



Condition Analysis and Forecasting in the Fashion Industry

Mariusz Czekala¹, Agnieszka Bukietynska¹, Marek Gurak², Jacek Jagodzinski³,
Jarosław Klosowski¹

¹Wrocław School of Banking, Finance and Management Department, Wrocław, Poland

²MG4-Limited Company, Wrocław, Poland

³Faculty of Electronics, Wrocław University of Science and Technology, Wrocław, Poland

Email address:

mariusz.czekala@wsb.wroclaw.pl (M. Czekala), agnieszka.bukietynska@wsb.wroclaw.pl (A. Bukietynska),
m.gurak@mg4.eu (M. Gurak), jacek.jagodzinski@pwr.edu.pl (J. Jagodzinski), j.klosowski@mongos.pl (J. Klosowski)

To cite this article:

Mariusz Czekala, Agnieszka Bukietynska, Marek Gurak, Jacek Jagodzinski, Jarosław Klosowski. Condition Analysis and Forecasting in the Fashion Industry. *International Journal of Economics, Finance and Management Sciences*. Vol. 7, No. 2, 2019, pp. 74-81.

doi: 10.11648/j.ijefm.20190702.14

Received: April 21, 2019; **Accepted:** May 23, 2019; **Published:** June 15, 2019

Abstract: The work considers the problem of demand for the clothing industry's goods. It shows how this problem is connected with the mathematical problem of the partition of the set. Investment decisions depend on a diagnosis based on forecasting demand in individual product groups. These groups are characterized by a number of features and even in the simplest situations (3 attributes) lead to computationally complex situations. In this situation, the recursive partitioning method can be used. This is a method related to the construction of classification trees (regression). These methods are widely used in natural, technical and economic sciences. The main direction of their applications is to support decision-making processes. The article shows how to support the construction of classification trees. The paper proposes a practical solution to the problem using the method of random partitions. The proposed method can be a complement to the recursive partitions method, or used in some situations instead. The submitted method is a practical proposal to avoid the problem of computational complexity. The numerical example shows how to replace a population of about 52 trillion by a sample of only 100. The applied method was justified by an example of a less numerous population, where the result could be verified empirically by reviewing all possibilities. Such verification is not practically possible in the case of 20 product profiles. Such a number generates a number of partitions amounting to almost 52 trillion. The article also presents the estimation of the calculation time. These results are useful from a practical point of view, although they are not optimal.

Keywords: Fashion, Clothing, Set Partition, Recursive Partitioning, Chuprov

1. Introduction - Clothing Industry, Optimal Decisions

There are more and more people interested in fashion and taking care of their own image. This applies to both women and men. On the one hand, clothing is an element of the image that emphasizes following fashion trends, emphasizing social and cultural status. On the other hand, we pay attention to functional properties of fashion products: convenience of use, characteristics of the material and its features, its impact on contact with the body. The impact on the natural environment and biodegradation is equally important for the younger generation.

The impact of fashion trends and the condition of the lingerie and clothing industry have a huge impact on the state of the global economy, on the number of jobs and the level of consumption. Charismatic head of the US Federal Reserve Bank Alan Greenspan using the MUI (Men's Underwear Index) index justified the Bank's strategy, based also on this indicator of forecasts regarding the economic prosperity or recession [1]. In the literature, there is still a discussion as to whether this index really reflects business trends or is just a "joke" of Greenspan. Is it possible to treat it globally or locally, even if it is related only to specific climatic zones and communities with defined average incomes? There are many points of view in relation to this index from those approvingly

explicitly disrespectful. However, no one is questioning the role of the clothing and underwear industry in the development of the global economy.

It is estimated that the revenues of the lingerie industry in 2019 will amount to approximately USD 287,787 million and an increase in 2019-2024 is expected in the range of 5% per annum. Global expenditures per capita for lingerie products are estimated at 38.96 USD. In 2019, the largest market in the industry will be the United States with a market worth USD 52,191 mln, followed by the largest lingerie markets: China USD 45,266 mln, India USD 27,949 mln Germany USD 13,478 mln and United Kingdom USD 13,160 mln. In 2023, it is expected to produce 48,243.14 mln items of underwear (although the range of underwear includes: women's underwear, men's underwear, bathrobes and bathing underwear with the exclusion of clothes for swimming, and T-shirts underwear shirts). In Poland the underwear industry in 2022 will reach a value of approximately EUR 10 billion. [2]

