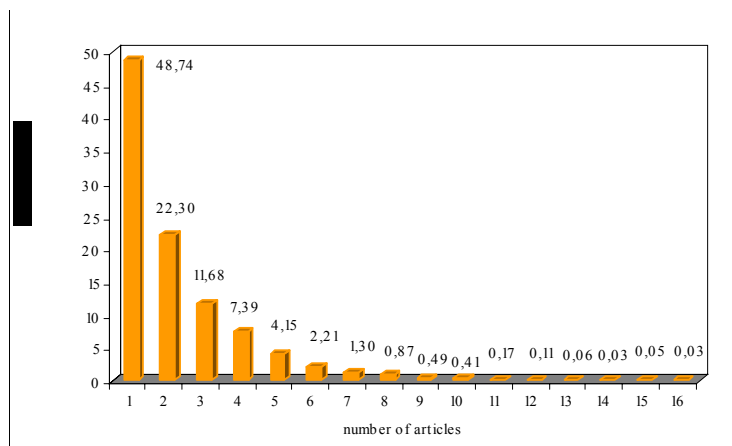
		Category 1	Category 2	Category 3	Category 4	Category 5	Category 6
Total	Category 1	---	32,52	42,58	16,32	2,82	1,47
	Category 2		---	1,84	0,98	0,12	0,00
	Category 3			---	0,49	0,12	0,25
	Category 4				---	0,00	0,49
	Category 5					---	0,00
	Category 6						---

So far we have studied the purchasing behaviour of consumers with respect to categories of ham. In this section we analyze the purchasing behaviour with respect to the individual articles bought. Table 36 and Graph 3 shows that the majority of consumers (71,04 % for all the customers) did not buy more than two articles along the analysed period, and almost half of them bought one article only (48,74%). Therefore, this is a first indication of the high degree of loyalty, not only to specific categories, but also to a specific article or brand within the category. To identify the most bought articles is a future task.

Table 36 Number of customers who only buy "x" articles

	Number of articles	Number of customers	% customers who buy "X" articles
	1	3180	48,74
	2	1455	22,30
	3	762	11,68
	4	482	7,39
	5	271	4,15
	6	144	2,21
	7	85	1,30
	8	57	0,87
	9	32	0,49
	10	27	0,41
	11	11	0,17
	12	7	0,11
	13	4	0,06
	14	2	0,03
	15	3	0,05
	16	2	0,03
Total	14494	6524	100

Graph 3 Distribution of customers according to the number of articles bought (% for total customers)

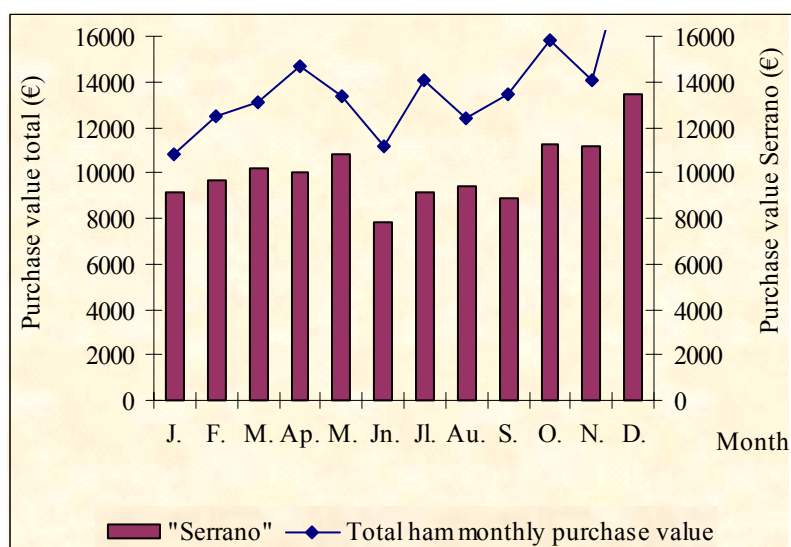


Seasonality of sales

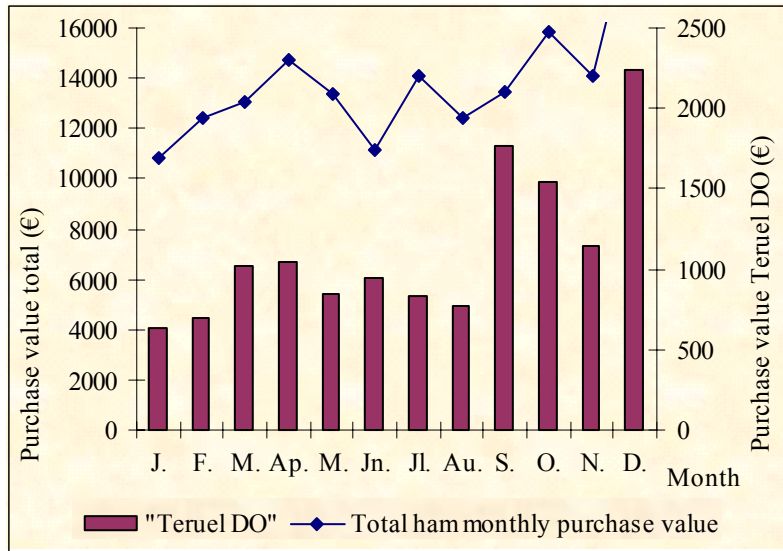
In this section, seasonality in purchases is investigated. In each graph, the monthly purchase value for each category of ham is displayed.

Graph 4 presents the monthly evolution of purchases on total ham and Serrano. With respect to the evolution along the year of the sales of cured ham, the minimum happens in January. In June and August there are also drops in sales, which recover in Autumn. Sales increase considerably in October, coinciding with the local festivity, and reach the maximum in December, corresponding with Christmas time. Serrano category and Teruel ham with DO follow clearly this pattern. The rest of categories do not show a remarkable seasonal behaviour (Graphs 4 to 8), with the exception of “STG Serrano”, which expresses an important sales increase in February) and “Iberian paleta no DO”, which shows a considerably sales increase in December, that means this category could follow “Serrano” seasonal trend.

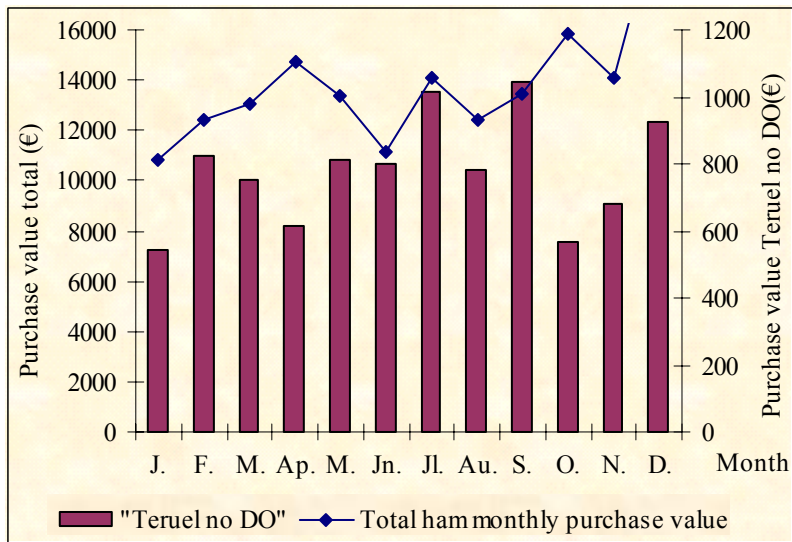
Graph 4 Monthly evolution of purchase value. Total ham and category 1 "Serrano"



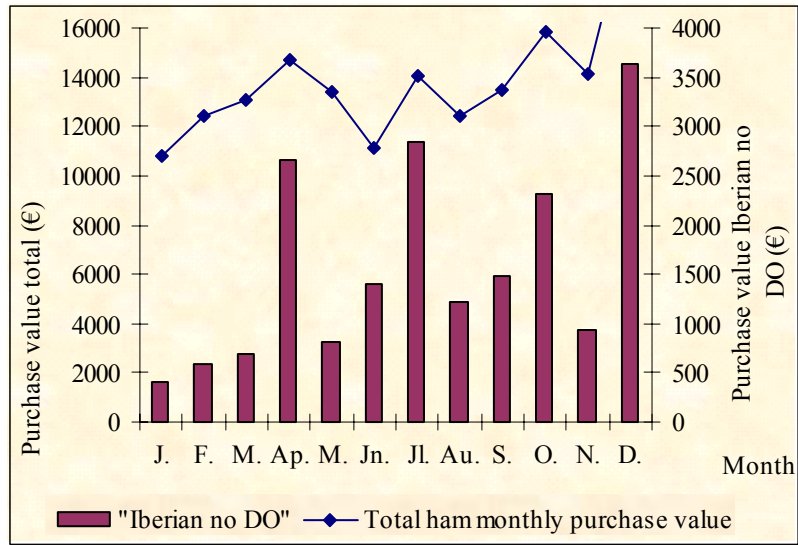
Graph 5 Monthly evolution of purchase value. Category 2 "Teruel DO"



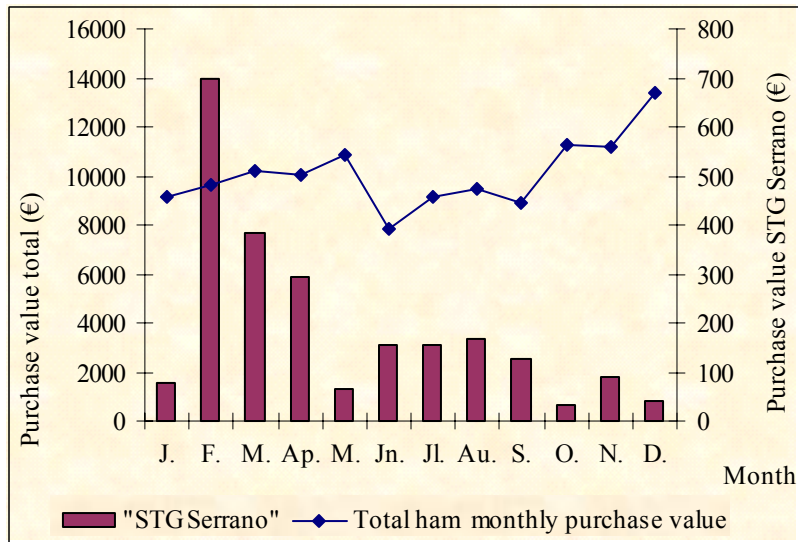
Graph 6 Monthly evolution of purchase value. Category 3 "Teruel no DO"



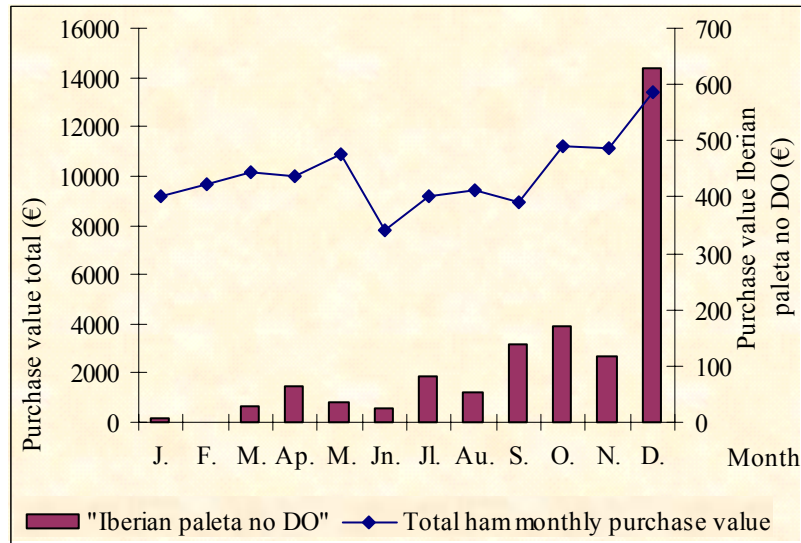
Graph 7 Monthly evolution of purchase value. Category 4 "Iberian no DO"



Graph 8 Monthly evolution of purchase value. Category 5 "TSG Serrano"

