Relational methodology

Ulisse Di Corpo and Antonella Vannini[1]

The discovery of attractors has widened science to the study of phenomena which cannot be caused, phenomena which we can observe, but not reproduce in a laboratory. Until now it was generally accepted, without any clear justification, that all phenomena were based on causal relations and that every aspect of reality could therefore be studied using the experimental method. The introduction of syntropy shows that, in addition to causable phenomena (entropic), non causable (syntropic) phenomena exist and that a new scientific methodology is needed.

Two methodologies allow the study of relations:

1. the methodology of differences, from which the experimental method takes form, but which allows to study only causal relations (entropic).
2. the methodology of concomitances, introduced by Stuart Mill in 1843, from which the relational methods takes form, which permits the study of any type of relation, causal and non-causal, allowing science to broaden to syntropic phenomena.

Until now, relational methodology has been used only in a very limited way, because:

- it was generally accepted that only causal relations existed;
- data-processing power was not available.

[1] syntropy@gmail.com

1. Relational methodology

In order to understand the difference between the methodology of differences and the methodology of concomitances, we start describing the experimental method when used to study living systems.

The experimental method studies differences between groups:

1. similar groups are formed (same mean values and variances);
2. only one element varies (treatment);
3. differences which arise between the two groups can be attributed only to the treatment.

For example, to test a drug, two similar groups (same average values, variances, etc.) are formed, all the other variables are kept under control and the drug is given only to one group. The differences observed between the two groups can be attributed only to the drug. The drug is the cause, the differences are the effects: cause-effect knowledge is produced.

While the experimental method has been used with great success in physics and chemistry, in medicine, biology, psychology and sociology it has lead to the treatment of life, humans and society as if they were mechanisms and has produced an enormous amount of single cause-effect relations, loosing the unified vision of life. Pharmacology is the most evident example: a huge amount of cause-effect relations, with no understanding of life.

Relational methodology, instead, studies concomitances. How does it work? It is very simple, data is gathered through a questionnaire or an observation grid, and information crossed.

In the following example sex and car accidents are crossed. We see that 20% of men have had few accidents compared to 70% of women, 80% of men have had many car accidents compared to 30% of women:

Accidents	Males	Females	Total
Few	50	105	155
	20%	**70%**	**39%**
Many	200	45	245
	80%	**30%**	**61%**
Total	250	150	400
	100%	**100%**	**100%**

It is difficult to understand why the relation men-accidents exists, but when we cross a third variables things change radically. For example, if the populations is divided in two groups, those who drive a lot and those who drive little, the relation male-accidents disappears:

Accidents	drive little		drive a lot	
	Male	Female	Male	Female
Few	70%	70%	20%	20%
Many	30%	30%	80%	80%
Total	100%	100%	100%	100%

This example shows that there is no difference in the number of accidents between males and female when the relation is checked with the variable “drive little, drive a lot”. The relation between sex and accidents is mediated by this third variable: males drive more than females and therefore are subject to a greater number of accidents:

Males -> drive a lot -> do more accidents

Three variable tables allow to check the boundaries of relations. For example, if we find a relation between treatment and healing, it is possible to control when the relation is valid. We could find that the relation is present only in certain age groups, sex, or other particular cases.

From these short examples it is possible to deduce some properties of the relational methodology:

- it allows the study of many relations at the same time, producing in this way global and also analytical information;

- it uses dichotomic variables, any information quantitative or qualitative can be transformed in one or more dichotomic variable;
- it uses "a-posteriori" controls, and allows to develop information which can reproduce the complexity of natural phenomena;
- it does not require controlled laboratory environment, homogeneous or randomized groups;
- it allows to use very different groups, and use information which has been gathered previously.

These properties of the relational methodology permit to work directly on the field, without using artificial settings and environments. Research activity becomes easy, accessible, cheap, and it allows to study any kind of phenomena.