Decision optimization in management in the fashion industry concerns many aspects. It applies to the sphere of designing, production, management of sales channels, brand management, advanced technologies used for the production of new materials, and recycling. The impact on the market of lingerie products such factors as: population, number of women, men, teenagers and children, business cycle, level of consumer spending, taxes and fiscal policy, tourism and international exchange. Over the last decades, much attention has been paid to analysing the fashion supply chain. It is associated with the rapid development of clothing brands. An overview of the applied operational research methods is presented in the paper of Wen, Choi and Chung. [3]

These factors can be predictably forecast in the perspective of several years ahead. Among the numerous market research, the Y generation deserves special attention. Producers and traders appreciate its function in the development of the market and recognise its specificity. The analysis and identification of trade profiles is presented by Gonthier and Lajante [4]. According to the authors, it is much more difficult to forecast the situation on the lingerie market taking into account such an elusive factor as fashion and fashion trends. Considering the size of the market, its globality, the availability of products and the competition between manufacturers, any mistake in estimating the sales volume of new collections can be of great importance to producers and distributors. If the chimerical fashion market recognises that a given product is not fashionable, it is not trendy, regardless of its usable value and price, the product may not find recognition in the eyes of potential customers. The problem of production and sales planning for the fashion industry is rarely described by researchers. Interesting publication on planning have been published by M Xia, W. K Wong. [5]

2. Presentation of the Problem¹

The problem considered in this work is a typical decision problem whose first stage is the diagnosis of the situation. There are not well-known studies where recursive partitioning methods or decision trees are used in the clothing industry, although they probably appear as internal studies. However, such solutions appeared in issues related to demand research. These methods have been applied to the car industry. In an interesting discussion of the partition method presented by Therneau and Atkinson, 34 variables concerning 111 cars were considered. [6] This interesting example, however, was based on data from the 90s of the previous century. These methods were also used in medicine. In the paper of Athey and Imbens the application of a recursive division to the analysis of groups of patients from the point of view of the effectiveness of treatment with specific methods is considered. [7]

The diagnosis will be presented through the analysis of sales reports. These reports contain, in addition to the characteristics of products (usually a dozen or so), also financial data on the number of items being traded and the margin. The article will present an analysis based on three features characterizing clothing products.

Table 1 presents a fragment of a report made available by one of the companies dealing in trade in underwear and clothing.

The three characteristics selected for analysis are collection, type and colour. The first two features come in two variants. The collection that appears in the analysis is the basic collection (B). It is characterized by the fact that it is directed to people who are mainly driven by utilitarian considerations. Collection A is a collection that has to meet the aesthetic needs of buyers, but also meeting usages. In fact, the collections appeared in four variants, and the partition into two categories took place in accordance with expert opinions.

The second feature includes the type of boxer briefs (MX) and panties briefs (MP). The third feature in sales reports is in 15 variants. The colours are presented in five groups. The white colour is marked C1, C2 is black. The remaining colours were grouped in three sets, naming them C3, C4 and C5. This partition was also made after consulting experts. The method proposed in the work can be used to search for the optimal partition of both colours and collections. As shown by the calculations presented in table 3, such a task is feasible in a relatively short period of time and therefore is not a good illustration of the proposed method. This would be the first step in the recursive partitioning method that consists of the selection of variables and with more variants of the feature, the proposed method can be helpful.

The starting point in the analysis is the contingency table created on the basis of a report on 10,984 transactions.

¹ This is a part of the Project implemented as part of the Program Operacyjny Inteligentny Rozwój 2014-2020 of Narodowe Centrum Badań i Rozwoju. This is a part of the Project title: "Demand forecasting system controlled by dynamic fashion trends for clothing, textile and lingerie enterprises". Number of application: POIR.01.01.01-00-0886/ 17-00.

Table 1. Transaction report (a fragment).

Transaction	Collection	Fabric	Types	Colour	Size	Quantity	Value (PLN)	Margin (PLN)
561	B	95% cotton, 5% elastane	MX	white	XXL	20	167.53	50.73
562	B	95% cotton, 5% elastane	MX	black	XXL	20	173.13	56.33
563	B	95% cotton, 5% elastane	MP	white	XL	10	78.47	33.47
564	B	95% cotton, 5% elastane	MP	black	L	6	47.08	16.18

Source: Data obtained from the MG4 company.

These data are presented in table 2. In the case in question, there are 20 variants (profiles) of the analysed products. These groups of products are marked with symbols M1 to M20.

Table 2. Contingency table - product profiles and decision function.