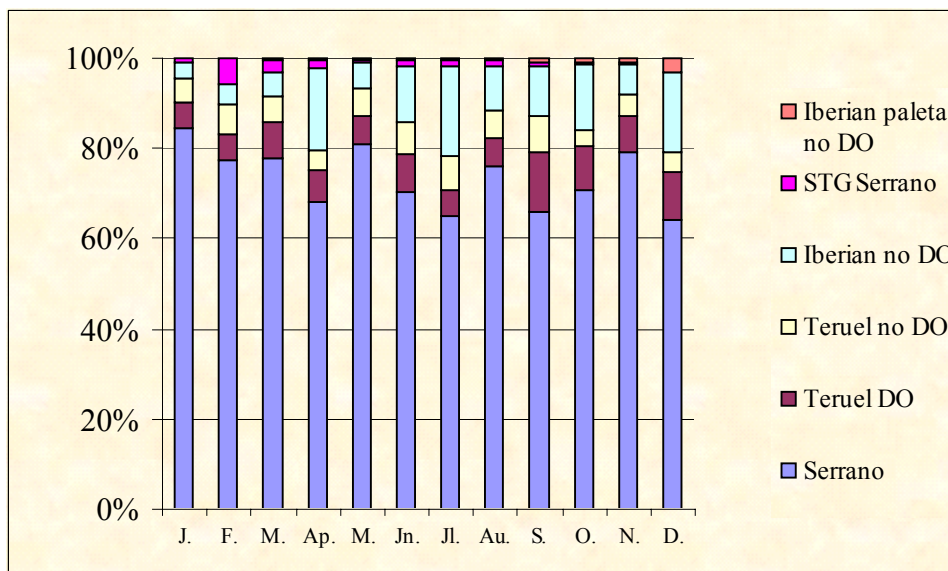


Graph 9 Monthly evolution of purchase value. Category 6 "Iberian paleta no DO"



Next, the monthly proportion of each category is calculated in order to ascertain if there is some seasonal substitution between categories. Results are shown in Graph 10. The market share for each category changes clearly across months. Focusing in Teruel ham, and more specifically, in DO Teruel, the market share is bigger in October, in coincidence with local festivities, followed by December, at Christmas season. The market share for Iberian ham and paleta, is also significantly higher in December. Therefore, and although with caution because we do not know the existence of promotions or special offers, it seems that the most expensive hams (Iberian and D.O. Teruel are more sold) at special festivities, and accordingly, consumed at special occasions.

Graph 10 Monthly percentage of weight of ham in the total expenditure of ham



Prices

In Table 37 some descriptive statistics of prices per category of ham are presented. This is a very preliminary analysis as further depuration of data is needed for getting more accurate prices (see Working Paper: “Scanner database: brands, formats and prices”). Anyway, this provides a first insight into prices. The first remarkable fact is the large dispersion of prices across categories, and even within. Obviously ham is not homogeneous, and that explain this diversity. But also, within the same category, the format of the sale is an important determinant. In the future we will develop a more detailed description based on the different formats. For instance, the price per kg of one leg is always much lower than the price per kg of sliced ham, because of the weight of the bone and the mixture of qualities across the leg. Nevertheless, the information in Table 37 is still useful to appreciate the different average prices per category and its distribution across observations. Thus, the average price for Serrano ham is 12.9€, while for PDO Teruel ham is 19.9, and Teruel no Do is 17.3€. Therefore, within the white breed ham categories, ham from Teruel is more expensive, while the Designation of Origen label adds value to the product. The average price of Iberian ham in the sample is 39€, but there is a maximum of 83€. Interestingly, there is more homogeneity within the white breed ham categories than within Iberian. For instance, half of the observations within Teruel ham, have a price around 23 and 13€, for the ham with and without DO, respectively, while only 5% of the observations of both categories have a price higher than 25€. Within Iberian ham, however, the median is around 42€, and this price goes up to 60€ in the 75% percentile, with 5% of the observations having a price higher than 83€. Note also that none of these Iberian hams carry the Designation of Origin.

Table 37 Descriptive statistics of price/kg per category

	Minimum price €/kg	Maximum price €/kg	Average price €/kg	Percentile 25	Percentile 50	Percentile 75	Percentile 95
Total	3,45	83,50	14,19	10,87	12,98	15,87	23,80
Category 1	3,45	29,40	12,91	10,82	12,45	14,91	18,76
Category 2	8,49	38,39	19,90	17,91	23,65	23,86	25,00
Category 3	11,99	25,11	17,28	13,19	13,20	23,69	25,06
Category 4	4,14	83,50	39,05	13,14	41,97	59,94	83,50
Category 5	4,73	16,66	7,53	5,57	6,60	8,70	16,40
Category 6	3,14	53,50	32,06	9,99	44,90	53,00	53,50

Note: Category 1 is Serrano ham; Category 2 is Teruel DO; Category 3 is Teruel no DO; Category 4 is Iberian no DO; Category 5 is TSG Serrano; and Category 6 is Iberian Paleta no DO. Consumers' segmentation

In order to get a deeper understanding of the typology of consumers, the total sample has been segmented according to a set of alternative criteria. These criteria are: the expenditure on cured ham; the number of purchases, the search for specific categories of ham or for a collective quality label. Then consumers have been classified in a number of segments. The criteria used and the different levels considered for segmentation are the following:

- a) Total purchase value for the whole year:
 - 1) < 30 €
 - 2) > 30 and ≤ 60 €
 - 3) > 60 and ≤ 90 €
 - 4) > 90 €

- b) Number of purchases per year. Roughly, it could be interpreted also as the frequency of purchase if purchases are uniformly spread along the year, whose corresponding values are shown in parentheses:
 - 1) ≥ 12 (at least once every month)
 - 2) < 12 and ≥ 9 (at least once every 1,5 months)
 - 3) < 9 and ≥ 6 (at least once every 2 months)
 - 4) < 6 and ≥ 4 (at least once every 3 months)
 - 5) < 4 (less than once every 3 months)

- c) Customers who buy only specific products, either in terms of categories or in terms of quality labels:
 - By categories:
 1. Only Category 1 (Serrano) is bought
 2. Only Categories 2(Teruel DO) or 3(Teruel no DO) are bought
 3. Only Categories 4(Iberian ham) or 6(Iberian paleta) are bought
 4. Only Category 5 (Serrano STG) is bought
 0. The rest of consumers or non exclusive consumers

 - By quality label:
 1. Purchasers of only ham with a Quality label: only Categories 2 and 5 are bought
 0. The rest of consumers

As a result of the previous segmentation, we get each consumer classified in a specific group. In this section, we show the proportions of each segment. Bivariate cross-tables of frequencies, for each hypermarket, are shown graphically in the Appendix. In Table 41 and 42, the bivariate cross-tabulation between clusters defined with respect to purchase value and number of purchases is shown, for both

hypermarkets Actur and Augusta, respectively. This is an important classification as we use this information as a way to prioritize the contact with consumers for the next analysis stages (sensorial test and survey). The highest concentration of customers correspond to the group that spends less than 30€ and has bought less than 4 times cured ham.

Table 38 Cross-tabulation between the clusters defined with respect to purchase value and number of purchases (both hypermarkets)

number of purchases	purchase value				total
	<30€	30-60€	61-90€	>90€	
>=12 times	7	103	55	112	277
<12 and >=9 times	35	88	26	27	176
<9 and >=6 times	185	117	34	38	374
<6 and >=4 times	439	76	40	80	635
<4 times	3788	469	216	147	4620
total	4454	853	371	404	6082

Table 39 Cross-tabulation between the clusters defined with respect to purchase value and number of purchases (%)

number of purchases	purchase value			
	<30€	30-60€	61-90€	>90€
>=12 times	0.16	12.08	14.82	27.72
<12 and >=9 times	0.79	10.32	7.01	6.68
<9 and >=6 times	4.15	13.72	9.16	9.41
<6 and >=4 times	9.86	8.91	10.78	19.80
<4 times	85.05	54.98	58.22	36.39
total	100.00	100.00	100.00	100.00

In Table 40 bivariate cross-tabulation between the clusters defined with respect to purchase value and loyalty to a specific category or exclusivity is shown. Interestingly, the more the customer spends, the more keen is on trying different categories of ham. Thus, for those customers who spend less than 30€ a year, most of it is devoted to buy Serrano ham only, while for those who spend more than 90€, half of the consumers buy different categories of ham, and the percentage which only buy Serrano reduces to 45%. The medium range expenditure cluster is formed by half of the customers who only buy Serrano, 11% by customers loyal to Teruel ham, with or without PDO, and 11% by consumers who only buy Iberian products.

Table 40 Cross-tabulation between the clusters defined with respect to purchase value and loyalty to a specific category (%)

Exclusivity	purchase value				Total
	<30€	30-60€	61-90€	>90€	
Only Serrano(TSG included)	81.7	75.3	50.1	45.3	76.5
Only Teruel ham	6.5	1.2	11.6	2.0	5.8
Only Iberian ham and paleta	1.4	1.5	11.1	7.2	2.4
The rest of consumers	10.4	22.0	27.2	45.5	15.4
Total	100	100	100	100	100

In Table 41 the bivariate cross-tabulation between clusters define with respect to expenditure and loyalty to products carrying any quality label is shown. The percentage of customers loyal to the quality label is higher in that group of consumers who spend between 61 and 90€. Only 1% of the customers who spend the most on cured ham do it exclusively on certified labelled ham.

Table 41 Cross-tabulation between the clusters defined with respect to purchase value and loyalty to the quality label (PDO/TSG) (%)

Quality label	purchase value				Total
	<30€	30-60€	61-90€	>90€	
Only Quality label	2.5	2.5	11.1	1.0	2.9
The rest of onsumers	97.5	97.5	88.9	99.0	97.1
Total	100.0	100.0	100.0	100.0	100.0

For future research (sensorial analysis) it would be interesting to have the participation of those consumers more active, given that these are the ones who spend more and purchase with a higher frequency. Accordingly, we would expect that these consumers are more experienced, know more about ham, but also, take the purchasing decision a greater number of times. Focusing on those who spend more than 30€ and have gone to the hypermarket to buy ham at least four times (an average of once every three months), we get six segments, formed by 792 consumers, split in the following way:

Ha_p: 103 customers with total purchase value between 31- 60 € and number of purchases ≥ 12 .
Ha_m: 56 customers with total purchase value between 61- 90 € and number of purchases ≥ 12 .
Ha_e: 112 customers with total purchase value higher than 90 € and number of purchases ≥ 12 .

Es_p: 74 customers with total purchase value between 31- 60 € and number of purchases between 4 and 6.
Es_m: 40 customers with total purchase value between 61- 90 € and number of purchases between 4 and 6.
Es_e: 80 customers with total purchase value higher than 90 € and number of purchases between 4 and 6.

Me_p: 201 customers with total purchase value between 31- 60 € and number of purchases ≥ 9 or number of purchases between 6 and 9.
Me_m: 59 customers with total purchase value between 61- 90 € and between 31- 60 € and number of purchases ≥ 9 or number of purchases between 6 and 9.
Me_e: 67 customers with total purchase value higher than 90 € and number of purchases ≥ 9 or number of purchases between 6 and 9.

FINAL REMARKS

In this working paper, we have tried to offer a quick look into the scanner data, providing some basic characteristics and descriptive statistics.

The main remarks that can be made after this preliminar analysis are the following. First, no highly significant differences have been found between hypermarkets. The only facts to highlight are that Augusta has a slight larger number of customers and represent more than half of the total sales of ham, and that Iberian ham, the most expensive category, is sold to a greater extent in this retail outlet.

Second, "Serrano" category predominates in number of customers as well as in purchase value. The average price of this category is clearly lower than for the rest of hams what seems to mean that price influences considerably consumers' choice, above other considerations as the region of origin, or the collective quality label.

Third, only a minority of customers buy Iberian ham, and even a lower proportion are devoted to Iberian only. Several factors explain this outcome: first, tradition. Zaragoza is located in an area of production of white ham and not Iberian, leading to a broader customer base for white ham. Second, Teruel ham in particular, is also produced in the region, what implies that, customers are not only used to white ham in general but also to ham produced in Teruel, in particular. And third, Iberian ham is minimum 2 to 3 times as expensive as white ham, even Teruel ham with Designation of Origin.

Fourth, the frequency of purchase is very low and also the expenditure on ham throughout the period under investigation. In fact, more than half of the customers spent less than 12 €, and bought cured ham less than twice during this period. A likely explanation for this low purchase frequency can be that cured ham is a product mainly consumed outside home, in bars and restaurants. Also, it is feasible that clients do not make every single purchase of ham in this hypermarket, using also other retail outlets. Therefore, it seems mandatory to investigate through the questionnaires where do customers usually buy cured ham, and which proportion the purchases at this retail outlet represent over the total ham purchases, otherwise we risk to classify customers as sporadic consumers when in fact they are usual consumers. And finally, even if the

client buys only in this distribution chain, it could happen that not every time he/she buys he/she uses the loyalty card.