A short example

In a study for the paediatric clinic of the University of Rome different therapies were compared in the treatment of the ITP syndrome. This syndrome is common in the acute form but rare in the chronic form. As an example, in 12 years the paediatric clinic of the University of Rome has treated only 15 cases. The limited number of patients lets us understand how difficult it would be to use the methodology of differences which requires similar and randomized groups. The children which are treated are very different in age, social conditions and treatments which have been used before the syndrome could be recognized (usually it takes over 6 months). The results obtained with the relational methodology showed that the commonly used therapy was not effective in the long term; the way patients responded to treatment was different compared to sex and age. A therapy which was accidentally used in a clinic in north Italy showed strong positive correlations. A comparison with the classical methodology of differences was also carried out using ANOVA (analysis of variance), and in this case no relations were observed: all the information in the data set was lost.

Conclusion

The relational methodology allows to reproduce the same information which is produced with the experimental method, but it also permits to study that which it is impossible to study with the experimental method.

While the method of differences uses parametric statistical tests (ANOVA – Analysis of Variance), which require quantitative data, the normal distribution of data and randomized groups, the method of concomitances uses non parametric statistical tests which work well on quantitative and qualitative data, do not require any distribution of the data set, and do not require similar or randomized groups. The relational methodology is generally considered simple to use and efficient in the results.

In applied research it is often difficult to have similar groups. If we take the example of rare illnesses we understand how difficult it is to use randomized or controlled groups. In these cases data analyses carried out with the relational methodology are able to identify relations which with the methodology of differences (ANOVA) disappear. Many studies have shown that the relational methodology is more efficient than ANOVA and that it presents no limits in its possible applications.

The methodology of differences (ANOVA) can be used efficiently only when groups have been correctly randomized and if the inner variance within groups is low, otherwise only the strongest and most obvious differences show up. In other words information. With ANOVA the risk of not seeing relations, when instead they exist, is extremely high. The basic criteria of science, repeatability, is lost, and each research study leads to different results. In science conclusions are reached on the basis of more than one study. Knowledge is scientific when it can be reproduced. Only information which can be reproduced proves that it was not a result of mere chance.

2. Statistical techniques

It is well known that statistics was born in the fields of biology, psychology, medicine and sociology, in order to study the human being, society and life in general, whereas mathematics was born in astronomy and physics and was used to study mechanical laws and deterministic (cause-effect) phenomena. The main difference between mathematics and statistics is that mathematics was created to study exact laws such as the laws of the physical entropical world, whereas statistics is a result of the study of probabilistic laws, typical of life and syntropical systems. The takeover of statistics by mathematicians underlines a deep confusion in concepts such as syntropy and entropy, and shows that life has been confused with a mechanical system.

Everyone can object that statistics uses mathematical tools such as sums, subtractions and so on, but in order to have an efficient statistical tool it is necessary to use simple mathematics.

In the previous examples we based our analysis on percentages and classifications. Even when many variables are used the analyses remain the same: percentages, comparisons, classifications, nothing too complex.

Until a few years ago computers were inaccessible and researchers had to look for short-cuts, trying to reduce the number of calculations introducing in this way mathematical short-cuts. We have therefore seen the birth of a huge amount of mathematical statistical techniques, which are as complex from a mathematical point of view as they are useless on an applied level. All these mathematical statistical techniques have made statistics become one of the most disliked subjects in social, psychological and life sciences.

Moving from statistics to mathematical statistics has also required data to be in a form which could be handled with these techniques. Mathematical statisticians need quantitative data and have therefore tried, in all possible ways, to translate qualitative data into quantitative. Many techniques have been proposed in order to translate qualitative information, which is typical of life (syntropic phenomena), into quantitative information, which is typical of physical (entropic) phenomena. From a mathematical point of view quantitative data is richer in information. When using quantitative data it is possible to calculate average values, deviations, variances and differences, whereas when using qualitative information it is only possible to count and cross table.

