

# KUNST SOCSTAT

Sociometric status

By Maria Thissen  
&  
Nol Bendermacher

RTOG-FSW  
Radboud university Nijmegen

April 2012

# 1 Content

1	Content .....	2
2	Program description .....	4
2.1	Method overview .....	5
2.2	Individual scores .....	6
2.3	Probability scores .....	7
2.3.1	Probability scores for the criteria and the impact scale .....	7
2.3.2	Computation of probability scores for the criteria and the impact scale .....	10
2.3.3	Computation of probability scores for the preference scale .....	11
2.4	Classifications .....	12
2.4.1	The standard score model .....	12
2.4.2	The probability model .....	13
2.4.3	The GRD model .....	13
2.5	Dyads, cliques and group characteristics .....	14
2.6	Two data types .....	15
2.6.1	Flags .....	15
2.6.2	Case numbers .....	16
2.7	Two data formats: fixed or free .....	17
2.7.1	Fixed format .....	17
2.7.2	Free format .....	18
3	Files .....	19
4	Installing <i>SocStat</i> on Windows .....	20
5	Running <i>SocStat</i> .....	21
6	Menu options .....	23
6.1	The main menu .....	23
6.1.1	T Title .....	23
6.1.2	D Data definition .....	23
6.1.3	C Classification models .....	23
6.1.4	A Additional results .....	23
6.1.5	W Write results to raw file .....	23
6.1.6	O Open settings file .....	23
6.1.7	S Save current settings .....	24
6.1.8	X eXecute SOCSTAT .....	25
6.1.9	Q Quit .....	25
6.2	Definition of the data .....	25
6.2.1	N Name(s) of the data file(s) .....	25
6.2.2	B Blanks to separate groups .....	26
6.2.3	T Type of data (Numbers/Flags) .....	26
6.2.4	I case Identification field .....	27
6.2.5	C data contain Case numbers .....	27
6.2.6	F Free field or Fixed field format .....	27
6.2.7	M Maximum nominations per criterion .....	27
6.2.8	G maximum Group size .....	27
6.2.9	D Data list .....	27
6.2.9.1	Data list for case numbers with fixed format .....	28
6.2.9.2	Data list for case numbers with free format .....	29
6.2.9.3	Data list for flags with fixed format .....	31

---

6.2.9.4	Flags with free format.....	33
6.2.10	S Set names of the criteria .....	33
6.2.11	A Assumed case names .....	34
6.2.12	E Echo first .. cases.....	35
6.3	The Classification Models.....	35
6.3.1	P Probability model .....	35
6.3.2	Z standard Z-score model .....	35
6.3.3	G GRD model.....	36
6.3.4	T Tresholds for prob & GRD... ..	36
6.3.5	A All N None.....	37
6.4	Additional results in the listing file .....	37
6.4.1	T Table of raw scores .....	37
6.4.2	F Frequency table of scores .....	37
6.4.3	M group Measures .....	37
6.4.4	I sociographical Info on some cases .....	38
6.4.5	C Cliques and symmetric choices.....	38
6.4.6	G antaGonistic nominations .....	39
6.4.7	P Plot of status by Z score model .....	39
6.4.8	A All N None.....	39
7	Results.....	40
7.1	Results in the listing file .....	40
7.2	Results in a raw text file .....	52
8	Literature.....	54

## 2 Program description

Sociometry is a label for a family of methods and techniques that are used for the study of relations among the members of a group. The term "sociometry" was originally introduced in 1953 by J.L. Moreno in his book "Who Shall Survive". Sociometric techniques have been used in various research fields, especially in social sciences. An introduction can be found in Borgatta (1978). In the past decades sociometric techniques regained popularity in the study of peer relations in childhood and adolescence. An overview is given by Newcomb, Bukowski and Pattee (in press). A good Dutch introduction in the use of sociometric techniques in peer relations research is presented by Cillessen and ten Brink (1991). The computer program *SocStat* provides a tool for the use of some sociometric techniques. In this introductory section an overview of possibilities and limitations of the program is given.

The basic sociometric information that *SocStat* can process, consists of nominations of group members. An example of the gathering of this kind of information is the question: "Which children in your group do you like most?" Sociometric ratings, sociometric ranking or pairwise comparison of group members cannot be processed by *SocStat*.

The basic characteristic of *SocStat* is the reorganization of a dataset of nominations *given* by a group of persons into a new dataset of nominations *received* by persons in the same group. Several special arrangements of these received nominations, like compound scores, and special subgroup memberships can be created. In addition, special dyads and cliques can be traced and some group characteristics can be computed.

Another special characteristic of *SocStat* is the simultaneous usage of two sociometric criteria or questions, usually one "positive" criterion (for example: acceptance) and one "negative" criterion (for example: rejection). In this manual, the use of the program is explained and demonstrated with criteria called *Like* and *Dislike*, but the program offers possibilities for the use of other criteria.

Usually, all members of a group participate in a sociometric interview procedure. The two sociometric questions (or "criteria") acceptance and rejection are probably the most widespread in sociometric research on peer relations. They are often labeled or denoted as *Like* and *Dislike*. However, *SocStat* also allows the use of other sociometric criteria like sociometric behavioral or personality nominations.

The combined use of the acceptance- and rejection-criteria offers the possibility to compute compound scores like *Impact* or *Preference* and to apply three special sociometric classification models. Technically, these compound scores and models can also be used with other sociometric criteria, but the validity of such applications may be questionable.

For reasons of convenience acceptance (*Like*) is sometimes called a "positive" criterion and rejection (*Dislike*) a "negative" criterion in this manual. This might suggest that the two criteria must be each other's opposite. Technically, this is not required. In the case of rejection and acceptance, the correlation between the two criterion variables is usually negative and only moderate to high, say  $-0.70$ , which is far from perfect. Such an imperfect negative correlation might be interpreted as an effect of measurement errors. However, in the case of rejection and acceptance the imperfect negative correlation is usually considered theoretically important. It is assumed that these two criteria are at least partially independent. This allows for the sociometric classification of special subgroups of children, like controversial children who receive both many like and many dislike nominations.

The choice of sociometric criteria is an important issue in the organization of a sociometric interview. Another important issue is the decision whether to use fixed or free numbers of nominations. Fixed numbers of nominations can be gathered by using questions such as: "Name the three children in your group you like most". If the number of nominations is free, one can use a formulation such as "Which children in your group do you like most?" The fixed-number-approach may sometimes force cases to nominate more or less group members than the free number approach. Consider, for example, a sociometric question about "best friends". In the fixed-number-approach, a person may be asked to name his three best friends in a group. In reality, this person may not have a best friend in that group, or only one, or maybe four. A further, more serious, disadvantage of the use of fixed numbers of nominations is that the number of mutual nominations is too limited or too large, which may limit valid tracing of dyads and cliques. A final disadvantage of fixed numbers of nominations is that possibilities for examination of differences between groups or between criteria are diminished. In general, the choice of the method, free or fixed, influences the probability of occurrence of numbers of received nominations. SocStat allows a compromise between the two approaches: the number of nominations is free, but there is a maximum. If less nominations than the maximum are given, the remaining values are marked as *missing values*.

Sometimes cases nominate themselves. Self-nominations are mostly not accounted for in sociometric interviews. However, the nature of the criterion may determine whether self-nominations should be allowed in the sociometric interview procedure. *SocStat* ignores these self-nominations.

Sometimes a child nominates the same group member on contrasting sociometric questions. For example, a child may nominate another child both as most liked and as most disliked. Such pairs of nominations are considered ambiguous by *SocStat* although the interpretation may depend on the nature of the criteria. *SocStat* issues warnings if ambiguous nominations are found.

## 2.1 Method overview

*SocStat* analyzes sociometric data and derives several measures from it. The original nominations stem from individual group members judging each other on two criteria. Often the group is a school class and the members are children. The judgments may be elicited by questions like "name the classmates you like most" and "name the children you like least". Each of these criteria contains information about individual children and about their relation with other children. *SocStat* calls the group members **cases**.

Like					Dislike				
0	1	0	1	0	0	0	1	0	1
0	0	1	1	1	1	0	0	0	0
1	0	0	0	0	0	1	0	1	1
0	1	0	0	1	0	0	0	0	0
0	0	1	0	0	0	0	0	1	0

Figure 1: Example of the two basic data matrices

The base of all computations is formed by two matrices in which the nominations are summarized. Figure 1 shows an example for a group with 5 cases. The rows of these matrices represent the nominating cases, the columns the nominated cases. A 1 indicates a nomination. In the example, case 1 *likes* cases 2 and 4 and *dislikes* cases 3 and 5 and so on.

From the responses on the two criteria several scores can be computed for each individual case:

- 1 the number of times a case is nominated as liked defines its raw *like* score.
- 2 the number of times a case is nominated as disliked (or least liked) defines its raw *dislike* score.
- 3 the total number of cases that nominate a case on either of the criteria defines its raw *impact* score.
- 4 the difference between a cases' *Like* and *Dislike* scores defines its raw *preference* score.

On the base of its *Like* and *Dislike* scores, a child can be classified into one of five sociometric status categories: popular, rejected, neglected, controversial or average. *Soc-Stat* allows three ways to perform this classification. They are called the *classification models*.

In addition, *SocStat* may derive some group related results:

- 1 an overview of cliques (subgroups in which all cases nominate each other) or dyads (mutual nominations on the same criterion).
- 2 a list of antagonistic nominations (*a* likes *b*, but *b* dislikes *a*).

## 2.2 Individual scores

From the data several single scores are derived, based on one criterion each, as well as some compound or "multiple" scores, based on the two criteria.

The raw score of a case is the total number of times it is nominated on a certain criterion. The raw score derived from the *Like* criterion is called the *raw like score*. The raw score derived from the *Dislike* criterion is called the *raw dislike score*.

In addition, some transformations on raw scores are performed. The aim of these transformations is to make the scores comparable between criteria and between groups.

The following transformations are computed:

- the *standard like scores*: also called *like-Z-scores*: the raw like score diminished by its group mean and then divided by its standard deviation.
- the *standard dislike scores*: analogous to the standard like score.
- the *probability of the like score*, given the distribution of the nominations. This probability is based on the generalized binomial distribution. A high score indicates a high social acceptance. The use of the generalized binomial distribution instead of the simple binomial distribution allows for a variable number of nominations per nominating case. A detailed description of the logic and computation of these probability scores is given in 2.3.
- the *probability of the dislike score*: analogous to the probability of the like score.

The compound scores are called "social impact" and "social preference":

- the *raw impact score* is the sum of the raw like score and the raw dislike score, corrected for ambiguous nominations. An ambiguous nomination contributes one point to the raw impact score, not two. This score is a measure for the degree to which a child is visible or standing out in the group, but it tells nothing about the positive or negative character of this visibility or 'outstandingness'.
- the *raw preference score* is the difference of the raw like score and the raw dislike score. It reflects the predominance of the like score over the dislike score.

Along with the two raw compound scores, some transformations are computed:

- the *standard impact score* (impact z-score) is the standardized sum of the standardized like score and the standardized dislike score.
- the *standard preference score* (preference z-score) is the standardized difference of the standardized like score and the standardized dislike score.
- the *probability of the impact score* (impact P-score) the probability of the impact score, based on the generalized binomial distribution.
- the *probability of the preference score* (preference score): the probability of the preference score based on the generalized multinomial distribution (with possible values -1, 0 and 1 for the differences between the individual nominations on the like and dislike criterion).

## 2.3 Probability scores

### 2.3.1 Probability scores for the criteria and the impact scale

The probability scores on the two criteria and on the impact scale are based on the generalized binomial distribution. The probability scores on the preference scale are based on the generalized multinomial distribution. In this section, we will explain the logic and the computational details of these approaches. As an illustration, we will use the nominations on the first criterion in the example of figure 1.

If a case would be nominated by its group mates randomly with a fixed probability  $p$ , the number of nominations it received would follow a binomial distribution with parameters  $g-1$  and  $p$ , where  $g$  is the group size and  $g-1$  is the number of 'trials'. However, in our situation each trial is performed by another nominating case with possibly another probability  $p$ . The resulting distribution is called the **generalized** binomial distribution.

	Nominated					freq.	$\hat{p}$	$\hat{q} = 1 - \hat{p}$
	-	1	0	1	0			
Nominating	-	1	0	1	0	2	0.50	0.50
	0	-	1	1	1	3	0.75	0.25
	1	0	-	0	0	1	0.25	0.75
	0	1	0	-	1	2	0.50	0.50
	0	0	1	0	-	1	0.25	0.75
Total	1	2	2	2	2	9	2.25	2.75
Expected	1.75	1.50	2.00	1.75	2.00			

Figure 2: Matrix of nominations and related information.

Figure 2 shows in its first part the five by five table of like-nominations in the example. The probability scores we want to find are the (cumulative) probabilities of the numbers of received nominations per case, that is the (cumulative) probabilities of the five first numbers in the row with label 'Total'.

Each actual score is the result of four trials, i.e. four nominations by the cases. First thing to do is to estimate the probabilities of a nomination by each of these cases. They can simply be estimated by the proportion of nominations given by the corresponding case: the number of nominations in the row divided by its maximum, that is by  $g-1$ . Figure 2 shows these estimated probabilities in the column  $\hat{p}$ . For practical reasons we define also  $\hat{q} = (1 - \hat{p})$ .

Once we have found the estimated probabilities  $\hat{p}$ , we can compute the expected values for the numbers of nominations *received*, by adding for each case all  $\hat{p}$  values except that of the case itself. For example, the expected raw score for the first case is:  $0.75 + 0.25 + 0.50 + 0.25 = 1.75$ . Figure 2 shows the expected values in its last row.

Knowing the probabilities  $\hat{p}$ , we can also find the probability of each *pattern* of received nominations. Note that these patterns are the **columns** of the matrix, not the rows. Since trials are supposed to be independent, a pattern of zeros and ones has as its probability the product of the corresponding  $\hat{q}$  and  $\hat{p}$  values. It should be noted that the probability of a certain pattern is different for different cases, since the groups of nominating cases are different. Figure 3 shows a list of all possible patterns and the corresponding probabilities for the first two cases.