Fifth, there is clearly a seasonal behaviour of sales. The sales increase during the months of October and December, corresponding with local festivities and Christmas season, respectively. Interestingly, the highest priced articles are more purchased at this time of the year.

Finally, the vast majority of customers only buy one category, showing a great degree of loyalty towards a specific category, being Serrano the most preferred. However, some of the customers combine two categories, and in this case Teruel ham is the most preferable combination.

It is important to explain that this document simply gathers the initial phase of the study and, therefore, it will be continued in order to improve the analysis of scanner data. Much work needs still to be done, because the information is large and complex, and accordingly there are many complementary ways to explore the data.

APPENDIX 1

In this appendix complementary analysis are shown. The first tables show some descriptive statistics on monthly purchase values on total ham across customers, without distinction between categories. Statistics include position measures, such as mean and deciles, and dispersion measures, such as minimum and maximum. Results for each of the hypermarkets and both of them are presented in Table 42, 43 and 44, respectively.

Table 42 Descriptive statistics of monthly purchase value, across customers.
Supermarket 4000

Month	Total	Min.	Máx.	Mean	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8	Decile9
January	5415,74	0,23	92,75	6,66	1,39	1,97	2,63	3,09	3,49	3,76	4,82	6,03	10,42
February	6103,90	0,38	99,45	8,14	1,72	2,54	3,20	3,50	3,80	4,45	6,04	7,31	15,04
March	6540,44	0,23	99,00	7,96	1,63	2,12	2,93	3,32	3,57	4,64	6,04	7,52	14,26
April	8000,47	0,47	144,02	9,82	1,69	2,52	2,99	3,48	3,84	4,63	6,43	7,79	25,63
May	6537,72	0,69	90,96	7,04	1,71	2,44	2,87	3,15	3,54	4,13	5,26	6,60	12,53
June	6087,58	0,39	152,16	6,84	1,77	2,33	2,72	3,02	3,41	3,96	4,86	6,11	10,05
July	7213,50	0,87	149,96	8,93	1,89	2,57	3,00	3,41	3,84	4,55	5,87	7,71	17,34
August	6367,33	0,38	225,37	7,54	1,93	2,59	3,12	3,45	3,74	4,35	5,69	7,15	12,92
September	7490,21	0,56	223,32	8,22	1,74	2,66	3,00	3,26	3,58	4,05	5,14	6,93	12,63
October	8380,01	0,49	209,35	8,79	1,74	2,50	2,96	3,38	3,54	4,11	5,17	7,08	17,43
November	7355,58	0,20	111,78	7,80	1,65	2,06	2,90	3,22	3,57	4,37	5,98	7,14	17,55
December	11315,52	0,14	252,01	11,35	1,65	2,17	2,87	3,30	3,69	4,56	6,54	8,71	37,77

Table 43 Descriptive statistics of monthly purchase value, across customers.
Hypermarket 4000

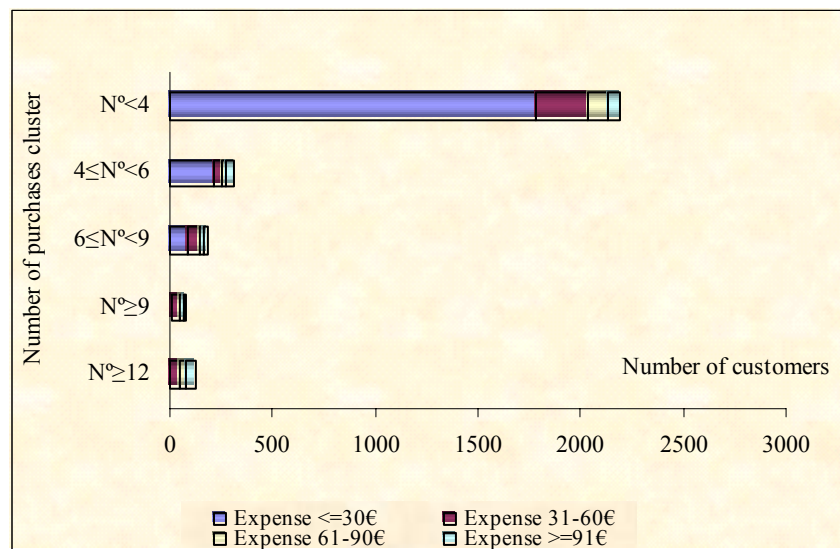
Month	Total	Min.	Máx.	Mean	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8	Decile9
January	5400,44	0,85	145,41	7,60	1,66	2,62	3,09	3,44	3,59	4,28	5,28	6,80	12,36
February	6353,32	0,70	118,29	8,89	2,07	2,80	3,29	3,54	3,90	4,57	6,16	7,99	20,50
March	6516,46	1,00	107,50	8,09	1,84	2,73	3,31	3,57	3,99	4,87	6,13	7,55	14,90
April	6700,02	0,63	114,48	10,42	1,96	2,92	3,47	3,76	4,24	5,02	6,79	9,25	26,83
May	6866,45	0,89	90,00	8,52	1,95	2,51	2,85	3,14	3,57	4,40	5,34	8,44	24,55
June	5073,41	0,67	152,16	7,48	2,10	2,69	2,96	3,28	3,58	4,08	5,21	7,32	12,79
July	6860,04	1,19	253,92	9,50	2,15	2,93	2,99	3,35	3,86	4,50	5,98	8,96	24,83
August	6057,91	1,22	126,21	7,52	2,23	3,00	3,34	3,60	4,01	4,50	5,08	6,82	12,26
September	5980,35	0,79	182,54	7,82	1,75	2,92	3,15	3,47	3,71	4,34	5,46	7,25	14,61
October	7480,30	0,78	95,00	8,84	2,05	2,78	3,07	3,45	3,58	4,30	5,61	7,31	27,87
November	6752,36	1,05	78,90	8,74	1,79	2,93	3,14	3,45	3,84	4,61	6,24	7,56	29,00
December	9576,96	0,85	117,00	11,23	1,94	2,81	3,05	3,45	3,81	4,95	6,63	9,90	38,99

Table 44 Descriptive statistics of monthly purchase value, across customers. Both hypermarkets

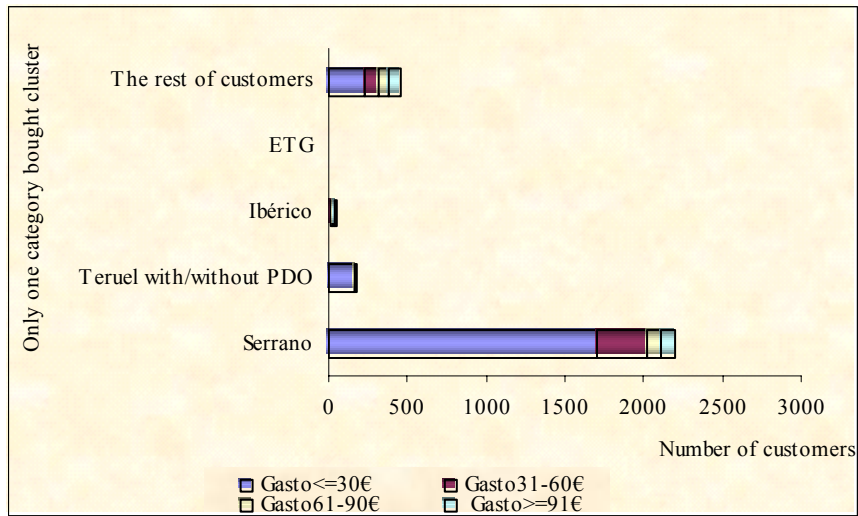
Month	Total	Mín.	Máx.	Mean	Decile1	Decile2	Decile3	Decile4	Decile5	Decile6	Decile7	Decile8	Decile9
January	10816,18	0,23	145,41	7,10	1,66	2,12	2,89	3,27	3,52	3,98	5,26	6,29	11,05
February	12457,22	0,38	118,29	8,50	1,75	2,79	3,24	3,51	3,86	4,53	6,11	7,62	16,65
March	13056,90	0,23	107,50	8,03	1,66	2,41	3,16	3,49	3,78	4,73	6,08	7,54	14,28
April	14700,49	0,47	144,02	10,08	1,78	2,64	3,24	3,60	4,02	4,88	6,55	8,20	26,49
May	13404,17	0,69	90,96	7,73	1,77	2,48	2,86	3,15	3,54	4,27	5,30	7,28	16,36
June	11160,99	0,39	152,16	7,12	1,84	2,53	2,83	3,16	3,50	4,05	5,05	6,62	10,84
July	14073,54	0,87	253,92	9,20	2,05	2,80	3,00	3,38	3,85	4,53	5,96	8,04	22,89
August	12425,24	0,38	225,37	7,53	1,99	2,82	3,26	3,51	3,88	4,50	5,27	7,00	12,63
September	13470,56	0,56	223,32	8,04	1,74	2,76	3,02	3,39	3,62	4,19	5,30	7,02	13,67
October	15860,31	0,49	209,35	8,82	1,82	2,61	3,02	3,41	3,57	4,17	5,32	7,15	22,44
November	14107,94	0,20	111,78	8,22	1,66	2,44	3,05	3,30	3,68	4,47	6,11	7,21	22,28
December	20892,48	0,14	252,01	11,29	1,66	2,46	2,99	3,37	3,76	4,74	6,62	9,12	38,89

In the next graphs, cross frequencies between alternative pairs of segmentation variables are shown, for each hypermarket.

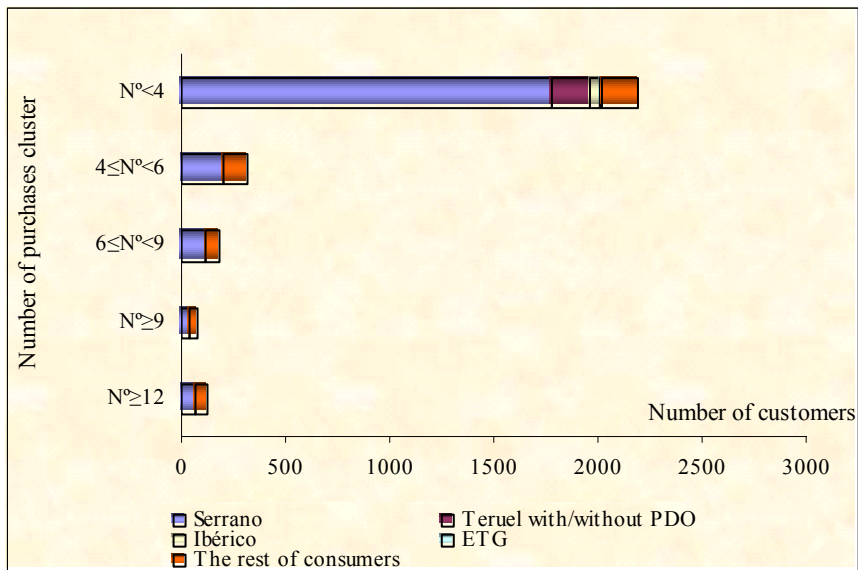
Graph 11 Cross-tabulation between the clusters defined with respect to purchase value and exclusive purchases of one ham category. Hypermarket 850.



