Stockholms universitet

Bachelor thesis

Department of Statistics

Statistical entropy methods with applications to the analysis of questionnaire data

Jonas Eriksson and Filip Langemar

15 ECT-Credits in Statistics III, Spring 2013
Supervisor: Ove Frank

Abstract

In this study the relatively new method of entropy analysis is used to analyze a dataset with 14 961 observations from 29 different cities all over the globe. The method can be applied to variables on both nominal and ordinal scale and enables the researcher to find complex relationships within a dataset. The aim of the study is to explain the practice of entropy analysis and evaluate its ability to analyze marketing research data. The authors find that the method could be of great use in the area of marketing research in the future.

Keywords: entropy analysis, marketing research, discrete variables, explorative research.

Table of content

1 Introduction 6
2 Methodology..7
2.1 Statistical entropy measures 7
2.1.1 Aggregation 7
2.1.2 Entropy properties 8
2.1.3 Univariate entropy 8
2.1.4 Bivariate entropy 9
2.1.5 Trivariate and higher order entropies 10
2.2 Multivariate entropy analysis 10
2.2.1 Entropy matrix 10
2.2.2 Joint entropy and dependencies...11
2.3.3 J-relationship graph 11
2.3.4 Functional relationships..12
2.3.5 Trivariate entropy analysis...13
2.3.6 Tetravariate and higher order entropies 13
3 Application...15
3.1 The data 15
3.1.1 Variables 15
3.1.2 Aggregation 15
3.1.3 Calculating entropies...17
3.2 The Analysis process 17
3.2.1 Entropy matrix 17
3.2.2 J-matrix..18
3.2.3 Critical J-value 19
3.2.4 J-relationship graph 19
3.3 Functional relationships..22
3.3.1 Bivariate functional relationships 23
3.3.2 Higher order functional relationships 24
3.3.3 Independent variables 25
3.3.4 Variable independence..25
3.3.5 Conditional independence 26
3.4 Component structures 27
3.5 Total structure...28
4 Conclusion 31
References list 32
Appendix A..33
Table 12. Calculations for trivariate entropies (Y6, Y7, Y12, Y14, Y15, Y23).......33
Table 13. Calculations for trivariate entropies (Y8, Y9, Y11, Y16, Y18) 34
Table 14. Calculations for tetravariate entropies 34

1 Introduction

There are many different areas in statistical analysis where discrete variables are present. Marketing research for instance, is an area where explorative analysis is common and where different discrete categorical variables related to preferences and purchasing habits often occur. Many of the established techniques for exploratory analysis, exploratory factor analysis for instance, does however analyze continuous variables. In addition, techniques developed to analyze discrete variables are often limited to finding simple linear relationships.

A new multivariate technique developed to analyze variables on both nominal and ordinal scale is the technique called *entropy analysis*. It originates in the discipline of information theory and is based on the measure of spread named *entropy*. One of the technique's strong suites is its ability to identify non-linear relationships between variables in a dataset and illustrate them in a so called *relationship graph*. The method enables the researcher to identify more complex relationships in a dataset, a feature that makes entropy analysis a potential player in the future of research. Using the technique and evaluating its application in different research areas is hence important for the future use and development of the technique.

This study aims to apply the entropy analysis technique on the marketing research data of the *Metropolitan Report* collected by the business intelligence firm United Minds and the media company Metro International. The study aims to give a general understanding of the technique and evaluate its ability to analyze marketing research data.

2 Methodology

2.1 Statistical entropy measures

Imagine a questionnaire with ten questions and assume that all questions have the range of 5, meaning they can take five different outcomes. This means that there are 5^{10} possible response patterns in a ten dimensional distribution. It is not comprehensible nor user friendly to present the data as a ten dimensional distribution and consequently, the data need to be simplified in some way. How to simplify the data and find which variables should be presented is what the method of using entropy is all about. All presented methodology in this thesis will be referred to the works by Frank (2011).

Instead of thinking of the variables as a ten dimensional distribution the method separates variables into different groups with strong relationships within the groups and as small relationships as possible among the groups. This can be done by using certain statistical measures based on entropies. Assume that the structure of the ten variables is transformed in to two components with three variables each, one component with two variables and two variables that are independent of all the others. There are now five independent components of variables. The gain from knowing the structure of the variables is that we now can treat the material as two three dimensional, one two dimensional and two one dimensional distributions. Doing so enables the distribution over the response pattern to be derived from five distributions with a total of only $2 \cdot 5^3 + 5^2 + 2 \cdot 5 = 285$ response patterns. This is a more efficient way to present and get an overall view of the data than with the original $5^{10} = 9\,765\,625$ response patterns.

The method of entropies could also simplify the data even more than already explained. If one variable is found to be explained by other variables and hence can be determined from these, the variable is redundant and can be dropped from the analysis. If the two variables would explain the third variable in the two components of three variables in the example above, this would imply that we can drop the two explained variables from the analysis. That would leave 8 variables that still have the ability to give the same information as the previous ten. The total response patterns we would be interested in are now reduced to $3 \cdot 5^2 + 2 \cdot 5 = 85$.

2.1.1 Aggregation

Another way of simplifying the data that in no way is unique to the method of entropy analysis, is aggregating response categories that are similar or hard to distinguish. If the response categories are to be aggregated, it is desirable to make the variable uniformly distributed as this will make relationships emerge clearer later in the analysis. In the previous example it might be possible for each variable to be reduced from having five outcomes to having three outcomes. This means that the final response patterns under investigation, after variables have been dropped, would have been reduced to $3 \cdot 3^2 + 2 \cdot 3 = 33$ outcomes. Thus 33 different response patterns could roughly explain the same information as the 9 765 625 possible patterns of the original data.

By aggregating response categories one has to take into account that information will be lost. This have to, on the other hand, be reflected in the light of that the data can be presented in a much easier way. The researcher will have to decide an appropriate balance between details and interpretability.

2.1.2 Entropy properties

The entropy, H, can be thought of as a measure of spread. The highest value the entropy of a certain variable is the logarithm of its range. The entropy will reach its maximum when the variable is uniformly distributed and the minimum entropy zero is obtained for a variable with only one outcome. This means that a variable with entropy close to its *log r* is close to uniform, a variable with entropy far from log range has few occurring outcomes and a variable with entropy zero has only one outcome.

The entropy measure is most common when dealing with a finite discrete variable on ordinal or nominal scale. The entropy can however, still be used for continuous variables if transformed into categories. Analysis with entropies does not assume any specific distribution for the variables, an assumption that is easily violated in other statistical methods. The entropy measure can also be used for all data levels and compare variables at different kinds of levels. The entropy also measures all kinds of relationships between variables and not only linear ones like correlation methods.

2.1.3 Univariate entropy

Consider a variable X that has response categories on ordinal scale which reach from *very good, good, average, bad* and *very bad*. The variable then has a range of five and the response alternatives of the variable X are denoted by $x = 1, 2,3,4,5$ corresponding to the five alternatives. If X is considered to be a random variable, the calculated probability $P(X = x) = p_x$ for each of the outcomes x is approximated by its relative frequency = $\frac{n_x}{n}$ among the responses from n individuals for$x = 1, 2, \ldots r$.

Once the probabilities are computed the univariate entropy is calculated using the formula:

$$H_X = \sum_{x=1}^{r} p_x \log_2 \frac{1}{p_x} \qquad for\ all\ p_x > 0$$ *Equation 1*

Note that entropies are normally calculated using the logarithm base two. The entropy will be non-negative since the inverted probability is > 1 and take values between:

$$0 \leq H_X \leq log_2(r)$$ *Inequality A*

The univariate entropies can first be examined to see if there are any of the variables that can be considered a constant or that is uniformly distributed. If any entropy is zero or close to zero there is no variability within the variable and can be treated as a constant and can therefore be excluded in the analysis. What counts as close to zero is up to the researcher to decide. Since the entropy is a measure of spread, one can interpret the entropy as the logarithm of the numbers of outcomes that would correspond to a flat distribution. By taking 2 to the power of the entropy, one can see how many of the actual outcomes that is used within the variable. Another measure is the relative entropy which is a measure extended from *Inequality A*:

$$Relative\ entropy\ \widetilde{H}_X = \frac{H_X}{log_2(r)}\ (when\ r > 1) \qquad \textit{Equation 2}$$

$$0 \leq \widetilde{H}_X \leq 1 \qquad \textit{Inequality B}$$

The relative entropy takes the value zero if the variable is a constant and the value 1 if the variable is uniformly distributed. Hence the relative entropy gives the same information as the *Inequality A* but does so in a standardized manner. The relative entropy can be useful to identify which variables that could be dropped from the analysis at this early stage of the analysis. If the value for a variable is close to zero, the variable can be dropped. It is however, always a good idea to refer to the absolute value of the univariate entropy before doing so. It is desirable to have a variable that is uniformly distributed or close to uniformly distributed since it will give clearer relationships between variables later in the analysis. The relative entropy can also be computed for higher ordered entropies to get an idea of how big part of the outcome space that is being used.

2.1.4 Bivariate entropy

Consider two variables X and Y with outcomes $x = 1, 2, \dots r$ and $y = 1, 2, \dots s$. The bivariate entropy is given by the formula:

$$H_{X,Y} = \sum_{x=1}^{r}\sum_{y=1}^{s} p_{x,y}\, log_2 \frac{1}{p_{x,y}} \quad for\ p_{x,y} > 0 \qquad \textit{Equation 3}$$

The bivariate entropy can also be illustrated using an example. Consider two dichotomous variables X and Y both that take the values 0 and 1. Since the variables only can take two values $r = 2$ and $s = 2$, the sample space for the bivariate entropy is $r \cdot s = 4$. The table shows the numbers $n_{x,y}$ of individuals with response $X = x$ and $Y = y$, for $x = 0\ , 1$ and $y = 0\ , 1$.

x \ y	0	1	
0	$n_{0,0}$	$n_{0,1}$	$n_{0.}$
1	$n_{1,0}$	$n_{1,1}$	$n_{1.}$
	$n_{.0}$	$n_{.1}$	$n_{..} = n$

The relative frequencies for the joint distribution between X and Y are used for estimating the probabilities $\frac{n_{x,y}}{n} = p_{x,y}$, which is used for calculating the bivariate entropy $H_{x,y}$. The univariate entropies H_x and H_y can be calculated using the frequencies in the corresponding marginal.

The bivariate entropy satisfies the inequalities:

$$H_X \leq H_{X,Y} \leq H_X + H_Y \qquad \textit{Inequality C}$$

With equality to the left if and only if Y is explained by X $(X \to Y)$ and equality to the right if and only if X and Y is independent$(X \perp Y)$.

2.1.5 Trivariate and higher order entropies

Trivariate, or higher order entropies, are computed in the same way as the bivariate. Consider three variables X, Y and Z that each has the corresponding range r, s, t where the range is denoted by $x = 1,2 \ldots r, y = 1,2 \ldots s$ and $z = 1,2 \ldots t$.

The formula for the trivariate entropy is:

$$H_{X,Y,Z} = \sum_x^r \sum_y^s \sum_z^t p_{x,y,z} * log_2\left(\frac{1}{p_{x,y,z}}\right) \quad for\ p_{x,y,z} > 0 \qquad \textit{Equation 4}$$

$$H_{X,Y} \leq H_{X,Y,Z} \leq H_{X,Y} + H_Z \qquad \textit{Inequality D}$$

With equality to the left if and only if Z is explained by the pair (X, Y), $((X, Y) \to Z)$ and equality to the right if and only if the pair (X, Y) is independent of Z, $((X, Y) \perp Z)$.

2.2 Multivariate entropy analysis

Entropy analysis becomes as most useful in multivariate analysis. Consider *m* variables $X_1, X_2 \ldots X_m$ where variable X_i has range r_i for $i = 1,2 \ldots m.$ If $X_i, X_j\ and\ X_k$ are three different variables from a set of m variables and $i \neq j \neq k \neq l$ then the univariate entropy is denoted H_i, the bivariate entropy $H_{i,j}$and the trivariate as $H_{i,j,k}$, and notations is continued in the same manner for higher order entropies. Also note that $H_{i,i}$ is the same as H_i and $H_{i,j}$ is the same as $H_{j,i}$.