Number	profile	name	Y = 0	Y = 1	sum
1	A MX C1	M1	14	24	38
2	A MX C2	M2	40	69	109
3	A MX C3	M3	308	1103	1411
4	A MX C4	M4	40	52	92
5	A MX C5	M5	37	65	102
6	A MP C1	M6	38	32	70
7	A MP C2	M7	320	241	561
8	A MP C3	M8	20	32	52
9	A MP C4	M9	900	690	1590
10	A MP C5	M10	790	620	1410
11	B MX C1	M11	24	20	44
12	B MX C2	M12	38	40	78
13	B MX C3	M13	440	1211	1651
14	B MX C4	M14	50	31	81
15	B MX C5	M15	50	32	82
16	B MP C1	M16	30	42	72
17	B MP C2	M17	208	341	549
18	B MP C3	M18	32	54	86
19	B MP C4	M19	683	869	1552
20	B MP C5	M20	512	842	1354
			4574	6410	10984

Source: Own calculations based on data obtained from the MG4 company.

The decision variable is variable y (discriminant or regression - depending on the adopted scale) in the case under consideration. In the example considered here accepts that the variable y assumes a value of 1 if the margin from a specific transaction exceeds PLN 50. Otherwise, the value of y is 0. This level has been set arbitrarily. This makes it possible to interpret the decision variable both as a discriminant variable and as a regression variable.

The goal is to build the model.

$$y = f(x_i) = \sum_{k=1}^K \alpha_k I(x_i \in S_k)$$

where I is the index function of the set. This approach differs from the approach known from the recursive partition method. This method considers variables and not transactions (objects).

The domain of the above function is the transactions space. Each transaction belongs to one of the disjoint sets S_k and only for this set the function $I(x_i \in S_k)$ takes the value 1. In the case where the model is a regression model to assess the homogeneity of individual spaces, the function presented

below is used. Despite the difference in methods, the way of determining parametre evaluations is similar. [8]

$$\sum_{x_i \in S_k} (y - \alpha_k)^2$$

The value α_k is determined from the formula (1).

$$\alpha_k = \frac{1}{N_k} \sum_{x_i \in S_k} y_i \quad (1)$$

Unfortunately, finding the optimal partition by reviewing all cases is not possible in practical issues. This is evidenced by the results of considerations regarding computational complexity.

3. Computational Complexity

In order to evaluate the performance of methods presented in article running times of the algorithms were tested. In the examined method, the most time-consuming task is to determine a set of partition of a set, the other calculations have a linear computational complexity in relation to the problem of generating subsets.

Therefore, in order to estimate the time of CPU computation, the generation times of SetPartitions function were examined. The procedure is part of the Combinatorica package from the Wolfram Mathematica 11.3 software [9]. The tests were carried out on an Intel Core i7-8550U @ 1.80GHz with 16GB RAM. The operating times of the CPU during generation sets consisting of n elements were examined. The results are given in tab. 1.

Attention should be paid to the generation of the solution for $n = 14$. There has been a significant change in the length of the calculation time. It turned out that up to $n = 13$ calculations were allocated mainly in RAM memory, while at $n = 14$ computational complexity were significantly increased, because the amount of RAM memory was not enough. In this case the Mathematica used hard disk as dynamic memory to allocation variables during CPU run time. That is why two estimates of duration time were made. In first case to estimate the parametres of the generating function only data up to $n = 13$ was used (sufficient RAM memory) and the second estimation including $n = 14$ (insufficient RAM memory).

Estimations were made using the least squares method assuming the following function class.

$$f(x) = ae^{bx} [s] \quad (2)$$

where: $a = 2.2596 \cdot 10^{-8}$, $b = 1.7455$ in sufficient RAM

memory case and $a = 1.2327 \cdot 10^{-8}$, $b = 1.8493$ in insufficient RAM memory case. For comparison, similar results for CPU timing (without estimation) could be found [10].