		First case	Second case
	score	probability	probability
0000	0	$\hat{q}_2 \times \hat{q}_3 \times \hat{q}_4 \times \hat{q}_5 = 0.0703125$	$\hat{q}_1 \times \hat{q}_3 \times \hat{q}_4 \times \hat{q}_5 = 0.1406250$
0001	1	$\hat{q}_2 \times \hat{q}_3 \times \hat{q}_4 \times \hat{p}_5 = 0.0234375$	$\hat{q}_1 \times \hat{q}_3 \times \hat{q}_4 \times \hat{p}_5 = 0.0468750$
0010	1	$\hat{q}_2 \times \hat{q}_3 \times \hat{p}_4 \times \hat{q}_5 = 0.0703125$	$\hat{q}_1 \times \hat{q}_3 \times \hat{p}_4 \times \hat{q}_5 = 0.1406250$
0011	2	$\hat{q}_2 \times \hat{q}_3 \times \hat{p}_4 \times \hat{p}_5 = 0.0234375$	$\hat{q}_1 \times \hat{q}_3 \times \hat{p}_4 \times \hat{p}_5 = 0.0468750$
0100	1	$\hat{q}_2 \times \hat{p}_3 \times \hat{q}_4 \times \hat{q}_5 = 0.0234375$	$\hat{q}_1 \times \hat{p}_3 \times \hat{q}_4 \times \hat{q}_5 = 0.0468750$
0101	2	$\hat{q}_2 \times \hat{p}_3 \times \hat{q}_4 \times \hat{p}_5 = 0.0078125$	$\hat{q}_1 \times \hat{p}_3 \times \hat{q}_4 \times \hat{p}_5 = 0.0156250$
0110	2	$\hat{q}_2 \times \hat{p}_3 \times \hat{p}_4 \times \hat{q}_5 = 0.0234375$	$\hat{q}_1 \times \hat{p}_3 \times \hat{p}_4 \times \hat{q}_5 = 0.0468750$
0111	3	$\hat{q}_2 \times \hat{p}_3 \times \hat{p}_4 \times \hat{p}_5 = 0.0078125$	$\hat{q}_1 \times \hat{p}_3 \times \hat{p}_4 \times \hat{p}_5 = 0.0156250$
1000	1	$\hat{p}_2 \times \hat{q}_3 \times \hat{q}_4 \times \hat{q}_5 = 0.2109375$	$\hat{p}_1 \times \hat{q}_3 \times \hat{q}_4 \times \hat{q}_5 = 0.1406250$
1001	2	$\hat{p}_2 \times \hat{q}_3 \times \hat{q}_4 \times \hat{p}_5 = 0.0703125$	$\hat{p}_1 \times \hat{q}_3 \times \hat{q}_4 \times \hat{p}_5 = 0.0468750$
1010	2	$\hat{p}_2 \times \hat{q}_3 \times \hat{p}_4 \times \hat{q}_5 = 0.2109375$	$\hat{p}_1 \times \hat{q}_3 \times \hat{p}_4 \times \hat{q}_5 = 0.1406250$
1011	3	$\hat{p}_2 \times \hat{q}_3 \times \hat{p}_4 \times \hat{p}_5 = 0.0703125$	$\hat{p}_1 \times \hat{q}_3 \times \hat{p}_4 \times \hat{p}_5 = 0.0468750$
1100	2	$\hat{p}_2 \times \hat{p}_3 \times \hat{q}_4 \times \hat{q}_5 = 0.0703125$	$\hat{p}_1 \times \hat{p}_3 \times \hat{q}_4 \times \hat{q}_5 = 0.0468750$
1101	3	$\hat{p}_2 \times \hat{p}_3 \times \hat{q}_4 \times \hat{p}_5 = 0.0234375$	$\hat{p}_1 \times \hat{p}_3 \times \hat{q}_4 \times \hat{p}_5 = 0.0156250$
1110	3	$\hat{p}_2 \times \hat{p}_3 \times \hat{p}_4 \times \hat{q}_5 = 0.0703125$	$\hat{p}_1 \times \hat{p}_3 \times \hat{p}_4 \times \hat{q}_5 = 0.0468750$
1111	4	$\hat{p}_2 \times \hat{p}_3 \times \hat{p}_4 \times \hat{p}_5 = 0.0234375$	$\hat{p}_1 \times \hat{p}_3 \times \hat{p}_4 \times \hat{p}_5 = 0.0156250$

Figure 3: Possible nomination patterns and their probabilities for the first two cases.

By combining patterns with the same score, figure 3 can be collapsed to a list of probabilities for the raw scores. Figure 4 shows the resulting probabilities for all cases. In each cell the second number (with gray background) is the cumulative probability, i.e. the probability of a score less than or equal to the score in the column header.

Case	Scores				
	0	1	2	3	4
1	0.0703125	0.3281250	0.4062500	0.1718750	0.0234375
Cum.	0.0703125	0.3984375	0.8046875	0.9765625	1.0000000
2	0.1406250	0.3750000	0.3437500	0.1250000	0.0156250
Cum.	0.1406250	0.5156250	0.8593750	0.9843750	1.0000000
3	0.0468750	0.2500000	0.4062500	0.2500000	0.0468750
Cum.	0.0468750	0.2968750	0.7031250	0.9531250	1.0000000
4	0.0703125	0.3281250	0.4062500	0.1718750	0.0234375
Cum.	0.0703125	0.3984375	0.8046875	0.9765625	1.0000000
5	0.0468750	0.2500000	0.4062500	0.2500000	0.0468750
Cum.	0.0468750	0.2968750	0.7031250	0.9531250	1.0000000
Overall	0.0750000	0.3062500	0.3937500	0.1937500	0.0312500
Cum.	0.0750000	0.3812500	0.7750000	0.9687500	1.0000000

Figure 4: Single and cumulative probabilities of specific raw scores per case.

The program will also compute the *overall* probabilities of the specific scores: they are just the means over all cases; in figure 4, these overall probabilities are reported in the last row. The probability scores as *SocStat* computes them are closely related to the cumulative probabilities in figure 4. For a case  $c$  with actual score  $x$  the cumulative score gives the probability that case  $c$  would obtain a score less than or equal to  $x$ .

However, there is still a small problem to be solved. If we define the probability score in this way, the probability score for the fourth case with raw score 2 (see figure 4) would be 0.8046875, suggesting a rather high score, but in fact 2 is the most probable score of all with a probability as high as 0.40625. If we would define the probability score as the probability that cases obtain a score *less than* his actual score, we run into a similar problem on the lower end of the scale: see for instance the score for case 2. The problem is the discreteness of the distribution, especially with small groups. *SocStat* solves the problem by defining the probability score as follows:

For a case with raw score  $x$  on a criterion or on the impact scale, the corresponding probability score is the probability of a score less than  $x$  plus **half** the probability of a score equal to  $x$ , with probabilities based on the generalized binomial distribution.

Case	Raw score	Probability score
1	1	0.2343750
2	2	0.6875000
3	2	0.5000000
4	2	0.6015625
5	2	0.5000000

Figure 5: Raw and probability scores.

Figure 5 shows the resulting probability scores for our example.

### 2.3.2 Computation of probability scores for the criteria and the impact scale

The enumeration of all possible nomination patterns as in figure 3 is not the most practical way to derive the probabilities scores based on the generalized binomial distribution. Instead, *SocStat* uses the following strategy:

Define  $\text{prob}(n,m)$  = the probability that the first  $n$  other cases nominate a case  $m$  times. A score of  $m$  nominations by the first  $n$  other cases can only be realized if the number of nominations by the first  $n-1$  other cases was  $m$  or  $m-1$ .

$$\text{Therefore } \text{prob}(n,m) = \text{prob}(n-1,m-1) \times \hat{p}_n + \text{prob}(n-1,m) \times \hat{q}_n.$$

With 0 cases the only possible score is 0 with probability 1.

With 1 case, the possible scores are 0 with probability  $\hat{q}_1$  and 1 with probability  $\hat{p}_1$

In general: when the  $n$ -th other case is added there are two ways to reach a score  $s$ :

- 1) the first  $n-1$  cases resulted in a score  $s-1$  and the  $n$ -th case assigns a 1
- 2) the first  $n-1$  cases resulted in a score  $s$  and the  $n$ -th case assigns a 0.

So with  $m$  cases the probability of score  $s$  can be computed as:

$$\text{prob}(n,m) = \text{prob}(n-1,m-1) \times \hat{p}_{n^*} + \text{prob}(n-1,m) \times \hat{q}_{n^*}$$

where  $n^*$  is the sequence number of the  $n$ -th case with the nominated case itself left out.

So for the score of case 1 the subsequent values of  $n^*$  are 0 (no cases), 2 (first case added), 3(second case added) and so on.

If we apply this algorithm to the probabilities of the scores for case 1, the computational steps are as depicted in figure 6.

Score	none	Cases added			
		case 2 $\hat{q} = 0.25$ $\hat{p} = 0.75$	case 3 $\hat{q} = 0.75$ $\hat{p} = 0.25$	case 4 $\hat{q} = 0.50$ $\hat{p} = 0.50$	case 5 $\hat{q} = 0.75$ $\hat{p} = 0.25$
0	1.0000000	0.2500000	0.1875000	0.0937500	0.0703125
1		0.7500000	0.6250000	0.4062500	0.3281250
2			0.1875000	0.4062500	0.4062500
3				0.0937500	0.1718750
4					0.0234375

Figure 6: Computation of score probabilities for one case.

### 2.3.3 Computation of probability scores for the preference scale.

The computation of the probability scores for the preference scale is based on an approach that is similar to that for the computation of the probability scores for the two criteria and for the impact scale. The only difference is that the cells of the preference matrix can have three different values: -1, 0 and 1. Figure 7 shows the preference matrix for the first matrix in figure 1.

		Preference					$\hat{p}$	$\hat{q}$	$\hat{r}$
		1	-	1	-1	1			
Nominating cases	1	-	1	-1	1	-1	0.00	0.50	0.50
	2	-1	-	1	1	1	0.00	0.25	0.75
	3	1	-1	-	-1	-1	0.00	0.75	0.25
	4	0	1	0	-	1	0.50	0.00	0.50
	5	0	0	1	-1	-	0.50	0.25	0.25

Figure 7: Preference matrix and row probabilities.

Consequently, the theoretical distribution is no longer a generalized binomial, but a generalized *multinomial* distribution. If the group size is  $g$ , possible values range from  $-g+1$  to  $g-1$ . For each nominating case, we estimate the probability of these three values from the preference matrix. We define the three estimates as:

$\hat{p}$  = estimated probability of 0,

$\hat{q}$  = estimated probability of -1

and  $\hat{r}$  = estimated probability of +1.

We define  $\text{prob}(n,m)$  = the probability that the first  $n$  cells of a cases column add up to  $m$ . The nominating cases are introduced one by one. Initially all probabilities are zero except that of score 0, which is 1. If a new case is added, a score  $x$  can only be reached if one of the following situations exist:

- the old score was  $x$  and the new case adds 0
- the old score was  $x-1$  and the new case adds 1
- the old score was  $x+1$  and the new case adds -1

So if a case is added with estimated probabilities  $\hat{p}$ ,  $\hat{q}$  and  $\hat{r}$ , the new probability  $\text{prob}(x)$  can be computed as:

$$\text{prob}(n-1,x-1) \times \hat{r} + \text{prob}(n-1,x) \times \hat{p} + \text{prob}(n-1,x+1) \times \hat{q}$$

The computation of the probabilities of the possible scores for case 1 in our example is shown in figure 8.

Score	Cases added				
	None	case 2 $\hat{r} = 0.75$ $\hat{p} = 0.00$ $\hat{q} = 0.25$	case 3 $\hat{r} = 0.25$ $\hat{p} = 0.00$ $\hat{q} = 0.75$	case 4 $\hat{r} = 0.50$ $\hat{p} = 0.50$ $\hat{q} = 0.00$	case 5 $\hat{r} = 0.25$ $\hat{p} = 0.50$ $\hat{q} = 0.25$
-4					0.0000000
-3				0.0000000	0.0234375
-2			0.1875000	0.0937500	0.0703125
-1		0.2500000	0.0000000	0.0937500	0.1484375
0	1.0000000	0.0000000	0.6250000	0.3125000	0.2578125
1		0.7500000	0.0000000	0.3125000	0.2578125
2			0.1875000	0.0937500	0.1484375
3				0.0937500	0.0703125
4					0.0234375

Figure 8: Computation of score probabilities of the preference scale for one case.

Finally the definition of the probability scores on the preference scale is as follows:

The probability score for a case with raw preference score  $x$  is the probability of a score less than  $x$  plus **half** the probability of a score  $x$ , with probabilities based on the generalized polynomial distribution.

## 2.4 Classifications

Group members can be classified into social status groups, based on the like and dislike scores. *SocStat* offers three classification systems: the *standard score model* (Coie, Dodge and Coppotelli, 1982; Coie and Dodge, 1983), the *probability model* (Newcomb and Bukowski, 1983) and a small modification of the latter: the *GRD model*\*.

### 2.4.1 The standard score model

The standard score model is based on the 4 standard scores ( $Z$ -scores) of like, dislike, preference and impact. The 5 status types are defined as follows:

- (a) the **popular** group consists of cases with a standard preference score greater than 1.0, a positive standard like score and a negative standard dislike score.
- (b) members of the **rejected** group have a standard preference score less than -1.0, a negative standard like score and a positive standard dislike score.
- (c) cases in the **neglected** group have a standard impact score less than -1.0, a negative standard like score and a negative standard dislike score.
- (d) the **controversial** group consists of cases with a standard impact score greater than 1.0, a positive standard like score and a positive standard dislike score.
- (e) the **average** group contains all individuals not belonging to one of the foregoing groups.

---

\* named after the former department "Groep Rekentechnische Dienstverlening" that developed the first version of *SocStat*.

### 2.4.2 The probability model

The probability model is based on the probabilities of the raw like score and the raw dislike score as well as the probability of the social impact score. The definition of an extremely high or low probability depends on the user-defined threshold level  $\alpha^*$ .

The 5 status types are defined as follows:

- (a) the **popular** group consists of cases with an extremely high like score (probability greater than or equal to  $1-\alpha$ ) and a raw dislike score less than or equal to the mean.
- (b) members of the **rejected** group have an extremely high dislike score (probability greater than or equal to  $1-\alpha$ ) and a raw like score less than or equal to the mean.
- (c) cases in the **neglected** group have an extremely low impact score (lower than  $\alpha$ ).
- (d) the **controversial** group consists of cases with a extremely high like score (probability greater than or equal to  $1-\alpha$ ) and a raw dislike score greater than the mean, and cases with an extremely high dislike score (probability greater than or equal to  $1-\alpha$ ) and a raw like score greater than the mean.
- (e) the **average** group contains all cases not belonging to one of the foregoing groups.

*SocStat* will show a plot of the impact scores versus the preference scores.

### 2.4.3 The GRD model

The GRD model is very similar to the probability model. Only the neglected status group is defined slightly different. By letting out the social impact score from the definition, the classification as well as the original like and dislike scores can be summarized in one simple two-dimensional scheme, as shown in figure 9.