2.2.1 Entropy matrix

The entropy matrix is the basis for all further analysis and most information for future analysis will be derived from it. The entropy matrix is a symmetric matrix since $H_{i,j}$ is the same as $H_{j,i}$. The matrix shows the univariate entropies in the diagonals and the bivariate entropies in the off diagonals. The entropy matrix is ordered to have the variable with highest entropy first and the rest in descending order. This is done to get an easier overview which variables that might explain another later in the analysis.

If there is *m* variables in the dataset there will be *m* univariate entropies in the diagonals and $\binom{m}{2}$ bivaraite entropies in the off diagonals. Once the entropy matrix is computed a more extensive analysis can be performed.

2.2.2 Joint entropy and dependencies

To visualize which variables that share a relationship between each other the joint entropy will be used and a joint entropy matrix will be formulated. The joint entropy is derived from the inequality $H_{i,j} \leq H_i + H_j$ that corresponds to the right side of *Inequality C*. If the difference is taken of the inequality a measure of association between the variables is obtained. Hence the formula for calculating the joint entropy is:

$$J_{i,j} = H_i + H_j - H_{i,j} \qquad \textit{Equation 5}$$

$$0 \leq J_{i,j} \leq \min\,(H_i, H_j) \qquad \textit{Inequality E}$$

The joint entropy will take values between zero and the smaller of the two univariate entropies. If it assumes the value zero, there is independence between the variables $\mathrm{X_i}$ and $\mathrm{X_j}$. If the joint entropy assumes the $\min\,(\mathrm{H_i}, \mathrm{H_j})$ value this means that $\mathrm{H_{i,j}} = \max(\mathrm{H_i}, \mathrm{H_j})$ and $\max(\mathrm{H_i}, \mathrm{H_j})$ and one variable is a function of the other variable as explained by *Inequality C.*

The joint entropies are presented into the joint entropy matrix which also is a symmetrical matrix with univariate entropies in the diagonals and the joint entropy between the variables in the off diagonals. The joint entropy matrix gives a quick overview of which variables share relationships and which variables are independent from each other.

2.3.3 J-relationship graph

From the J-matrix, it can now be illustrated visually which variables that share relationships with other variables, by constructing a J-relationship graph. By choosing a convenient critical value of J, variable pairs that have a J-value larger than the critical value, are plotted with lines between each other to illustrate their relationship. This will give a map that shows which variables that should be further analyzed. Choosing the critical value for J to use is up to researcher. The ideal case is when clear components appear and all variables within each group are connected together. Each component should preferably consist of quite few variables.

If the value of J is very small many relationships between the variables is also small and the components usually grow very big. With decreasing J-value the components grow larger and larger and when J equals zero, all variables are connected together. As stated earlier, the ideal case is a set of small and isolated components containing strong relationships. This is obviously not always the case and what critical J-value to choose depends on the structure of the groups that appear. A tool in choosing the critical J-value can be found by creating a histogram for all J-values in the joint entropy matrix. The histogram might reveal if there is any gap between two values that would indicate a natural critical J-value.

When the critical value of J is to be determined, one should keep in mind that variables can be explained from other variables. If a variable can be explained from other variables it might be dropped from further analysis and the structure of the J-relationship graph is thereby changed. A component may for instance, be split to two smaller components if a variable that connects these components can be dropped. Examining functional relationships and determining the critical value of J is a simultaneous process where both processes benefit from each other.

2.3.4 Functional relationships

To get a measure of influence for $H_{i,j}$ on H_i there is two different matrices that can be used, both derived from the entropy matrix. They present the same information but the difference is that one presents it in absolute values and the other in relative values.

Consider the left side of inequality C again:

$$H_i \leq H_{i,j}$$

If the inequality is equal it means that X_i is a function of X_j. If equal, the difference between them are zero and the ratio will be one. This gives the two different formulas for finding the influence among the variables.

$$\overrightarrow{A}_{i,j} = H_{i,j} - H_i \qquad \textit{Equation 6.1}$$

And

$$A_{i,j} = \frac{H_i}{H_{i,j}} \qquad \textit{Equation 6.2}$$

Which share the attributes of:

$$0 \leq \overrightarrow{A}_{i,j} \leq H_i \qquad \textit{Inequality D.1}$$

$\overrightarrow{A}_{i,j}$ takes the value zero for functional dependency

$$0 \leq A_{i,j} \leq 1 \qquad \textit{Inequality D.2}$$

$A_{i,j}$ assumes the value 1 if X_i explains X_j, meaning, X_j is a function of X_i $(X_i \rightarrow X_j)$.

With two univariate entropies for every bivariate entropy $\left(A_{i,j} = \frac{H_i}{H_{i,j}}\ ; A_{j,i} = \frac{H_j}{H_{i,j}}\right)$ there are two A-values for each variable pair and $A_{i,j}$ are not the same as $A_{j,i}$. The first variable in the notation for every A-value is hence the variable explaining the other. The values are to be presented in the A-matrix. The A-matrix will not be symmetric and it has to be examined for the highest/lowest values on the matrix at the whole. If a high/low value is found the other part of the pair should be examined as well. This gives a clue if X_i explains X_j, or if X_j explains X_i or if the explanation is equally strong in both directions suggesting they have a mutual functional relationship. Since the A-values give directional functional relationships the direction can be included in the J-relationship graph, which will be of help in further analysis. If the A-value is 1 or close to 1 the variable explained can be seen as redundant.

Note that if the matrix is computed using the absolute values of *Equation 6.1* Values close to zero indicate that a functional relationship exists compared to using the relative measure where a value close to one indicate functional relationship.

2.3.5 Trivariate entropy analysis

For trivariate entropy analysis the methodology is very similar to the methodology of the bivariate entropies. Because of the extra variable added in the equation, there are more relationships to be examined. There are a total of $\binom{m}{3}$ trivariate entropies in the whole dataset and this quickly makes the analysis very time consuming if all trivariate sets were to be analyzed. Therefore the J-relationship graph could be consulted to give a hint of which variables to further examine. If three variables for example are linked together, there is some form of relationship between them and they should therefore be further analyzed. If the univariate and bivariate entropies can explain the data there is no need for trivariate or higher order entropies. Another reason for not calculating trivariate, and higher order entropies, may be the limitation of the observations. If three variables each have three different outcomes there are $3^3 = 27$ different outcomes. A rule of thumb is that there should be at least 5 possible observations for each outcome. In this case there would have to be at least $27 \cdot 5 = 135$ observations total. This can quickly decrease the possibility of higher order entropy research especially if there are variables with a vast number of outcomes.

Each identified trivariate relationship generates nine inequalities to analyze.

$$H_{i,j,k} \geq H_{i,j} \qquad \textit{eq if } (X_i, X_j) \rightarrow X_k \qquad 3\binom{m}{3} \qquad \textit{Inequality E.1}$$

The inequalities of E.1 examines whether any of the two variables explains the third. Hence it is a trivariate A-value. The difference, or ratio, of the inequality is examined in the same manner as with the bivariate A-values to see if any functional relationships are present. It is just an expanded version of the $\overrightarrow{A}_{i,j}$ presented by *Equation 6.1* or the ratio measure of *Equation 6.2.*

$$H_{i,j,k} \leq H_{i,j} + H_k \qquad \textit{eq if } (X_i, X_j) \perp X_k \qquad 3\binom{m}{3} \qquad \textit{Inequality E.2}$$

Inequalities E.2 can be thought of as a trivariate J-value, $J_{(i,j),k}$, which measures the degree of association between the pair (X_i, X_j) on (X_k). The difference equals the magnitude of the relationship between the two variables in the variable pair and the third variable. If the difference is zero they are independent.

$$H_{i,j,k} \leq H_{i,k} + H_{j,k} - H_k \qquad \textit{eq if } (X_j \perp X_k | X_i) \qquad 3\binom{m}{3} \qquad \textit{Inequality E.3}$$

Inequality E.3 measures the conditional structure between the variables. The set of variables that best explains the conditional structure assumes a difference close to zero. The measure can be thought of as a J-value for conditional structures, $J_{j,k|i}$.

2.3.6 Tetravariate and higher order entropies

Tetravariate and higher ordered entropies are calculated in the same manner as with bi- and trivariate entropies. There will be $\binom{m}{4}$ possible combinations of tetravariate entropies in the data set and each of them can be analyzed by 29 different formulas. As higher order relationships are examined the complexity of the analysis increases as well. The analysis examines the same relationships as the nine formulas presented for the trivariate entropies (*Inequality E.1-E.3*) but the inequalities have become 29 instead of 9 because of the extra variable.

$H_{i,j,k,l} \geq H_{i,j,k}$	*eq if* $(X_i, X_j, X_k) \rightarrow X_l$	$4\binom{m}{4}$	*Inequality D.1*
$H_{i,j,k,l} \leq H_{i,j,k} + H_l$	*eq if* $(X_i, X_j, X_k) \perp X_l$	$4\binom{m}{4}$	*Inequality D.2*
$H_{i,j,k,l} \leq H_{i,j} + H_{k,l}$	*eq if* $(X_i, X_{j,}) \perp (X_k, X_l)$	$\binom{m}{4}\frac{1}{2}\binom{4}{2}$	*Inequality D.3*
$H_{i,j,k,l} \leq H_{i,j,l} + H_{k,l} - H_l$	*eq if* $(X_i, X_j) \perp X_k \mid X_l$	$\binom{m}{4}\binom{4}{2}2$	*Inequality D.4*
$H_{i,j,k,l} \leq H_{i,j,k} + H_{i,j,l} - H_{i,j}$	*eq if* $X_k \perp X_l \mid (X_i, X_j)$	$\binom{m}{4}\binom{4}{2}$	*Inequality D.5*

There is no given stop for how far the researcher can go in analyzing higher ordered entropies. What determines when there is no matter searching for relationships in higher order entropies depends on different factors. If bi- and trivariate entropies can explain the data, there is no idea to go for higher order entropy in the analysis. There might also be restrictions from the sample size. Since there should roughly be at least five possible observations for each outcome this criteria quickly gets violated if sample size is small and high order entropies are calculated.

The researcher should also take into account the analysis burden for high order entropies. When pentavariate entropies are examined the analysis has been reduced to only check for functional relationships and independence and analysis for conditional independence have been dropped for simplicity reasons. This means that there are $\binom{m}{5}$ possible combinations of pentavariate entropies and 10 calculations for each set, with analysis performed for functional dependencies and for independence, equivalent to *inequalities D.1* and *D.2*.

3 Application

3.1 The data

Data used for analysis are contributed by the business intelligence firm United Minds. They were collected the year 2011 in collaboration with the media company Metro International and were used to create The Metropolitan Report, a magazine with the purpose of helping businesses understand urban consumers around the world. The survey investigated the lifestyles, opinions, values and habits of urban citizens. Data were collected through an online survey in 29 large cities worldwide and Metro International provided the panelists. The sample contained 14 961 respondents with approximately 500 respondents in each city. Some of the cities included were Beijing, Montreal, Paris, Sydney and Copenhagen.

The respondents were self-recruited and the sample is hence not random. This limits the extent to which conclusions from the analysis can be drawn on the population. The main purpose of the study is however not to make inference but to give a general understanding of the technique and evaluate its ability to analyze marketing research data.

3.1.1 Variables

The total number of variables in the dataset was originally 261. In order to illustrate the entropy method, a reduced set of variables was considered. In the process of selecting variables, all multiple-select questions were first removed together with variables with missing observations. All variables in the survey were part of question blocks in which all questions aimed to explore a certain aspect. All questions belonging to the same block were answered sequentially and had the same range and set of answer alternatives. In order to make the analysis more comprehensible, the variables used should preferably belong to the same blocks. For this reason, question blocks that earlier contained variables with missing observations had to be removed from the dataset.

The remaining number of question blocks was 10 with a total of 97 variables. Each block examined different types of aspects such as life satisfaction, attitudes towards cosmetic surgery, cultural habits, media habits, use of technology and reading habits regarding the Metro newspaper. To limit the extent of the analysis, the 4 blocks that presumably would give the best relationships were selected. The question blocks chosen were (1) the basic demographic background variables, (2) life satisfaction, (3) work related opinions and habits and (4) cultural habits. The selection process had limited the number of variables to 30, which is a practical and comprehensible amount for the analysis.