All the attempts of the mathematical statisticians to make quantitative what is qualitative have failed (even if they are still used), while there have never been problems in translating quantitative and qualitative information into dichotomic variables. For example if we measure blood pressure, values can be translated in one or more dichotomic variables: over/under 140, over/under 150, within a set range of values or outside a range. Cross tabling these dichotomic variables allows to study relations and which cut-off point are significant. With quantitative/mathematical data analysis it is only possible to study linear relations, while with dichotomic variables it is possible to study any kind of relation. Often in living systems (such as society, biology and medicine) the system is stable until certain condition are reached and then suddenly it changes. Relations are not linear but are often abrupt, sudden. It is impossible to study these relations, typical of life systems, with quantitative/mathematical tools. Using dichotomic variables it is possible to study linear and non-linear relations.

Data analysis

Dichotomic variables are those which can only have two values, for example: yes/no, true/false, 0/1; they are similar to the bit used in computers which can be set only to 0 or 1 and for this reason they are often called “bit of information”. Any information, quantitative or qualitative, can be translated in one or more dichotomic variables. At this level it is therefore possible to cross any kind of information and to study any kind of relation.

The basic operation which is performed with dichotomic variables is counting. Counting produces frequencies on which it is possible to calculate percentages. When dichotomic variables are cross tabled it is possible to apply sums, subtractions, divisions and multiplications. In this way from frequencies we arrive at more complex analysis such as connection tables (correlation tables) and factor analysis.

With dichotomic variables it is possible to produce:

- **frequency distributions:** which permit the study of the distribution of the values of the variables;
- **cross tables:** which are produced crossing the distribution of two variables allowing the study of the relation between the two variables, thanks to the comparison of row and column percentages;
- **relations tables:** for each pair of dichotomic variables it is possible to calculate the strength of the relation using Chi Square and other statistical indexes. By sorting the relation values it is possible for each variable to produce a profile in which variables are put in order on the basis of the strength of the relation. As every researcher knows cross tables can produces thousands of pages which are often difficult to read and interpret. Relation tables synthesize this huge amount of information in a few pages which are easy and fast to read and interpret.
- **factor analysis:** relation indexes are values which express the relation between couples of variables. When the interest is to reach a global perspective of the phenomena it is necessary to analyze the relations together. Factor analysis generates new variables which use the correlations of the original variables as coordinates in such a way that it is possible to represent the variables in a multidimensional space and study how they cluster together.
- **factorial points:** the clusters of variables can be translated into new variables (factorial points) which are used to select subjects in homogeneous groups.

- frequency distributions

Age years	n.
13	2
14	56
15	161
16	183
17	194
18	134
19	72

Frequency distributions provide a picture of the distribution of the variable. Each line of the table corresponds to a dichotomic variable. For example the variable 13 years can have the values yes/no, the same can be said for the dichotomic variable 14 years,...

Frequency distributions permit to describe the sample population used in the study (age, sex, level of education). When the sample is representative of the population it is possible to weight each unit in order to obtain an estimate of the distribution of the variable in the population.

- cross tables

Accidents	Males	Females	Total
Few	50	105	155
	20%	***70%***	***39%***
Many	200	45	245
	80%	***30%***	***61%***
Total	250	150	400
	100%	***100%***	***100%***

Cross tables are produced crossing the distribution of two variables, thanks to the comparison of the percentage values. In this example 39% of the 400 interviewed have had few accidents, while 61% have had many. If no relation exists, between sex and accidents, we would expect the same percentages in the columns “males” and in the column “females”. In the example we see, instead, that only 20% of males have had few accidents, compared to 39% of the total which is the expected value, and 70% of the females; 80% of males have had many accidents compared to 61% of the total (which is the expected value) and 30% of the females. It is therefore possible to state that a relation between sex and car accidents exists. Grater is the difference between observed and expected percentages, strongest is the relation. It is possible to study the strength and significance of relations using statistical tests such as Chi Square, r phi, f of Fischer, etc...

- Relations

Many statistical tests permit to study the significance and strength of the relations. One of the most widely used is the Chi Square.