The categories are defined as follows:

- (a) the **popular** group consists of cases with an extremely high like score (probability greater than or equal to  $1-\alpha$ ) and a raw dislike score less than or equal to the mean, and cases with a extremely low dislike score (probability less than or equal to  $\alpha$ ) and a raw like score greater than the mean.
- (b) members of the **rejected** group have an extremely high dislike score (probability greater than or equal to  $1-\alpha$ ) and a raw like score less than or equal to the mean, or they have an extremely low like score (probability less than or equal to  $\alpha$ ) and a raw dislike score greater than the mean.
- (c) the **neglected** group consists of cases with an extremely low like score (probability less than or equal to  $\alpha$ ) and a raw dislike score less than or equal to the mean, and cases with an extremely low dislike score (probability less than or equal to  $\alpha$ ) and a raw like score less than or equal to the mean.

---

\* We avoid the word *significance level*, since there is no a priori hypothesis to be tested.

- (d) the **controversial** group consists of cases with a extremely high like score (probability greater than or equal to  $1-\alpha$ ) and a raw dislike score greater than the mean, and cases with a extremely high dislike score (probability greater than  $1-\alpha$ ) and a raw like score greater than the mean.
- (e) the **average** group contains all cases not belonging to one of the foregoing groups.

*SocStat* will show the connection between the like and the dislike scores and the classification in a 4 by 4 frequency table where the cells can be attributed to the five categories in a very straight forward way as shown in figure 9. Note that the diagonals represent the scales impact and preference.

		Scale Dislike				
		Low	Exp	High		
Scale	Highest Prefer.	High	Popular	Popular	Controversial	Controversial
	Exp.	Like	Popular	Average	Average	Controversial
	Like	Low	Neglected	Average	Average	Rejected
	Lowest Impact	Lowest Prefer.	Neglected	Neglected	Rejected	Rejected
Legend:		High: > 1 - threshold level Low: < threshold level Exp.: expected value (group mean)				

Figure 9: Summary table of the GRD model.

### 2.5 Dyads, cliques and group characteristics

If two cases nominate each other on a criterion, the nominations are called *symmetric*. *SocStat* has (limited) possibilities for tracing such pairs of cases. This feature may for instance be used to trace mutual friends. The relevant sociometric question then could be: "Which children in your group are your best friends?"

The simultaneous use of two contrasting criteria allows for the tracing of *antagonistic* pairs, for instance a pair of cases in which the first case likes (and not dislikes) the second and the second dislikes (and not likes).

*SocStat* calls a group of cases all nominating each other a *clique*. The theoretical number of cliques is very large, but if their actual number is not too large, the program can report them. However, the validity of the clique report may be questioned if the number of nominations is restricted to a small number.

The program computes four group characteristics:

- Group expansiveness: the number of nominations made in the group divided by its maximum, i.e. by the total number of off-diagonal cells in the matrix.
- Group cohesion: the number of symmetric pairs of nominations divided by the maximum possible number of pairs.
- Group integration index:  $1/(x+1)$  where  $x$  is the total number of items not nominated at all.
- Group cohesion index: the number of symmetric nominations (each symmetric pair adds 2) divided by the number of unidirectional nominations.

## 2.6 Two data types

As said before, the data consist of nominations on two criteria: a positive criterion (like) and a negative one (dislike). Therefore, the data must provide information for two matrices, where the first one describes the responses on the positive criterion and the second one the responses on the negative criterion.

This information can be given in two alternative ways: (1) flags or (2) case numbers.

### 2.6.1 Flags

If the data are coded as flags, each row in the data contains the nominations of one case on both criteria in the form of zeros and ones. The columns represent the same cases, so the matrices are always square. A one in a cell indicates that the row case nominates the column case, a zero that it does not.

With flags the input file contains the basic matrices directly as they are used by the program. Yet, there are still a few possible variations. Figure 10 shows the simplest form for the example data in figure 1.

0	1	0	1	0	0	0	1	0	1
0	0	1	1	1	1	0	0	0	0
1	0	0	0	0	0	1	0	1	1
0	1	0	0	1	0	0	0	0	0
0	0	1	0	0	0	0	0	1	0

Figure 10: Simple input with flags.

In this form, the file contains nothing but the 0/1 flags on both criteria. However, it is possible to add some information: case numbers and/or case identifications. A *case number* is a numerical code (a whole number) to identify a case; a *case identification* is an alphanumeric string (not longer than 18 characters) naming the case. Figure 11 shows the example data with both identifications added.

1	John	0	1	0	1	0	0	0	1	0	1
2	David	0	0	1	1	1	1	0	0	0	0
3	Mary	1	0	0	0	0	0	1	0	1	1
4	Bill	0	1	0	0	1	0	0	0	0	0
5	Tony	0	0	1	0	0	0	0	0	1	0

Figure 11: Enriched input with flags.

If no case numbers are specified in the data, the program automatically assigns numbers from 1 to the group size.

The program assumes that the rows and columns match perfectly: the first case and the first column represent the same case and so on. Therefore, their order must be arranged very carefully in the input file. For the computation of the group measures it is also important that all cases are represented in the data, even these cases that do not give or receive any nomination, must be included.

### 2.6.2 Case numbers

With this form of data, nominations are coded as a list of case numbers. There is a maximum number of nominations that a case can give on each criterion. Figure 12 shows a file with a maximum of 3 nominations per row.

2	4	0	3	5	0
3	4	5	1	0	0
1	0	0	2	4	5
2	5	0	0	0	0
3	0	0	4	0	0

Figure 12: Simple input with numbers.

In this simple form, the cases receive automatically case numbers from 1 to the group size. If a case gives less than the maximum number of nominations, the remaining positions must be filled with *missing* values. These values can be any number outside the range of case numbers. So figure 12 shows that case 1 nominates cases 2 and 4 on the like criterion and cases 3 and 5 on the dislike criterion.

Like the flag data, these number data may be enriched with case numbers and/or case identification strings. The case numbers may for example contain a code for the group. Figure 13 gives an example where the first digit of the case numbers might be a group code:

11	John	12	14	0	13	15	0
12	David	13	14	15	11	0	0
13	Mary	11	0	0	12	14	15
14	Bill	12	15	0	0	0	0
15	Tony	13	0	0	14	0	0

Figure 13: Enriched input with numbers.

If the input contains case numbers, the nominations must refer to these numbers. Nominations that are not contained in the list of case numbers are treated as missing values. In figure 13, one can see that John nominates two cases on the first criterion (David and Bill) and two on the second (Mary and Tony). The two zeros in the first row are interpreted as *missing values*. Likewise, any number outside the set [11,12,13,14,15] would be interpreted as a *missing value*.



The case identification strings will be used to make the output in the listing file more understandable. The program offers the opportunity to specify default case identifications to be used if the data do not contain such information at all or if for some cases the corresponding field is blank.

## 2.7 Two data formats: fixed or free

For both types of data, flags and numbers, there are two formats that may be used: fixed or free format.

### 2.7.1 Fixed format

In a *fixed* format, each number has an exact position: for each value it is exactly defined in which line it is given (within the row) and also where it is to be found exactly in that line. In addition, the number of lines per case is fixed.

```

◦11◦John◦◦◦12◦14◦◦0
◦13◦15◦◦0
◦12◦David◦◦◦13◦14◦15
◦11◦◦0◦◦0
◦13◦Mary◦◦◦11◦◦0◦◦0
◦12◦14:15
◦14◦Bill◦◦◦12◦15◦◦0
◦◦0◦◦0◦◦0
◦15◦Tony◦◦◦13◦◦0◦◦0
◦14◦◦0◦◦0

```

Figure 14: Number data in a fixed format (spaces replaced by ◦).

Figure 14 shows an example of a file that obeys to a fixed format. In this figure, spaces are replaced by the symbol ◦ in order to make it easier to count the positions of the numbers. The format can be described as follows:

1. case identification string: line 1, columns 5-10
2. case number: line 1, columns 1-3
3. first nomination on like: line 1, columns 11-13
4. second nomination on like: line 1, columns 14-16
5. third nomination on like: line 1, columns 17-19
6. first nomination on dislike: line 2, columns 1-3
7. second nomination on dislike: line 2, columns 4-6
8. third nomination on dislike: line 2, columns 7-9

It should be noted that the case identification string is by definition to be found in the first line.

The data must very strictly correspond to this format: if for example the first nomination in the first line is shifted one position to the left, the program would read it as 120 (12◦) in stead of 12 (◦12).

### 2.7.2 Free format

In a *free* format, the exact positions of the values may be different from row to row, although their order must be always the same.

If the data are case numbers, the format may enforce line numbers or not. If so, a line may contain additional (and superfluous) information after the indicated numbers.

If the data are flags, there can only be one line per row. (The reason is that the program uses the number of values in the first line to determine the group size.)

Figure 15 shows number data in free format. It should be noted that even in free format data the case identification string has a fixed position in the first line of each row.

11	John	12	14	0	13	15	0
12	David	13	14	15	11	0	0
13	Mary	11	0	0	12	14	15
14	Bill	12	15	0	0	0	0
15	Tony	13	0	0	14	0	0

Figure 15: Number data in a free format.

**In general, the use of free format is highly recommended since it is less sensitive to irregularities in the data and errors in the specifications.**

There are, however, a few situations where a fixed format is unavoidable:

- If the data contain values that should be skipped.
- If not all values in the data are separated by spaces, commas or tabs.
- If the data contain case numbers but these do not precede the nominations.
- If the data contain case numbers and these overlap with the case identification field.
- If some values are left out, i.e. left completely blank: in a fixed format situation they will be seen as zeros and interpreted as missing values (unless there is a case with case number 0), but in a free format the program will just take the next number and thereby mix up the columns.

## 3 Files

There are five file types that may be important to this program:

### **data files:**

The files that contain the data to be analyzed. One program run can analyze several files and each file can contain several groups. The chosen options and data definitions will apply to all these files and groups.

### **settings files:**

These files are used to save the options as they are specified by a user. A settings file contains all information about the analyses to be performed, including the description of the data, but not the data themselves. Many settings files may be available. By default, their extension is `'.setsoc'`.

### **listing files:**

A listing file contains the main results of an analysis in a nice layout for human readers. There will be one listing file for each data file. The name will borrow its first part from the data file and end on `'.LST'` (or on `'1.LST'`, `'2.LST'`, ... and so on). You can inspect it by any editor, but in order to have an orderly layout you must view it in a small non-proportional font like Courier New 9. Listing files are stored in the folder in which the corresponding input files are found.

### **raw output files:**

Depending on the chosen options, the program may produce one file with “raw” output for each input file. These raw files are meant to be input to other programs. Their names will take their first part from the data files and end with `'.OUT'` (or `'1.OUT'`, `'2.OUT'`, ... and so on). SocStat shows in each listing file the syntax that you need in order to open the corresponding raw output file from the statistical package SPSS.

### **plot files:**

If the program produces any plots (graphical representations) the listing files will contain coarse versions of the plots, but the program will also produce more refined bitmaps. The names of these bitmap files will take their first part from the data files and, on Windows machines, they will end with `'.BMP'` (or `'1.BMP'`, `'2.BMP'`, ... and so on).

## 4 Installing *SocStat* on Windows

The installation of the program is very simple:

1. Copy the file *SocStat.exe* to any place on your hard disk. Optionally you may make shortcuts on the task bar and/or on the desktop.
2. After the first time you have used the program, double click on the listing file. Windows will ask you to select the program to be used when opening the file. Select a simple text editor like Notepad or WordPad.
3. After the first time you have saved the program settings, double click on the settings file. Windows will ask you to select the program to be used when opening the file. Select the program *SocStat.exe* or any shortcut to it.

That is all: from now on, you can start the program by double clicking the exe-file, one of its shortcuts or a settings file and you can open a listing file by double clicking it.

## 5 Running *SocStat*

To run *SOCSTAT* you must double click on its executable file (for Windows: *Soc-Stat.exe*) or, if you have used the program before, on one of its settings files (for instance *Current.setsoc*). The first thing you will see then, is the main window, as shown in figure 16.

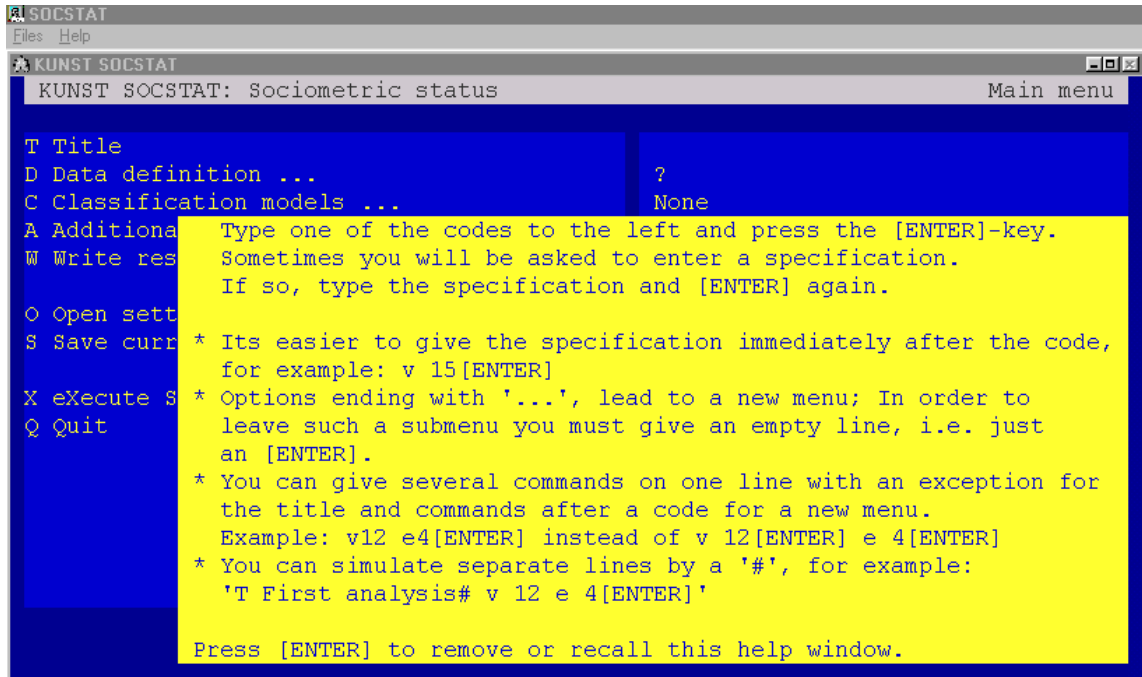


Figure 16: The main window.

On the screen the light part with the text 'Type one ...' is a yellow text window. Windows like that contain hints and explanations. If you have read the text (or do not need it), you can press the Enter-key and the yellow window will disappear (see figure 17).