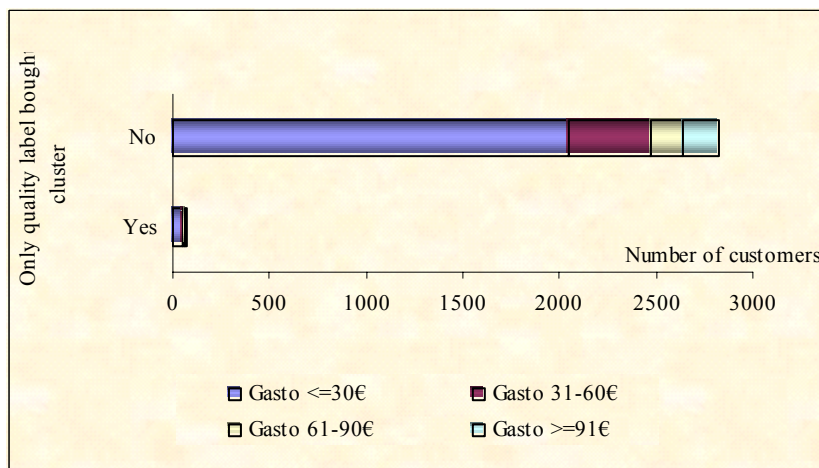
Graph 12 Cross-tabulation between the clusters defined with respect to number of purchases and exclusive purchases of one ham category. Hypermarket 850



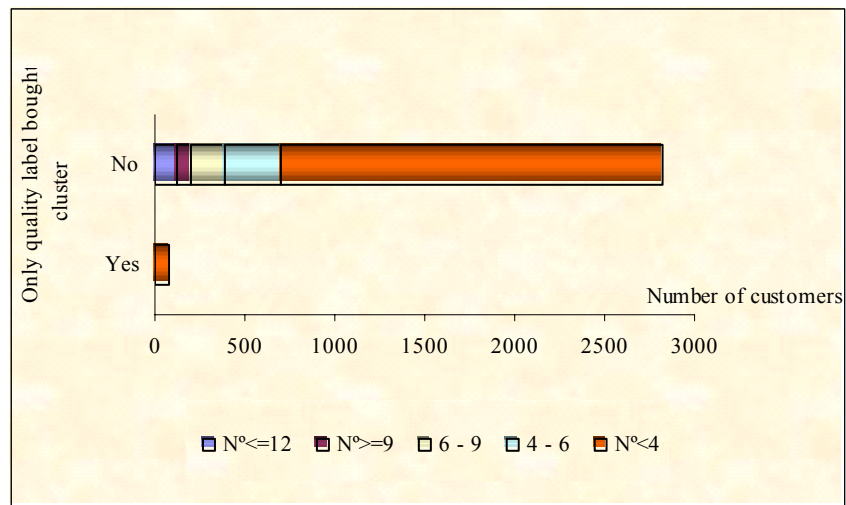
Graph 13 Cross-tabulation between the clusters defined with respect to number of purchases and exclusive purchases of one ham category. Hypermarket 850



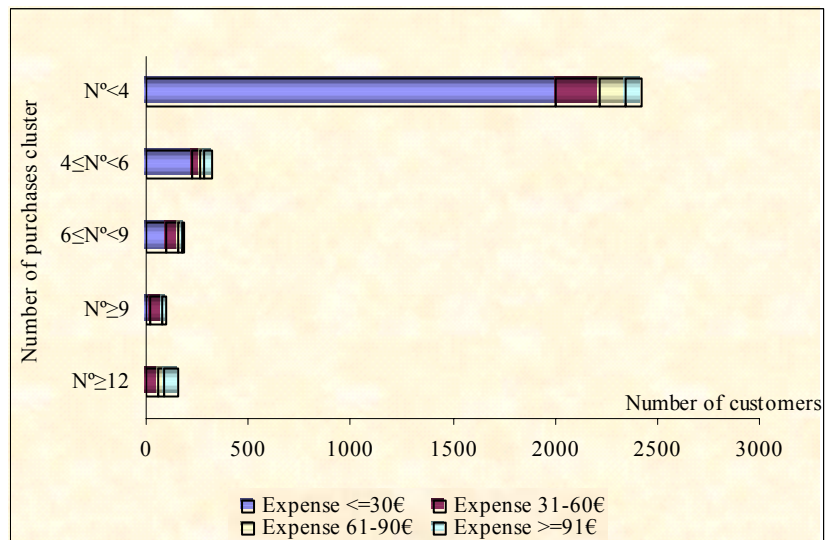
Graph 24 Cross tabulation between the clusters defined with respect to purchase value and exclusive purchases of quality labels. Hypermarket 850



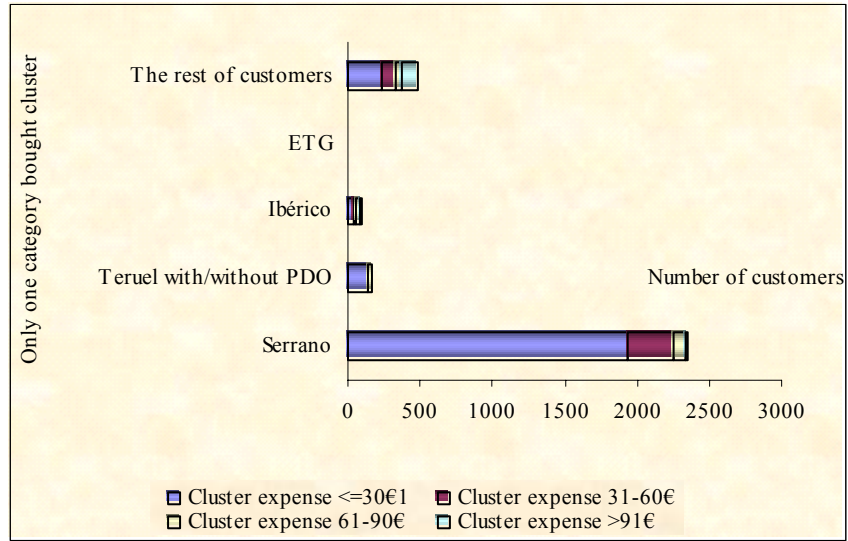
Graph 15 Cross-tabulation between the clusters defined with respect to number of purchases and exclusive purchases of quality labels. Hypermarket 850



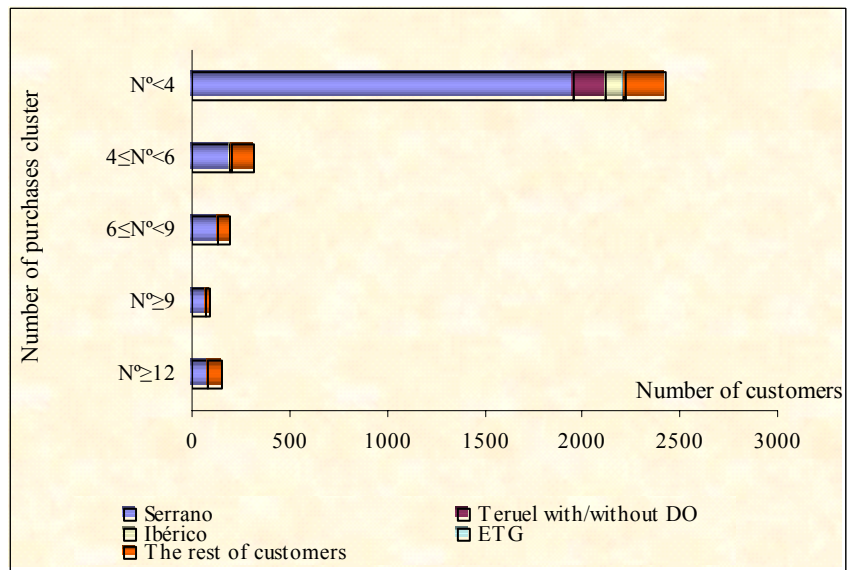
Graph 16 Cross-tabulation between the clusters defined with respect to purchase value and number of purchases. Hypermarket 4000



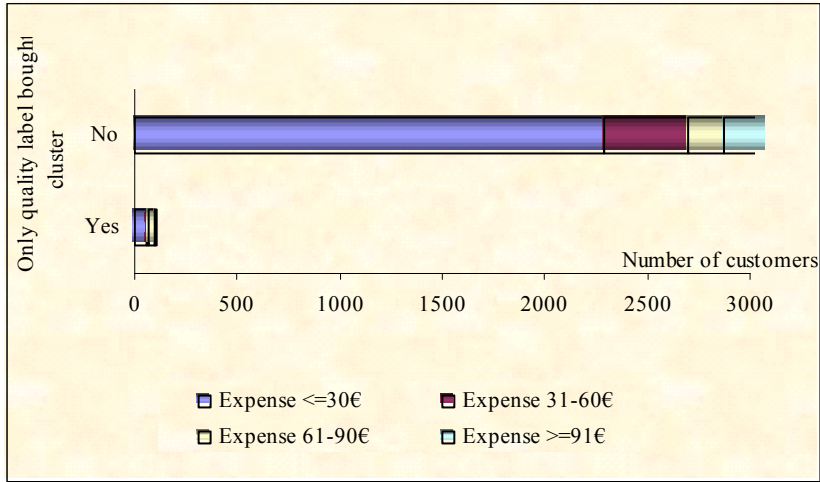
Graph 17 Cross-tabulation between the clusters defined with respect to purchase value and exclusive purchases of one category. Hypermarket 4000



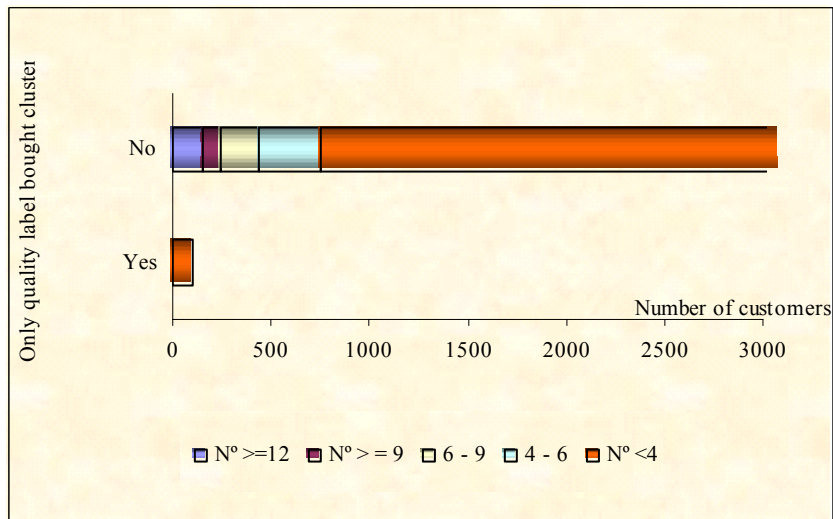
Graph 18 Cross-tabulation between the clusters defined with respect to number of purchases and exclusive purchases of one category. Supermarket 4000



Graph 19 Cross-tabulation between the clusters defined with respect to purchase value and exclusive purchases of quality labels. Supermarket 4000



Graph 20 Cross-tabulation between the clusters defined with respect to number of purchases and exclusive purchases of quality labels. Supermarket 4000



APPENDIX 2.

GAUSS Programmes:

LISTA*.GAU: lists of clients, articles,...

DESCRI*.GAU: Descriptive statistics

```
@-----@
LISTA0.GAU          21 julio 2004

Primeras pruebas con datos de scanner de carrefour
Datos definitivos, después primera depuración

Objetivo: obtener listado de clientes por hyper e identificar los presentes
en ambos hyper
-----@

new;
loadm dat[19599,19]=C:\ANA\typic\scanner\scan_num.txt;
dat = trimr(dat,1,0);
year = dat[.,1]; mes = dat[.,2]; centro=dat[.,3]; cliente=dat[.,4];
articulo =dat[.,5]; subfamil=dat[.,6]; gramunid=dat[.,7]; seccion=dat[.,8];
formato=dat[.,9];jampal=dat[.,10]; categori=dat[.,11]; sample=dat[.,12];
ventas=dat[.,13]; unidad=dat[.,14]; eurounid=dat[.,15]; precio=dat[.,16];
pata=dat[.,17]; mayor1=dat[.,18]; patas2 =dat[.,19];

/* listado de clientes por centro */
/* ACTUR */

cli_act = selif(cliente,centro.==850);
" numero de observaciones en Actur = " rows(cli_act);
cli_act = sortc(cli_act,1);
list_act = cli_act[1,];

i=2;
do while i<= rows(cli_act);
    if cli_act[i,1].ne cli_act[i-1,1]; list_act=list_act|cli_act[i,1]; endif;
i=i+1;
endo;

format /m1 /rd 2,0;
cactur = rows(list_act); " numero clientes Actur " cactur;

/* AUGUSTA */
cli_aug = selif(cliente,centro.==4000); " numero de observaciones en Augusta = " rows(cli_aug);
cli_aug = sortc(cli_aug,1);
list_aug = cli_aug[1,];

i=2;
do while i<= rows(cli_aug);
    if cli_aug[i,1].ne cli_aug[i-1,1]; list_aug=list_aug|cli_aug[i,1]; endif;
i=i+1;
endo;

format /m1 /rd 2,0;
```

```

caugusta = rows(list_aug); " numero clientes Augusta " caugusta;

list_both = zeros(1,1);
i=1;
do while i<= caugusta;
    j=1;
    do while j<= cactur;
        if list_aug[i,1] .== list_act[j,1];    list_both= list_both|list_aug[i,1]; endif;
        j=j+1;
    endo;
    i=i+1;
endo;

cboth = rows(list_both); "numero de clientes en ambos Hiper " cboth;

output file = c:\ana\typic\scanner\salidas\clientes reset;
screen off;
lista = ones(cactur,1).*850~list_act | ones(caugusta,1).*4000~list_aug| ones(cboth,1).*4850~list_both;
ncol = cols(lista);
mask = ones(1,ncol);
fmt = ("*.*1f"~20~0)+ZEROS(ncol,1);
CALL PRINTFM(lista,mask,fmt);

```

@-----
LISTA1.GAU **28 julio 2004**

Objetivo: características de los clientes: nº actos de compra, frecuencia de compra, nº categorías compradas,...