Chi Square studies the difference between observed and expected frequencies. The stronger the relation (concomitance) between the two variables the wider is the differences between observed and expected frequencies. The value of Chi Square, when applied on dichotomical variables, can vary between 0 (no relation) and n (the cases in the sample) for maximum relation. Chi Square values equal or over 3.48 have a probability of 5% of happening by chance, while Chi Square values equal or over 6.635 have a probability of

1% of happening by chance. It is common to consider significant relations which have a possibility of happening by chance inferior to 1%.

The difference between observed and expected variables can have two directions: a positive relation means that one variable is present when the other one is also present (yes/yes or no/no), a negative relation means that one variable is present when the other is absent (yes/no or no/yes). In the example which follows Chi Square values are paired with r-phi which varies between -1 to +1, where 0 means no correlation, +1 maximum positive correlation, while -1 maximum negative correlation.

Relations obtained by the dichotomic variable "I feel depressed" in a questionnaire study which involved 974 high school students

I feel depressed:

```
       ChiQ   rPhi          %Yes     %No

    974.00  1.000 (100.00%/  0.00%) I feel depressed
    507.08  0.722 ( 85.42%/ 12.96%) I feel anxious
    229.19  0.485 ( 69.37%/ 20.60%) I feel useless
    209.18  0.463 ( 78.04%/ 31.94%) I feel lonely
    189.70  0.441 ( 72.14%/ 27.78%) I am pessimist
    189.15  0.441 ( 65.50%/ 21.30%) Displeased
    179.18  0.429 ( 71.59%/ 28.47%) Unsatisfied
    173.57  0.422 ( 71.59%/ 29.17%) Un-happy
    169.88  0.418 ( 71.59%/ 29.63%) I don't trust myself
  ......
```

We see that the highest Chi Square value is 974 which coincides with the number of subjects used in this research; r-phi is equal to +1.000 which tells us that the correlation is the highest possible. This correlation obtained the maximum possible value because "I feel depressed" was crossed with itself.

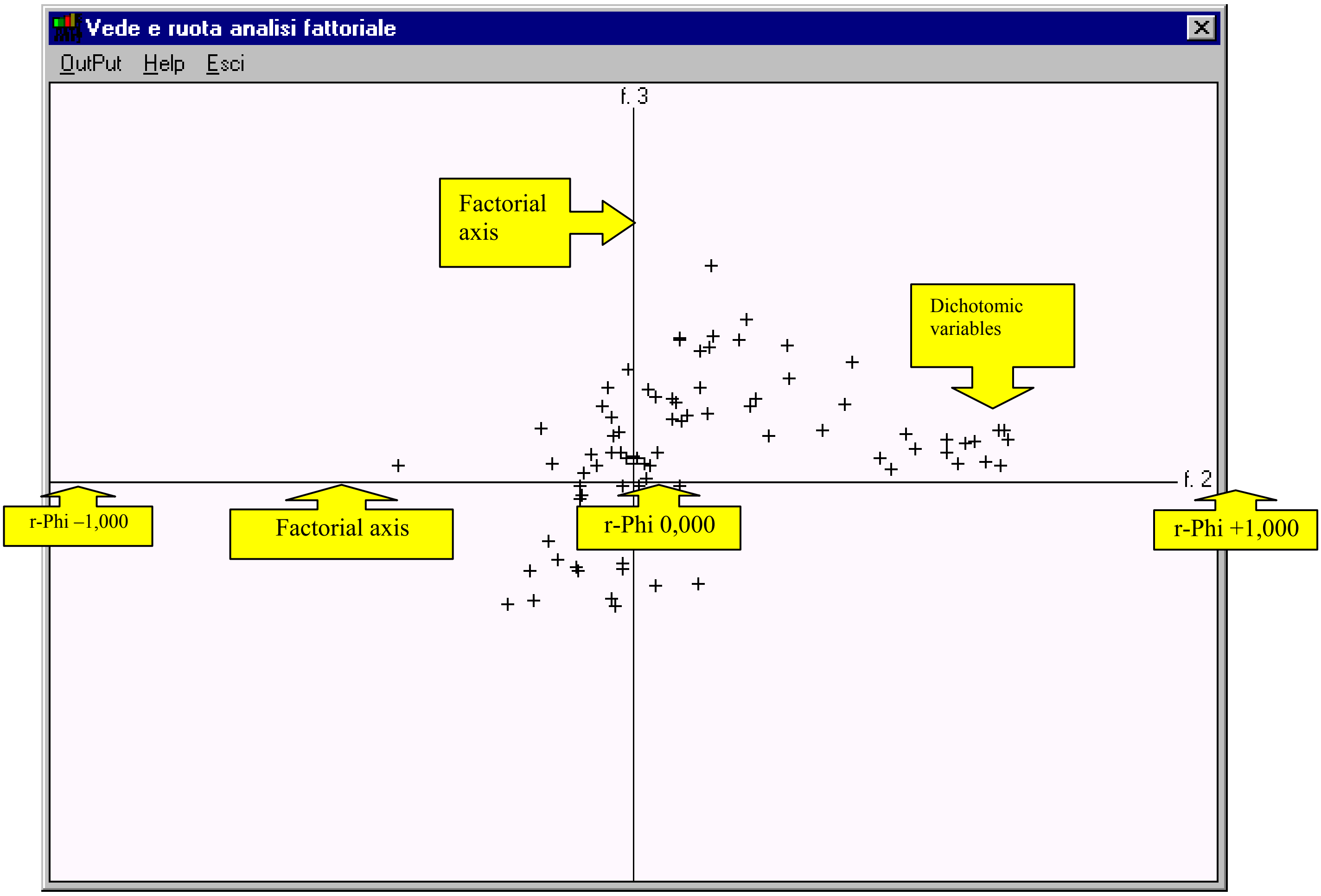
In the second line we see that "I feel depressed" is related to "I feel anxious"; 85% of the students which have answered "I feel depressed" have also answered "I feel anxious", while only 12% of those which have not answered "I feel depressed" have said "I feel anxious".