Table 3. Computational complexity.

n	sufficient RAM memory			insufficient RAM memory	
	CPU Time	CPU Time		CPU Time	CPU Time
	[s]	[year]		[s]	[year]
8	0,031	3,142E-17	test	0,031	3,142E-17
9	0,109	1,100E-16		0,109	1,100E-16
10	0,594	5,971E-16		0,594	5,971E-16
11	3,641	3,661E-15		3,641	3,661E-15
12	24,938	2,507E-14		24,938	2,507E-14
13	163,984	1,649E-13	estimated	163,984	1,649E-13
14	927,0	2,940E-05		2220,0	6,859E-05
15	5310,8	1,684E-04		13748,1	4,359E-04
16	30425,1	9,648E-04		87377,8	2,771E-03
17	174303,8	5,527E-03		555342,1	0,018
18	998578,6	0,032		3529554,2	0,112
19	5,721E+06	0,181		2,243E+07	0,711
20	3,277E+07	1,039		1,426E+08	4,521
21	1,878E+08	5,954		9,061E+08	28,734
25	2,023E+11	6,414E+03		1,479E+12	4,688E+04
30	1,248E+15	3,958E+07		1,533E+16	4,862E+08

4. Proposal for a Solution

One of the most popular methods to solve the problem is the hill climbing strategy. This method was used in the problem of recursive partitions. [11]. An example of the application of the models considered here comes from the food industry. The paper of Mugridge and Wang presents the use of decision trees to choose assortment on the example of a cafeteria [12]. The authors are not aware of the use of the presented methods to select the assortment in clothing industry. The cafeteria example, however, shows the potential and possibilities of these methods. Similarly to the method proposed here, it does not give an optimal solution. This method allows the selection of variables, but only gives the local optimum. The essence of the presented proposal is to draw partitions in order to find a suboptimal solution. Such a proposal is favoured by a friendly size distribution characterizing the quality of the partition tested on a relatively small number of cells in the contingency table. These quantities are mainly quantities based on chi square statistics. As we know, the large values of chi squared correspond to a larger relationship between the variables (groups). In addition to this measure, other statistic indicators can be used, such as p-value or the Chuprov indicator. In

problems strictly related to making business decisions, it would be desirable to use measures corresponding to economic efficiency. However, this issue will be skipped in this article. Determining the sampling scheme requires analysis of the distribution of the partition of the set when the number of components is 20. This is consistent with the data presented in table 2. This is part of a more general problem and has been considered in combinatorics many times and is related to Stirling numbers the second kind [13]. It is proved that the number of partitions of the k element set on r classes is the number of Stirling $S(k, r)$ of the second type. A recurrent relationship exists for these numbers:

$$S(k, 1) = 1, S(k, k) = 1$$

$$S(k + 1, r + 1) = S(k, r) + (r + 1)S(k, r + 1)$$

Bell numbers are defined as sums of Stirling numbers in the following way [14].

$$B_n = \sum_{k=0}^n S(n, k)$$

Stirling numbers of the second type and Bell numbers for selected values of n are shown in table 4.

Table 4. Distributions of partitions.

Bell numbers	n \ r	1	2	3	4	5	6	...
52	5	0,00037	0,005547	0,009246	0,003698	0,00037	0	
115975	10	7,43E-11	3,8E-08	6,94E-07	2,54E-06	3,16E-06	...	
51724158235372	20	3,74E-28	1,96E-22	2,17E-19	1,69E-17	2,8E-16	...	
846749014511809000000000	30	1,39E-48	7,49E-40	4,79E-35	6,7E-32	1,08E-29	...	

Source: Own calculation.

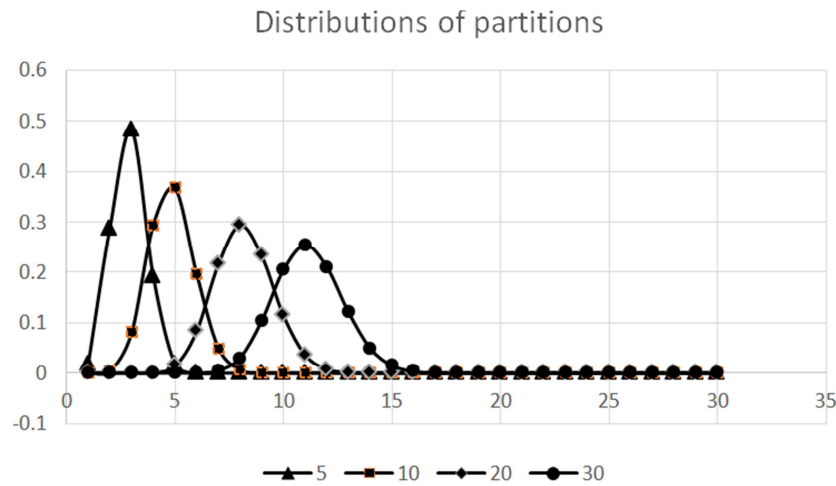


Figure 1. Distributions of partitions.