3.1.2 Aggregation

The variables did originally have a variety of ranges from 2 to 7 (with an exception for the variable *city* with a range of 29). The aggregation procedure was conducted on the basis of the frequency distribution, the underlying logic of the answer alternatives and the homogeneity of answers alternatives within the question blocks. The answers with a low frequency were to be merged together to create bigger representation within the categories. In doing so however, the underlying logic of the categories were not to conflict with each other. The question “how often do you go to the cinema?” for instance, had categories ranging from “every day” to “less often than once every year”. The answers “every day” and “less often than once every year” could not be merged in this case since the underlying logic of the answers was conflicting.

Name	Questions/Statement	Answser alternative
Y1	Which city do you live in?	{1 to 29 - 29 Cities }
Y2	I often socialize with my colleagues off work	{1 - Disagree, 2 - Neutral, 3 - Agree}
Y3	My salary is the most important measure of success in my career	{1 - Disagree, 2 - Neutral, 3 - Agree}
Y4	I often work somewhere other than my office/workplace	{1 - Disagree, 2 - Neutral, 3 - Agree}
Y5	I often work outside my actual working hours	{1 - Disagree, 2 - Neutral, 3 - Agree}
Y6	Go to a pub or bar	{1 - At least once a month, 2 - Between once a month and once a year, 3 - Less often}
Y7	Go to a sports event	{1 - At least once a month, 2 - Between once a month and once a year, 3 - Less often}
Y8	Your career/work	{ 1 - Dissatisfied, 2 - Neutral, 3 - Satisfied}
Y9	Your appearance/body	{ 1 - Dissatisfied, 2 - Neutral, 3 - Satisfied}
Y10	Go to theatres, art galleries or other cultural institutions	{1 - At least once a month, 2 - Between once a month and once a year, 3 - Less often}
Y11	I feel I have a good work/life balance	{1 - Disagree, 2 - Neutral, 3 - Agree}
Y12	Go to a night club	{1 - At least once a month, 2 - Between once a month and once a year, 3 - Less often}
Y13	Have breakfast on the go	{1 - At least once a month, 2 - Between once a month and once a year, 3 - Less often}
Y14	Go to the cinema	{1 - At least once a month, 2 - Between once a month and once a year, 3 - Less often}
Y15	Go to a music concert	{1 - At least once a month, 2 - Between once a month and once a year, 3 - Less often}
Y16	Your health	{ 1 - Dissatisfied, 2 - Neutral, 3 - Satisfied}
Y17	Spend time in a park or public garden (in season)	{1 - At least once a month, 2 - Between once a month and once a year, 3 - Less often}
Y18	Your life as a whole	{ 1 - Dissatisfied, 2 - Neutral, 3 - Satisfied}
Y19	I always keep an eye open for new job/educational opportunities	{1 - Disagree, 2 - Neutral, 3 - Agree}
Y20	Go to a fast food restaurant	{1 - At least once a month, 2 - Between once a month and once a year, 3 - Less often}
Y21	My work is a part of who I am, not just a way to make money	{1 - Disagree, 2 - Neutral, 3 - Agree}
Y22	Have dinner at a restaurant	{1 - At least once a month, 2 - Between once a month and once a year, 3 - Less often}
Y23	Visit a café or coffee shop	{1 - At least once a month, 2 - Between once a month and once a year, 3 - Less often}
Y24	Go shopping (not groceries)	{1 - At least once a month, 2 - Between once a month and once a year, 3 - Less often}
Y25	Gender	{1 - Female, 2 - Male}
Y26	Age	{1 - 18-34, 2 - 35-49}
Y27	Kids in household	{1 - Kids, 2 - No kids}
Y28	I believe formal education is important to	{1 - Disagree, 2 - Neutral, 3 - Agree}
Y29	Marital status	{1 - Single, 2- Not single}

Table 3.1, Variable list, Aggregated and selected variables.

In addition, the aggregation process should also aim to maintain the homogeneity of answer alternatives within the question blocks and all the variables in one block should ideally have the same set of answer alternatives. An exception was the block containing background variables, where it was not possible. The process of finding new sets of answer alternatives are more directed to finding the ones suiting the question blocks rather than the individual questions. After the aggregation process was completed, all variables had range 3 except for the four background variables with range 2 and the variable *city* that still had range 29.

The frequency distribution of the variable *work status* had very little variation and the majority of the respondents had answered the work status *Employed*. Therefore this variable was excluded from further analysis. The variables remaining after the selection- and aggregation process is presented together with their range in *Table 3.1*.

3.1.3 Calculating entropies

Software used in performing the entropy analysis is SAS, SPSS and Microsoft Excel. The dataset was delivered to the authors in the form of a SPSS-file and the authors also had experience of working with SPSS. Hence, the preceding process of selecting variables and aggregating answer alternatives have mainly been performed in SPSS. In calculating the entropies, SAS has been used. After calculating the entropies most of the analysis has been executed in Excel.

3.2 The Analysis process

3.2.1 Entropy matrix

Initially, the univariate entropies were calculated and sorted based on their size and renamed from Y1 to Y29, hence, $H_1 \geq H_2 \geq \cdots \geq H_{29}$. The bivariate entropies were then calculated and the entropy matrix was created. The entropy matrix, presented in *Table 3.2,* is the foundation that most analysis starts from and is presented below.

	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y13	Y14	Y15	Y16	Y17	Y18	Y19	Y20	Y21	Y22	Y23	Y24	Y25	Y26	Y27	Y28	Y29
Y1	**486**																												
Y2	636	**158**																											
Y3	638	314	**158**																										
Y4	638	310	313	**157**																									
Y5	637	308	311	293	**155**																								
Y6	632	308	312	310	308	**154**																							
Y7	634	309	310	308	307	303	**153**																						
Y8	635	309	309	308	307	307	305	**153**																					
Y9	634	308	309	308	306	305	304	297	**152**																				
Y10	632	306	308	306	304	299	299	302	301	**151**																			
Y11	631	303	305	304	302	303	301	290	297	299	**149**																		
Y12	625	300	303	301	300	277	293	298	296	291	294	**146**																	
Y13	623	300	301	299	297	296	294	296	295	293	292	286	**144**																
Y14	623	299	301	299	298	293	291	295	294	281	291	283	285	**143**															
Y15	623	296	298	295	294	286	283	292	291	272	288	276	281	274	**140**														
Y16	621	295	296	295	293	292	291	284	273	289	283	284	282	281	278	**138**													
Y17	615	291	292	290	289	287	285	287	285	278	283	278	276	275	270	273	**135**												
Y18	616	290	291	290	288	288	287	267	271	283	273	279	277	276	273	258	267	**134**											
Y19	615	288	290	286	285	287	286	286	284	283	280	278	276	276	273	271	267	266	**133**										
Y20	612	290	290	289	288	286	284	285	284	283	281	277	272	273	272	271	267	267	266	**133**									
Y21	615	286	288	286	284	286	285	277	282	282	275	278	276	275	272	269	266	263	263	265	**132**								
Y22	607	288	289	288	287	280	282	284	283	278	279	273	274	267	268	270	266	265	265	258	264	**132**							
Y23	605	283	286	284	282	271	279	280	279	272	276	267	268	265	263	266	259	261	260	258	259	252	**128**						
Y24	587	262	262	261	259	258	257	257	256	253	253	249	248	245	243	243	238	238	238	236	237	234	230	**105**					
Y25	585	258	257	256	254	253	250	253	251	251	249	245	244	243	240	238	234	234	233	233	232	232	228	204	**100**				
Y26	585	257	258	256	255	253	253	253	251	251	249	243	243	242	240	238	234	234	233	232	232	232	228	205	200	**100**			
Y27	582	257	257	255	254	252	252	251	251	250	248	244	242	242	239	237	233	232	232	232	231	231	227	203	199	194	**99**		
Y28	577	252	253	252	250	250	249	248	247	247	243	242	240	239	236	234	230	229	227	229	225	228	224	201	196	196	195	**96**	
Y29	573	248	248	246	244	243	243	242	241	240	238	235	234	233	230	228	224	222	223	223	222	222	218	194	190	185	178	186	**90**

Table 3.2, Entropy matrix, or H-matrix, rounded to integer per cent.

The matrix contains 29 univariate entropies and $\binom{29}{2} = 406$ bivariate entropies. The values in the H-matrix represent the bivariate entropies with the univariate entropy on the diagonals. From the entropy matrix the J-matrix, using *Equation 5*, is easily calculated which is the next step in the analysis.

3.2.2 J-matrix

The J-matrix presents the J-values between two variables. The diagonal values display the univariate entropy just as the entropy matrix does and are hence not displayed. As the J-value is a sort of measure of association between variables, high or low values are of interest at since this indicates that variables share strong relationships or are independent. Note that J-value does not give any information about the nature of the relationship, only that it is present or not. The matrix is presented with values in integer percent, which is sufficient accuracy in this case. The J-values are calculated from the original H-matrix and the J-values have then been rounded down, some values are consequently not consistent between the H- and the J-matrix.

	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y13	Y14	Y15	Y16	Y17	Y18	Y19	Y20	Y21	Y22	Y23	Y24	Y25	Y26	Y27	Y28
Y2	7																											
Y3	5	1																										
Y4	3	4	1																									
Y5	3	4	1	18																								
Y6	7	3	0	0	0																							
Y7	5	2	0	1	0	4																						
Y8	2	1	1	1	0	0	0																					
Y9	2	1	0	0	0	0	0	7																				
Y10	3	2	0	1	1	5	5	1	1																			
Y11	3	3	2	1	1	0	0	11	3	0																		
Y12	6	4	0	1	0	23	6	0	1	5	0																	
Y13	6	1	0	1	1	1	2	0	0	1	0	3																
Y14	5	2	0	0	0	4	5	0	0	12	0	6	2															
Y15	3	2	0	1	0	8	9	0	0	19	0	9	2	8														
Y16	3	1	0	0	0	0	0	7	17	0	4	0	0	0	0													
Y17	5	1	0	0	0	2	3	0	0	6	0	2	2	3	4	0												
Y18	3	1	0	0	0	0	0	19	13	1	9	0	0	0	1	14	0											
Y19	3	3	1	4	2	0	0	0	1	1	1	0	0	1	0	1	0	0										
Y20	6	1	1	0	0	1	2	0	0	0	0	1	4	3	0	0	0	0	0									
Y21	3	3	2	2	3	0	0	8	1	1	6	0	0	0	0	1	0	3	2	0								
Y22	11	2	0	0	0	6	3	1	0	5	1	4	2	8	4	0	1	1	0	6	0							
Y23	8	3	0	1	1	11	3	0	0	6	0	6	3	6	5	0	3	0	1	3	1	8						
Y24	3	0	0	0	0	1	1	0	0	2	0	1	0	2	1	0	1	0	0	2	0	3	2					
Y25	0	0	0	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0				
Y26	0	0	0	0	0	1	0	0	0	0	0	3	0	1	0	0	0	0	0	0	0	0	0	0	0			
Y27	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4		
Y28	5	1	1	1	1	0	0	0	0	0	2	0	0	0	0	1	0	1	2	0	3	0	0	0	0	0	0	
Y29	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	4	10	0

Table 3.3, Joint entropy matrix, or J-matrix, rounded to integer percent

As can be seen in the J-matrix, the strongest relationship is found between variables Y6 and Y12 with the J-value of 23 %. The lowest J-value is found between variables Y26 and Y18 implying that these variables are independent from each other. The J-matrix is used for creating the J-relationship graph, which presents the structure of the variables and will work as an aid in the future analysis.