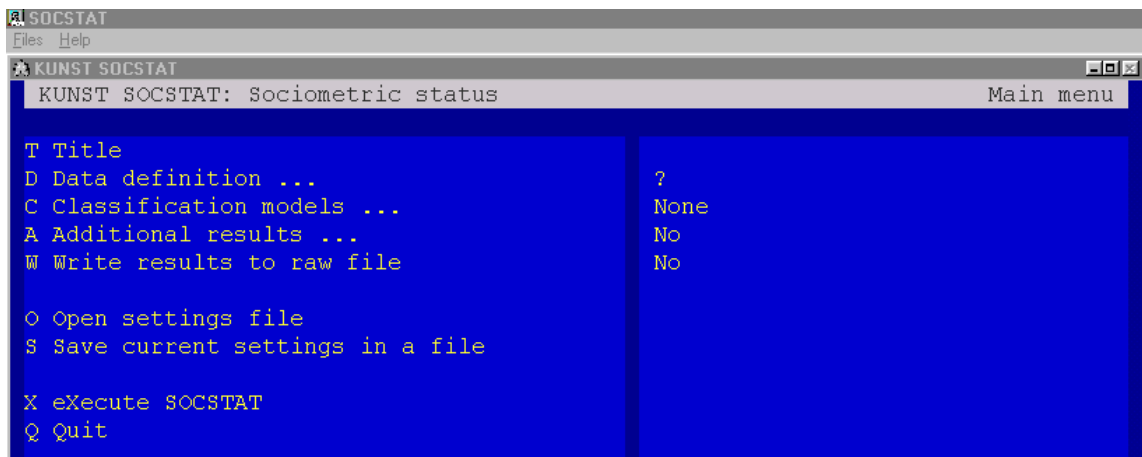


Figure 17: The main window without the help information.

Now you can see the entire main window. The left part is the main menu. It consists of a list of options, each preceded by a one-character code. To select an option, you must type the code, followed by the information you want to give and then the Enter-key. From now on, we will indicate the Enter-key as: `Enter`. You may for instance type: `T Analyzing first sessionEnter` to define the title that will appear as a header in the listing files.

If you do not know what the meaning of an option is, you may just enter its code and `Enter`. There will appear a question on the screen and, if helpful, a yellow window with more information. Press `Enter` to remove the yellow window and then give the specification belonging to the code, or ignore the yellow window and give the specification at once. **Do not repeat the one-character code itself!**

The right part of the window gives a short review of the options, as they are currently set. In the example of figure 17, you can see the following:

- No title line is defined.
- The input data are not yet specified. The question mark indicates that this information is mandatory. The three dots at the end of the option text mean that the option leads to a submenu: a new menu with its own options.
- As to now, no classification are requested.
- No additional results will be reported.
- No raw output files will be produced.

If the text of a menu item ends with three dots, like 'D Data definition ...' in figure 17 the option leads to a submenu, i.e. a new menu with new options. To return from such a submenu you must simply press `Enter` without any text. If the screen shows a yellow help-window, you will have to type two `Enter`s: one to remove the yellow window and one to leave the submenu.

## 6 Menu options

### 6.1 The main menu

The *main* menu (see figure 17) contains the following options:

#### 6.1.1 T Title

This option allows you to specify a header to be used in the listing files.

#### 6.1.2 D Data definition ...

If you type **D** the *main* menu will be replaced by the *data definition* menu. This menu allows you to define the input data. It will be discussed in 6.2.

#### 6.1.3 C Classification models ...

This option leads to the *classification models* menu where you can define the models you wish to use for the classification into status types. It will be treated in 6.3.

#### 6.1.4 A Additional results ...

This option leads to a new menu that allows you to order additional information in the listing file. It will be treated in 6.4.

#### 6.1.5 W Write results to raw file ...

By this option you can order output in a “raw” text file. It will be discussed in 7.2.

#### 6.1.6 O Open settings file

If you have ever saved the options for *SocStat* or if you received a settings file from someone else, you can retrieve the options from the settings file. If you type **O**, a file-selector box will appear on the screen that allows you to select the settings file. Figure 18 shows such a file selector box.

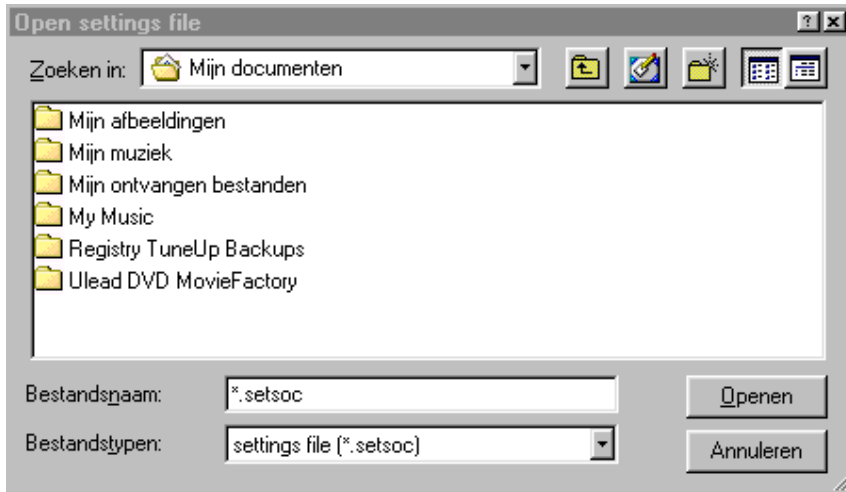


Figure 18: Retrieving information from a settings file.

By default settings files from *SocStat* have the extension `'.setsoc'`. It may be handy to save your current settings before collecting new ones. The program may remind you of that. After you have collected information from a settings file, its name will be visible on the upper right part of the main window.

### 6.1.7 S Save current settings

If you want to save the options and specifications that you have made so far, you can enter `S` `Enter`. If you do so a file-selector box will appear that allows you to specify the place and the name of the file to which the settings must be written (see figure 19).



Figure 19: Saving information to a settings file.



### 6.1.8 X eXecute SOCSTAT

If you have specified all options, you can type `x``Enter` to start the computations. The program will check if all obliged options are specified and if there are no inconsistencies. If everything is right, the computations will start. If the program has been correctly installed and runs without problems, it will, when it is finished, automatically open the last (or only) listing file it has produced. If it fails to do so, you can open it yourself by any text editor like *WordPad*, *Notepad* or *Word*. In order to have a nicely outlined text, you must select a small non-proportional font like Courier New 9.

### 6.1.9 Q Quit

The option `Q` is a kind of emergency exit. If you choose it, *SocStat* will halt without performing any calculations and without producing any output files.

## 6.2 Definition of the data

From the *main* menu, you may type the option `D``Enter` to enter the *data definitions* menu. After you have filled out the options in this menu you must press `Enter` to return to the *main* menu. If there is a yellow window on the screen, you must type two `Enter`s: one to remove the yellow window and one to return to the main menu.

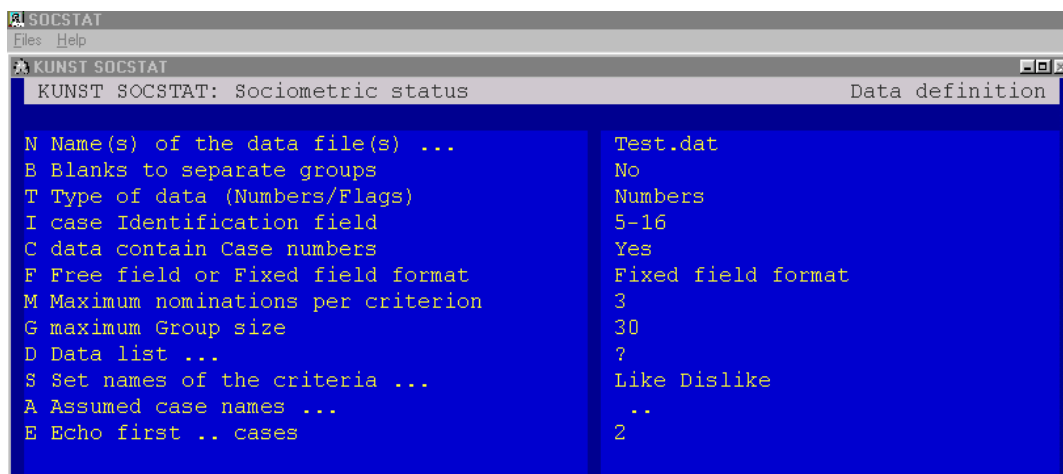


Figure 20: The data definition menu.

In the *data definition* menu (see figure 20) the following options can be chosen:

#### 6.2.1 N Name(s) of the data file(s) ...

If you choose this option, a file-selector box will appear that enables you to select one or more data files. Figure 21 contains an example.

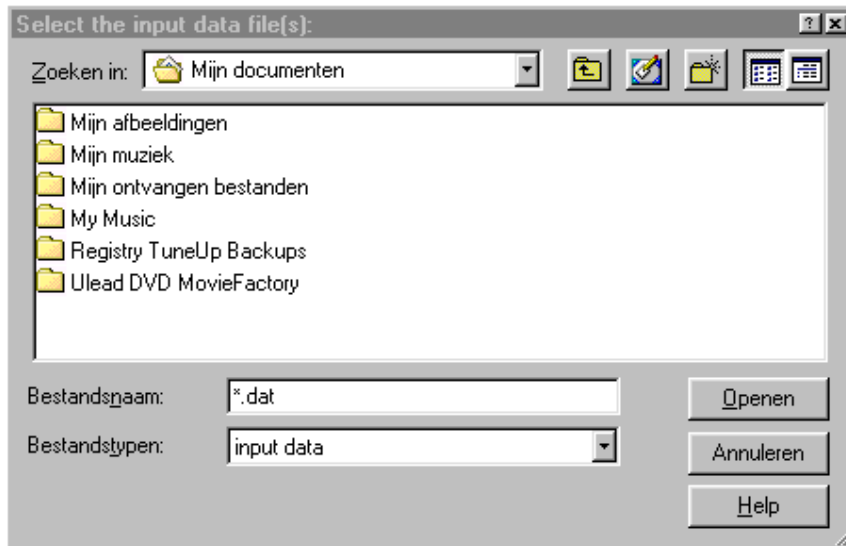


Figure 21: Choosing one or more input data files.

Each of the files may contain data from one or more groups. If for example the data are collected from several schools and within each school from several classes, you may prepare one file for each school and one group for each class. The program will then produce a separate listing file for each school and within each listing file separate blocks for the classes.

Note however that all the groups will be analyzed according to the same options and must have the same data organization.

### 6.2.2 B Blanks to separate groups

This option is needed if a file may contain more than one group. It is used to mark the boundaries between the groups. If it is active, the program will recognize the boundaries between tests by an empty line, or, to be more precise, a data row with spaces in a specific field of its first line. The given number is the starting position of the field; the length of the field is always 4. If you are using fixed format input with more than one line per profile, the empty profile must also contain more lines. The specification of the option comes down to giving the starting position, usually: `b1`. You can cancel the option by typing `b0`.

If a case identification field is defined, the content of that field may be used to specify some information, for instance about the next group.

### 6.2.3 T Type of data (Numbers/Flags)

By this option, you define the type of data: numbers or flags. Each time you type `t` the option switched back and forth between *numbers* and *flags*. See 2.6 for an explanation.

#### 6.2.4 I case Identification field

If the data contain case identification strings (see 1.6), you can use this option to define their position in the data. If a row contains more than one line (it seldom will), this field can only occur in the first line. Specify the beginning and ending position (counting the characters in the line) with a dash in-between, for instance: `i 1-10`. To indicate that there is no case identification field, type: `i 0`. You can set this same option also from the *data list* as discussed in 5.2.9.

#### 6.2.5 C data contain Case numbers

You can use this option to specify whether the input files contain case numbers or not. The option is a switch: each time you type `c`, the it switches from yes to no or back.

#### 6.2.6 F Free field or Fixed field format

This option defines the format of the data as explained in 1.7.

If you choose a fixed format, the input files must strictly fit into the format you specify. Then the actual specification takes place by way of the option *Data list*, to be discussed in 5.2.9.

Each time you type `F` the choice between fixed and free format switches.

#### 6.2.7 M Maximum nominations per criterion

This option applies only if the data contain numbers (not flags). If so, the option is obligatory. It specifies the maximum number of nominations a case can make on one criterion. That must also be the number of values in the data for each of the criteria. If a case gives fewer nominations, missing values must fill the resulting gaps. In the example data of figures 12 and 13, the value would be 3.

#### 6.2.8 G maximum Group size

With this option you can specify the maximum group size. The maximum applies to all groups and all files to be analyzed by the program run to be performed. The value is restricted to the range 1-150. The default value is 50. If the data are flags and a fixed format is used, this value has a direct influence on the number of flag positions that each data row must contain. See 5.2.9.3 for more details.

#### 6.2.9 D Data list

If you choose this option, a new window will open with a layout that differs from the usual menu layout. Its layout depends on the chosen format type (fixed or free) and the type of data (flags or numbers).

### 6.2.9.1 Data list for case numbers with fixed format

```

KUNST SOCSTAT: Sociometric status
Data list: Columns
Case id.: columns ... -...;
Name Type Line Columns      Tot. lines per case 1
Name Type Line Columns
1 CaseNr C 1 ... - ...
2 Like1 L 1 ... - ...
3 Like2 L 1 ... - ...
4 Like3 L 1 ... - ...
5 Dislike1 D 1 ... - ...
6 Dislike2 D 1 ... - ...
7 Dislike3 D 1 ... - ...

You can use the following options:

C to define the positions of variables
I to define the position of a case-id.
T to define the total number of lines per case
L to define line numbers within a case

After this character you may add further
specifications or press [Enter] for more help

Press [Enter] to remove this help window

C = Columns; I=case Id.; T = Total lines; L = Line numbers
Give C, I, T, L or continue with column definitions:

```

Figure 22: A data list for case numbers in fixed format.

If the data consist of case numbers, you must describe 2 times the maximum number of possible nominations per case. Figure 22 shows such a data list for data in fixed format when the data consist of 3 nominations on the Like and 3 nominations on the Dislike criterion. If the data contain a case number, that number is always the first in the list. That does not mean that the case number itself must precede the nominations in the input file(s). In figure 22, the position of the variable called `CaseNr` could for example be 50-56, while the nominations were in positions 4-21.

Now the data list contains room for 7 variables. You can see from the character in the column 'Type' that the second through the fourth variable define the scores on 'Like' (L) and the variables 5 to 7 are the scores on 'Dislike' (D).

Now you can type one of the indicated characters (C, I, T or L) followed by a sequence number or a range of sequence numbers and then followed by the corresponding information. The possibilities can best be clarified by some examples. We will define the data format according to the data in figure 14.

You type: `c1 1-3`

Thereby you specify that the first column (the value of variable 1: `CaseNr`) is in positions 1 through 3 (right adjusted).

From now on, you can leave out the code `c`; it will be assumed as long as you do not select another code.

You type: `2-4 11-19`

This means that variables 2, 3 and 4 (`Like1` through `Like3`) occupy positions 11 through 19, so variable 2 is in position 11-13, variable 3 is in positions 14-16 and variable 4 in positions 17-19.