Due to the fact that in the case of 30 subsets, the above table would contain 30 columns, only the first five probabilities were presented. The remaining calculations will be illustrated in figure 1.

5. Partition Drawing

According to the data in table 2, the considered case concerns 20 subsets. This means that the number of possible subsets is almost 52 trillion. The analysis presented in the

section on computational complexity shows that the best partition cannot be achieved within a reasonable time. Therefore, we will use the proposed method. A sample size of 100 (sufficient for illustration of the solution) will be drawn from all 51,724,158,235,372 classes of subsets. The draw will be carried out in two stages. In the first of them, the number of elements of each partition included in the sample will be drawn. The partition distribution presented in table 5 should be used here. The results of this draw are presented in table 5.

Table 5. The results of the draw.

Number of elements	5	6	7	8	9	10	11
Subsets drawn	1	6	18	32	25	14	4
Cumulative subsets drawn	1	7	25	57	82	96	100
Probability	0.014	0.083	0.215	0.293	0.232	0.114	0.036

Source: Own calculation using Excel.

The table above shows that subsets with a small number of components (up to 4) and large (over 12) were not drawn. This is a result in line with expectations, because the probabilities of such events were significantly lower than 0.01.

The second draw step was made using the R program [15]. The "sample" function was used. Thanks to this function, the indexes of sets that have to be part of a given partition have been randomly selected. Collection numbering was in accordance with the data in table 2. In case the randomly selected partition did not occur in the results of the draw, an additional draw was made to determine the missing index in the sequence of indicators. Some of the results of this draw are presented in Table 6.

Table 6. Indexes of sets.

Indexes of sets	Number of partition
1 1 3 2 1 / 4 5 4 3 2 / 1 4 5 3 2 / 4 2 2 4 5	1
6 5 4 4 2 / 4 5 5 6 4 / 2 4 1 4 3 / 4 3 4 2 6	2
1 3 4 5 1 / 6 1 5 3 3 / 6 2 7 1 3 / 7 6 4 5 5	17
7 4 5 5 6 / 7 7 5 5 5 / 2 4 4 1 3 / 5 6 7 1 3	18
4 2 2 1 2 / 2 5 1 7 9 / 9 1 2 5 7 / 9 6 3 8 1	78
6 4 8 4 3 / 4 5 4 2 7 / 4 1 9 5 5 / 6 5 7 5 6	79

Indexes of sets	Number of partition
1 10 5 2 8 / 9 3 10 9 11 / 3 9 7 9 4 / 11 11 6 11 3	99
1 4 8 3 2 / 5 6 2 4 11 / 9 7 9 5 9 / 9 2 10 8 6	100

Source: Own calculations using the R program.

From table 6 one can read the composition of the randomly selected partition. For convenience, the sets indices are given in 5-number portions. For example, the first partition consists of five subsets. The first subset is M1 + M2 + M5 + M11, the second M4 + M10 + M15 + M17 + M18 etc.

6. Results

For selected subsets, one should calculate the values of selected goodness of fit criteria. Calculate and compare these values for 100 contingency tables. Table 7 will present two contingency tables, which is related to the proposed suboptimal solution.

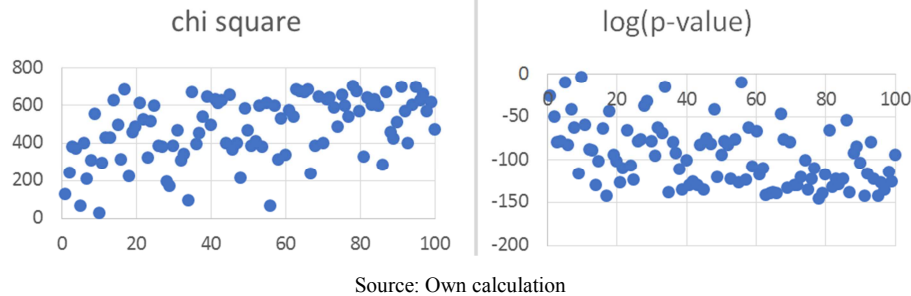
The contingency tables were calculated for all drawn compositions. Next, the values of chi square, p-value, ϕ^2 and Chuprov coefficient T^2 were calculated [16]. For convenience, the Chuprov coefficient will be multiplied by

10,984 (number of observations).

Table 7. Contingency tables-compositions 17 and 78.