3.2.3 Critical J-value

When the J-relationship graph is created, a critical J-value must be determined. In choosing the critical J-value, the frequency distribution histogram could be used. The J-values in the dataset are distributed as follow:

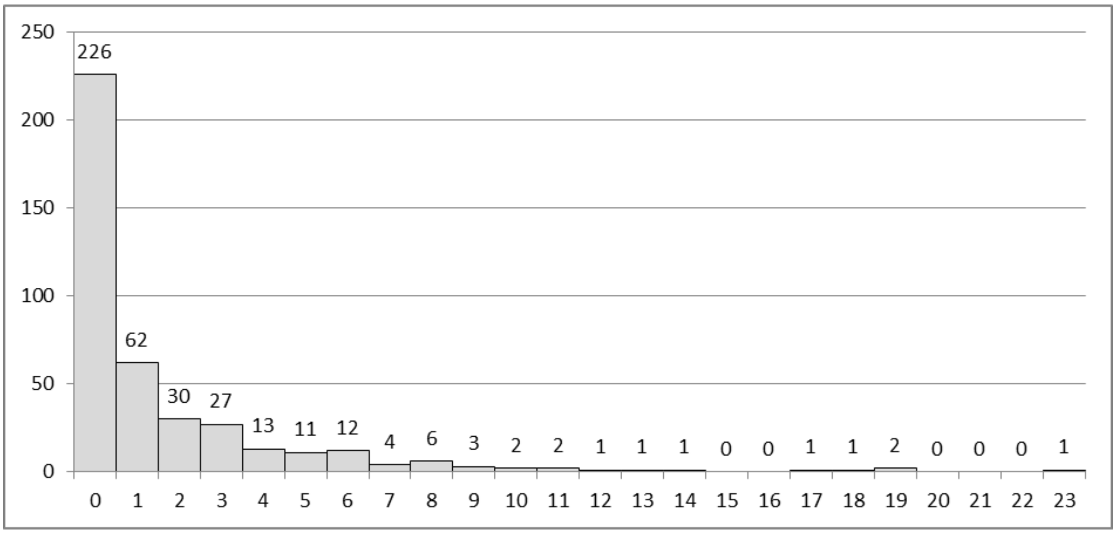

Table 3.4, J-value frequency histogram

The histogram gives some important information. For example are there 226 relationships with J-value smaller than 1 in the dataset meaning there are 226 variable pairs with variables that are independent or close to independent from each other. A clear cut in the histogram could indicate a good critical J-value. The histogram for the data does not however give a clear indication for a good critical value but suggests that it might lie somewhere between 6 and 9 %. The J-relationship graph will have to be plotted for different critical J-values to see which one that seems to explain the data best.

3.2.4 J-relationship graph

From the J-matrix, the J-relationship graph is created with variables shown as vertices with the variable name inside. Relationships between variables are then illustrated with lines between the vertices that are related to each other. The J-relationship graph visualizes the structure in the data and clusters the variables related to each other into components. The critical J-value is chosen so that it generates the best J-relationship graph and different J-values are hence tested. The J-relationship graphs for J-values 6 to 9 are presented in *Figure 3.1* to 3.*4*. The relationships added at each J-value are illustrated with dotted lines.

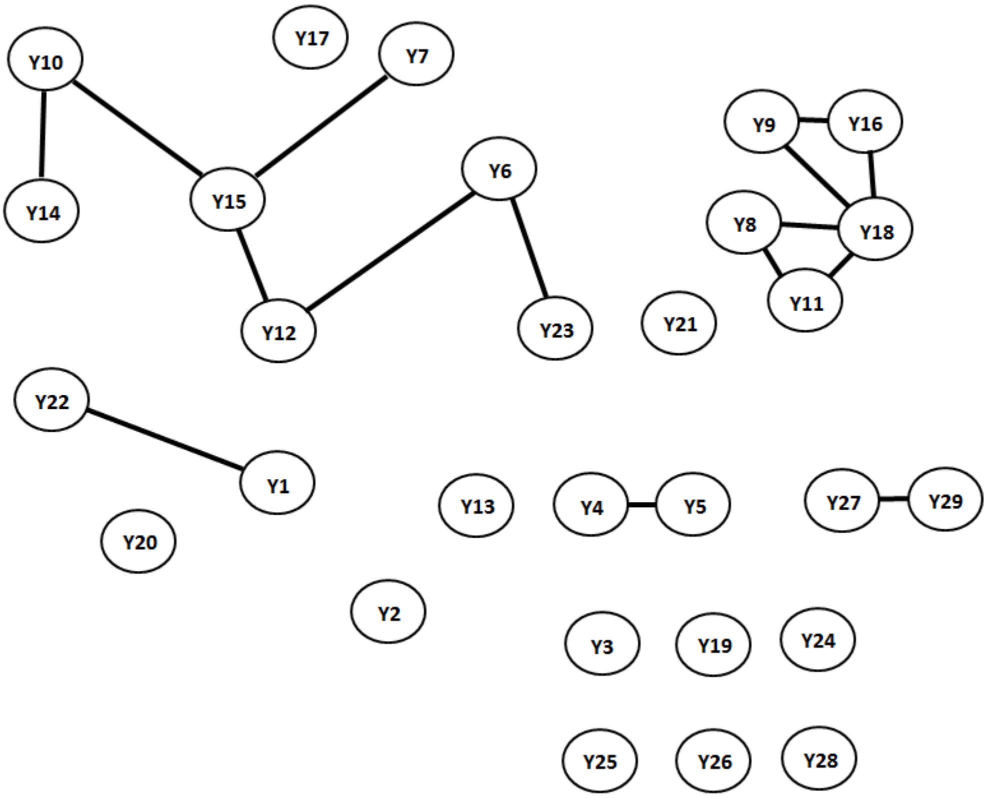

Figure 3.1, Relationship graph for J ≥ 9

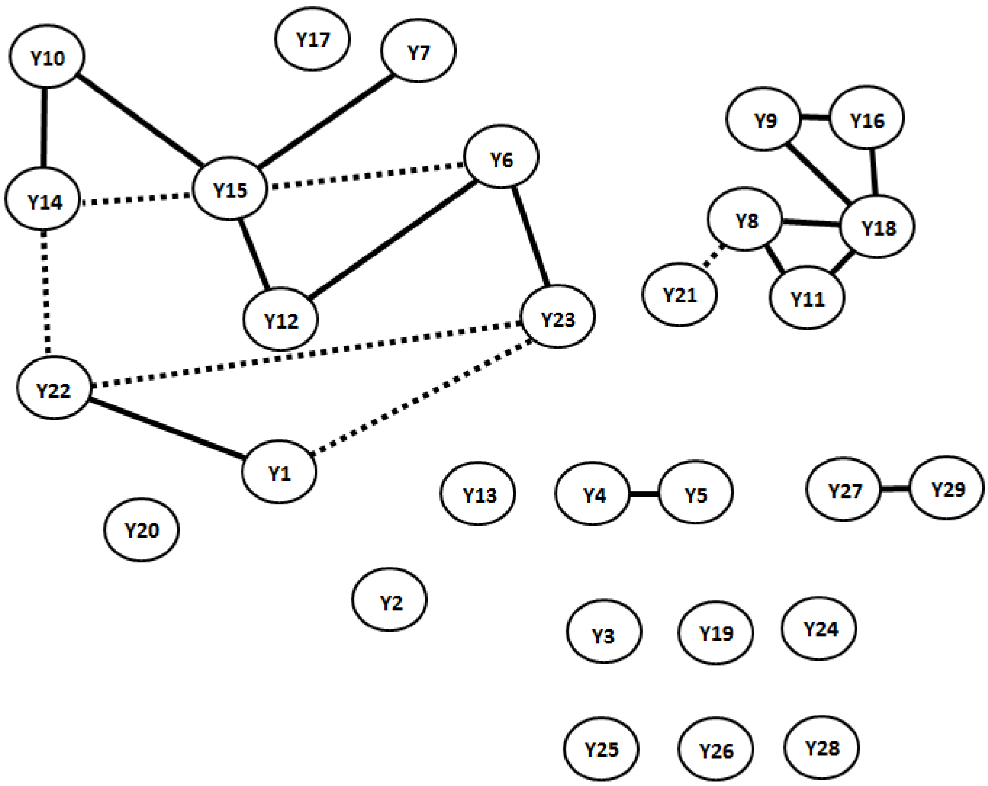

Figure 3.2, Relationship graph for J ≥ 8

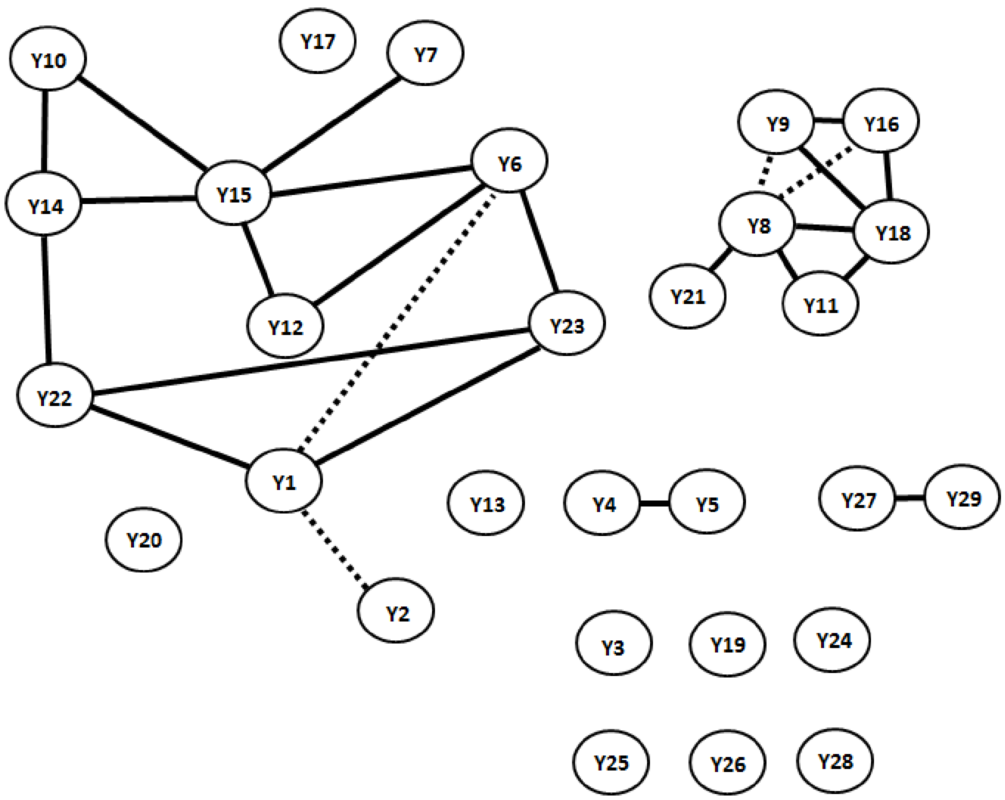

Figure 3.3, Relationship graph for J ≥ 7

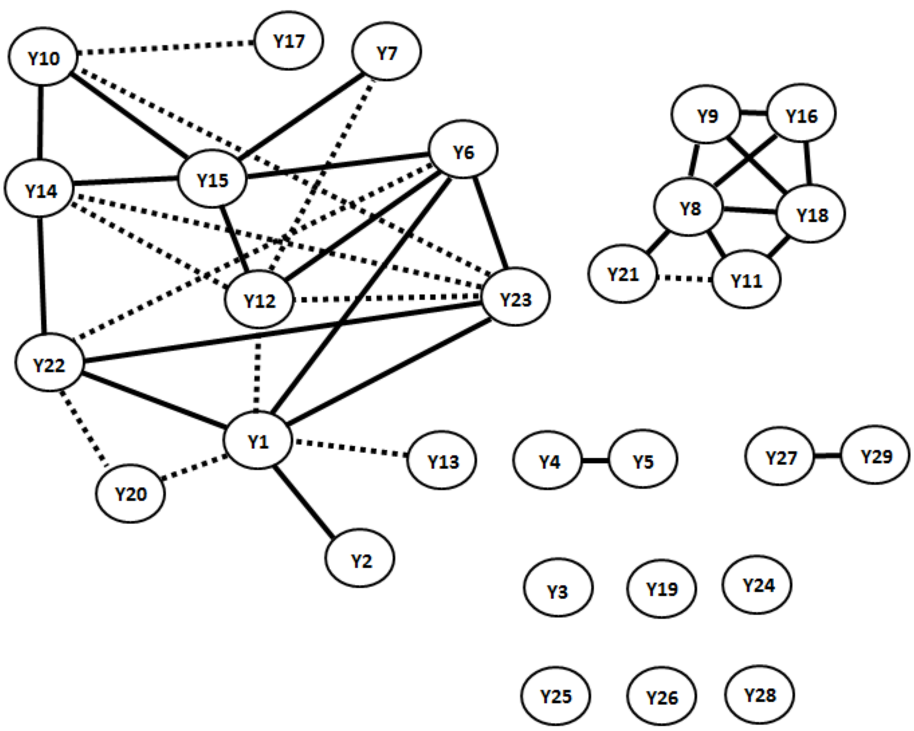

Figure 3.4, Relationship graph J ≥ 6

The number of relationships entered for additional values on J can also be viewed in *Table 3.5*. The aim is to get strong relationships within the components without adding relationships with new variables or having the components interconnect. It is also preferable not to get too big components as the purpose is to simplify the data. It is therefore better to choose a model with two components with two variables in each that have strong relationships, than one component with four variables that have weak relationships. As can be seen in *Figure 3.1* (J≥9) the relationships within the components are quite strong and the number of variables within the components is also quite small. In *Figure 3.2* (J≥8), one can see that relationships between the components start to emerge and in *Figure 3.3* (J≥7) and 3*.4* (J≥6) the components grow even larger in size. Consequently, the critical J-value selected is ≥9.