-----@

```

new;
/* leo datos totales, y ordeno por cliente */
loadm dat[19599,19]=C:\ANA\typic\scanner\scan_num.txt; dat = trimr(dat,1,0);
dat = sortc(dat,4);
year = dat[.,1]; mes = dat[.,2]; centro=dat[.,3]; cliente=dat[.,4];

```

```

@
articulo =dat[.,5];
subfamil=dat[.,6]; gramunid=dat[.,7]; seccion=dat[.,8];
formato=dat[.,9];jampal=dat[.,10]; categori=dat[.,11]; sample=dat[.,12];
ventas=dat[.,13]; unidad=dat[.,14]; eurounid=dat[.,15]; precio=dat[.,16];
pata=dat[.,17]; mayor1=dat[.,18]; patas2 =dat[.,19];
@

```

```

/* leo listado de clientes y los ordeno */
loadm dat1[6525, 2] = c:\ana\typic\scanner\clientes.txt; dat1 = trimr(dat1,1,0);
dat1 = sortc(dat1,2); hiper = dat1[.,1]; cllista = dat1[.,2];

```

```

/* leo listado de articulos: Seccion Subfamilia categoria      formato      articulo sample; y los
ordeno por categ y art. */

```

```

loadm dat2[95, 6] = c:\ana\typic\scanner\articulo.txt; dat2 = trimr(dat2,1,0);
let v = 3 5; dat2 = sortmc(dat2,v);
artlist = dat2[.,5]; catlist = dat2[.,3];

```

```

nclient = rows(dat1); nart = rows(dat2);
tot = zeros(nclient,nart*4+5);
compra = zeros(nclient,3+3*6);

```

```

i=1;
do while i<= rows(clista);
  dat0 = selif(dat,cliente.==clista[i,1]); /* selecciono 1 cliente */
  compra[i,1] = clista[i,1];
  compra[i,2] = sumc(dat0[:,13]); /* valor de compras del cliente i */
  compra[i,3] = rows(dat0); /* numero actos de compra del cliente i, a lo largo del
    año */
  j=1;
  do while j<= 6;
    dat01 = selif(dat0, dat0[:,11].==j); /* selecciono datos totales para cliente i,
      y categoria j */
    if scalmiss(dat01).==0;

      artic = selif(dat2,catlist.==j); /* selecciono categoria j en listado de
        articulos */
      compra[i,3*j+1] = j;
      compra[i,3*j+2] = sumc(dat01[:,13]); /* valor compras de categoria j */
      compra[i,3*j+3] = rows(dat01); /* numero de actos de compra del cliente i,
        de la categoria j */
    else;
      compra[i,3*j+1]=j;
      compra[i,3*j+2:3*j+3] = zeros(1,2);
    endif;
    j=j+1;
  endo;
i=i+1;
endo;

compra = sortc(compra,2);

format /m1 /rd 4,2;
compra1 = compra[:,1:10];
compra2 = compra[:,11:cols(compra)];

/* troceo en varios ficheros para poder leer bien la salida en excel */
output file = c:\ana\typic\scanner\salidas\compras1 reset;
screen off;
ncol = cols(compra1);
mask = ones(1,ncol);
let fmt = {
  ".*1f" 16 0, ".*1f" 4 2, ".*1f" 6 2,
  ".*1f" 4 0, ".*1f" 6 2, ".*1f" 4 0,
  ".*1f" 4 0, ".*1f" 6 2, ".*1f" 4 0,
  ".*1f" 4 0};
CALL PRINTFM(compra1,mask,fmt);

output file = c:\ana\typic\scanner\salidas\compras2 reset;
screen off;
ncol = cols(compra2);
mask = ones(1,ncol);

let fmt = { ".*1f" 6 2, ".*1f" 4 0,
  ".*1f" 4 0, ".*1f" 6 2, ".*1f" 4 0,
  ".*1f" 4 0, ".*1f" 6 2, ".*1f" 4 0,
  ".*1f" 4 0, ".*1f" 6 2, ".*1f" 4 0};
CALL PRINTFM(compra2,mask,fmt);

```


@-----
LISTA2.GAU

28 julio 2004

Objetivo: hacer un listado de los articulos

-----@
new;
loadm dat[19599,19]=C:\ANA\typic\scanner\scan_num.txt; dat = trimr(dat,1,0);
year = dat[:,1]; mes = dat[:,2]; centro=dat[:,3]; cliente=dat[:,4];
articulo =dat[:,5]; subfamil=dat[:,6]; gramunid=dat[:,7]; seccion=dat[:,8];
formato=dat[:,9];jampal=dat[:,10]; categori=dat[:,11]; sample=dat[:,12];
ventas=dat[:,13]; unidad=dat[:,14]; eurounid=dat[:,15]; precio=dat[:,16];
pata=dat[:,17]; mayor1=dat[:,18]; patas2 =dat[:,19];

designac = seccion~subfamil~categori~formato~articulo~sample;
/* " seccion subfamil categori formato articulo sample" */
v = { 1, 2, 3, 4, 5};
designac = sortmc(designac,v);
list_art = designac[1,:];

i=2;
do while i<= rows(designac);
 if designac[i,5].ne designac[i-1,5]; list_art=list_art|designac[i,:]; endif;
i=i+1;
endo;

format /m1 /rd 2,0;
nartic = rows(list_art);
" numero articulos " nartic;

output file = c:\ana\typic\scanner\salidas\artics reset;
screen off;
ncol = cols(list_art);
mask = ones(1,ncol);
fmt = ("*. *lf"~12~0)+ZEROS(ncol,1);
CALL PRINTFM(list_art,mask,fmt);

@-----
LISTA3.GAU

28 julio 2004

Objetivo: características de los clientes: nº actos de compra, frecuencia de compra, ...

Para los clientes de ambos centros, se calcula cuanto gastan en cada uno

-----@
new;
/* leo datos totales, y ordeno por cliente */

loadm dat[19599,19]=C:\ANA\typic\scanner\scan_num.txt; dat = trimr(dat,1,0);
dat = sortc(dat,4);
year = dat[:,1]; mes = dat[:,2]; centro=dat[:,3]; cliente=dat[:,4];

/* leo listado de clientes y los ordeno, 1º por centro, 2º por codigo cliente */

loadm dat1[6525, 2] = c:\ana\typic\scanner\clientes.txt; dat1 = trimr(dat1,1,0);

/* let v = 3 2; dat1 = sortmc(dat1,v); */

dat1 = sortc(dat1,2);

```

dat1 = selif(dat1,dat1[:,1].==4850); /* selecciono clientes que compran en los dos hiper */
clista = dat1[:,2];

/* leo listado de articulos: Seccion Subfamilia categoria      formato      articulo sample; y los
ordenado por categ y art. */

loadm dat2[95, 6] = c:\ana\typic\scanner\articulo.txt; dat2 = trimr(dat2,1,0);
let v = 3 5; dat2 = sortmc(dat2,v);
artlist = dat2[:,5]; catlist = dat2[:,3];

nclient = rows(dat1); nart = rows(dat2);
compra = zeros(nclient,3+3*6);

i=1;
do while i<= rows(clista);
    dat0 = selif(dat,centro.==850 .and cliente.==clista[i,1]); /* selecciono 1 cliente */
    compra[i,1] = clista[i,1];
    compra[i,2] = sumc(dat0[:,13]); /* valor de compras del cliente i */
    compra[i,3] = rows(dat0); /* numero actos de compra del cliente i, a lo largo del
                                año */
    j=1;
    do while j<= 6;
        dat01 = selif(dat0, dat0[:,11].==j); /* selecciono datos totales para cliente i,
                                                y categoria j */
        if scalmiss(dat01).==0;
            artic = selif(dat2,catlist.==j); /* selecciono categoria j en listado de
                                                articulos */
            compra[i,3*j+1] = j;
            compra[i,3*j+2] = sumc(dat01[:,13]); /* valor compras de categoria j */
            compra[i,3*j+3] = rows(dat01); /* numero de actos de compra del cliente i,
                                                de la categoria j */
        else;
            compra[i,3*j+1]=j;
            compra[i,3*j+2:3*j+3] = zeros(1,2);
        endif;
        j=j+1;
    endo;
i=i+1;
endo;

compra = sortc(compra,2);
compra1 = ones(rows(compra),1).*850~compra[:,1:10];
compra2 = compra[:,11:cols(compra)];

output file = c:\ana\typic\scanner\salidas\compras1 reset;
screen off;
ncol = cols(compra1);
mask = ones(1,ncol);
let fmt = {
    "* *1f" 4 0,
    "* *1f" 16 0, "* *1f" 6 2, "* *1f" 4 0,
    "* *1f" 3 0, "* *1f" 6 2, "* *1f" 4 0,
    "* *1f" 3 0, "* *1f" 6 2, "* *1f" 4 0,
    "* *1f" 3 0};
CALL PRINTFM(compra1,mask,fmt);

output file = c:\ana\typic\scanner\salidas\compras2 reset;
screen off;
ncol = cols(compra2);
mask = ones(1,ncol);

```

```

let fmt = { "%.1f" 6 2, "%.1f" 4 0,
            "%.1f" 3 0, "%.1f" 6 2, "%.1f" 4 0,
            "%.1f" 3 0, "%.1f" 6 2, "%.1f" 4 0,
            "%.1f" 3 0, "%.1f" 6 2, "%.1f" 4 0 } ;
CALL PRINTFM(compra2,mask,fmt);

/* repito ordenes, para obtener los datos de las clientes de ambos super, pero referidos a Augusta */
compra = zeros(nclient,3+3*6);

i=1;
do while i<= rows(clista);
    dat0 = selif(dat,centro.==4000 .and cliente.==clista[i,1]); /* selecciono 1 cliente */
    compra[i,1] = clista[i,1];
    compra[i,2]= sumc(dat0[:,13]); /* valor de compras del cliente i */
    compra[i,3]=rows(dat0); /* numero actos de compra del cliente i, a lo largo del
                            año */
    j=1;
    do while j<= 6;
        dat01 = selif(dat0, dat0[:,11].==j); /* selecciono datos totales para cliente i,
                                                y categoria j */
        if scalmiss(dat01).==0;
            artic = selif(dat2,catlist.==j); /* selecciono categoria j en listado de
                                                articulos */

            compra[i,3*j+1] = j;
            compra[i,3*j+2] = sumc(dat01[:,13]); /* valor compras de categoria j */
            compra[i,3*j+3] = rows(dat01); /* numero de actos de compra del
                                                cliente i, de la categoria j */

        else;
            compra[i,3*j+1]=j;
            compra[i,3*j+2:3*j+3] = zeros(1,2);
        endif;
        j=j+1;
    endo;
i=i+1;
endo;

compra = sortc(compra,2);
compra11 = ones(rows(compra),1).*4000~compra[:,1:10];
compra21= compra[:,11:cols(compra)];

output file = c:\ana\typic\scanner\salidas\comp1 reset;
screen off;
ncol = cols(compra11);
mask = ones(1,ncol);
let fmt = {
            "%.1f" 4 0,
            "%.1f" 16 0, "%.1f" 6 2, "%.1f" 4 0,
            "%.1f" 3 0, "%.1f" 6 2, "%.1f" 4 0,
            "%.1f" 3 0, "%.1f" 6 2, "%.1f" 4 0,
            "%.1f" 3 0};
CALL PRINTFM(compra11,mask,fmt);

output file = c:\ana\typic\scanner\salidas\comp2 reset;
screen off;
ncol = cols(compra21);
mask = ones(1,ncol);

let fmt = { "%.1f" 6 2, "%.1f" 4 0,

```

```

"*.*f" 3 0, "*.*f" 6 2, "*.*f" 4 0,
"*.*f" 3 0, "*.*f" 6 2, "*.*f" 4 0,
"*.*f" 3 0, "*.*f" 6 2, "*.*f" 4 0} ;
CALL PRINTFM(compra21,mask,fmt);