17	S1	S2	S3	S4	S5	S6	S7	S8
0	421	38	1780	340	1255	270	470	4574
1	361	40	1411	1157	1795	393	1253	6410
	782	78	3191	1497	3050	663	1723	10984

78	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
0	610	863	32	14	370	208	950	683	844	4574
1	966	2480	54	24	272	341	722	869	682	6410
	1576	3343	86	38	642	549	1672	1552	1526	10984



Source: Own calculation

Figure 2. Chi square statistics and logarithm of p-values for the sample.

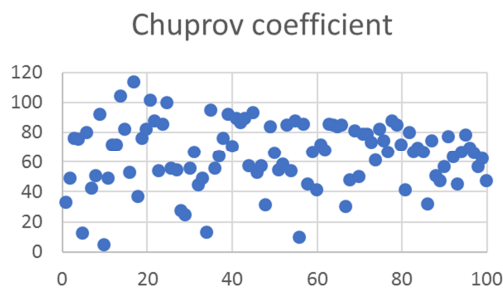


Figure 3. Chuprov T^2 times 10,984.

The fourth figure shows the ratio of T^2 Chuprov multiplied by a constant of 10,984. The analysis of the data from the calculations of the mentioned quantities leads to the conclusion that the best partition in terms of the value of chi square statistics is the partition of 78. The following indicators were calculated for this partition: 4 2 2 1 2/2 5 1 7 9/9 1 2 5 7/9 6 3 8 1. This means that we have a nine-element partition with the following components: M4 + M8 + M12 + M20, M2 + M3 +

$$\hat{y} = 0.613IS1 + 0.742IS2 + 0.628IS3 + 0.632IS4 + 0.424IS5 + 0.621IS6 + 0.432IS7 + 0.560IS8 + 0.447IS9$$

And for partition 17:

$$\hat{y} = 0.462IS1 + 0.513IS2 + 0.442IS3 + 0.773IS4 + 0.589IS5 + 0.593IS6 + 0.727IS7$$

It should be remembered that the components in both cases have different definitions and different numbers in the contingency tables. In the formulas above (ISn), means the set indicator. The above equations can be a tool supporting decision making. Based on the data in table 2, it can be calculated that an average of 58.35% of transactions are transactions with a satisfactory margin. In the case of partition 17 (the highest measure of Chuprov in the sample), the four subsets have a better than average score, and the best

M5 + M6 + M13, M18, M1, M7 + M14, M17, M9 + M15, M19, M10 + M11 + M16. Also this partition has the lowest p-value and in this respect is the best in the sample. The last conclusion is not, however, logical equivalence due to differences in the number of degrees of freedom. In terms of the criterion φ^2 the result is the same as in the case of chi square because these quantities are proportional. The chi square for this partition is 701.28, and the value of $\log(p\text{-value}) = -145.42$. The value of $\varphi^2 = 0.064$.

Taking into account the Chuprov criterion, composition 17 proved to be the best composition. For this partition, indicators were issued: 1 3 4 5 1/ 6 1 5 3 3/ 6 2 7 1 3/ 7 6 4 5 5. This means that we have a seven-element partition with the following components: M1 + M5 + M7, M12, M2 + M9 + M10 + M15, M3 + M18, M4 + M8 + M19 + M20, M6 + M11 + M17, M13 + M16. For this partition, the value of $T^2 = 0.01031$, that is $10,984T^2 = 113,26$. The above results lead to discrimination (regression) functions with the following form. One should use the formula (1). For partition 78:

is the S4 set. For partition 17, the least preferred group is the subset of S3. The above solutions can be the basis for making decisions, but they may also be a stage in the construction of more complex models. According to the idea of recursive partitioning, one can use the proposed method for the second time, etc. for each of the disjoint subsets one get. In other partitions, one can use different attributes.

7. Conclusions

The proposal of subset draws can and should be used also for the goodness of fit other than those used above. The chi-square criterion is the basis for dependency testing, but it is not necessarily a good measure of dependence. However, the distribution of the empirical values of chi-square (and p-value) inclines to the recognition that the method based on randomization in very complex computational problems can give practically useful results. Two graphs prepared using the Mathematica program [9] for the case $n = 10$ can testify to

this. Then (table 5) the appropriate number of Bell is 115, 975 and in this case it is possible to compute all chi square values and p-value values. The chart is based on the data considered in this article (table 4), but without taking into account the dichotomous attribute - collection. Hence, the number of possible categories is 10. As can be seen from the data from table 3, finding the value for drawing such figures for the case of $n = 20$ would take from one year to about 4.5 years, depending on the assumptions regarding computer memory.