J-value	Number of relationsships added	Cumulative number of relationships
23	1	1
19	2	3
18	1	4
17	1	5
14	1	6
13	1	7
12	1	8
11	2	10
10	2	12
9	3	15
8	6	21
7	4	25
6	12	37
5	11	48
4	13	61
3	27	88
2	30	118
1	62	180
0	226	406

Table 3.5, Added relationships for each J-value

The data have now already been greatly simplified since each component can be treated individually compared to a 29 dimensional distribution as before. The J-relationship graph will help determine which variables that should be analyzed for higher order relationships. For example, the component containing Y8, Y9, Y11, Y16 and Y18 shows potential for having relationships when higher ordered entropies are calculated. One can see that Y8, Y9, Y11 and Y16 all share strong relationships with Y18 and this implies that these variables may be a good start for further analysis.

3.3 Functional relationships

The information provided by the J-relationship graph shows which variables that could share functional relationships. The A-matrix for these variables indicates that there are in general weak functional relationships between the variables, with an exception for Y1 and

Y22, where the A-values for $A_{1,22} = 80$ and $A_{22,1} = 22$. This tells us that there is a functional relationship from Y1 to Y22 where Y22 is explained by Y1 from having the lower A-value. Since Y22 is explained by Y1, Y22 could be considered redundant.

	Y1	Y4	Y5	Y6	Y7	Y8	Y9	Y10	Y11	Y12	Y14	Y15	Y16	Y18	Y22	Y23	Y27	Y29
Y1		25	24	24	24	24	24	24	24	23	23	22	22	22	22	21	17	16
Y4	76		53	50	50	50	49	49	49	48	48	47	47	46	46	45	39	36
Y5	76	53		50	50	50	50	49	49	49	48	48	47	46	46	45	39	37
Y6	77	51	50		51	50	50	50	49	53	49	49	47	46	47	47	39	37
Y7	77	51	50	51		50	50	50	49	50	49	49	48	47	47	46	39	37
Y8	76	51	50	50	50		51	50	51	49	49	48	49	50	47	46	39	37
Y9	77	51	51	50	50	51		50	50	49	49	48	51	49	47	46	40	37
Y10	77	51	51	51	51	50	50		50	50	51	52	48	47	48	47	40	37
Y11	77	52	51	51	51	53	51	50		50	49	49	49	49	47	46	40	38
Y12	78	52	52	56	52	51	51	52	51		51	51	49	48	48	48	41	38
Y14	78	52	52	53	53	52	52	54	51	52		51	49	48	49	48	41	39
Y15	78	53	53	54	54	52	52	55	52	53	52		50	49	49	49	41	39
Y16	78	53	53	53	53	54	56	52	53	51	51	50		52	49	48	42	39
Y18	79	54	54	54	54	57	56	53	54	52	52	51	54		50	49	43	40
Y22	80	54	54	55	54	54	54	54	53	53	54	52	51	51		51	43	40
Y23	80	55	55	57	55	54	54	55	54	55	54	53	52	51	53		44	41
Y27	83	61	61	61	61	61	60	60	60	60	59	59	58	58	57	56		50
Y29	85	64	63	63	63	63	63	63	62	62	62	61	61	60	60	59	56	

Table 3.6, A-matrix for dependent variables

3.3.1 Bivariate functional relationships

The functional relationships between variables can be examined using frequency tables. The tables would show a dominating value in the cells for each row of the explanatory variable explaining how one variable explains the other. This can be seen in the bivariate frequency distribution for variables Y1 and Y22 in *Table 3.7* where distribution reveals a determining pattern between the variables.

Frequency distribution for Y1 and Y22					
		Y22			
		1	2	3	Total frequency
Y1 Row %	1	63%	4%	33%	**514**
	2	64%	7%	29%	**513**
	3	92%	2%	6%	**500**
	4	48%	11%	41%	**508**
	5	23%	30%	47%	**506**
	6	62%	6%	32%	**447**
	7	29%	11%	60%	**507**
	8	88%	2%	10%	**521**
	9	60%	9%	31%	**504**
	10	73%	6%	22%	**506**
	11	64%	5%	31%	**519**
	12	57%	9%	33%	**502**
	13	64%	7%	28%	**503**
	14	61%	7%	31%	**518**
	15	73%	4%	24%	**506**
	16	29%	32%	38%	**570**
	17	73%	6%	22%	**501**
	18	61%	7%	32%	**533**
	19	31%	20%	49%	**503**
	20	39%	7%	54%	**622**
	21	63%	5%	32%	**512**
	22	65%	9%	26%	**511**
	23	59%	10%	31%	**459**
	24	56%	6%	39%	**502**
	25	47%	10%	43%	**502**
	26	43%	10%	47%	**533**
	27	26%	28%	46%	**527**
	28	63%	7%	30%	**516**
	29	69%	3%	27%	**596**

Table 3.7, Bivariate frequency distribution for Y1 and Y22

3.3.2 Higher order functional relationships

The previous example examined the strongest functional bivariate relation found in the data. If no fully functional relations have been found between two variables one can analyze the trivariate, or higher, relationships that exists. The J-relationship graph is used to determine which variables that might share higher ordered functional relationships. One of the strongest trivariate functional relationships found in the data was the degree of explanation (Y8,Y9) had on Y18. The frequency distribution for this relationship is shown in *Table 3.8*.

The A-value for the degree of explanation on Y18 from the pair (Y8, Y9) is 74 %. The pair (Y8,Y16) has an almost equally strong relationship on Y18 as the pair (Y8,Y9) which indicates that Y16 might be included for a higher ordered relationship analysis on Y18. Hence, the three variables Y8, Y9, Y16 probably explain Y18 even better.

Trivariate frequency distribution for Y8, Y9 and Y18						
			Y18			
			1	2	3	Total
Y8, Y9	1 , 1	Frequency	782	358	206	**1346**
		Row %	58%	27%	15%	
	1 , 2	Frequency	352	463	334	**1149**
		Row %	31%	40%	29%	
	1 , 3	Frequency	183	310	470	**963**
		Row %	19%	32%	49%	
	2 , 1	Frequency	235	365	306	**906**
		Row %	26%	40%	34%	
	2 , 2	Frequency	182	1021	768	**1971**
		Row %	9%	52%	39%	
	2 , 3	Frequency	89	418	1209	**1716**
		Row %	5%	24%	70%	
	3 , 1	Frequency	127	221	538	**886**
		Row %	14%	25%	61%	
	3 , 2	Frequency	80	348	1392	**1820**
		Row %	4%	19%	76%	
	3 , 3	Frequency	60	301	3843	**4204**
		Row %	1%	7%	91%	
	Total	Frequency	**2092**	**3808**	**9070**	**14969**

Table 3.8, Trivariate frequency distribution for Y8, Y9 and Y18

3.3.3 Independent variables

In addition to the five multivariable components, many components contain only one variable and are independent or approximately independent from the other variables. These variables, illustrated in *Figure 3.1*, have no relationships with other variables when the critical J-value is 9. One can further examine the independence between variables using the frequency table were no value should be dominating for each row if the variables share no relationship.

3.3.4 Variable independence

In *Table 3.9* the bivariate frequency table between the two variables Y18 and Y26 is shown. The J-value for this pair is 0 and one can also see in the relationship graph in *Figure 3.1* that there is no relationship among these variables. In the *Table 3.9*, the row percentage between the answer alternatives are almost equal indicating no relationship present, confirming the independence in the J-relationship graph to be accurate.

	Bivariate frequency distribution for Y18 and Y26				
			Y26		
			1	2	Total
Y18	1	Frequency	1061	1029	**2090**
		Row %	51%	49%	
	2	Frequency	1937	1868	**3805**
		Row %	51%	49%	
	3	Frequency	4608	4458	**9066**
		Row %	51%	49%	
	Total	Frequency	**7606**	**7355**	**14961**

Table 3.9, Bivariate frequency distribution for Y18 and Y26

If a model is good, variables in different components should have no or a relationship between each other. *Table 3.10* shows the bivariate frequency table of variables Y6 and Y8, two variables that are part of different components and therefore should share no or a very weak relationship.

	Bivariate frequncy distribution for Y6 and Y8					
			Y8			
			1	2	3	Total
Y6	1	Frequency	1484	2001	3191	**6676**
		Row %	22%	30%	48%	
	2	Frequency	934	1218	1578	**3730**
		Row %	25%	33%	42%	
	3	Frequency	1040	1374	2141	**4555**
		Row %	23%	30%	47%	
	Total	Frequency	**3458**	**4593**	**6910**	**14961**

Table 3.10, Bivariate frequency distribution for Y6 and Y8

3.3.5 Conditional independence

The conditional structure in the components is examined by analyzing the trivariate entropies. If the value attained from using *Inequality E.3* is close to zero, there is a conditional independence within the triad. By knowing the conditional structure the data can be further simplified and easier interpreted.

			Y6=1				Y6=2				Y6=3			
			y23				y23				y23			
			1	2	3	Total	1	2	3	Total	1	2	3	Total
Y7	1	Frequency	1678	70	226	**1974**	268	116	142	**526**	508	69	214	**791**
		Row %	85%	4%	11%		51%	22%	27%		64%	9%	27%	
	2	Frequency	1586	166	312	**2064**	789	781	651	**2221**	1084	254	840	**2178**
		Row %	77%	8%	15%		36%	35%	29%		50%	12%	39%	
	3	Frequency	2197	113	328	**2638**	482	225	276	**983**	989	139	458	**1586**
		Row %	83%	4%	12%		49%	23%	28%		62%	9%	29%	
Total		Frequency	5461	349	866	**6676**	1539	1122	1069	**3730**	2581	462	1512	**4555**

Table 3.11, Conditional frequency distribution for Y7 and Y23 on Y6

The conditional frequency table for Y6, Y7 and Y23 is illustrated in *Table 3.11* where Y7 should to be independent from Y23 given Y6. This relationship assumes the value 0.01 in *Table 12, Appendix A* and is hence the strongest conditional independence within the components. As can be seen in *Table 3.11,* Y7 and Y23 seems to be independent in each given tables for Y6, and it is thereby evident that Y7 and Y23 are independent given Y6.