You type: L5 2

Now you have entered a different code. The code **L** indicates that you are defining line numbers. You specify that variable 5 is contained in the second line of a row. Line numbers must be in ascending order. Therefore, the program will also adjust the line numbers of variable 6 and 7 to 2. From now on, you may leave the code **L** out, until you switch to another code.

You type: C5-7 1-9

The code **C** indicates that you are going to define columns. You specify that variables 5 through 7 (**Dislike1** through **Dislike3**) are in positions 1-9 (of the second line): **Dislike1** in 1-3, **Dislike2** in 4-6 and **Dislike3** in 7-9.

You type: I5-10

With the option **I** you define where the case identification string is to be found. It will always be in the first line. Unlike when a free format is used, with a fixed format this identification field may overlap with any other information. Note that this option can also be set from the option **I** in the data menu.

You type: T2

If you don't use the option **T**, the program will assume that the number of lines for each case is equal to the line number of the last row in the data list. If there are more lines in a row (apparently with unused information), you must specify so by the option **T**. In this example the specification **T2** is correct, but superfluous.

You cannot type: F  or B

But if there would be more than 32 variables, they would not fit at once on the data list screen. Therefore, you would get the possibility to scroll **f**orward and **b**ackward with these options **F** and **B**.

If you have finished the data list, you can go back to the data menu by entering an empty line (just ). If there is a yellow window, you must enter two empty lines ( ): one to remove the yellow window and one to return to the *data* menu.

### 6.2.9.2 Data list for case numbers with free format

If the data can be read in free format, you probably do not need to specify a data list. The only situation in which you must use it is if you must specify line numbers or if the data rows end with additional lines, that have to be skipped.

Figure 23 shows an example where the data consist of 3 nominations on the Like and 3 nominations on the Dislike criterion. If the data contain a case number, that number must precede the nominations.

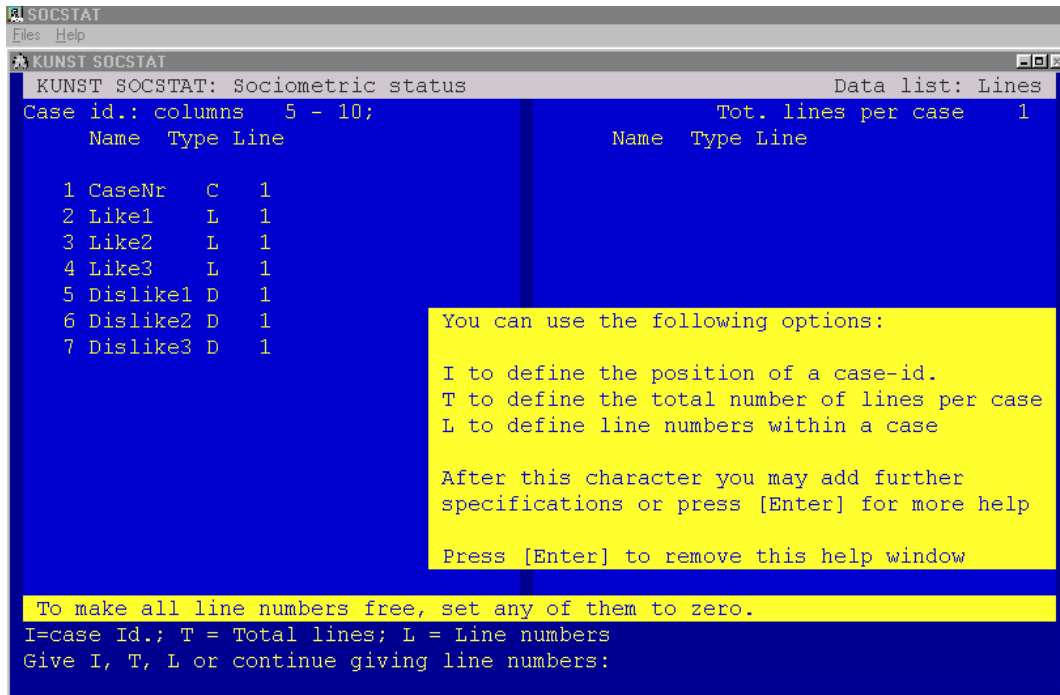


Figure 23: A data list for case numbers in free format.

In order to specify that the nominations on the Dislike criterion are to be found in the second line of a data row, you can type:

L5 2`Enter`  
 or: 5 2`Enter`

The code **L** indicates that you are defining line numbers, but at entry of the data list, that is already the default code. Line numbers must be in ascending order. Therefore, the program will also change the line numbers of variable 6 and 7 to 2.

You may type: I5-10`Enter`

With the option **I** you define where the case identification string is to be found. It will always be in the first line. If a free format is used, identification field must not overlap with any other information. Note that this option can also be set from the option **I** in the data menu.

You may type: T2`Enter`

If you don't use the option **T**, the program will assume that the number of lines for each case is equal to the line number of the last row in the data list. If there are more lines in a row (apparently with unused information), you must specify so by the option **T**. In this example the specification **T2** is superfluous.

You cannot type: F`Enter` or B`Enter`

But if there would be more than 32 variables, they would not fit at once on the data list screen. Therefore, you would get the possibility to scroll forward and backward with the options **F** and **B**.



If the data contain a case number, the position of that number will always be the first in the list. That does not mean that the case number itself must precede the nominations in the input file(s). In figure 25 for example, the position of the variable `CaseNr` could have been 74-76, after the 30 positions for the Dislike flags.

Now the data list contains room for 61 variables. You can see from the character in the column 'Type' that the second through the 31st variable define the scores on 'Like' (L) and the variables 32 through 61 the scores on 'Dislike' (D).

To scroll **f**orward and **b**ackward through the list, you can type `FEnter` or `BEnter`:

You can type one of the indicated characters (C, I, T or L) followed by a sequence number or a range of sequence numbers and then followed by the corresponding information. The possibilities can best be clarified by some examples. We will define the data format according to the data in figure 25.

You type: `c1 1-3Enter`

Thereby you specify that the first column (the score on variable 1: `CaseNr`) is in positions 1 through 3 (right adjusted).

From now on, you can leave out the code `c`; it will be assumed as long as you do not select another code.

You type: `2-31 12-41Enter`

This means that the 30 flags for the criterion Like occupy positions 12 through 41, so they take one position each.

You could type: `L32 2Enter`

to specify that the flags for the second criterion could be found in a second line below those of Like, but since they actually are on the first line, you must not do so in this example! The code `L` indicates that you are defining line numbers. You specify that variable 32 is contained in the second line of a row. Line numbers must be in ascending order. Therefore, the program will also change the line numbers of variable 33 and following to 2. From now on, you may leave the code `L` out, until you switch to another code.

You could type: `32-61 1Enter`

to correct the foregoing specification: all flags of a case are on a single line.

You type: `C32-61 43-72Enter`

This means that the 30 flags for the criterion Dislike occupy positions 43 through 72, so they also take one position each.

You type: `I5-10Enter`

With the option `I` you define where the case identification string is to be found. It will always be in the first line. Unlike when a free format is used, with a fixed format this identification field may overlap with any other information. Note that this option can also be set from the option `I` in the data menu.



You type: `T1`

If you don't use the option `T`, the program will assume that the number of lines for each case is equal to the line number of the last row in the data list. If there are more lines in a row (apparently with unused information), you must specify so by the option `T`. In this example the specification `T1` is correct, but superfluous.

You type: `F`

By this option you scroll forward through the data list.

You type: `B`

By this option you scroll back through the data list.

If you have finished the data list, you can go back to the data menu by entering an empty line (just ). If there is a yellow window on the screen, you must enter two empty lines ( ): one to remove the yellow window and one to return to the *data* menu.

#### 6.2.9.4 Flags with free format

If flag data can be read in free format, no data list is needed. For each group the program will count the number of values in the first line and derive the group size from it. The first half of the numbers (after a possible group number) will be interpreted as flags for the first criterion, the last half as flags for the second.

#### 6.2.10 S Set names of the criteria ...

In the foregoing text, we called the two criteria Like and Dislike. However, you can change the names to be used on the screen and in the listing file. If you type `S` from the data menu, a new window appears, as shown in figure 26.

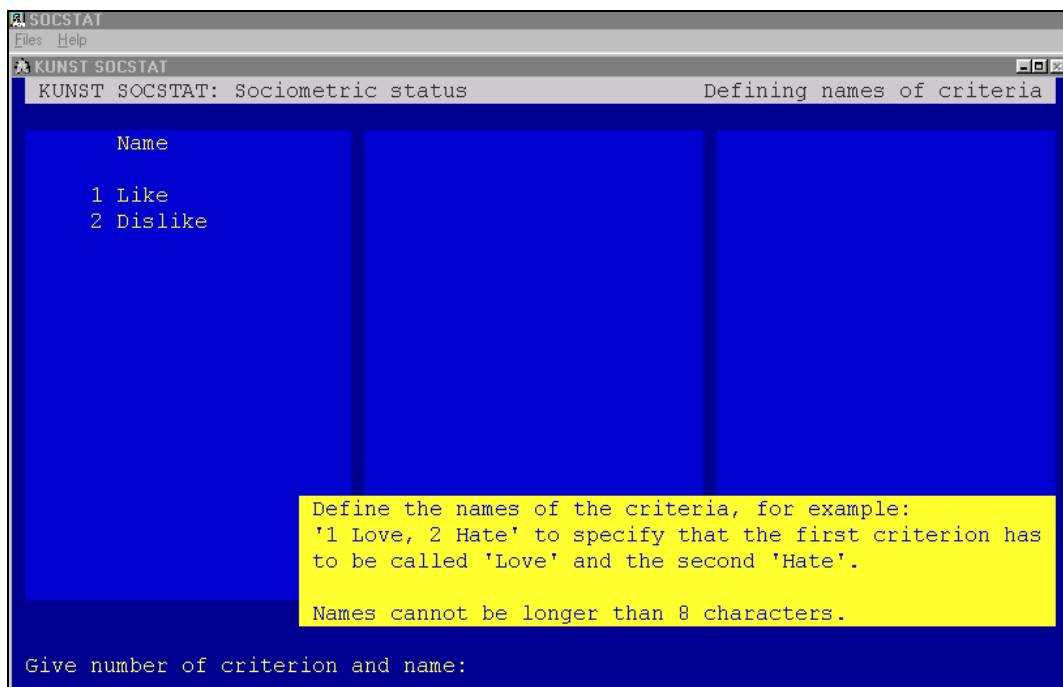


Figure 26: Changing the names of the criteria.

Now you can type a one or a two followed by a new name. For instance:

```
1 AcceptEnter
2 RejectEnter
```

or at once: 1 Accept 2 Reject`Enter`

The new names cannot be longer than eight characters.

After you have changed the names, you must type `Enter` to return to the data menu. If there is a yellow window on the screen, you must type two `Enter`s: one to remove the yellow window and one to return to the data menu.

### 6.2.11 A Assumed case names ...

As explained in 1.6 the rows in a data file may contain case information strings, to be used when reporting results. However, if the data do not contain such information, or if for any case the string is empty, the program may substitute a default string. If you type `Aenter` from the data menu, a new window appears, as shown in figure 27.

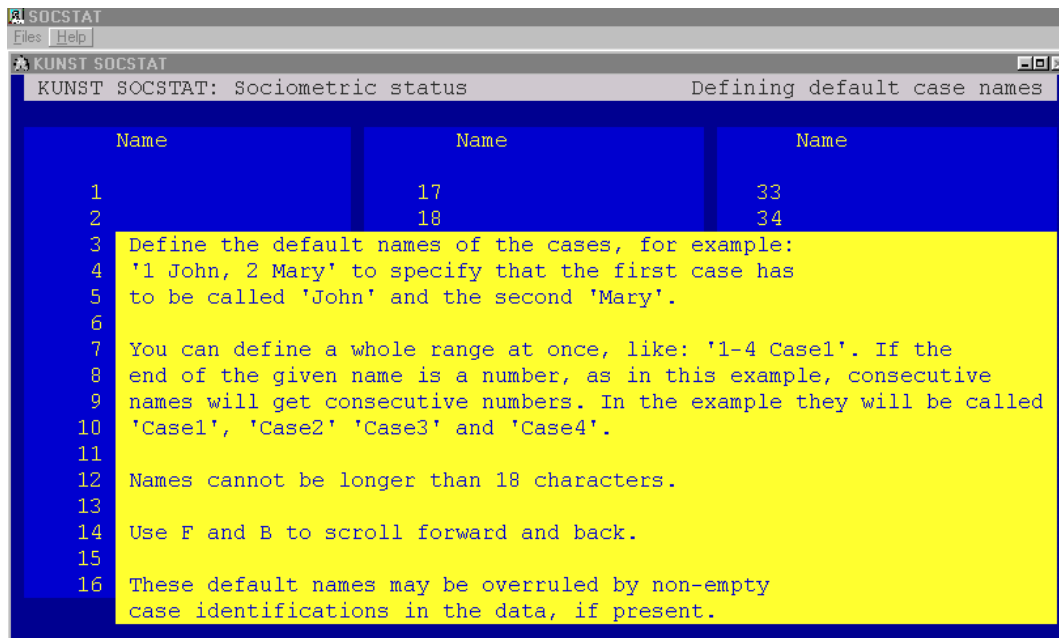


Figure 27: Defining default case names.

The length of the list is defined by the maximum group size. The names cannot be longer than 18 characters. Define the case names by typing a sequence number followed by a name (embed the name in quotes if it contains spaces). If you use a range of sequence numbers, you can define them all at once: if the given name ends with a number, consecutive names will get consecutive numbers.

```
Some examples: 1 Child_1 2 Child2 3 Child3Enter
                1-50 Child1Enter
```

In the first example you specify three names for the first three cases of a group, but in the second example you define 50 names for the first 50 cases at once: the first case will be named `Child1`, the second `Child2`, the third `Child3` and so on, ending with `Child50` as the name for the 50th case. Generally, if a name ends with a number, the number is automatically increased for subsequent names.

If you have finished giving names, you can go back to the data menu by entering an empty line (just `Enter`). If a yellow window is visible, you must enter two empty lines (`Enter Enter`): one to remove the yellow window and one to return to the *data* menu.

### 6.2.12 E Echo first .. cases

In the data menu you can type: `e ##Enter` to specify that the first ## cases of each analysis must be shown in the listing file. This may help you to check if the input specifications and the data match correctly.

## 6.3 The Classification Models

From the main menu, you may type the option `CEnter` to enter the classification menu, as shown in figure 28.