When the dimension of the sample (number of subjects) is high very weak relations can become significant. The highest is the number of subjects the lowest is the background noise, and very weak relations can be taken into account.

Relation tables cross each dichotomic variable with all the other dichotomic variable of the study. In this way it is possible to produce a profile of each dichotomic variable. Dichotomic variables behave in a similar way to bit used in computers, this property allows to device software which produce instant data analysis even when handling millions of subjects (records) and thousands of variables.

- *factor analysis*

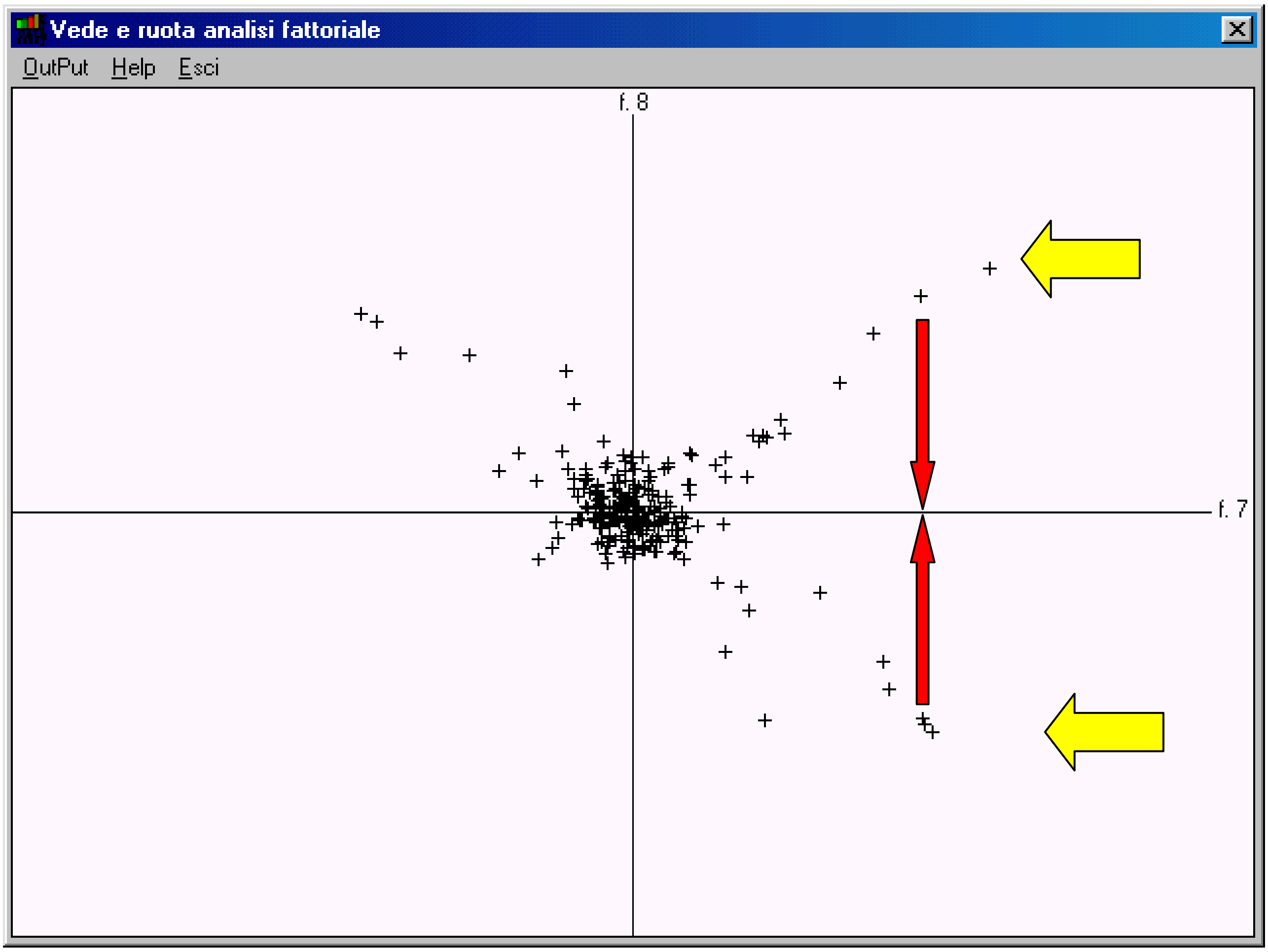
Factor analysis, which is the further logical development is capable of revealing patterns and structures responsible for the observed single relations. In short, the purpose of factor analysis is to find a new set of variables, fewer in number than the original variables, which express that which is common among the original variables.