3.4 Component structures

The component with the 7 variables *Y6, Y7, Y10, Y12, Y14, Y15* and *Y23* had $\binom{7}{3} = 35$ different combinations of trivariate entropies, presented in *Table 12, Appendix A*. Among these, the strongest functional relationships one can observe are on Y15 from Y10 and Y14 (1.25), and on Y23 from Y6 and Y12 (1.25). The calculated A-values for these relationships are though only 74 % and 73 % but although these relationships are quite weak, it gives a hint of what kind relationships one could find in the tetravariate analysis. The output from the tetravariate calculations are presented in *Table 14, Appendix A* where one can see that the relationships calculated is not too strong neither with the smallest value of 1.11 which gives the A-value 75 %. This suggests a very weak functional relationship between the triad Y6, Y10 and Y14, and the single variable Y23. After analyzing the pentavariate entropies it can be seen that variables Y7, Y10, Y12 and Y14 have a functional relationship on Y15 ($H_{6,7,10,12,14,15,23}$=6.696). The calculated A-value for those variables relation on Y15 is 84 % which is a much higher value than for the tri- and tetravariate entropies and the result is presented in *Figure 3.5.*

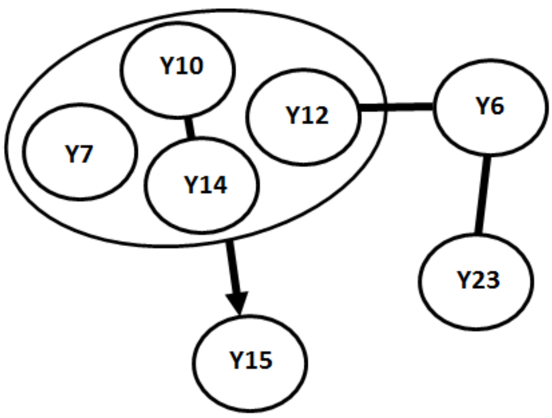

Figure 3.5, Component structure explaining Y15

Even though the data at the present state is smaller than the original dataset, one can simplify the data even further using higher order relationships. The best way of doing so is by finding functional relations so that variables can be dropped and hence change the structure of the component. If the researcher decides that the A-value of 84 % for the relationship illustrated in *Figure 3.5* is high enough, Y15 can be dropped. This would transform the component into one 3-dimensional component, one 2-dimensional and one single variable component. If this component were to be treated as a seven dimensional distribution the total response pattern would have been $3^7 = 2187$. Dropping Y15 changes the response patterns to having $3^3 + 3^2 + 3 = 39$ outcomes and has consequently simplified the data greatly.

In the component containing variables Y8, Y9, Y11, Y16 and Y18 there are 5 variables with $\binom{5}{3} = 10$ different combinations of trivariate entropies, presented in *Table 13, Appendix A*. The strongest trivariate functional relationship present is the quite weak relationship on Y18 from Y8 and Y9 (1.32) with an A-value of 74%. When analyzing higher ordered entropies the relationship amongst the variables Y8, Y9, Y11, Y16 and Y18 gets more evident, especially the tetravariate degree of explanation on Y18 from Y8, Y9 and Y16 (1.02). When calculating the pentavariate entropies ($H_{8,9,11,16}$=*6.479*) the A-value is 85 % for Y8, Y9, Y11, Y16 degree of explanation on Y18 (relationships is illustrated in *Figure 3.6*).

Since there are no more variables in the component, reaching for higher order relationships is not possible. If decided that an A-value of 85% is enough to consider the relationship as functional, the variable Y18 can be dropped in the same manner as earlier. Doings so would also have major consequences for the structure inside the component. If it were to be treated as a five dimensional distribution the total response pattern would have been$3^5 = 243$. But if Y18 is dropped, this splits the original component into two 2-dimensional components. The two 2-dimensional components would then together explain the original 243 possible response patterns with just $3^2 + 3^2 = 18$. The total structure and the underlying meaning of the variables will be further discussed in the forthcoming section.

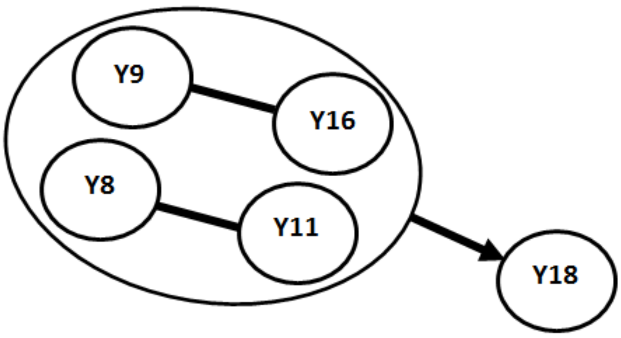

Figure 3.6, Component structure explaining Y18

3.5 Total structure

In *Figure 3.7* the total structure of the entropies is presented along with the variable's explanation. The components are also named according to their content.

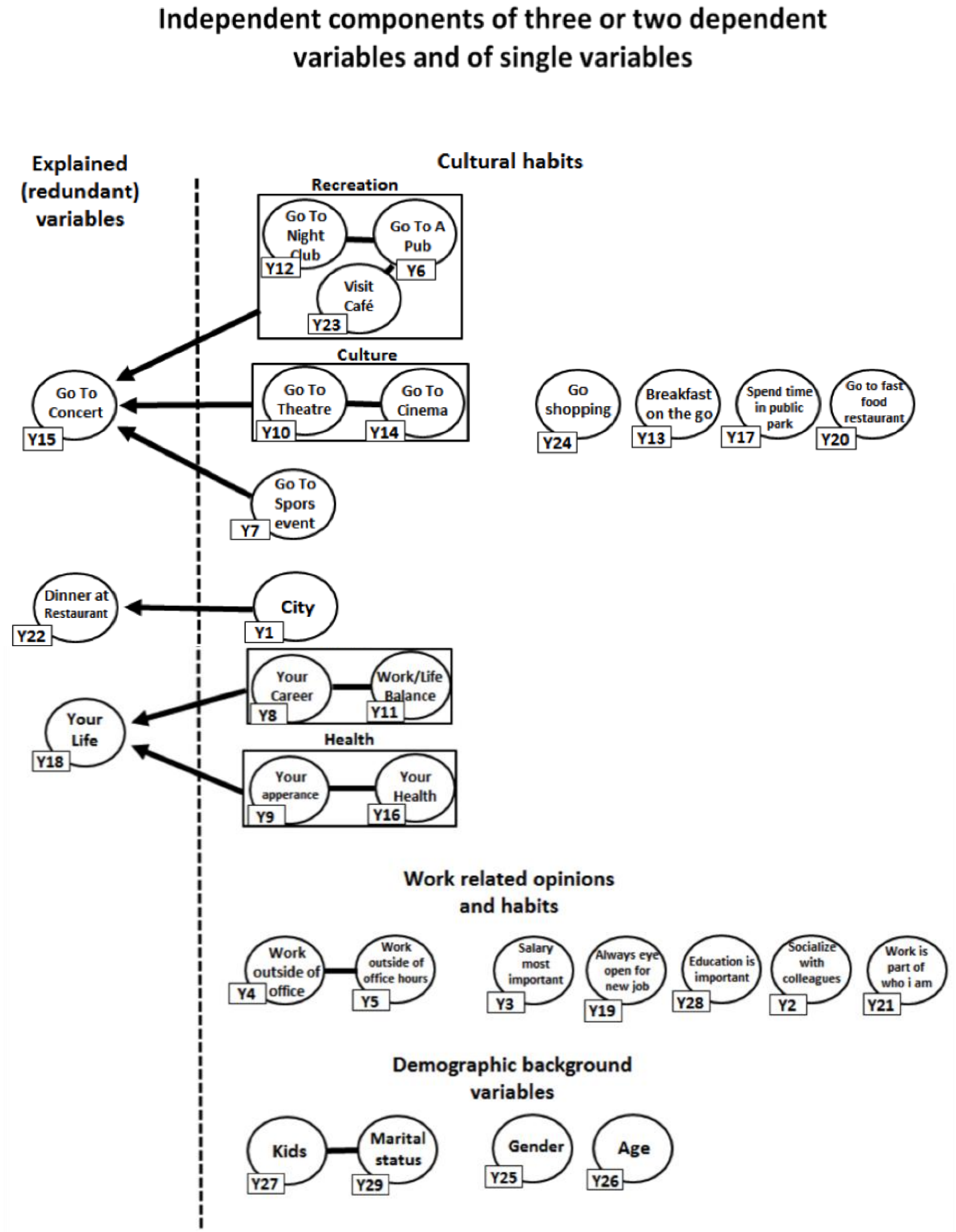

Figure 3.7, Total structure of components

The explained variables that can be considered redundant are shown to the left in *Figure 3.7*. Other variables are presented to the left of the dotted line and with the single independent variables to the right.

The relationships within the components tend to follow an intuitive pattern. One could for example assume that a respondent's satisfaction regarding his or her career would be related to the satisfaction regarding his or her work/life-balance, as the illustration shows. An exception is the component *Eat out*, which suggests that the city a person lives in explains how often that person eats dinner at restaurants. The two functional relationships between the components *Culture and Recreation* suggest that one can tell if a person goes to see concerts frequently or not, from his or her habits regarding visits to other cultural facilities, sport events and recreational habits. One can also predict a person's life satisfaction from his or her satisfaction regarding career and health.

The illustration also shows a bivariate relationship between how often a person work odd hours and how often he or she work from other places than the office. As one could have guessed, there is also a strong relationship between a person's marital status and parental status since people that get children tend to be married first.

The variables concerning how often one goes to concerts, eats at a restaurant and life satisfaction could from this model be considered redundant. These variables could therefore be left out from future use of the model since we can estimate these simply by knowing the variables explaining them. Presenting the data in this way compromises it into one 3-variable component, five 2-variable components and 13 single variable components and makes their relationships easy to grasp.

4 Conclusion

The aim of the study has been to give a general understanding of the method of entropy analysis and evaluate its ability to analyze marketing research data. The process of using entropy analysis consists of several important steps and it is in many ways a creative process that leads to attaining the final component structures. The result is however of great use when it comes to analyzing discrete variables and explaining the relationships between variables in a dataset. In this study, the data turned out to reveal several interesting and complex discoveries, some of which might have been hard to find using a conventional analysis technique.

The way in which to present relationships in relationship graphs is also of great help. This feature of the entropy analysis could make the method of great benefit in the area of marketing research where much effort is put into translating discoveries into comprehensible models. A relationship graph is also an instrument that can be developed in the future, both in its esthetical form and in the aspect of presenting more complex relationships.

Even though entropy analysis could reveal interesting findings, the analysis process is with the tools available today, quite time consuming. This could threaten the attractiveness of entropy analysis and developing software that facilitates the analysis is vital for the future use of the method.

In addition to these conclusions, there are also several practical aspects that need to be addressed in order to improve the practice of entropy analysis.

- The J-matrix used in this study has been in absolute values. If the J-values had been standardized this would have changed the J-values and consequently the rest of the analysis. The use of standardized J-values or absolute J-values needs further attention in future research.
- How to choose the critical J-value is now a quite subjective practice that determines which variables that is to be included or excluded in the components. By choosing a critical J-value one decides what variables are to be thought of as independent and which variables that are considered to be further examined. A more elaborate method for picking the best critical J-value could be of great use.
- The process of picking the critical J-value also usually comes with an information loss. At the present state the researcher is unable to know if the information loss is motivated by how much the data gets simplified. The use of measures that would calculate the information loss and how much simpler a data becomes would be of interest.
- It would be desirable if there was a measure that could help determine if a variable should be dropped or not. A measure like this could be based on how strong the relationship are towards the explained variable and how much simpler the structure would become. The use of helping measures is an area that could benefit from future research.

References list

Frank, O. *Statistical information tools for multivariate discrete data. Modern Mathematical Tools and Techniques in Capturing Complexity, edited by L.* Pardo, N. Balakrishnan and M. Angeles Gil, Springer-Verlag, Berlin Heidelberg, 177-190, 2011.

Frank, O. *Tid för nya metoder i tillämpad statistik.(Time for new methods in applied statistics.) Qvintensen*, 2, 14-16, 2011.

Appendix A

Table 12. Calculations for trivariate entropies (Variables Y6, Y7, Y12, Y14, Y15, Y23)

The variables being explained are sorted in ascending order and presented in the left column. The first row gives the inequality calculated and the second row explains the relationship among the variables in each inequality.