```

@-----
LISTA4.GAU 28 julio 2004

Objetivo: comprobacion de que el fichero compras.txt es correcto;
para los clientes de ambos hiper, la cifra de totales aparece solo cuando hiper=4850.

```

-----@
new;
/* leo datos totales, y ordeno por cliente */
/* Primera comprobacion sobre : loadm dat[6528,22]=C:\ANA\typic\scanner\salidas\compras.txt; */
/* segunda comprobacion */

loadm dat[6525,22]=C:\ANA\typic\scanner\salidas\compras.txt;
/* orden variables: Cliente HYPERvalor compras actos compra
Categoría 1 "Serrano" Valor compras en categoría j Actos compra en categoría j
Categoría 2 "Teruel DO" Valor compras en categoría j Actos compra en categoría j
Categoría 3 "Teruel noDO" Valor compras en categoría j Actos compra en categoría j
Categoría 4 "IbnoDO" Valor compras en categoría j Actos compra en categoría j
Categoría 5 "ETG" Valor compras en categoría j Actos compra en categoría j
Categoría 6 "Pal_IbnoDO" Valor compras en categoría j Actos compra en categoría j */

cliente = dat[.,1]; hyper = dat[.,2]; valtot = dat[.,3];
clidoble = selif(cliente,hyper.==4850);

format /m1 /rd 20,0;
i=1;
do while i<=rows(clidoble);
    client_i = selif(cliente, cliente.==clidoble[i,.]);
    if rows(client_i).>=4; "comprobar cliente i = " clidoble[i,1];
    endif;
i=i+1;

endo;

```

@-----
 DESCRIO.GAU 3 Agosto 2004
 Objetivo: Estadísticas descriptivas para los clientes de cada hyper, y ambos,
 por categorías de jamón, sobre valor de compras
 -----@

```

new;
/* leo datos totales, y ordeno por cliente */

loadm dat[6525,22]=C:\ana\typic\scanner\salidas\compras.txt;
/* orden variables: Cliente HYPERvalor compras actos compra
Categoría 1 "Serrano" Valor compras en categoría j Actos compra en categoría j
Categoría 2 "Teruel DO" Valor compras en categoría j Actos compra en categoría j
Categoría 3 "Teruel noDO" Valor compras en categoría j Actos compra en categoría j
Categoría 4 "IbnoDO" Valor compras en categoría j Actos compra en categoría j
Categoría 5 "ETG" Valor compras en categoría j Actos compra en categoría j
Categoría 6 "Pal_IbnoDO" Valor compras en categoría j Actos compra en categoría j */

hyper = dat[:,2];

/* cifra de compras, por hyper: valor total, min, max, media */

e = seqa(0.10,0.10,9); /* e = {0.025, 0.5, 0.975}; */

descri = zeros(4,4+rows(e));
valor = selif(dat[:,3],hyper.==850);
descri[1,1] = sumc(valor);descri[1,2] = minc(valor);descri[1,3] = maxc(valor);descri[1,4] = meanc(valor);
if rows(valor).*minc(e).<1 .or rows(valor).*maxc(e).>rows(valor)-1;
descri[1,5:4+rows(e)] = ones(1,rows(e)).*999999;
else;
descri[1,5:4+rows(e)] = quantile(valor,e);
endif;

valor = selif(dat[:,3],hyper.==4000);
descri[2,1] = sumc(valor);descri[2,2] = minc(valor);descri[2,3] = maxc(valor);descri[2,4] = meanc(valor);
if rows(valor).*minc(e).<1 .or rows(valor).*maxc(e).>rows(valor)-1;
descri[2,5:4+rows(e)] = ones(1,rows(e)).*999999;
else;
descri[2,5:4+rows(e)] = quantile(valor,e);
endif;

valor = selif(dat[:,3],hyper.==4850);
descri[3,1] = sumc(valor);descri[3,2] = minc(valor);descri[3,3] = maxc(valor);descri[3,4] = meanc(valor);
if rows(valor).*minc(e).<1 .or rows(valor).*maxc(e).>rows(valor)-1;
descri[3,5:4+rows(e)] = ones(1,rows(e)).*999999;
else;
descri[3,5:4+rows(e)] = quantile(valor,e);
endif;

valor = selif(dat[:,3],hyper.==850 .or hyper.==4000);
descri[4,1] = sumc(valor); descri[4,2] = minc(valor); descri[4,3] = maxc(valor);descri[4,4] =
meanc(valor);
if rows(valor).*minc(e).<1 .or rows(valor).*maxc(e).>rows(valor)-1;
descri[4,5:4+rows(e)] = ones(1,rows(e)).*999999;
else;
descri[4,5:4+rows(e)] = quantile(valor,e);
endif;
format /m1 /rd 6,2;
let titc = 850 4000 4850;

" valor de compras para el total de jamon

```

```

hyper total  min max media  1 2 3 4 5 6 7 8 9 10
-----
";
tite~descri[1:3,.];
"total  " descri[4,.];

categ = dat[.,5 8 11 14 17 20];
compras = dat[.,6 9 12 15 18 21];

i=1;
do while i<=6;
  descri = zeros(4,4+rows(e));
  valor = selif(compras[.,i], hyper.==850 .and compras[.,i]>0);
  descri[1,1] = sumc(valor);      descri[1,2] = minc(valor); descri[1,3] = maxc(valor);
  descri[1,4] = meanc(valor);

  if rows(valor).*minc(e)<1 .or rows(valor).*maxc(e)>rows(valor)-1;
    descri[1,5:4+rows(e)] = ones(1,rows(e)).*999999;
  else;
    descri[1,5:4+rows(e)] = quantile(valor,e)';
  endif;

  valor = selif(compras[.,i],hyper.==4000 .and compras[.,i]>0); descri[2,1] = sumc(valor);
  descri[2,2] = minc(valor); descri[2,3] = maxc(valor); descri[2,4] = meanc(valor);

  if rows(valor).*minc(e)<1 .or rows(valor).*maxc(e)>rows(valor)-1;
    descri[2,5:4+rows(e)] = ones(1,rows(e)).*999999;
  else;
    descri[2,5:4+rows(e)] = quantile(valor,e)';
  endif;

  valor = selif(compras[.,i],hyper.==4850 .and compras[.,i]>0); descri[3,1] = sumc(valor);
  descri[3,2] = minc(valor); descri[3,3] = maxc(valor); descri[3,4] = meanc(valor);

  if rows(valor).*minc(e)<1 .or rows(valor).*maxc(e)>rows(valor)-1;
    descri[3,5:4+rows(e)] = ones(1,rows(e)).*999999;
  else;
    descri[3,5:4+rows(e)] = quantile(valor,e)';
  endif;

  valor = selif(compras[.,i],hyper.==850 .and compras[.,i]>0 .or hyper.==4000 .and
compras[.,i]>0 );
  descri[4,1] = sumc(valor); descri[4,2] = minc(valor); descri[4,3] = maxc(valor); descri[4,4] =
meanc(valor);

  if rows(valor).*minc(e)<1 .or rows(valor).*maxc(e)>rows(valor)-1;
    descri[4,5:4+rows(e)] = ones(1,rows(e)).*999999;
  else;
    descri[4,5:4+rows(e)] = quantile(valor,e)';
  endif;

  format /m1 /rd 6,2;
  let tite = 850 4000 4850;
  " valor de compras para la categoria i = " i;
  " hyper total  min max media  1 2 3 4 5 6 7 8 9 10
-----
"; tite~descri[1:3,.]; "total  " descri[4,.];
i=i+1;
endo;

```

@-----
 DESCR101.GAU 4 Agosto 2004
 para los clientes de ambos hiper, la cifra de totales aparece solo cuando hiper=4850.
 Objetivo: Estadísticas descriptivas de los actos de compra

```

-----@
new;

loadm dat[6525,22]=C:\ana\typic\scanner\salidas\compras.txt;

hyper = dat[:,2];

/* numero de compras, por hiper: valor total, min, max, media */
e = seqa(0.10,0.10,9); /* e = {0.025, 0.5, 0.975}; */

descri = zeros(4,4+rows(e));
valor = selif(dat[:,4],hyper.==850);
descri[1,1] = sumc(valor);descri[1,2] = minc(valor);descri[1,3] = maxc(valor);descri[1,4] = meanc(valor);
    if rows(valor).*minc(e).<1 .or rows(valor).*maxc(e).>rows(valor)-1;
        descri[1,5:4+rows(e)] = ones(1,rows(e)).*999999;
    else;
        descri[1,5:4+rows(e)] = quantile(valor,e)';
    endif;

valor = selif(dat[:,4],hyper.==4000);
descri[2,1] = sumc(valor);descri[2,2] = minc(valor);descri[2,3] = maxc(valor);descri[2,4] = meanc(valor);
    if rows(valor).*minc(e).<1 .or rows(valor).*maxc(e).>rows(valor)-1;
        descri[2,5:4+rows(e)] = ones(1,rows(e)).*999999;
    else;
        descri[2,5:4+rows(e)] = quantile(valor,e)';
    endif;

valor = selif(dat[:,4],hyper.==4850);
descri[3,1] = sumc(valor);descri[3,2] = minc(valor);descri[3,3] = maxc(valor);descri[3,4] = meanc(valor);
    if rows(valor).*minc(e).<1 .or rows(valor).*maxc(e).>rows(valor)-1;
        descri[3,5:4+rows(e)] = ones(1,rows(e)).*999999;
    else;
        descri[3,5:4+rows(e)] = quantile(valor,e)';
    endif;

valor = selif(dat[:,4],hyper.==850 .or hyper.==4000);
descri[4,1] = sumc(valor); descri[4,2] = minc(valor); descri[4,3] = maxc(valor);descri[4,4] =
meanc(valor);
    if rows(valor).*minc(e).<1 .or rows(valor).*maxc(e).>rows(valor)-1;
        descri[4,5:4+rows(e)] = ones(1,rows(e)).*999999;
    else;
        descri[4,5:4+rows(e)] = quantile(valor,e)';
    endif;
format /m1 /rd 6,2;
let titc = 850 4000 4850;

" valor de compras para el total de jamon
hyper total  min  max  media  1 2 3 4 5 6 7 8 9 10
-----
";
titc~descri[1:3,:];
"total  " descri[4,:];

/* estadísticas descriptivas por categorías */
categ = dat[:,5 8 11 14 17 20];

```

```

compras = dat[.,6 9 12 15 18 21];
frec = dat[.,7 10 13 16 19 22]; /* frecuencia de compras */

i=1;
do while i<=6;
    descri = zeros(4,4+rows(e));
    valor = selif(frec[.,i], hyper.==850 .and frec[.,i]>0);
    descri[1,1] = sumc(valor); descri[1,2] = minc(valor); descri[1,3] = maxc(valor);
    descri[1,4] = meanc(valor);

    if rows(valor).*minc(e)<1 .or rows(valor).*maxc(e)>rows(valor)-1;
        descri[1,5:4+rows(e)] = ones(1,rows(e)).*999999;
    else;
        descri[1,5:4+rows(e)] = quantile(valor,e)';
    endif;

    valor = selif(frec[.,i],hyper.==4000 .and frec[.,i]>0);
    descri[2,1] = sumc(valor); descri[2,2] = minc(valor); descri[2,3] = maxc(valor);
    descri[2,4] = meanc(valor);

    if rows(valor).*minc(e)<1 .or rows(valor).*maxc(e)>rows(valor)-1;
        descri[2,5:4+rows(e)] = ones(1,rows(e)).*999999;
    else;
        descri[2,5:4+rows(e)] = quantile(valor,e)';
    endif;

    valor = selif(frec[.,i],hyper.==4850 .and frec[.,i]>0);
    descri[3,1] = sumc(valor); descri[3,2] = minc(valor); descri[3,3] = maxc(valor);
    descri[3,4] = meanc(valor);

    if rows(valor).*minc(e)<1 .or rows(valor).*maxc(e)>rows(valor)-1;
        descri[3,5:4+rows(e)] = ones(1,rows(e)).*999999;
    else;
        descri[3,5:4+rows(e)] = quantile(valor,e)';
    endif;

    valor = selif(frec[.,i],hyper.==850 .and frec[.,i]>0 .or hyper.==4000 .and frec[.,i]>0 );
    descri[4,1] = sumc(valor); descri[4,2] = minc(valor); descri[4,3] = maxc(valor);
    descri[4,4] = meanc(valor);

    if rows(valor).*minc(e)<1 .or rows(valor).*maxc(e)>rows(valor)-1;
        descri[4,5:4+rows(e)] = ones(1,rows(e)).*999999;
    else;
        descri[4,5:4+rows(e)] = quantile(valor,e)';
    endif;

    format /m1 /rd 6,2;
    let titc = 850 4000 4850;

    " frecuencia de compras para la categoria i = " i;
    " hyper total   min max media   1 2 3 4 5 6 7 8 9 10

    " ;
    titc~descri[1:3,.];
    "total " descri[4,.];

    i=i+1;
endo;

```


Objetivo: histogramas: frecuencia y valor de compras.