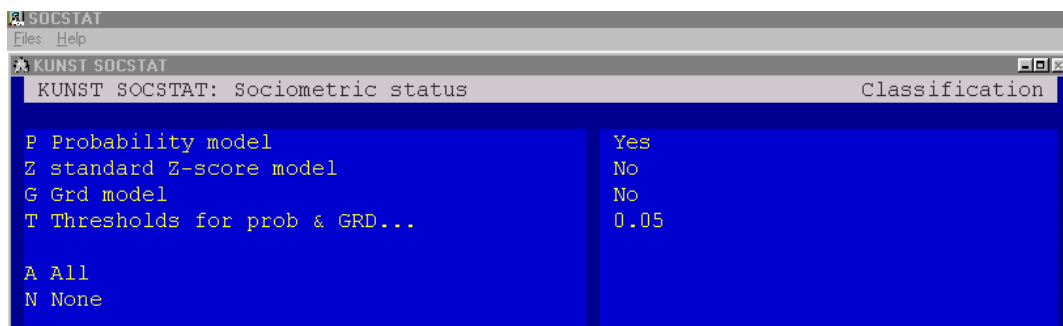


Figure 28: The classification menu.

In this menu you can choose one or more models to be used for classification into status types. If you do not choose any of the three models, no classification will take place.

The following options can be chosen:

### 6.3.1 P Probability model

By typing `pEnter,` you switch the option from *yes* to *no* or the other way around. If it is *yes*, the sociometric status according to the probability model is computed. The probability model is based on the probability scores on the two criteria as well as the probability of the impact score. The definition of an extremely high or low probability depends on the user-defined threshold called alpha. You may define up to five alpha-values. Socstat will apply the probability model for each of these alphas. See also 1.3.2.

### 6.3.2 Z standard Z-score model

By typing `zEnter,` you switch the option from *yes* to *no* or the other way around. If it is *yes*, the sociometric status according to the standard Z-score model is computed. The standard score model is based on the 4 standard scores (Z-scores) of like, dislike, preference and impact. See also 1.4.1.

### 6.3.3 G GRD model

By typing `g[Enter]`, you switch the option from `yes` to `no` or the other way around. If it is `yes`, the sociometric status according to the GRD model is computed. The model uses user defined threshold values (alphas) to define what are extremely high or low probability scores. You may define up to five alpha-values. Socstat will apply the classification procedure for each of these alphas. See also 1.4.3.

### 6.3.4 T Thresholds for prob & GRD...

If you type `s[Enter]`, the classification menu will be replaced by the alpha menu (see figure 29). This menu allows you to define up to 5 significance levels to be used in the probability model or the GRD model.

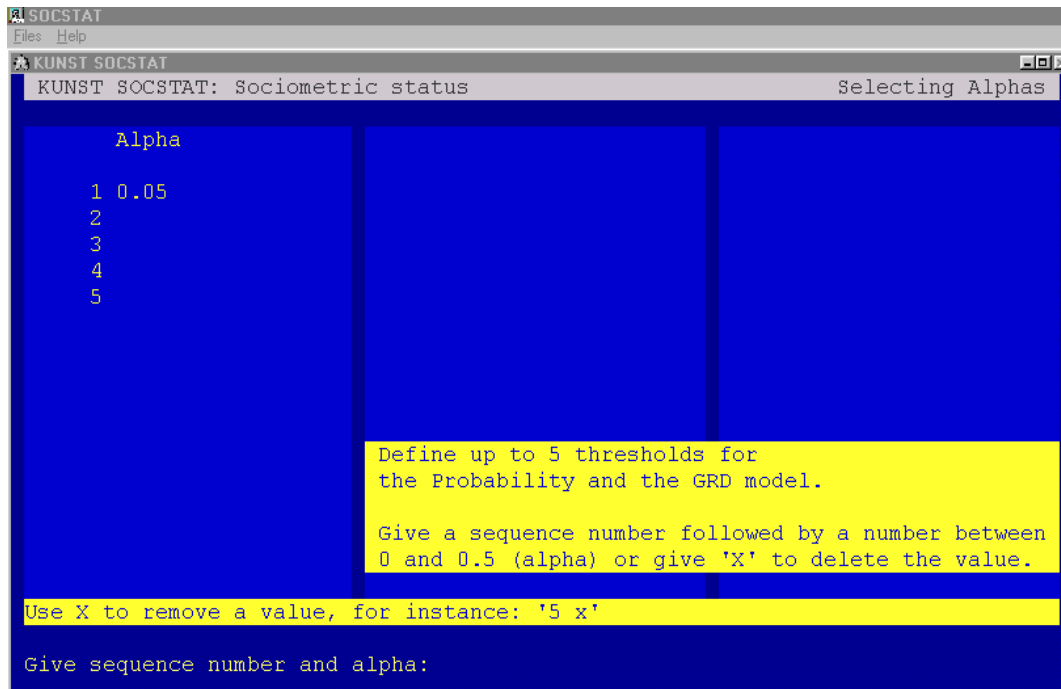


Figure 29: Choosing thresholds for the probability and GRD models.

Up to 5 threshold values (alpha) can be given, to be used by the probability model as well as by the GRD model. These values influence the assignment of cases to the classification categories. The classification procedure will be repeated for each of the significance levels.

To choose a value, give a sequence number followed by a threshold.

If you want to remove a threshold from the list, you must enter the sequence number followed by `x` instead of a value.

If you do not specify any alpha, a level of 0.05 will be used by default.

If you have finished the definition of alphas, you can go back to the classification menu by entering an empty line (just `[Enter]`). If there is a yellow window on the screen, you must enter two empty lines (`[Enter]` `[Enter]`): one to remove the yellow window and one to return to the classification menu.

### 6.3.5 A All N None

If you type a , options P, Z and G will be switched to yes.

If you type n , options P, Z and G will be switched to no all together.

## 6.4 Additional results in the listing file

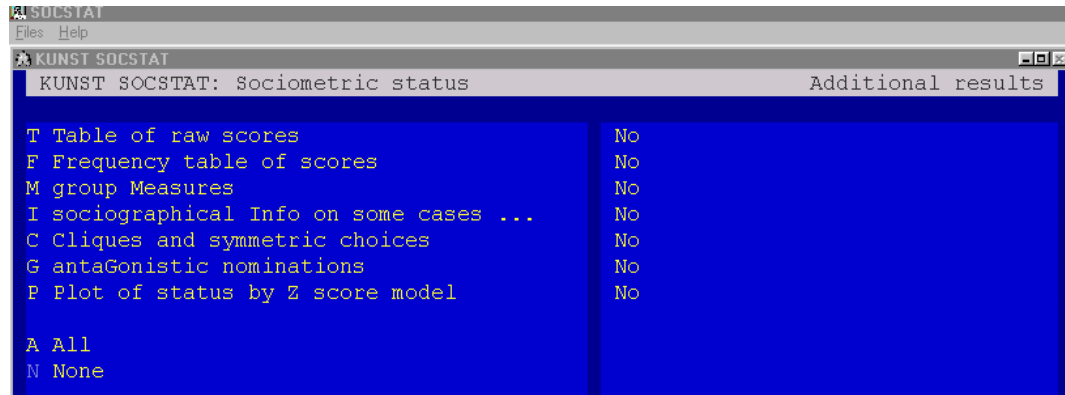


Figure 30: The additional results menu.

Entering A  in the main menu leads to the additional results menu (see figure 30). The options in this menu offer the opportunity to require additional results next to the default output. All options but one are switches with values no/yes. Change their values by typing the one-character code and pressing .

This menu contains the following options:

### 6.4.1 T Table of raw scores

If this option is set, the listing file will show the nominations on both criteria in the form of 0/1 flags with row sums and proportions added. The proportion for a case is its row sum divided by one less than the group size (self nominations are not counted).

### 6.4.2 F Frequency table of scores

If this option is set, the listing will contain the frequencies of the raw scores (the raw score is the number of times a case is nominated) with the following statistics: sum, mean, standard deviation, variance, minimum, maximum and mode.

### 6.4.3 M group Measures

If this option is set, the listing file will contain the group measures as mentioned in 2.5.

#### 6.4.4 I sociographical Info on some cases ...

Typing `i``Enter` leads to a new window where you can select cases on which you want detailed information, like whom the case is nominating and by whom it is nominated (see figure 31). The numbers in the list are sequence numbers of the cases within any group. The column names may contain default case names as described in 5.2.11. On the screen, selected cases are marked by an arrow. To select or deselect a case you just have to give its number and `Enter`. If it was selected it will be deselected now. If it was deselected, it will be selected now. You can also use a range of items. What will happen depends on the first item in the range. All items take the switched value of the first item. For example `3-5``Enter`.

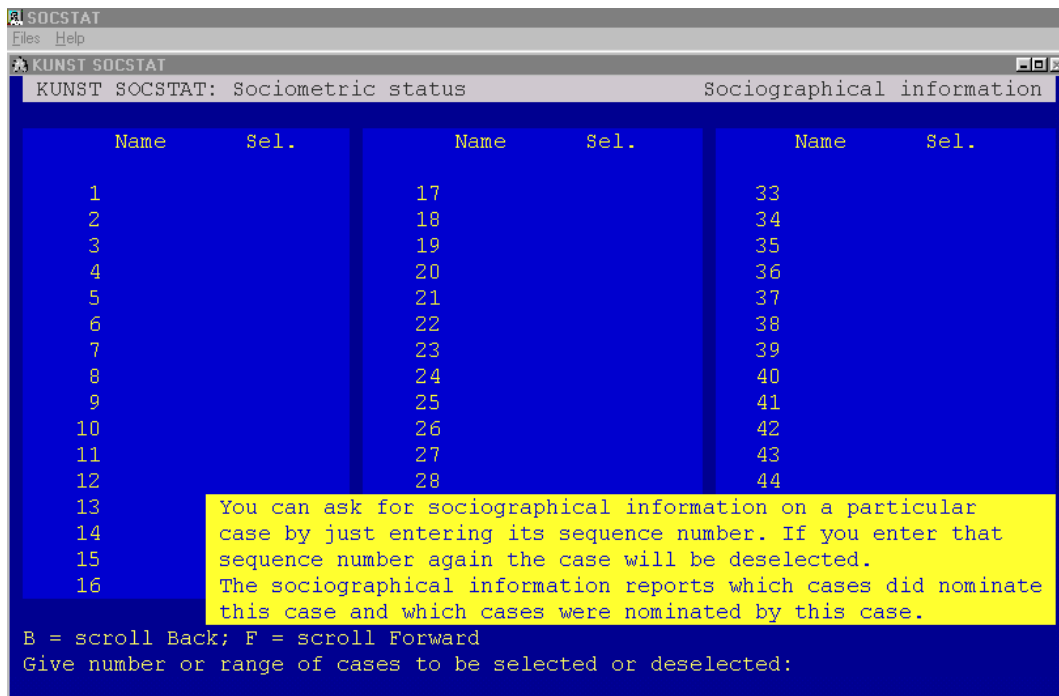


Figure 31: Asking additional sociographical information for some cases.

To go back to the additional results menu just press `Enter` once or twice.

#### 6.4.5 C Cliques and symmetric choices

If this option is set, the listing file will contain a list of cliques for each criterion. A clique is a group of cases all nominating each other on the criterion. There are however three restrictions to this analysis:

- The analysis is restricted to the first 60 cases of each group.
- The analysis may fail to find all clusters due to stack overflow. If so, it will detect and report this failure.
- The analysis will never report more than 100 cliques.

If for any of these reasons not all clusters are reported, the program will show an overview of all symmetric choices. A symmetric choice means that two cases nominate each other on the criterion.

**6.4.6 G antaGonistic nominations**

If this option is set, the listing file will contain a list of antagonistic nominations. An antagonistic nomination means that A nominates B on the first criterion, but B nominates A on the second. For example, A *likes* B, but B *dislikes* A. Of course, the adjective 'antagonistic' is only correct when the two criteria have opposite meanings.

**6.4.7 P Plot of status by Z score model**

If this option is set, the listing file will contain a graphical representation of the impact z-scores versus the preference z-scores. A picture in higher resolution will be stored in a separate bitmap file.

**6.4.8 A All  
N None**

If you type a , all options will be switched to *yes*.

If you type n , all options will be switched to *no*.

## 7 Results

SocStat will produce one or more output files. We will discuss them now.

### 7.1 Results in the listing file

For each input file, there will be a listing file, containing the main results for all groups in the file. Its precise content depends on the specifications by the user and the input data. The lines in a listing have a length of 80 characters or less. View (and print) these files with a small non-proportional font like Courier New 9.

We will use the data of figure 13 to illustrate the results.

The listing will start with a text similar to the one in figure 32.

```
View (and print) this file with a small non-proportional font like Courier 9.
Lines contain 80 characters or less.
-----
* * * * * 05-12-2005 13:35:11
      SOCSTAT: KUNST program for Sociometric Status
              Interactive version 1.20, November 2005
* * * * *
The operation and the accuracy of the program are not guaranteed.
```

Figure 32: Start of the listing file.

After this introductory text follows an overview of the options as they are chosen by the user. Figure 33 shows an example.

```
Example from figure 13
=====
1. Summary of user specifications and input data.
=====
The data consist of numbers of nominated cases.
They will be read from file:
  C:\Kunst2005\SocStat\SocStatTest\Figuur13.dat
The maximum group size is 10.
From the input all cases will be echoed.
Classification will take place according to:
  - the standard score model
  - the probability model
  - the GRD model
Thresholds alpha with probability model and/or GRD model are:
0.05
In addition to the standard results this listing will contain:
  - a table of the raw scores
  - frequency tables of the scores
```



- group measures
- cliques  
(Note that cliques are only meaningful with a free number of possible nominations per case.)
- a list of antagonistic nominations
- a plot of the status according to the Z-score model
- Sociographical information on cases with the following sequence numbers:  
1        3

The following information will be stored in a separate file with the name:  
C:\Kunst2005\SocStat\SocStatTest\Figuur63.out

	in columns
- Group number	1- 5
- Case identification	6- 23
- Case number	24- 33
- Like score	34- 36
- Dislike score	37- 39
- Raw impact score	40- 42
- Raw preference score	43- 45
- Probability of Like	46- 52
- Probability of Dislike	53- 59
- Probability of impact	60- 66
- Probability of preference	67- 73
- Z-score of Like	74- 80
- Z-score of Dislike	81- 87
- Z-score of impact	88- 94
- Z-score of preference	95-101
- Status type according to the Z-score model	102-103
- Status type based on probability model with alpha = 0.050	104-105
- Missing value (9)	106-107
- Missing value (9)	108-109
- Missing value (9)	110-111
- Missing value (9)	112-113
- Status type based on GRD model with alpha = 0.050	114-115
- Missing value (9)	116-117
- Missing value (9)	118-119
- Missing value (9)	120-121
- Missing value (9)	122-123
- Number of symmetric nominations on Like criterion	124-126
- Number of symmetric nominations on Dislike criterion	127-129

Missing elements are recorded as 9 or 9.999.