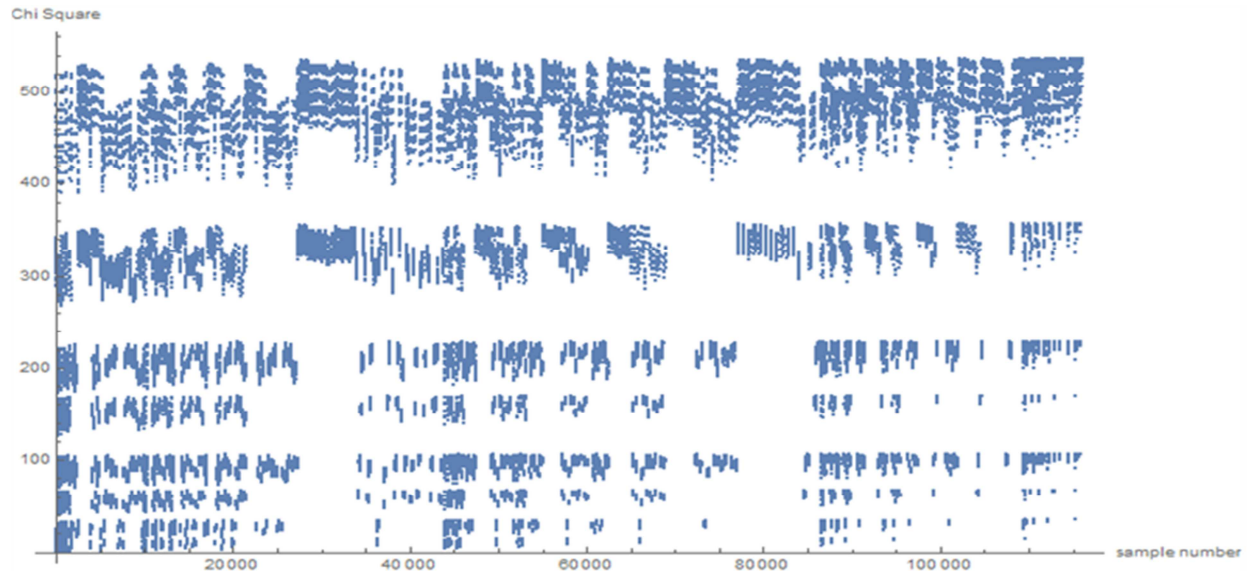


Figure 4. Chi-square values for all 115,975 compositions (case of $n=10$).

The maximum value of this statistic was around 536, and the maximum of the 100 elements around 532. Both values were significant.