INPUT: compras.txt (incluye informacion sobre compras, para cada cliente, agregadas por categoria para el total del periodo)

-----@

```
new;
library pgraph;
xlabel (" purchase (euro) along the period ");
ylabel (" number of customers ");

loadm dat[6525,22]=C:\ana\typic\scanner\salidas\compras.txt;
hyper = dat[.,2];

/* valor de compras total, todas categorias */
e = seqa(6,6,12); /* intervalos de valor de compras para los que computar el numero de clientes */
descri = zeros(rows(e),5);
des_compl = zeros(1,5); /* ultimo intervalo, para valor > el ultimo valor en e */
nclient = zeros(1,5);
descri[.,1]=e;

title("Total ham: Actur ");
valor = selif(dat[.,3],hyper.==850);
f = counts(valor,e);
histf(f,e);
descri[.,2]=f;
des_compl[.,2]= rows(valor)-sumc(f);
nclient[.,2]=rows(valor);

title("Total ham: Augusta ");
valor = selif(dat[.,3],hyper.==4000);
f = counts(valor,e);
histf(f,e);
descri[.,3]=f;
des_compl[.,3]= rows(valor)-sumc(f);
nclient[.,3]=rows(valor);

title ("Total ham: Customers of Actur and Augusta ");
valor = selif(dat[.,3],hyper.==4850); f = counts(valor,e); histf(f,e); descri[.,4] = f; des_compl[.,4]=
rows(valor)-sumc(f); nclient[.,4]=rows(valor);

title ("Total ham: Actur and Augusta ");
valor = selif(dat[.,3],hyper.==850 .or hyper.==4000);
f = counts(valor,e); histf(f,e); descri[.,5] = f; des_compl[.,5]= rows(valor)-sumc(f);
nclient[.,5]=rows(valor);

format /m1 /rd 4,0;
" intervalo   Actur   Augusta   Ambos   Total
-----";
descri;
des_compl;
" total number of customers " nclient;
```

Objetivo: histogramas: frecuencia de compra (mismo programa que descri1.gau, diferente variable).
INPUT: compras.txt

-----@

```
new;
library pgraph;
xlabel (" number of purchases ");
ylabel (" number of customers ");

loadm dat[6525,22]=C:\ana\typic\scanner\salidas\compras.txt;

hyper = dat[.,2];

/* numero de actos de compra, todas categorias */
e = seqa(1,1,20); /* intervalos de veces que se acude al hiper a comprar jamon para los que computar el
numero de clientes */
descri = zeros(rows(e),5);
des_compl = zeros(1,5); /* ultimo intervalo, para valor > el ultimo valor en e */
nclient = zeros(1,5);
descri[.,1]=e;

title("Total ham: Actur ");
valor = selif(dat[.,4],hyper.==850);
f = counts(valor,e);
histf(f,e);
descri[.,2]=f;
des_compl[.,2]= rows(valor)-sumc(f);
nclient[.,2]=rows(valor);

title("Total ham: Augusta ");
valor = selif(dat[.,4],hyper.==4000);
f = counts(valor,e);
histf(f,e);
descri[.,3]=f;
des_compl[.,3]= rows(valor)-sumc(f);
nclient[.,3]=rows(valor);

title ("Total ham: Customers of Actur and Augusta ");
valor = selif(dat[.,4],hyper.==4850);
f = counts(valor,e);
histf(f,e);
descri[.,4] = f;
des_compl[.,4]= rows(valor)-sumc(f);
nclient[.,4]=rows(valor);

title ("Total ham: Actur and Augusta ");
valor = selif(dat[.,4],hyper.==850 .or hyper.==4000); f = counts(valor,e); histf(f,e); descri[.,5] = f;
des_compl[.,5]= rows(valor)-sumc(f); nclient[.,5]=rows(valor);

format /m1 /rd 4,0;
" n° purchases          Actur   Augusta   Ambos   Total
-----"
descri;
des_compl; " total number of customers " nclient;
```

Objetivo: Numero de clientes por categoria

INPUT: compras.txt

-----@

```
new;
library pgraph;
loadm dat[6525,22]=C:\ana\typic\scanner\salidas\compras.txt;

hyper = dat[:,2];
nclient = zeros(3,8);
let tit = {850, 4000, 4850};
nclient[1:3,1]= tit;
let vcompra = {3,6, 9, 12, 15, 18, 21};
let ncompra = {4,7, 10, 13, 16, 19,22};

i = 1;
do while i<=3;
    tit1 = tit[i,:];
    j = 1;
    do while j<=7;
        tit2 = vcompra[j,:];
        tit3 = ncompra[j,:];
        compra = selif(dat,hyper.==tit1 .and (dat[:,tit2].>0 .and dat[:,tit3].>0));
        nclient[i,j+1]=rows(compra);
        j=j+1;
    endo;
    i=i+1;
endo;

format /m1 /rd 4, 0;
nclient;
sumc(nclient[1:2,:]);

/* valor de compras total, todas categorias */
e = seqa(6,6,12); /* intervalos de valor de compras para los que computar el numero de clientes */
descri = zeros(rows(e),5);
des_compl = zeros(1,5); /* ultimo intervalo, para valor > el ultimo valor en e */
nclient = zeros(1,5);
descri[:,1]=e;

title("Total ham: Actur ");
valor = selif(dat[:,3],hyper.==850);
f = counts(valor,e);
histf(f,e);
descri[:,2]=f;
des_compl[:,2]= rows(valor)-sumc(f);
nclient[:,2]=rows(valor);

title("Total ham: Augusta ");
valor = selif(dat[:,3],hyper.==4000);
f = counts(valor,e);
histf(f,e);
descri[:,3]=f;
des_compl[:,3]= rows(valor)-sumc(f);
nclient[:,3]=rows(valor);
```

```

title ("Total ham: Customers of Actur and Augusta ");
valor = selif(dat[:,3],hyper.==4850); f = counts(valor,e); histf(f,e); descri[:,4] = f;
des_compl[:,4]= rows(valor)-sumc(f); nclient[:,4]=rows(valor);

```

```

title ("Total ham: Actur and Augusta ");
valor = selif(dat[:,3],hyper.==850 .or hyper.==4000); f = counts(valor,e); histf(f,e);
descri[:,5] = f; des_compl[:,5]= rows(valor)-sumc(f); nclient[:,5]=rows(valor);

```

```

format /m1 /rd 4,0;
" intervalo   Actur   Augusta   Ambos   Total
-----"
descri;
des_compl;
" total number of customers " nclient;

```

@-----
DESCR13.GAU 5 agosto 2003

Objetivo: Numero de categorias compradas por los clientes;
clientes que solo compran una categoria y cual es

INPUT: compras.txt

-----@

```

new;
library pgraph;

loadm dat[6525,22]=C:\gauss6.0\compras\compras.txt;
hyper = dat[:,2];
let vcompra = {6, 9, 12, 15, 18, 21};
let ncompra = {7, 10, 13, 16, 19,22};

/* creo fichero de datos que contiene la identificacion del individuo: codigo cliente e hyper; y 6 variables
dummy,
una por cada categoria, que indica si el cliente compra (1) o no (0) dicha categoria; tiene el mismo
numero de filas que el fichero compras */

ident = dat[:,1:2];
dum = zeros(rows(dat),6);
i=1;
do while i<=rows(dat);
    j = 1;
    do while j<=6;
        tit2 = vcompra[j,.]; tit3 = ncompra[j,.];
        dum[i,j] = dat[i,tit2]>0 .and dat[i,tit3]>0; /* doble comprobacion, con valor y con actos
compra */
    de
        j=j+1;
    endo;
    i=i+1;
endo;

dum1 = sumc(dum'); /* las suma indica el numero de categorias compradas */

datos = ident~dum~dum1;
output file = c:\gauss6.0\scanner\category reset;
screen off;
ncol = cols(datos);
mask = ones(1,ncol);

```

```

fmt = {"*.*1f" 12 0,      "*.*1f" 6 0,      "*.*1f" 3 0,      "*.*1f" 3 0,      "*.*1f" 3 0,      "*.*1f"
3 0,
      "*.*1f" 3 0,      "*.*1f" 3 0,      "*.*1f" 3 0};
call printfm(datos,mask,fmt);

/* clientes exclusivos de una categoría */
" Supermercado 850: Actur ";
dat1 = selif(dat,hyper.==850);

dum = zeros(rows(dat1),6);
i=1;
do while i<=rows(dat1);
    j = 1;
    do while j<=6;
        tit2 = vcompra[j,.]; tit3 = ncompra[j,.];
        dum[i,j] = dat1[i,tit2]>0 .and dat1[i,tit3]>0;
        j=j+1;
    endo;
i=i+1;
endo;

dum1 = sumc(dum');
j=1;
do while j<=6;
    dum2 = selif(dum, dum[.,j].==1 .and dum1.==1);
    if scalmiss(dum2).==0; /* no hay nulos */
        " numero clientes exclusivos de categoria j= " J;
        rows(dum2);
    else;
        " no hay clientes exclusivos de la categoria j= " J;
    endif;
j=j+1;
endo;

" Supermercado 4000: Augusta ";
dat1 = selif(dat,hyper.==4000);

dum = zeros(rows(dat1),6);
i=1;
do while i<=rows(dat1);
    j = 1;
    do while j<=6;
        tit2 = vcompra[j,.]; tit3 = ncompra[j,.];
        dum[i,j] = dat1[i,tit2]>0 .and dat1[i,tit3]>0;
        j=j+1;
    endo;
i=i+1;
endo;

dum1 = sumc(dum');
j=1;
do while j<=6;
    dum2 = selif(dum, dum[.,j].==1 .and dum1.==1);
    if scalmiss(dum2).==0; /* no hay nulos */
        " numero clientes exclusivos de categoria j= " J;
        rows(dum2);
    else;
        " no hay clientes exclusivos de la categoria j= " J;
    endif;
j=j+1;

```

```
endo;
```

```
" Supermercado 4850: clientes de Actur y Augusta ";  
dat1 = selif(dat,hyper.==4850);
```

```
dum = zeros(rows(dat1),6);  
i=1;  
do while i<=rows(dat1);  
    j = 1;  
    do while j<=6;  
        tit2 = vcompra[j,]; tit3 = ncompra[j,];  
        dum[i,j] = dat1[i,tit2]>0 .and dat1[i,tit3]>0;  
        j=j+1;  
    endo;  
    i=i+1;  
endo;
```

```
dum1 = sumc(dum');  
j=1;  
do while j<=6;  
    dum2 = selif(dum, dum[:,j].==1 .and dum1.==1);  
    if scalmiss(dum2).==0; /* no hay nulos */  
        " numero clientes exclusivos de categoria j= " J;  
        rows(dum2);  
    else;  
        " no hay clientes exclusivos de la categoria j= " J;  
    endif;  
    j=j+1;  
endo;
```

```
" Total ";  
dat1 = selif(dat,hyper.==850 .or hyper.==4000);
```

```
dum = zeros(rows(dat1),6);  
i=1;  
do while i<=rows(dat1);  
    j = 1;  
    do while j<=6;  
        tit2 = vcompra[j,]; tit3 = ncompra[j,];  
        dum[i,j] = dat1[i,tit2]>0 .and dat1[i,tit3]>0;  
        j=j+1;  
    endo;  
    i=i+1;  
endo;
```

```
dum1 = sumc(dum');  
j=1;  
do while j<=6;  
    dum2 = selif(dum, dum[:,j].==1 .and dum1.==1);  
    if scalmiss(dum2).==0; /* no hay nulos */  
        " numero clientes exclusivos de categoria j= " J;  
        rows(dum2);  
    else;  
        " no hay clientes exclusivos de la categoria j= " J;  
    endif;  
    j=j+1;  
endo;
```

Objetivo: Descripción del número de categorías compradas, e identificación.
Qué categorías compran los clientes que compran 2?