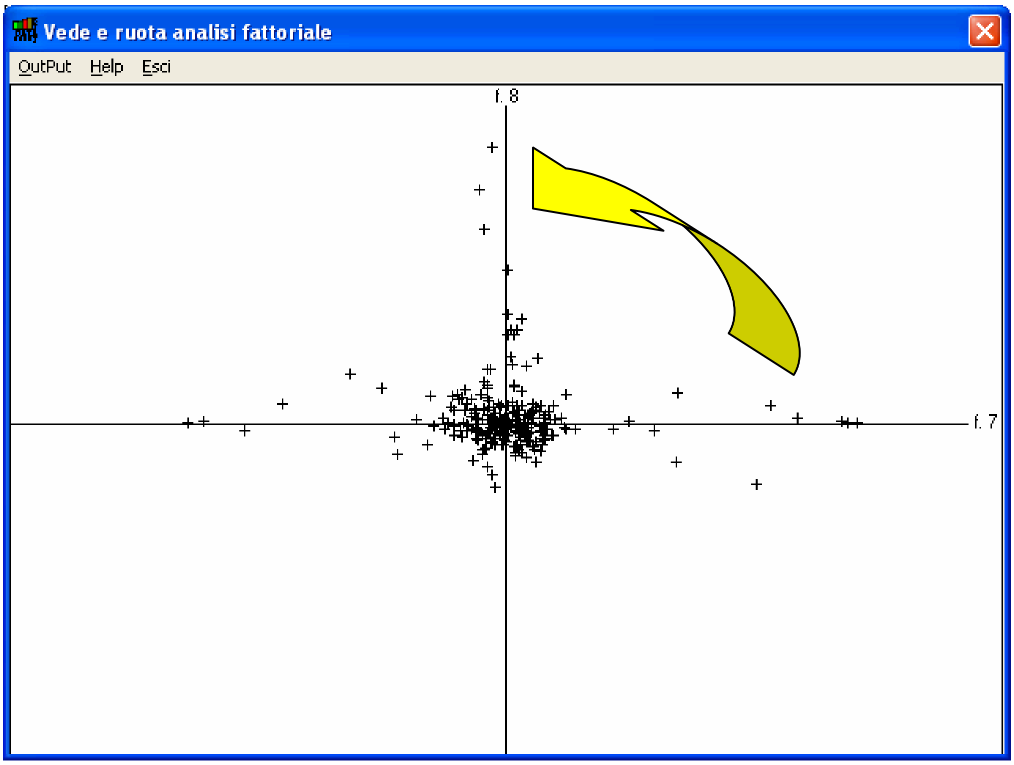


In medium-large size studies thousands of significant relations are obtained. When analyzing this large amount of relations it is difficult to find the underlying structures. Factor analysis solves this problem by creating new variables that are not correlated (factorial axes) and for which we know the correlation with the original variables. As has already been said, factorial axes have the property of not being correlated and each axis intersects the other axes at point 0, at a right angle. This results in a Cartesian space that has as many dimensions as there are factorial axes. The correlation between the original variables and the factorial axes are used as factorial coordinates. In this way, by intersecting any two factorial axes, it is possible to place every variable on a diagram:

The representation of the original variables on factorial planes permits to delineate structures of variables which are coherent among themselves. These structures are called "factors", as they identify (in general) the "conceptual factors" which describe the phenomena under study.

The geometrical view of factor axes introduces an important topic of factor analysis which is rotation. As it is possible to see in the next example the structures detected by the two yellow arrows at the right are not correlated (they form a 90° angle) but their projection on the horizontal axis coincide (red arrows). In the same way variables that are seen close together on the factorial plane might be at a 90° angle (therefore not correlated) when using another factorial axis. If the factorial planes are rotated in such a way that factorial structures tend to coincide with factorial axis the risk of seeing together variables which are not correlated is reduced and usually solved. It is therefore necessary to rotate factor axis before reading factor results. Only after rotation the reading of factor results can be considered scientific, otherwise the results will vary each time depending on the position of the axes.


Vede e ruota analisi fattoriale
OutPut Help Esci
f. 8
f. 7


Vede e ruota analisi fattoriale
OutPut Help Esci
f. 8
f. 7

Factorial points

Factors analysis produces profiles, lists of dichotomic variables which are correlated to each axis. Using each profile it is possible to calculate how similar each subject is to the factor. These similarity values, or factorial points, vary from 0% to 100%, where 0% means absence of similarity while 100% total identity with the profile. These values can be treated as variables and used in new data analysis or to select subjects, producing in this way groups which are homogeneous for each factor. It is interesting to note that a single subject can obtain high values on more than one profile and therefore take part in more than one group.

3. How to prepare a questionnaire or a form

Preparing a questionnaire/form is a very important moment. The outcome of the work depends on how well the form/questionnaire is designed.

Relations exist among all natural variables; in a scientific study it is therefore important to tests more than one hypothesis. If a study includes only one hypothesis, and does not consider other possibilities, the researcher might reach statistical significance even if the tested hypothesis is practically insignificant. This problem is common with ANOVA studies. With ANOVA it is possible to reach statistical significance increasing the number or the similarity of the subjects used in the experiment. This happens also with CORAN studies. However correlations add the information of the strength of the relation. Using thousands of subjects, even variables which have an infinitesimal amount of common variance can reach statistical significance. It is therefore important to set “strength” values under which relations, even if statistically significant, are not considered.