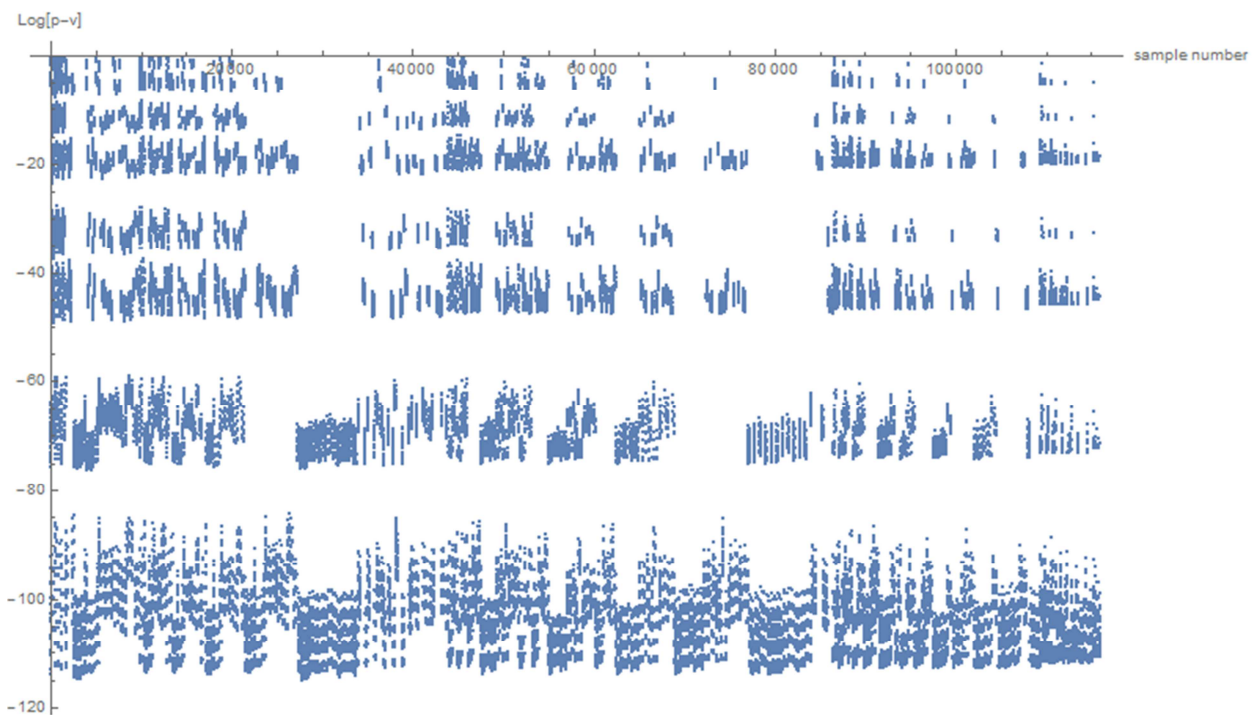


Figure 5. Logarithm of p-values for all 115,975 compositions (case of $n=10$).

Relying on the chi square criterion, on average, prefers a large number of subsets. Therefore, there are a number of proposals for dependency measures in the literature. One of them is the Chuprov measure. Among the best-known measures, the Goodman-Kruskal coefficient can be mentioned, the probability of incorrect classification, the Gini coefficient or the Ripley deviance. A dozen measures of heterogeneity were collected in the book of Gatnar [17]. The discussions conducted there, however, mainly concerned the recursive partitioning method. They can be used instead of or in addition to the chi square measure. The selection of criteria is a separate issue. It depends on the issue under consideration. At the stage of building the model, statistical criteria seem to be important. In a decision-making situation, other criteria related to economic indicators may be more important while maintaining the minimum relevance (significance) requirements for goodness of fit. From a practical point of view, it will be desirable to examine the quality of partitions after considering the financial criteria.

References

- [1] Ylan Q. Mui, Washington Post. Monday, August 31, 2009.
- [2] www.statista.com/outlook/90040000/100/underware/worldwide 10th of Apr 2019.
- [3] Wen X., Choi T.M., Chung S.H., Fashion retail supply chain management: A review of operational models, *International Journal of Production Economics* Volume 207, January 2019, pp. 34-55.
- [4] Gonthier J., Lajante M., Generation Y and online fashion shopping: Orientations and profiles *Journal of Retailing and Consumer Services*, Volume 48, May 2019, pp. 113-121.
- [5] Xia M., Wong W. K., A seasonal discrete grey forecasting model for fashion retailing. *Knowledge-Based Systems*, Vol. 57, Feb 2014, pp. 119-126.
- [6] Therneau T. M., Atkinson J., Mayo Foundation, *An Introduction to Recursive Partitioning Using the RPART Routines*, 2019.
- [7] Athey S., Imbens G., Recursive partitioning for heterogeneous causal effects, *PNAS*, July 5, 2016 113 (27) pp. 7353-7360
- [8] Walesiak M., Gatnar E., *Statystyczna Analiza Danych*, PWN, Warszawa, 2012.
- [9] Wolfram Research, Inc., *Mathematica*, Version 11. 3, 2019.
- [10] Djokic B., Short Note: A Fast Iterative Algorithm for Generating Set Partitions, *The Computer Journal* 32 (3), 1989, pp. 281-282.
- [11] Breiman L., Bagging Predictors, *Machine Learning*, 24, 1996, pp. 123-140.
- [12] Mugridge J., Wang Y. (2019) Applying Decision Tree in Food Industry – A Case Study. In: Wang K., Wang Y., Strandhagen J., Yu T. (eds) *Advanced Manufacturing and Automation VIII. IWAMA 2018. Lecture Notes in Electrical Engineering*, vol 484. Springer, Singapore.
- [13] Flachsmeier J., *Kombinatorik*, VEB Berlin 1969.
- [14] Lipski W., Marek W., *Analiza kombinatoryczna*, PWN, Warszawa 1986.
- [15] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2019.
- [16] Yule G. U., Kendall M. G., *An Introduction to the Theory of Statistics*, Charles Griffin & Co. Ltd., London, 1958.
- [17] Gatnar E., *Nieparametryczna metoda dyskryminacji i regresji*, PWN, Warszawa, 2001.