To read this file in SPSS, you could use the following syntax:

```
DATA LIST FILE='C:\Kunst2005\SocStat\SocStatTest\Figuur63.out' FIXED
  / GroupNum(F5.0) Ident(A18) CaseNum(F10.0)
  Like Dislike Impact Pref(4F3.0)
  ProbLike ProbDisl ProbImp ProbPref(4F7.5)
  ZLike ZDislike ZImpact ZPref(4F7.3) ZStatus(F2.0)
  PStat1 TO PStat5(5F2.0) GRDStat1 TO GRDStat5(5F2.0)
  NSymLike NSymDisl(2F3.0).
MISSING VALUES ZLike TO ZPref (9.999),ZStatus TO GRDStat5 (9).
VAR LABELS Like      'Raw score Like'
             / DisLike 'Raw score Dislike'
             / ProbLike 'Probability Like'
             / ProbDisl 'Probability Dislike'
             / ProbImp  'Probability of Impact'
             / ProbPref 'Probability of Preference'
             / ZLike    'Z-scores of Like'
             / ZDislike 'Z-scores of Dislike'
             / ZImpact  'Z-scores of Impact'
             / ZPref    'Z-scores of Preference'
             / ZStatus  'Status type according to Z-score model'
             / PStat1   'Status according to prob. model, alpha = TRIM(Buf(1))'
             / PStat2   'Status according to prob. model, alpha = TRIM(Buf(2))'
             / PStat3   'Status according to prob. model, alpha = TRIM(Buf(3))'
             / PStat4   'Status according to prob. model, alpha = TRIM(Buf(4))'
             / PStat5   'Status according to prob. model, alpha = TRIM(Buf(1))'
```

```

/ GRDStat1 'Status according to GRD model, alpha = TRIM(Buf(1))'
/ GRDStat2 'Status according to GRD model, alpha = TRIM(Buf(2))'
/ GRDStat3 'Status according to GRD model, alpha = TRIM(Buf(3))'
/ GRDStat4 'Status according to GRD model, alpha = TRIM(Buf(4))'
/ GRDStat5 'Status according to GRD model, alpha = TRIM(Buf(5))'
/ NSymLike 'Number of symmetric nominations on Like'
/ NSymDisl 'Number of symmetric nominations on Dislike'.
VALUE LABELS ZStatus TO GRDStat5
      1 'Popular' 2 'Average' 3 'Neglected'
      4 'Controversial' 5 'Rejected'.
EXECUTE.

```

Figure 33: Overview of the chosen options.

Now follows a description of the input data, similar to the text in figure 34.

```

The input data:
-----

The case identification will be taken from columns 5 to 9 of the first line.
The first number in a row will be interpreted as a case number.

The following values will be read for each case:

Seq.num.   Name                               Line
-----
      1     CaseNr                             1

The nominations on the criterion called Like:

Seq.num.   Name                               Line
-----
      2     Like1                               1
      3     Like2                               1
      4     Like3                               1

The nominations on the criterion called Dislike:

Seq.num.   Name                               Line
-----
      5     Dislike1                             1
      6     Dislike2                             1
      7     Dislike3                             1

```

Figure 34: Description of the data.

Next comes a report of the data as they are found in the input file. The extensiveness of the report depends on the option Echo in the data menu. Figure 35 gives an example.

```

=====
2. Processing the data.
=====

Reading the data:
-----

The cases:

Case 1 (John), with case number 11:
Like      12 14 0
Dislike   13 15 0

```

```

Case 2 (David), with case number 12:
  Like      13 14 15
  Dislike   11  0  0

Case 3 (Mary), with case number 13:
  Like      11  0  0
  Dislike   12 14 15

Case 4 (Bill), with case number 14:
  Like      12 15  0
  Dislike    0  0  0

Case 5 (Tony), with case number 15:
  Like      13  0  0
  Dislike   14  0  0

The group size is 5.

The following values are treated as missing nominations:
0

```

Figure 35: Report of the input data.

After the report of the input data follow several blocks of information on the first criterion, starting with a table of raw scores (if the corresponding option was chosen). Figure 36 shows an example.

```

=====
2.1 Information on the criterion Like.
=====

Table of raw scores:
-----

Legend:  Sum = Number of nominations
         Prop = Sum divided by maximum

Cells:   0 = Not nominated
         1 = Nominated

Like:
Nominating cases      Sum Prop      Nominated cases
                               12 14
                               11 13 15

11 John      2 0.5000   0 1 0 1 0
12 David     3 0.7500   0 0 1 1 1
13 Mary      1 0.2500   1 0 0 0 0
14 Bill      2 0.5000   0 1 0 0 1
15 Tony      1 0.2500   0 0 1 0 0

```

Figure 36: The table of raw scores on the first criterion.

This table (see figure 36) shows the basic matrix of nominations. The column with the header `Sum` gives the number of nominations given by each row case. The column under the header `Prop` gives the proportion of nominations, i.e. the number of nominations divided by its maximum. Since self-nominations are not taken into account, the maximum is one less than the group size. In figure 36 `David` nominates 3 cases, which is a proportion 0.75 of the maximum (4).

Depending on the chosen options, the table of raw scores may be followed by a frequency table of the scores, as in figure 37.

The frequencies of scores on the criterion Like:

Legend: N = Like score = number of times nominated  
 F = Number of cases with score N  
 P = Overall probability to obtain score N  
 PP = Cumulative overall probability of score N

	N	F	P	PP
	0	0	0.07500	0.03750
	1	1	0.30625	0.22813
	2	4	0.39375	0.57813
	3	0	0.19375	0.87188
	4	0	0.03125	0.98438
Mean score:	1.80000			
Stand. Dev.:	0.44721			
Variance:	0.20000			
Minimum:	1			
Maximum:	2			
Mode:	2			

Figure 37: The frequency table of the raw scores on the first criterion.

A case's raw score is the number of times it has been nominated. In the example of figure 37 it can be seen from the columns with headers *N* and *F*, that there are no cases that are not nominated at all, there is one case that is nominated once and there are 4 cases that are nominated twice.

The column with header *P* gives the probability of the possible raw scores, derived from the generalized binomial distribution. The column with header *PP* shows the overall probability scores, i.e. the sum of the probabilities of all lower scores plus half the probability of the corresponding score. In figure 37 for example, the overall probability score corresponding to the raw score 2 is  $0.07500 + 0.30625 + 0.39375 / 2 = 0.57813$ . A detailed description of the logic behind these probability scores is given in 1.3

The listing file will always show the standardized scores. Figure 38 contains an example.

The standardized scores on the criterion Like:

Legend: N = Number of times nominated  
 Exp = Expected value for Like  
 It is the sum of all Prop minus the case's own Prop  
 (Prop is the total number of nominations made by the case divided by the maximum)  
 Sym = Number of symmetric nominations

Case	N	Exp	Z-score	Prob	Sym
11 John	1	1.75000	-1.78885	0.23438	0
12 David	2	1.50000	0.44721	0.68750	1
13 Mary	2	2.00000	0.44721	0.50000	0
14 Bill	2	1.75000	0.44721	0.60156	1
15 Tony	2	2.00000	0.44721	0.50000	0

Figure 38: The standardized scores on the first criterion.

The column with header `N` gives for each case its raw score, that is the number of nominations it has received. The column with header `Exp` shows the expected number of nominations based on the proportions of nominations (as shown in figure 36) given by the other cases. In figure 38, for instance, one can see that John receives 1 nomination, whereas the expected value is 1.75. The numbers under the header `Z-score` show the standardized scores, i.e. the raw scores diminished by their mean and divided by their standard deviation. The column with header `Prob` shows the probability scores based on the generalized binomial distribution. For a detailed explanation see 1.3. Under the header `Sym` the table gives the number of symmetric nominations, that is the number of cases that not only nominate this case, but also are nominated by it.

Depending on the chosen options, the listing may continue with an overview of cliques and symmetric nominations. A clique is a group of cases that nominate each other and are nominated by each other. The list is restricted to the first 100 cliques found. In very rare cases, the algorithm may fail to find all cliques. If more than 100 cliques are found or if the algorithm fails the program will compensate for its deficiency by giving a list of all symmetric nominations. For our running example, the output will be as in figure 39.

```

Symmetric choices on the criterion Like:
-----

The following cliques are found:

      1:          12  David
           14  Bill
  
```

Figure 39: A list of cliques.

The listing will always contain the values of the four group measures as described in 1.5. Figure 40 gives an example of the report in the listing file.

```

Some group measures on the criterion Like:
-----

Group expansiveness      =  0.4500  number of nominations in the group
                           divided by the total number of off-
                           diagonal cells in the matrix
Group cohesion           =  0.1000  number of symmetric nominations divided
                           by the maximum possible number of pairs
Group integration index  =  1.0000  1/(x+1) where x is the number of cases
                           not nominated at all
Group cohesion index    =  0.2857  number of symmetric nominations divided
                           by the number of unidirectional
                           nominations
  
```

Figure 40: A list of group measures.

Depending on the chosen options the listing may contain extended sociographical information on one or more specific cases. Note that the selection of these cases is based on their *sequence* numbers, even if the data contain case numbers. If more than one group is analyzed the information will be given for all cases with the specified sequence numbers within their group. The information consists of a list of all cases nominating the specified case and of all cases nominated by it. Figure 41 shows an example of the output.

```

Sociographical information on the criterion Like:
-----
Case 11 John:
           Nominates   Is nominated by
12 David      X           .
13 Mary       .           X
14 Bill       X           .
15 Tony       .           .

Case 13 Mary:
           Nominates   Is nominated by
11 John       X           .
12 David      .           X
14 Bill       .           .
15 Tony       .           X

```

Figure 41: Detailed sociographical information on specific cases.

After the results for the first criterion, the listing file gives similar results for the second criterion (again depending on the chosen options):

- A table of raw scores
- A frequency table of the scores
- The standardized scores
- Cliques and symmetric nominations
- The four group measures
- Sociographical information on specific cases

Since these results are similar to those already reported above, we will not describe them again.

The listing will always contain a list of antagonistic pairs. A pair a-b is called antagonistic if a nominates b on one criterion, but b nominates a on the other, for instance if John likes Bill, but Bill dislikes John. Figure 42 shows the very short report for our running example.

```

Antagonistic nominations:
-----

There are no antagonistic nominations.

```

Figure 42: Report of antagonistic nominations.

After the results for the two criteria, the listing will show the findings for the compound score *impact*. The impact matrix is the sum of the two criterion matrices, corrected for controversial nominations. It contains a 1 in each cell for which either of the two criterion matrices contains a one. Therefore, if case A nominates case B on both criteria, the impact matrix will receive a 1, not a 2, in cell (A,B). The listing will contain the following information on the impact-scores:

- A table of raw scores
- A frequency table of the scores
- The standardized scores

Finally the listing will contain information on the compound score preference. The preference matrix is the difference between the two criterion matrices: the first criterion minus the second one. The preference matrix can contain three different values: 0, -1 and +1. The information consists again of the following three parts:

- A table of raw scores
- A frequency table of the scores
- The standardized scores

After the information on the two criteria and on the scales impact and preference, the listing contains the matrix of correlations between all scores: the raw scores on the two criteria, the impact and preference scores, the four corresponding z-scores and the four corresponding probability-scores. Figure 43 shows an example, based on our example data.

```

=====
3. Correlations between the scores.
=====

```

	1	2	3	4	5	6	7
	NLike	NDislike	Impact	Pref.	ZLike	ZDislike	Z-impact
2 NDislike	0.4082						
3 Impact	0.8018	0.8729					
4 Pref.	0.4082	-0.6667	-0.2182				
5 ZLike	1.0000	0.4082	0.8018	0.4082			
6 ZDislike	0.4082	1.0000	0.8729	-0.6667	0.4082		
7 Z-impact	0.8018	0.8729	1.0000	-0.2182	0.8018	0.8729	
8 Z-pref.	0.4082	-0.6667	-0.2182	1.0000	0.4082	-0.6667	-0.2182
9 PLike	0.8877	0.2472	0.6363	0.4776	0.8877	0.2472	0.6363
10 PDislike	0.3641	0.8475	0.7494	-0.5502	0.3641	0.8475	0.7494
11 P-impact	0.8729	0.8018	0.9915	-0.0891	0.8729	0.8018	0.9915
12 P-pref.	0.5057	-0.3160	0.0634	0.7289	0.5057	-0.3160	0.0634

	8	9	10	11
	Z-pref.	PLike	PDislike	P-impact
9 PLike	0.4776			
10 PDislike	-0.5502	-0.0079		
11 P-impact	-0.0891	0.7132	0.6913	
12 P-pref.	0.7289	0.8059	-0.5964	0.1621

Figure 43: Correlations between the scores.

Finally, the listing file contains information on the sociometric statuses according to the chosen classification models. Figure 44 shows results for the standard score model.

Example from figure 13

```

=====
4. The standard score model.
=====

4.1 Sociometric status by standard score model:
=====

Legend: Status type distribution

1 Popular:      Z Pref > 1  and Z Like    > 0 and Z Dislike < 0
5 Rejected:     Z Pref < -1 and Z Like    < 0 and Z Dislike > 0
3 Neglected:   Z Imp  < -1 and Z Like    < 0 and Z Dislike < 0
4 Controversial: Z Imp  > 1  and Z Like    > 0 and Z Dislike > 0
2 Average:      all others

Case              Z-scores
                  Like  Dislike  Impact  Pref.  Status type
11 John           -1.789 -0.730 -1.434 -0.730 3 Neglected
12 David           0.447 -0.730 -0.239  1.095 1 Popular
13 Mary           0.447 -0.730 -0.239  1.095 1 Popular
14 Bill           0.447  1.095  0.956 -0.730 2 Average
15 Tony           0.447  1.095  0.956 -0.730 2 Average

Frequencies of status types:
-----
Type              Freq    %
1 Popular         2     40.0
2 Average         2     40.0
3 Neglected      1     20.0
4 Controversial  0      0.0
5 Rejected        0      0.0
Total            5 100.000

```

Figure 44: Social status according to the standard score model.

With the standard score model, the program also produces a plot of the impact and the preference scores. Figure 45 shows an example. Note that the numbers in the plot are *sequence* numbers, even if the data contain case numbers! The plot is also stored in a separate bitmap file.