		Inequality D.1			Inequality D.2			Inequality D.3		
Variables	Entropy	$(X_1X_2) \rightarrow X_3$	$(X_1X_3) \rightarrow X_2$	$(X_2X_3) \rightarrow X_1$	$(X_1X_2)\ \ X_3$	$(X_1X_3)\ \ X_2$	$(X_2X_3)\ \ X_1$	$(X_2\ \ X_3) \mid X_1$	$(X_1\ \ X_3) \mid X_2$	$(X_1\ \ X_2) \mid X_3$
y6 * y10 * y12	**4,196**	1,20	1,43	1,29	0,26	0,08	0,25	0,02	0,20	0,02
y6 * y10 * y14	**4,279**	1,29	1,35	1,47	0,15	0,16	0,08	0,10	0,02	0,03
y6 * y10 * y15	**4,162**	1,17	1,30	1,45	0,23	0,21	0,10	0,15	0,04	0,02
y6 * y10 * y23	**4,126**	1,13	1,41	1,40	0,15	0,09	0,14	0,04	0,08	0,03
y6 * y12 * y14	**4,121**	1,36	1,19	1,29	0,08	0,26	0,25	0,03	0,01	0,20
y6 * y12 * y15	**4,043**	1,28	1,18	1,28	0,12	0,28	0,26	0,04	0,02	0,18
y6 * y12 * y23	**3,922**	1,16	1,21	1,25	0,12	0,25	0,29	0,01	0,06	0,18
y6 * y14 * y15	**4,186**	1,26	1,32	1,43	0,14	0,11	0,12	0,06	0,07	0,04
y6 * y14 * y23	**4,060**	1,13	1,35	1,41	0,15	0,08	0,13	0,04	0,09	0,02
y6 * y15 * y23	**4,011**	1,15	1,30	1,38	0,13	0,10	0,16	0,02	0,08	0,05
y7 * y6 * y10	**4,449**	1,42	1,46	1,46	0,09	0,08	0,08	0,04	0,03	0,03
y7 * y6 * y12	**4,221**	1,19	1,30	1,46	0,27	0,25	0,08	0,20	0,03	0,01
y7 * y6 * y14	**4,376**	1,35	1,47	1,45	0,09	0,08	0,09	0,03	0,04	0,03
y7 * y6 * y15	**4,278**	1,25	1,44	1,42	0,15	0,10	0,12	0,05	0,07	0,02
y7 * y6 * y23	**4,188**	1,16	1,40	1,48	0,12	0,14	0,06	0,09	0,01	0,03
y7 * y10 * y12	**4,346**	1,36	1,42	1,44	0,10	0,09	0,10	0,04	0,05	0,03
y7 * y10 * y14	**4,268**	1,28	1,36	1,45	0,16	0,15	0,08	0,10	0,03	0,02
y7 * y10 * y15	**4,141**	1,15	1,31	1,42	0,25	0,20	0,11	0,15	0,06	0,01
y7 * y10 * y23	**4,190**	1,20	1,41	1,47	0,08	0,10	0,07	0,05	0,02	0,04
y7 * y12 * y14	**4,260**	1,33	1,35	1,43	0,10	0,11	0,10	0,04	0,03	0,04
y7 * y12 * y15	**4,164**	1,24	1,33	1,40	0,16	0,13	0,13	0,06	0,06	0,03
y7 * y12 * y23	**4,129**	1,20	1,34	1,45	0,08	0,12	0,08	0,05	0,01	0,05
y7 * y14 * y15	**4,153**	1,24	1,32	1,41	0,16	0,11	0,13	0,06	0,07	0,03
y7 * y14 * y23	**4,114**	1,20	1,33	1,46	0,08	0,10	0,07	0,05	0,01	0,04
y7 * y15 * y23	**4,052**	1,22	1,27	1,42	0,06	0,13	0,11	0,03	0,01	0,08
y10 * y12 * y14	**4,183**	1,27	1,37	1,36	0,16	0,09	0,15	0,03	0,09	0,03
y10 * y12 * y15	**4,064**	1,15	1,35	1,30	0,25	0,11	0,20	0,05	0,15	0,01
y10 * y12 * y23	**4,082**	1,17	1,36	1,32	0,11	0,10	0,19	0,04	0,13	0,12
y10 * y14 * y15	**3,997**	1,18	1,28	1,25	0,22	0,15	0,26	0,03	0,13	0,06
y10 * y14 * y23	**3,998**	1,18	1,28	1,35	0,10	0,16	0,16	0,03	0,03	0,09
y10 * y15 * y23	**3,914**	1,20	1,19	1,28	0,08	0,21	0,22	0,02	0,03	0,16
y12 * y14 * y15	**4,076**	1,25	1,32	1,33	0,15	0,12	0,13	0,05	0,06	0,03
y12 * y14 * y23	**4,004**	1,18	1,33	1,35	0,10	0,10	0,10	0,04	0,04	0,04
y12 * y15 * y23	**3,951**	1,19	1,28	1,32	0,09	0,12	0,14	0,03	0,04	0,07
y14 * y15 * y23	**3,936**	1,19	1,29	1,31	0,09	0,11	0,13	0,03	0,04	0,06

Table 13. Calculations for trivariate entropies (Variables Y8, Y9, Y11, Y16, Y18)

		Inequality E.1			Inequality E.2			Inequality E.3		
Variables	**Entropy**	(X1X2) → X3	(X1X3) → X2	(X2X3) → X1	(X1X2) X3	(X1X3) X2	(X2X3) X1	(X2 X3)|X1	(X1 X3)|X2	(X1 X2)|X3
y8 * y9 * y11	**4,321**	1,36	1,42	1,36	0,13	0,09	0,17	0,02	0,09	0,05
y8 * y9 * y16	**4,147**	1,18	1,31	1,42	0,20	0,21	0,11	0,13	0,03	0,04
y8 * y9 * y18	**4,030**	1,06	1,26	1,32	0,27	0,25	0,21	0,18	0,13	0,11
y8 * y9 * y21	**4,202**	1,24	1,44	1,38	0,09	0,08	0,15	0,00	0,07	0,06
y8 * y11 * y16	**4,189**	1,29	1,35	1,36	0,09	0,13	0,17	0,02	0,05	0,09
y8 * y11 * y18	**4,010**	1,11	1,34	1,28	0,22	0,15	0,25	0,03	0,13	0,06
y8 * y11 * y21	**4,108**	1,21	1,34	1,36	0,11	0,15	0,17	0,03	0,05	0,08
y8 * y16 * y18	**3,894**	1,06	1,22	1,32	0,28	0,16	0,21	0,09	0,13	0,02
y8 * y16 * y21	**4,072**	1,24	1,31	1,38	0,09	0,08	0,15	0,00	0,07	0,06
y8 * y18 * y21	**3,905**	1,23	1,14	1,28	0,09	0,20	0,25	0,01	0,06	0,17
y9 * y11 * y16	**4,160**	1,19	1,43	1,33	0,19	0,06	0,19	0,02	0,15	0,02
y9 * y11 * y18	**4,099**	1,13	1,39	1,36	0,20	0,10	0,15	0,06	0,11	0,01
y9 * y11 * y21	**4,218**	1,25	1,39	1,47	0,07	0,09	0,05	0,06	0,01	0,03
y9 * y16 * y18	**3,861**	1,13	1,15	1,28	0,21	0,24	0,23	0,07	0,06	0,09
y9 * y16 * y21	**4,031**	1,30	1,21	1,34	0,02	0,18	0,18	0,01	0,01	0,16
y9 * y18 * y21	**3,998**	1,28	1,17	1,37	0,04	0,16	0,15	0,02	0,01	0,13
y11 * y16 * y18	**3,960**	1,13	1,23	1,38	0,21	0,16	0,10	0,12	0,06	0,01
y11 * y16 * y21	**4,086**	1,25	1,34	1,39	0,07	0,05	0,10	0,01	0,05	0,03
y11 * y18 * y21	**3,982**	1,25	1,23	1,23	0,08	0,10	0,25	0,01	0,16	0,19
y16 * y18 * y21	**3,861**	1,29	1,17	1,23	0,04	0,17	0,15	0,02	0,01	0,14

Table 14. Calculations for tetravariate entropies

The variables being explained are sorted in ascending order and presented in the left column. The first row gives the inequality calculated and the second row explains the relationship among the variables in each inequality.

Variables	**Tetra**	Inequality D.1				Inequality D.2			
1 * 2 * 3 * 4	**Entropy**	(X2X3X4) → X1	(X1X3X4) → X2	(X1X2X4) → X3	(X1X2X3) → X4	(X2X3X4) 1	(X1X3X4) 2	(X1X2X4) 3	(X1X2X3) 4
y6 * y10 * y12 * y14	**5,46**	1,28	1,34	1,18	1,27	0,26	0,17	0,28	0,17
y6 * y10 * y12 * y15	**5,34**	1,27	1,29	1,17	1,14	0,27	0,21	0,28	0,26
y6 * y10 * y12 * y23	**5,32**	1,24	1,40	1,19	1,12	0,30	0,11	0,27	0,16
y6 * y10 * y14 * y15	**5,43**	1,43	1,24	1,27	1,15	0,11	0,27	0,17	0,25
y6 * y10 * y14 * y23	**5,39**	1,39	1,33	1,26	1,11	0,15	0,18	0,17	0,17
y6 * y10 * y15 * y23	**5,29**	1,37	1,28	1,16	1,13	0,17	0,23	0,24	0,16
y6 * y12 * y14 * y15	**5,35**	1,27	1,16	1,31	1,23	0,27	0,29	0,13	0,17
y6 * y12 * y14 * y23	**5,25**	1,24	1,19	1,32	1,12	0,30	0,27	0,11	0,16
y6 * y12 * y15 * y23	**5,18**	1,23	1,17	1,26	1,14	0,31	0,29	0,14	0,14
y6 * y14 * y15 * y23	**5,31**	1,37	1,30	1,25	1,12	0,17	0,14	0,15	0,16
y10 * y12 * y14 * y15	**5,32**	1,24	1,32	1,26	1,14	0,26	0,13	0,18	0,26
y10 * y12 * y14 * y23	**5,33**	1,33	1,34	1,25	1,15	0,18	0,12	0,18	0,13
y10 * y12 * y15 * y23	**5,23**	1,28	1,31	1,15	1,16	0,23	0,15	0,26	0,12
y10 * y14 * y15 * y23	**5,17**	1,23	1,25	1,17	1,17	0,27	0,18	0,23	0,11
y12 * y14 * y15 * y23	**5,24**	1,30	1,29	1,23	1,16	0,16	0,15	0,17	0,12
y8 * y9 * y11 * y16	**5,49**	1,33	1,30	1,34	1,17	0,20	0,22	0,15	0,22
y8 * y9 * y11 * y18	**5,36**	1,26	1,35	1,33	1,04	0,27	0,17	0,16	0,30
y8 * y9 * y11 * y21	**5,53**	1,31	1,42	1,32	1,20	0,22	0,10	0,16	0,12
y8 * y9 * y16 * y18	**5,16**	1,30	1,27	1,13	1,02	0,22	0,25	0,25	0,32
y8 * y9 * y16 * y21	**5,38**	1,35	1,31	1,18	1,23	0,18	0,21	0,21	0,09
y8 * y9 * y18 * y21	**5,26**	1,26	1,35	1,06	1,23	0,27	0,16	0,28	0,10
y8 * y11 * y16 * y18	**5,22**	1,26	1,33	1,21	1,03	0,27	0,16	0,17	0,31
y8 * y11 * y16 * y21	**5,40**	1,31	1,32	1,29	1,21	0,22	0,16	0,10	0,12
y8 * y11 * y18 * y21	**5,21**	1,23	1,31	1,10	1,20	0,29	0,18	0,23	0,12
y8 * y16 * y18 * y21	**5,12**	1,26	1,22	1,05	1,23	0,27	0,17	0,29	0,10
y9 * y11 * y16 * y18	**5,24**	1,28	1,38	1,14	1,08	0,24	0,11	0,25	0,26
y9 * y11 * y16 * y21	**5,41**	1,32	1,38	1,19	1,25	0,19	0,11	0,20	0,08
y9 * y11 * y18 * y21	**5,34**	1,36	1,34	1,12	1,24	0,16	0,15	0,22	0,08
y9 * y16 * y18 * y21	**5,14**	1,28	1,14	1,11	1,28	0,24	0,24	0,23	0,04
y11 * y16 * y18 * y21	**5,20**	1,34	1,22	1,11	1,24	0,15	0,17	0,22	0,08