INPUT: category.txt

/* orden variables: Cliente HYPER compra_cat1 compra_cat2 compra_cat3
compra_cat4 compra_cat5 compra_cat6 num_cat*/

-----@
new;

```
loadm category[6525,9] = c:\gauss6.0\compras\categorym.txt;  
hyper = category[.,2]; numcat = category[.,9];  
tit = {850, 4000, 4850};
```

```
/* numero de clientes que compran m categorías, m= 1 hasta 6 */
```

```
k=1;  
do while k<=3;  
    j=tit[k,.]; " hypermercado " j;  
    m = 1;  
    do while m<=6;  
        dat = selif(category,hyper.==j .and numcat.==m);  
        if scalmiss(dat).==0;  
            " numero de clientes que compran " m " categorías " rows(dat);  
        else;  
            " 0 clientes compran " m " categorías";  
        endif;  
        m=m+1;  
    endo;  
k=k+1;  
endo;
```

```
/* qué categorías compran, los clientes que compran 2 categorías */
```

```
k=1;  
do while k<=3;  
    j=tit[k,.]; " hypermercado " j;  
    dat = selif(category, hyper.==j .and numcat.==2);  
    dat1 = dat[.,3:8];  
    m=1;  
    do while m<=5;  
        w=m+1;  
        do while w<=6;  
            "numero de clientes que compran categoría " m "y " w "es " ;  
            dat2 = selif(dat1,dat1[.,m].==1 .and dat1[.,w].==1);  
            if scalmiss(dat2).==0; rows(Dat2);  
            else; " 0 " ;  
        endif;  
        w=w+1;  
    endo;  
    m=m+1;  
    endo;  
k=k+1;  
endo;
```

@-----@
DESCR132.GAU 9 agosto 2003

Objetivo: Descripcion del numero de articulos comprados dentro de cada categoria.

INPUT: clientes.txt ; scan_num.txt
/* orden variables: Cliente_HYPER compra_cat1 compra_cat2 compra_cat3
compra_cat4 compra_cat5 compra_cat6 num_cat

```
-----@
new;
loadm dat[19599,19]=C:\gauss6.0\compras\scan_num.txt; dat = trimr(dat,1,0);
year = dat[.,1]; mes = dat[.,2]; hyper=dat[.,3]; cliente=dat[.,4];
articulo =dat[.,5]; subfamil=dat[.,6]; gramunid=dat[.,7]; seccion=dat[.,8];
formato=dat[.,9];jampal=dat[.,10]; categori=dat[.,11]; sample=dat[.,12];
ventas=dat[.,13]; unidad=dat[.,14]; eurounid=dat[.,15]; precio=dat[.,16];
pata=dat[.,17]; mayor1=dat[.,18]; patas2 =dat[.,19];

loadm artic[93,6]=c:\gauss6.0\compras\articulo.txt; artic = trimr(artic,1,0);
/* orden: seccion, familia, categoria, formato, articulo, sample */
artic = sortc(artic,3);
artic1 = zeros(rows(artic)*2,2);
i=1;
do while i<=rows(artic);
    artic1[2*i-1:2*i,1] = artic[i,3]|artic[i,3];
    artic1[2*i-1:2*i,2] = artic[i,5]|artic[i,5];
i=i+1;
endo;

loadm category[6525,9]=c:\gauss6.0\compras\category.txt; category =
trimr(category,1,0);
/* orden: cliente, hyper, compra1, compra2, compra3, compra4, compra5, compra6,
numcat */

m = 4850;
category = selif(category,category[.,2].==m);
category = sortc(category,1);

ident = zeros(rows(category),2);
salida = zeros(rows(category),rows(artic)*2);

i=1;
do while i<=rows(category);
    ident[i,1] = category[i,1]; /* cliente */
    ident[i,2]= category[i,2]; /* hyper */
    j=1;
    do while j<=6;

        art_list = selif(artic[.,5], artic[.,3].==j);

        p=1;
        do while p<=rows(art_list);
            art = art_list[p,];
            dat1 = selif(dat, cliente.==category[i,1] .and categori.==j .and articulo.==art);
            if scalmiss(dat1).==0;
                if j.==1;
                    salida[i, 2*p-1]=sumc(dat1[.,13]);
                    salida[i,2*p]= rows(dat1);
                elseif j.==2;
                    salida[i,104+2*p-1]=sumc(dat1[.,13]);
            end;
            p=p+1;
        end;
    j=j+1;
end;
end;
```



```

        salida[i,104+2*p]= rows(dat1);
    elseif j.==3;
        salida[i,112+2*p-1]=sumc(dat1[:,13]);
        salida[i,112+2*p]= rows(dat1);
    elseif j.==4;
        salida[i,118+2*p-1]=sumc(dat1[:,13]);
        salida[i,118+2*p]= rows(dat1);
    elseif j.==5;
        salida[i,160+2*p-1]=sumc(dat1[:,13]);
        salida[i,160+2*p]= rows(dat1);
    elseif j.==6;
        salida[i,166+2*p-1]=sumc(dat1[:,13]);
        salida[i,166+2*p]= rows(dat1);
    endif;
    p=p+1;
    endo;
    j=j+1;
    endo;
    i=i+1;
    endo;

```

```

ident = zeros(2,cols(ident))|ident;
salida = articl|salida;
salida1 = salida[:,1:50];
salida2 = salida[:,51:100];
salida3 = salida[:,101:150];
salida4 = salida[:,151:cols(salida)];

```

```

output file = c:\gauss6.0\scanner\identif2 reset;
screen off;
ncol = cols(ident);
mask = ones(1,ncol);
let fmt = {"*.*1f" 15 0, "*.*1f" 5 0};
call printfm(ident,mask,fmt);

```

```

output file = c:\gauss6.0\scanner\sala9 reset;
screen off;
ncol = cols(salida1);
mask = ones(1,ncol);
let fmt = {"*.*1f" 14 2};
fmt = ones(ncol,3).*fmt;
call printfm(salida1,mask,fmt);

```

```

output file = c:\gauss6.0\scanner\sala10 reset;
screen off;
ncol = cols(salida2);
mask = ones(1,ncol);
let fmt = {"*.*1f" 14 2};
fmt = ones(ncol,3).*fmt;
call printfm(salida2,mask,fmt);

```

```

output file = c:\gauss6.0\scanner\sala11 reset;
screen off;
ncol = cols(salida3);
mask = ones(1,ncol);
let fmt = {"*.*1f" 14 2};
fmt = ones(ncol,3).*fmt;
call printfm(salida3,mask,fmt);

```



```

ident[i,1] = category[i,1]; /* cliente */
ident[i,2]= category[i,2]; /* hyper */

art = 494059;
dat2 = selif(dat1, dat1[:,4].==category[i,1] .and dat1[:,5].==art);
if scalmiss(dat2).==0;
    salida[i,1]=sumc(dat2[:,13]);
    salida[i,2]= rows(dat2);
endif;

art = 497678;
dat3 = selif(dat1, dat1[:,4].==category[i,1] .and dat1[:,5].==art);

if scalmiss(dat3).==0;
    salida[i,3]=sumc(dat3[:,13]);
    salida[i,4]= rows(dat3);
endif;

i=i+1;
endo;

ident = zeros(2,cols(ident))|ident;
salida = articl|salida;
sal1 = ident~salida;

output file = C:\ana\typic\scanner\salidas\sals03 reset;
screen off;
ncol = cols(sal1);
mask = ones(1,ncol);
let fmt0= {"*.*1f" 15 0};
let fmt1 = {"*.*1f" 14 2};
let fmt = {"*.*1f" 15 0, "*.*1f" 5 0, "*.*1f" 15 2, "*.*1f" 15 2, "*.*1f" 15 2, "*.*1f" 15 2};
call printfm(sal1,mask,fmt);

end;

```

@-----
DESCMES.GAU 05 agosto 2004
Objetivo: estadísticas descriptivas por meses
-----@

```

new;
/* leo datos totales, y ordeno por meses */
/* Primera comprobacion sobre : loadm dat[19599,19]=C:\gauss6.0\compras\scan_num.txt; */

loadm dat[19599,19]=C:\gauss6.0\compras\scan_num.txt; dat=trimr(dat,1,0);

dat = sortc(dat,2);
year = dat[:,1];mes = dat[:,2];centro =dat[:,3];cliente=dat[:,4];categori=dat[:,11];ventas =dat[:,13];

@/* cifra de compras del total de centros, para el mes: valor total, min, max, media,deciles */
e = seqa(0.10,0.10,9);
descri = zeros(12,4+rows(e));
"totalmes i min max media decil1 2 3 4 5 6 7 8 9";
i=1;
do while i<=12;

    valor = selif(dat[:,13], mes.==i .and (centro.==850 .or centro.==4000));

```

```

descri[i,1] = sumc(valor);descri[i,2] = minc(valor);descri[i,3] = maxc(valor);
descri[i,4] = meanc(valor);
    if rows(valor).*minc(e).<1 .or rows(valor).*maxc(e).>rows(valor)-1;
        descri[i,5:4+rows(e)] = ones(1,rows(e)).*999999;
    else;
        descri[i,5:4+rows(e)] = quantile(valor,e)';
    endif;
format/m1/rd 6,2;

descri[i,];
i=i+1;
endo;

/* cifra de compras del centro 850, para el mes: valor total, min, max, media,deciles */
e = seqa(0.10,0.10,9);
descri = zeros(12,4+rows(e));
"850mes i min max media decil1 2 3 4 5 6 7 8 9";
i=1;
do while i<=12;
    valor = selif(dat[,13], mes.==i .and centro.==850);

    descri[i,1] = sumc(valor);descri[i,2] = minc(valor);descri[i,3] = maxc(valor);
    descri[i,4] = meanc(valor);
        if rows(valor).*minc(e).<1 .or rows(valor).*maxc(e).>rows(valor)-1;
            descri[i,5:4+rows(e)] = ones(1,rows(e)).*999999;
        else;
            descri[i,5:4+rows(e)] = quantile(valor,e)';
        endif;
    format/m1/rd 6,2;

descri[i,];
i=i+1;
endo;

/* cifra de compras del centro 4000, para el mes: valor total, min, max, media,deciles */
e = seqa(0.10,0.10,9);
descri = zeros(12,4+rows(e));
"4000mes i min max media decil1 2 3 4 5 6 7 8 9";
i=1;
do while i<=12;
    valor = selif(dat[,13], mes.==i .and centro.==4000);

    descri[i,1] = sumc(valor);descri[i,2] = minc(valor);descri[i,3] = maxc(valor);
    descri[i,4] = meanc(valor);
        if rows(valor).*minc(e).<1 .or rows(valor).*maxc(e).>rows(valor)-1;
            descri[i,5:4+rows(e)] = ones(1,rows(e)).*999999;
        else;
            descri[i,5:4+rows(e)] = quantile(valor,e)';
        endif;
    format/m1/rd 6,2;

descri[i,];
i=i+1;
endo;@

```

```

/* cifra de compras, para el mesj de las categoria1,para
el mesj de las categoria2,etc: valor total */

descri = zeros(6,1);

i=1;
do while i<=6;
  j=1;
  do while j<=12;
    valor = selif(dat[.,13], mes.==j .and (categori.==i .and centro.==4000));

    descri[i,1] = sumc(valor);
    format/m1/rd 6,2;
    "categoriai " descri[i,1];
    j=j+1;
  endo;
  "4000categori";
  i=i+1;
endo;

/* cifra de compras, para el mes1 de las jcategorias,para
el mes2 de las jcategorias,etc: valor total */

descri = zeros(6,1);

i=1;
do while i<=6;
  j=1;
  do while j<=12;
    valor = selif(dat[.,13], mes.==j .and (categori.==i .and centro.==850));

    descri[i,1] = sumc(valor);
    format/m1/rd 6,2;

    "categoriai " descri[i,1];
    j=j+1;
  endo;
  "850mesj";
  i=i+1;
endo;
/* cifra de compras, para el mes1 de las jcategorias,para
el mes2 de las jcategorias,etc: valor total */

descri = zeros(6,1);

i=1;
do while i<=6;
  j=1;
  do while j<=12;
    valor = selif(dat[.,13], mes.==j .and (categori.==i .and (centro.==4000
    .or centro.==850)));
    descri[i,1] = sumc(valor);
    format/m1/rd 6,2;
    "categoriai " descri[i,1];
    j=j+1;
  endo;
  "totalmesj";
  i=i+1;
endo;

```