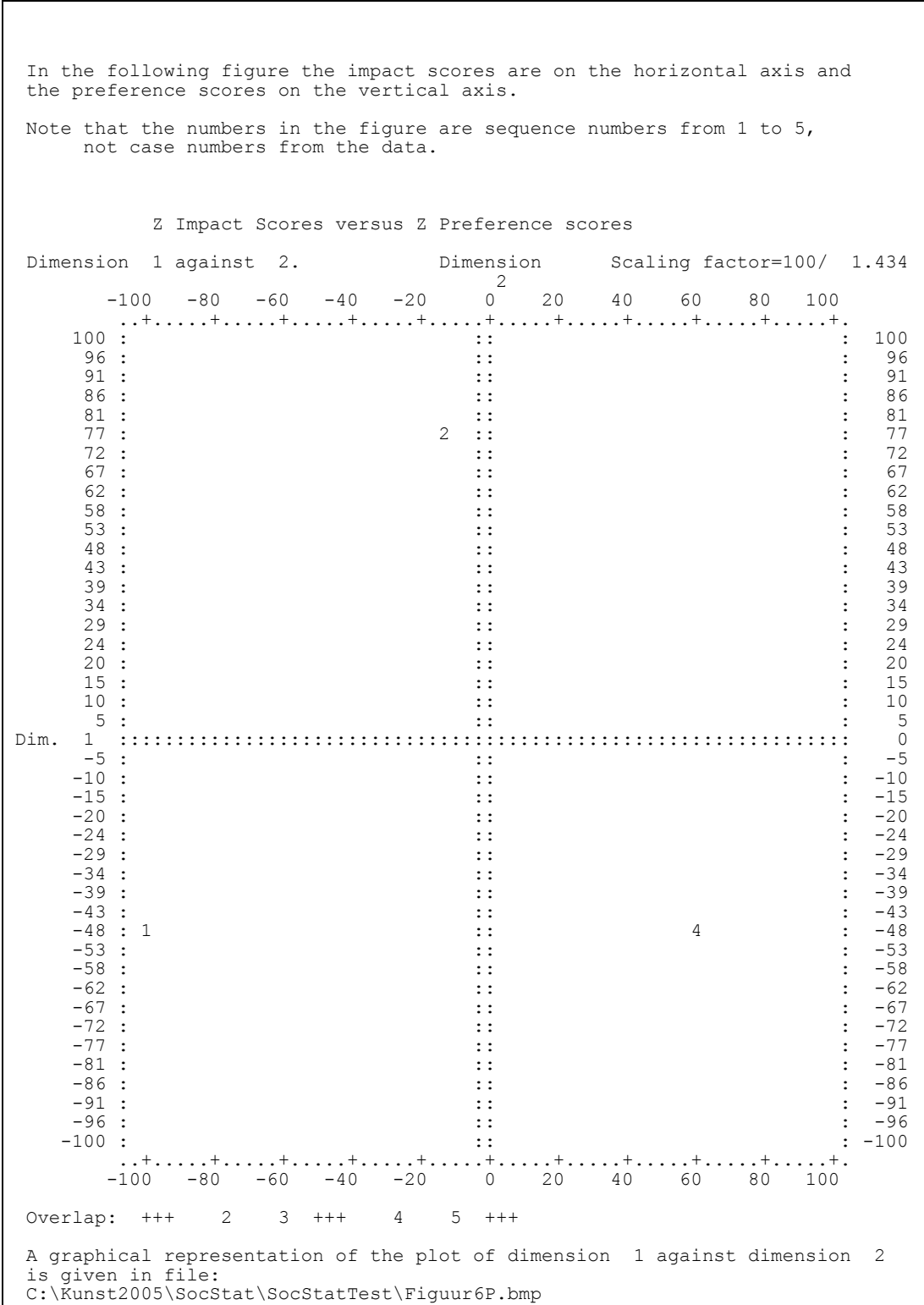


Figure 45: Plot of impact and preference scores.

Figure 46 shows the classification according to the probability model applied to our example data. If the options specify several threshold values, the results will be repeated for subsequent thresholds.

Example from figure 13

```

=====
5. The probability model.
=====

Legend: Status type distribution

1 Popular          Prob Like      > 1-alpha  and  N Dislike  < Exp
5 Rejected         Prob Dislike   > 1-alpha  and  N Like     < Exp
4 Controversial    Prob Like      > 1-alpha  and  N Dislike   > Exp
                  or Prob Dislike  > 1-alpha  and  N Like     > Exp
3 Neglected       Prob (Impact) < alpha
2 Average          all others

=====

5.1 Sociometric status by probability model (alpha = 0.05):
=====

      Case          N      Like      Dislike      Status type
      Name          N      Prob.      N      Prob.
-----
11 John            1    0.234      1    0.398    2 Average
12 David           2    0.688      1    0.297    2 Average
13 Mary            2    0.500      1    0.516    2 Average
14 Bill            2    0.602      2    0.602    2 Average
15 Tony            2    0.500      2    0.703    2 Average

Frequencies of status types:
-----
Type          Freq      %
-----
1 Popular          0      0.0
2 Average          5     100.0
3 Neglected       0      0.0
4 Controversial    0      0.0
5 Rejected         0      0.0
Total            5     100.0

```

Figure 46: Social status according to the probability model.

Figure 47 shows the classification results according to the GRD model for the example data. If the options specify several threshold values, the results will be repeated for subsequent thresholds.

Example from figure 13

6. The GRD model.

Legend: Status type distribution

1 Popular	Prob Like	> 1-alpha	and N Dislike	< Exp
	or Prob Dislike	< alpha	and N Like	> Exp
5 Rejected	Prob Dislike	> 1-alpha	and N Like	< Exp
	or Prob Like	< alpha	and N Dislike	> Exp
4 Controversial	Prob Like	> 1-alpha	and N Dislike	> Exp
	or Prob Dislike	> 1-alpha	and N Like	> Exp
3 Neglected	Prob Like	< alpha	and N Dislike	< Exp
	or Prob Dislike	< alpha	and N Like	< Exp
2 Average	all others			

6.1 Sociometric status by GRD model (alpha = 0.05):

Case	Like		Dislike		Status type
	N	Prob	N	Prob	
11 John	1	0.234	1	0.398	2 Average
12 David	2	0.688	1	0.297	2 Average
13 Mary	2	0.500	1	0.516	2 Average
14 Bill	2	0.602	2	0.602	2 Average
15 Tony	2	0.500	2	0.703	2 Average

Frequencies of status types:

Type	Freq	%
1 Popular	0	0.0
2 Average	5	100.0
3 Neglected	0	0.0
4 Controversial	0	0.0
5 Rejected	0	0.0
Total	5	100.0

Figure 47: Social status according to the GRD model.

If the GRD model is applied, the listing will also contain a summary table, as in figure 48. The horizontal axis shows the first criterion with three cutting points: the alpha-threshold, the expected value and the (1-alpha) point. The vertical axis shows the second criterion with the corresponding cutting points. These two classifications determine the status according to the GRD model. The table gives the frequencies in the cells of the resulting four by four table and, at the end of the listing, for each case the cell to which it belongs. Note that the cases in the upper left cell have the highest score on the preference scale and the cases in the lower right cell have the lowest preference scores. On the other hand, the cases in the upper right cell have the highest and the cases in the lower left cell the lowest preference scores.

Summary table of GRD classification (alpha = 0.05):

---

Legend: High = > 1-alpha  
 Low = < alpha  
 Exp = Expected value

Scale Dislike

		Scale Dislike				
		Low	Exp	High		Highest Impact
Highest Pref.						Total
Scale Like	High	Popular 0	Popular 0	Controve 0	Controve 0	0
	Exp	Popular 0	Average 1	Average 1	Controve 0	2
		Neglecte 0	Average 2	Average 1	Rejected 0	3
	Low	Neglecte 0	Neglecte 0	Rejected 0	Rejected 0	0
Lowest Impact						Lowest Pref.
Total		0	3	2	0	5

Overview of the cases in the table:

---

Case	Row	Column
11 John	3	2
12 David	2	2
13 Mary	3	2
14 Bill	2	3
15 Tony	3	3

===== Normal end of analysis =====

Figure 48: Summary table of the GRD classification.

### 7.2 Results in a raw text file

In the main menu, the option 'w Write results to raw file' offers the opportunity to store some results in a simple text file in order to analyze them by any other program. There will be one such file for each input file. It will contain one line for each case with the following information:

<b>Content</b>	<b>position</b>
Group number .....	1- 5
Case identification string (possibly blank) .....	6- 23
Case number .....	24- 33
Raw score on first criterion.....	34- 36
Raw score on second criterion.....	37- 39
Raw impact score.....	40- 42
Raw preference score.....	43- 45
Probability score on first criterion .....	46- 52
Probability score on second criterion .....	53- 59
Probability score on impact .....	60- 66
Probability score on preference .....	67- 73
Z-score on first criterion* .....	74- 80
Z-score on second criterion* .....	81- 87
Z-score on impact*.....	88- 94
Z-score on preference*.....	95-101
Status type according to the Z-score model.....	102-103
Status type based on probability model with first alpha** .....	104-105
Status type based on probability model with second alpha** .....	106-107
Status type based on probability model with third alpha** .....	108-109
Status type based on probability model with fourth alpha** .....	110-111
Status type based on probability model with fifth alpha** .....	112-113
Status type based on GRD model with first alpha** .....	114-115
Status type based on GRD model with second alpha** .....	116-117
Status type based on GRD model with third alpha** .....	118-119
Status type based on GRD model with fourth alpha** .....	120-121
Status type based on GRD model with fifth alpha** .....	122-123
Number of symmetric nominations on first criterion .....	124-126
Number of symmetric nominations on second criterion.....	127-129

The listing file will show this list of columns positions and a complete syntax to create an SPSS data file with the corresponding information. See figure 33 for an example.

In order to create the SPSS-file you must do the following:\*\*\*

- 1 Select the syntax in the listing file and copy it to the clipboard.
- 2 Start SPSS, and open a syntax window (File menu → New → Syntax).
- 3 Paste the syntax into the syntax file.
- 4 Choose from the Run-menu the option All.

Now select from the Window-menu the Data Editor and you see the data.

The codes for the status types have the following meanings:

1 = popular, 2 = average, 3 = neglected, 4 = controversial, 5 = rejected

---

\* Possibly missing value (9.999)

\*\* Possibly missing value (9)

\*\*\* Under Windows first make sure that the decimal symbol is set to a period and the grouping symbol to a comma (Settings → Configuration → Country settings → Number).

## 8 Literature

- Borgatta, E.F., (1978) *Sociometry*. In W.H. Kruskal & J.M. Tanur (eds.). International encyclopedia of statistics, 2, 1031-1036. New York: The Free Press.
- Brink ten, P.W.M., (1985) *De gegeneraliseerde binomiale verdeling als alternatief voor de sociometrische status berekening volgens het probabiliteitsmodel*. Dept. of Developmental Psychology, University of Nijmegen, The Netherlands.
- Bukowski, W.M., & Hoza, B., (1989) *Popularity and friendship. Issues in theory, measurement and outcome*. In T.J. Berndt & G.W. Ladd (Eds.). Peer relationships in child development. New York: Wiley.
- Cillessen, A.H.N., & Brink ten, P.W.M., (1991) *Vaststelling van relaties met leeftijdgenoten*. Pedagogische Studiën, 68, 1-14.
- Coie, J.D. & Dodge, K.A., (1983) *Continuities and Changes in Children's Social Status: A Five-Year Longitudinal Study*. Merrill-Palmer Quarterly, 29, 3, 262-282.
- Coie, J.D., Dodge, K.A., & Coppotelli, H., (1982) *Dimensions and types of social status: A cross-age perspective*. Developmental Psychology, 18, 557-570.
- Moreno, J.L., *Who shall survive?: foundations of sociometry, group psychotherapy and sociodrama*, Beacon House Inc., Beacon, N.Y., 1953.
- Newcomb, A.F., & Bukowski, W.M., (1983) *Social impact and social preference as determinants of children's peer group status*. Developmental Psychology, 19, 856-867.
- Newcomb, A.F., & Bukowski, W.M., (1984) *A longitudinal study of the utility of social impact and social preference classification schemes*. Child Development, 55, 1434-1447.
- Newcomb, A.F., Bukowski, W.M., & Pattee, L., (in press), *Children's peer relations: A meta-analytic review of popular, rejected, neglected, controversial and average sociometric status*. Psychological Bulletin, in press.

## Index

acceptance .....	4
additional results .....	23; 37
ambiguous nominations .....	5
antagonistic nominations ...	6; 15; 39; 46
assumed case names .....	34
average .....	6; 12; 13; 14; 53
case .....	5
case identification	15; 16; 26; 29; 30; 32
case numbers .....	15; 16; 27; 28; 32
classification models .....	6; 23; 35; 47
cliques .....	6; 14; 15; 38; 45
cohesion .....	15
cohesion index .....	15
column positions .....	28; 29; 32
compound scores .....	4
controversial .....	6; 12; 13; 14; 53
correlations .....	47
data definitions .....	23; 25
data file .....	19
data format .....	17; 27
data list .....	27; 29
data types .....	15; 26
dislike score .....	6
dyads .....	6; 14
echoing the data .....	35; 42
executing SocStat .....	25
expansiveness .....	15
file name(s) .....	25
file types .....	19
fixed format .....	17; 28; 31
flags .....	15; 31; 33
free format .....	18; 29; 33
frequencies of scores .....	37
frequency table .....	44
generalized binomial distribution .....	7
generalized multinomial distribution ..	11
GRD model .....	13; 36; 50
group characteristics .....	14
group measures .....	37
group separator .....	26
groups .....	26
impact .....	4; 6; 7; 46
impact score .....	7
installation .....	20
integration index .....	15
like score .....	6
line number .....	29; 30; 32
listing file .....	19; 40
maximum group size .....	27
maximum nominations per case .....	27
menu options .....	23
names of criteria .....	33
neglected .....	6; 12; 13; 53
one-character code .....	22
overall probabilities .....	9
plot .....	39; 48
plot file .....	19
popular .....	6; 12; 13; 53
preference .....	4; 6; 12; 47
probability model .....	13; 35; 49
probability of dislike score .....	7
probability of impact score .....	7
probability of like score .....	7
probability of preference score .....	7
probability score .....	7; 10; 12; 44
quitting .....	25
raw dislike score .....	6
raw file .....	23
raw impact score .....	7
raw like score .....	6
raw output file .....	19
raw preference score .....	7
raw scores .....	37; 43
raw text file .....	52
rejected .....	6; 12; 13; 53
rejection .....	4
retrieving settings .....	23
returning from a submenu .....	22
running SocStat .....	21
saving settings .....	24
scrolling .....	29; 30; 32; 33
self-nominations .....	5
separating groups .....	26
settings file .....	19; 23; 24
significance level .....	13
social impact .....	7
social preference .....	7
sociographical information .....	38; 45
sociometric status .....	47
sociometric status categories .....	6
sociometry .....	4
SPSS syntax .....	53
SPSS-file, creating .....	53
standard dislike score .....	7
standard impact score .....	7
standard like score .....	7

---

standard preference score.....	7	symmetric nominations .....	45
standard score model.....	35; 47	total number of lines.....	29; 30; 33
standardized scores .....	44; 45	treshold value .....	36
status types .....	12; 53	yellow window .....	21
submenu .....	22	z-score .....	7
summary table .....	51	z-score model .....	35
symmetric choices.....	38		