With experimental studies only one hypothesis at a time is tested, and if the results are not positive, for example if the effect of a drug does not appear, the researchers may increase the number of laboratory animals and their similarity to force statistical significance. In this way it is possible to prove any cause-effect relation. It is needless to say that this way of producing information has no scientific value.

Using relational methodology variables are usually divided into two groups: those which describe the phenomena under investigation, and those which might correlate to it. For example, if the research topic is cancer, variables will include data which describe the type of cancer, its physiological parameters and history, and data which describe what is believed to be in connection with cancer, for example diet, environment, family life, and so on.

When preparing a form or a questionnaire it is, therefore, always important to put together different hypotheses. More hypotheses are compared the more the study becomes interesting, rich and useful.

4. Qualitative data, subjective answers and the social mask

Despite the social mask that each one uses, relational methodology works well with qualitative and subjective information; it is therefore possible to ask questions directly regarding personality, social environment, quality of life, etc. For example we can ask: *"do you feel alone?"* [ ] very much, [ ] little, [ ] not at all

Everyone knows that to this kind of question nearly everyone answers using a social mask. Even if the person feels alone, unsatisfied, depressed, he will probably try to present a more positive image, not only to others but also to himself.

Psychology has tried to solve this problem by developing psychological tests, which are designed to get information about these traits in an indirect way. Secondary questions which are connected and correlated to the trait which is being measured, and which the person does not distort with a social mask, are used. In this way it is possible to obtain an indirect estimation of anxiety, depression and other psychological characteristics.

Even if data is distorted by the effects of social masks and cannot be used in a quantitative way (average values or frequency distributions), it can be used successfully with correlation analysis. This happens because the mask is coherent: it is not used only on one variable, but on all the variables. For example if a subject says that he does not feel depressed, when in fact he is depressed, he will also say that he does not feel anxiety, when instead he feels it. The relation between depression and anxiety still shows up, even if the subjects are using masks.

As an example, in the case of a study in which 200 subjects have been interviewed, and asked if they felt depression or anxiety, the following results could be obtained:

	Depressed	Not depressed	Total
Anxiety	15	3	18
No anxiety	2	***180***	182
Total	17	183	200

Because of the social mask the answers are mainly grouped in the cell "no anxiety" and "no depression". Depression and anxiety come out as correlated (concomitance of the answer No). If psychological tests are used which provide an objective measure of depression and anxiety the result would shift to:

	Depressed	Not depressed	Total
Anxiety	***158***	10	168
No anxiety	2	30	32
Total	160	40	200

the answers are mainly grouped in the cell “yes anxiety” and “yes depression”. Because the relation is studied as presence of concomitances, depression and anxiety remain correlated (concomitances of the answer Yes)

This example shows that if a relation exists between variables, this relation shows up even if the data has been masked. This happens because the mask is used in a coherent way on all the answers. This last consideration shows that when asking someone to answer a questionnaire or a form where there is the risk that the social mask might distort the answers, it is important that the subject responds without any change in the time or setting, so that the use of the mask remains coherent over all the answers which have been given.

The possibilities given by relational methodology to analyze data distorted by social masks and qualitative data are an innovation in the field of research.

5. The sample

It is important to note that:

- when the aim is that of quantifying, a representative sample is necessary;
- when the aim is that of studying correlations, a diversified sample is needed.

If our aim is that of studying what explains drug addiction, a sample in which half of the population is drug addict and half is not drug addict will be used. The comparison of these two samples will tell us what is linked, correlated, to drug addiction. If the study would be carried out only on drug-addicts we could study quantitative information, but not what characterizes drug-addicts different from non-drug addicts.

In other words, in order to study correlations it is necessary to maximize the variance of the sample. The greater is the variance, the greater is the possibility to see correlations. When the sample is homogeneous correlations tend to disappear.