Variables	Tetra	Inequality D.3						Inequality D.4					
1 * 2 * 3 * 4	Entropy	(X1X2)	(X3X4)	(X1X3)	(X2X4)	(X1X4)	(X2X3)	(X1X2)	X3 | X4	(X1X3)	X2 | X4	(X2X3)	X1 | X4
y6 * y10 * y12 * y14	5,46	0,36		0,12		0,38		0,21		0,04		0,22	
y6 * y10 * y12 * y15	5,34	0,42		0,14		0,43		0,19		0,02		0,19	
y6 * y10 * y12 * y23	5,32	0,35		0,17		0,30		0,20		0,04		0,19	
y6 * y10 * y14 * y15	5,43	0,31		0,22		0,25		0,08		0,07		0,03	
y6 * y10 * y14 * y23	5,39	0,25		0,26		0,13		0,10		0,11		0,04	
y6 * y10 * y15 * y23	5,29	0,33		0,30		0,14		0,19		0,16		0,06	
y6 * y12 * y14 * y15	5,35	0,16		0,34		0,34		0,04		0,20		0,19	
y6 * y12 * y14 * y23	5,25	0,17		0,36		0,29		0,05		0,21		0,19	
y6 * y12 * y15 * y23	5,18	0,21		0,35		0,29		0,09		0,22		0,20	
y6 * y14 * y15 * y23	5,31	0,25		0,20		0,15		0,10		0,07		0,06	
y10 * y12 * y14 * y15	5,32	0,33		0,25		0,22		0,09		0,04		0,07	
y10 * y12 * y14 * y23	5,33	0,23		0,15		0,21		0,12		0,06		0,11	
y10 * y12 * y15 * y23	5,23	0,31		0,16		0,26		0,20		0,08		0,16	
y10 * y14 * y15 * y23	5,17	0,27		0,20		0,30		0,18		0,11		0,21	
y12 * y14 * y15 * y23	5,24	0,22		0,17		0,18		0,11		0,08		0,09	
y8 * y9 * y11 * y16	5,49	0,31		0,14		0,31		0,10		0,05		0,12	
y8 * y9 * y11 * y18	5,36	0,34		0,25		0,28		0,07		0,03		0,08	
y8 * y9 * y11 * y21	5,53	0,19		0,19		0,21		0,10		0,08		0,14	
y8 * y9 * y16 * y18	5,16	0,38		0,39		0,24		0,11		0,11		0,03	
y8 * y9 * y16 * y21	5,38	0,28		0,28		0,12		0,19		0,19		0,10	
y8 * y9 * y18 * y21	5,26	0,34		0,24		0,22		0,25		0,15		0,18	
y8 * y11 * y16 * y18	5,22	0,25		0,35		0,28		0,03		0,07		0,07	
y8 * y11 * y16 * y21	5,40	0,20		0,19		0,20		0,08		0,10		0,13	
y8 * y11 * y18 * y21	5,21	0,31		0,21		0,29		0,20		0,12		0,21	
y8 * y16 * y18 * y21	5,12	0,34		0,24		0,22		0,25		0,15		0,18	
y9 * y11 * y16 * y18	5,24	0,31		0,23		0,31		0,10		0,02		0,10	
y9 * y11 * y16 * y21	5,41	0,25		0,07		0,25		0,18		0,05		0,18	
y9 * y11 * y18 * y21	5,34	0,26		0,12		0,22		0,18		0,08		0,14	
y9 * y16 * y18 * y21	5,14	0,22		0,27		0,26		0,19		0,23		0,22	
y11 * y16 * y18 * y21	5,20	0,26		0,23		0,12		0,19		0,15		0,09	

Variables	Tetra	Inequality D.4											
1 * 2 * 3 * 4	Entropy	(X1X2)	X4 | X3	(X1X4)	X2 | X3	(X2X4)	X1 | X3	(X1X3)	X4 | X2	(X1X4)	X3 | X2	(X3X4)	X1 | X2
y6 * y10 * y12 * y14	5,46	0,10		0,11		0,03		0,04		0,22		0,21	
y6 * y10 * y12 * y15	5,34	0,16		0,16		0,03		0,07		0,23		0,21	
y6 * y10 * y12 * y23	5,32	0,09		0,05		0,07		0,09		0,21		0,25	
y6 * y10 * y14 * y15	5,43	0,16		0,14		0,06		0,06		0,04		0,05	
y6 * y10 * y14 * y23	5,39	0,11		0,05		0,10		0,10		0,04		0,09	
y6 * y10 * y15 * y23	5,29	0,10		0,04		0,09		0,09		0,05		0,11	
y6 * y12 * y14 * y15	5,35	0,08		0,23		0,22		0,07		0,06		0,03	
y6 * y12 * y14 * y23	5,25	0,09		0,21		0,25		0,09		0,04		0,07	
y6 * y12 * y15 * y23	5,18	0,09		0,19		0,23		0,08		0,04		0,08	
y6 * y14 * y15 * y23	5,31	0,11		0,05		0,09		0,10		0,06		0,12	
y10 * y12 * y14 * y15	5,32	0,17		0,07		0,14		0,16		0,11		0,21	
y10 * y12 * y14 * y23	5,33	0,07		0,06		0,05		0,06		0,12		0,12	
y10 * y12 * y15 * y23	5,23	0,07		0,05		0,04		0,05		0,16		0,17	
y10 * y14 * y15 * y23	5,17	0,06		0,09		0,08		0,05		0,14		0,15	
y12 * y14 * y15 * y23	5,24	0,07		0,06		0,06		0,06		0,08		0,09	
y8 * y9 * y11 * y16	5,49	0,18		0,18		0,08		0,05		0,11		0,12	
y8 * y9 * y11 * y18	5,36	0,21		0,13		0,15		0,16		0,12		0,19	
y8 * y9 * y11 * y21	5,53	0,06		0,06		0,10		0,10		0,13		0,14	
y8 * y9 * y16 * y18	5,16	0,18		0,08		0,15		0,18		0,08		0,15	
y8 * y9 * y16 * y21	5,38	0,08		0,04		0,10		0,08		0,04		0,10	
y8 * y9 * y18 * y21	5,26	0,06		0,02		0,08		0,08		0,14		0,19	
y8 * y11 * y16 * y18	5,22	0,16		0,12		0,19		0,21		0,13		0,15	
y8 * y11 * y16 * y21	5,40	0,10		0,12		0,14		0,06		0,06		0,10	
y8 * y11 * y18 * y21	5,21	0,09		0,09		0,10		0,06		0,14		0,18	
y8 * y16 * y18 * y21	5,12	0,06		0,02		0,07		0,08		0,14		0,19	
y9 * y11 * y16 * y18	5,24	0,11		0,07		0,07		0,17		0,21		0,20	
y9 * y11 * y16 * y21	5,41	0,06		0,07		0,02		0,01		0,16		0,16	
y9 * y11 * y18 * y21	5,34	0,05		0,06		0,02		0,02		0,13		0,12	
y9 * y16 * y18 * y21	5,14	0,01		0,10		0,10		0,03		0,08		0,07	
y11 * y16 * y18 * y21	5,20	0,05		0,02		0,06		0,07		0,08		0,11	

Variables	Tetra	Inequality D.4			Inequality D.5	
1 * 2 * 3 * 4	Entropy	(X2X3) X4 | X1	(X2X4) X3 | X1	(X3X4) X2 | X1	X3 X4 |(X1X2)	X2 X4|(X1X3)
y6 * y10 * y12 * y14	5,46	0,12	0,04	0,11	0,02	0,09
y6 * y10 * y12 * y15	5,34	0,18	0,05	0,16	0,03	0,14
y6 * y10 * y12 * y23	5,32	0,05	0,03	0,05	0,01	0,03
y6 * y10 * y14 * y15	5,43	0,17	0,12	0,21	0,02	0,11
y6 * y10 * y14 * y23	5,39	0,06	0,12	0,12	0,02	0,02
y6 * y10 * y15 * y23	5,29	0,04	0,16	0,17	0,01	0,02
y6 * y12 * y14 * y15	5,35	0,09	0,08	0,06	0,05	0,03
y6 * y12 * y14 * y23	5,25	0,05	0,06	0,04	0,03	0,01
y6 * y12 * y15 * y23	5,18	0,03	0,06	0,05	0,02	0,01
y6 * y14 * y15 * y23	5,31	0,05	0,07	0,09	0,01	0,03
y10 * y12 * y14 * y15	5,32	0,07	0,05	0,08	0,02	0,04
y10 * y12 * y14 * y23	5,33	0,06	0,05	0,07	0,02	0,03
y10 * y12 * y15 * y23	5,23	0,05	0,06	0,09	0,01	0,03
y10 * y14 * y15 * y23	5,17	0,04	0,04	0,05	0,01	0,03
y12 * y14 * y15 * y23	5,24	0,05	0,07	0,08	0,02	0,03
y8 * y9 * y11 * y16	5,49	0,14	0,03	0,14	0,01	0,12
y8 * y9 * y11 * y18	5,36	0,11	0,04	0,09	0,03	0,08
y8 * y9 * y11 * y21	5,53	0,04	0,05	0,02	0,03	0,01
y8 * y9 * y16 * y18	5,16	0,13	0,18	0,17	0,05	0,04
y8 * y9 * y16 * y21	5,38	0,01	0,13	0,13	0,00	0,00
y8 * y9 * y18 * y21	5,26	0,01	0,09	0,09	0,01	0,00
y8 * y11 * y16 * y18	5,22	0,11	0,10	0,04	0,08	0,03
y8 * y11 * y16 * y21	5,40	0,03	0,02	0,05	0,01	0,03
y8 * y11 * y18 * y21	5,21	0,04	0,04	0,06	0,01	0,03
y8 * y16 * y18 * y21	5,12	0,01	0,10	0,09	0,01	0,01
y9 * y11 * y16 * y18	5,24	0,12	0,08	0,07	0,06	0,05
y9 * y11 * y16 * y21	5,41	0,06	0,03	0,07	0,01	0,05
y9 * y11 * y18 * y21	5,34	0,07	0,08	0,11	0,01	0,04
y9 * y16 * y18 * y21	5,14	0,03	0,09	0,07	0,02	0,01
y11 * y16 * y18 * y21	5,20	0,02	0,13	0,13	0,01	0,01

Variables	Tetra	Inequality D.5			
1 * 2 * 3 * 4	Entropy	X2 X3|(X1X4)	X1 X4|(X2X3)	X1 X3|(X2X4)	X1 X2|(X3X4)
y6 * y10 * y12 * y14	5,46	0,01	0,01	0,19	0,03
y6 * y10 * y12 * y15	5,34	0,01	0,01	0,17	0,14
y6 * y10 * y12 * y23	5,32	0,02	0,05	0,17	0,12
y6 * y10 * y14 * y15	5,43	0,06	0,03	0,01	0,29
y6 * y10 * y14 * y23	5,39	0,08	0,07	0,01	0,30
y6 * y10 * y15 * y23	5,29	0,14	0,07	0,03	0,26
y6 * y12 * y14 * y15	5,35	0,02	0,02	0,01	0,21
y6 * y12 * y14 * y23	5,25	0,02	0,05	0,01	0,28
y6 * y12 * y15 * y23	5,18	0,04	0,05	0,02	0,24
y6 * y14 * y15 * y23	5,31	0,05	0,07	0,04	0,26
y10 * y12 * y14 * y15	5,32	0,02	0,11	0,06	0,11
y10 * y12 * y14 * y23	5,33	0,02	0,03	0,08	0,20
y10 * y12 * y15 * y23	5,23	0,05	0,03	0,13	0,12
y10 * y14 * y15 * y23	5,17	0,02	0,02	0,12	0,11
y12 * y14 * y15 * y23	5,24	0,04	0,03	0,05	0,16
y8 * y9 * y11 * y16	5,49	0,01	0,03	0,09	0,19
y8 * y9 * y11 * y18	5,36	0,01	0,10	0,06	0,24
y8 * y9 * y11 * y21	5,53	0,02	0,05	0,07	0,16
y8 * y9 * y16 * y18	5,16	0,09	0,12	0,02	0,30
y8 * y9 * y16 * y21	5,38	0,13	0,07	0,03	0,15
y8 * y9 * y18 * y21	5,26	0,08	0,06	0,12	0,05
y8 * y11 * y16 * y18	5,22	0,01	0,10	0,01	0,29
y8 * y11 * y16 * y21	5,40	0,02	0,05	0,05	0,17
y8 * y11 * y18 * y21	5,21	0,03	0,04	0,13	0,07
y8 * y16 * y18 * y21	5,12	0,09	0,06	0,12	0,05
y9 * y11 * y16 * y18	5,24	0,01	0,05	0,09	0,21
y9 * y11 * y16 * y21	5,41	0,02	0,01	0,15	0,09
y9 * y11 * y18 * y21	5,34	0,05	0,01	0,11	0,13
y9 * y16 * y18 * y21	5,14	0,06	0,01	0,06	0,09
y11 * y16 * y18 * y21	5,20	0,12	0,04	0,05	0,11