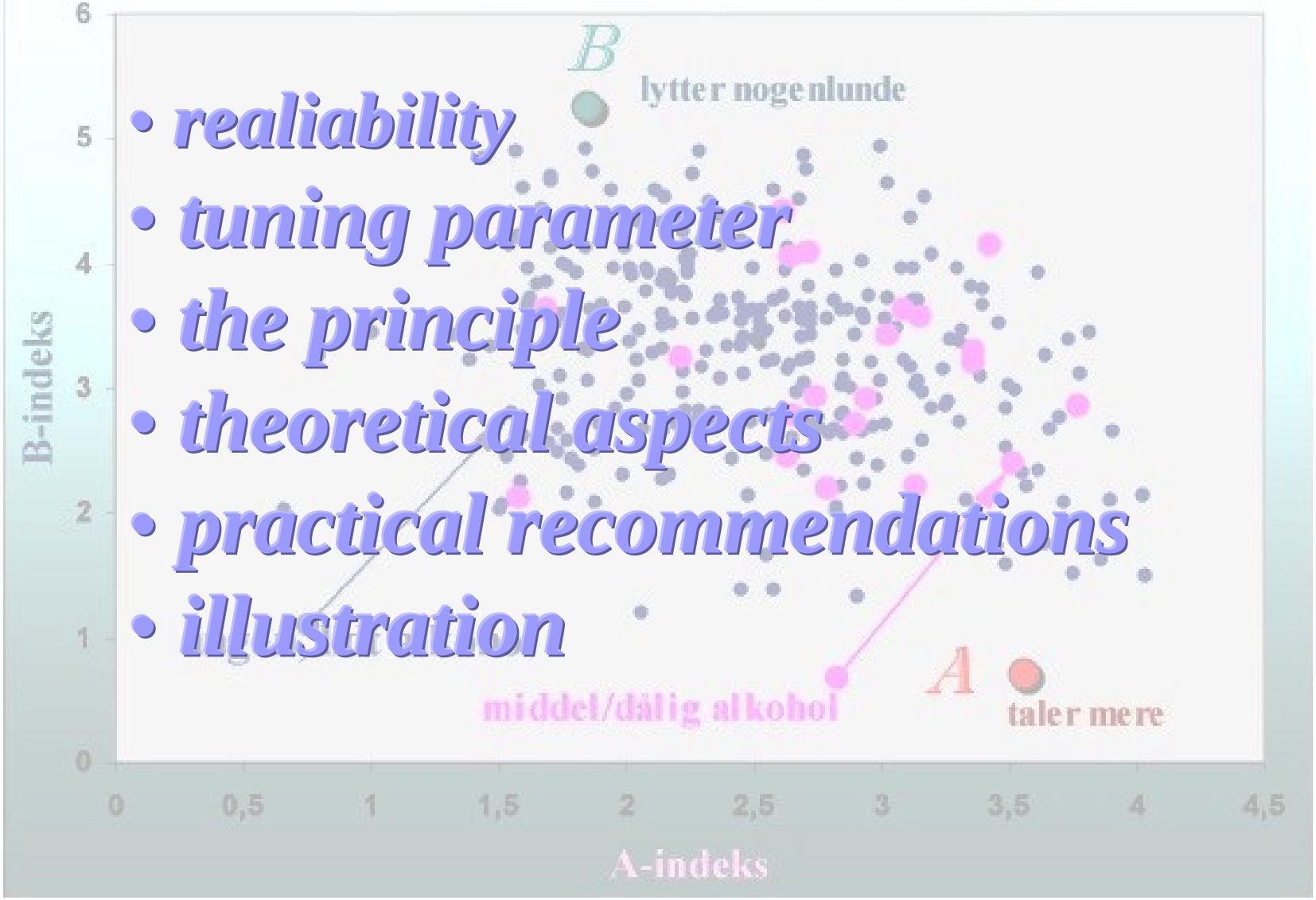
Maximum Principle for Survey Data Analysis

- *realiability*
- *tuning parameter*
- *the principle*
- *theoretical aspects*
- *practical recommendations*
- *illustration*



Food preferences investigation

Table 1	Dairy	Grain	Greenery	Fish	Meat	Total
Respond. nr.1		X	X			2
Respond. nr.2	X	X		X	X	4
Respond. nr.3			X	X		2
Respond. nr.4	X	X		X	X	4
Respond. nr.5			X	X		2
Respond nr. 6	X	X	X	X	X	5
Respond. nr.7		X	X			2
Total	3	5	5	5	3	21
%	43%	71%	71%	71%	43%	

Food preferences investigation

Table 2	Dairy	Grain	Greenery	Fish	Meat	Total
Respond. nr.2	X	X		X	X	4
Respond. nr.4	X	X		X	X	4
Respond nr. 6	X	X	X	X	X	5
Total	3	3	1	3	3	13

Table 3	Dairy	Grain	Fish	Meat	Total
Respond. nr.2	X	X	X	X	4
Respond. nr.4	X	X	X	X	4
Respond nr. 6	X	X	X	X	4
Total	3	3	3	3	13

Some theoretical aspects

For some time I'm working on a problem of sampling a set of K observations (cases) from a large data set with N >> K cases so that the selected observations are as "different as possible". In more mathematical terms, I'm interested in locating those K cases which will result in a (not necessarily Euclidean) distance matrix in which the smallest off-diagonal entry d_ij is as large as possible.

I have developed an algorithm which seems to work very well and generates sets which are either optimal or close to optimality without computing the entire distance matrix. However, I'm thinking more and more that this maybe a known problem to people who work in Cluster Analysis, MDS, or classification. I wonder if anybody on this list could point me to some references about this search problem.

Thanks, Wolfgang Hartmann

Some theoretical aspects

Suppose that respondents $N=\{1,...,i,...n\}$ participate in the survey. Let x, $x\in 2^N$, are those who expressed their preferences towards certain questions $M=\{1,...,j,...,m\}$. We lose no generality in looking at list M as at the profile – negative or positive. Let a Boolean table $W=\|a_{ij}\|_n^m$ reflect the survey result of respondents' preferences; $a_{ij}=1$ if respondent i prefers the answer j, $a_{ij}=0$ if not. Also all lists 2^M of answers $y\in 2^M$ within a profile M have been examined. Let an index $\delta_{ij}^k=0$, $i\in x, j\in y$ if $\sum_{j\in y} a_{ij}<k$, otherwise $\delta_{ij}^k=1$, e.g. $\sum_{j\in y} a_{ij}\geq k$, where k is our tuning parameter. We can calculate an index $F_k(H)$ using a subtable H on crossing entries of the rows x and columns y in the original table W. Let the number of 1-entries $\delta_{ij}^k\cdot a_{ij}=1$ in each column within the range y determines the index $F_k(H)$ by further selection of a column with the least – the minimum number $F_k(H)$ from the list y.

In order to find a reliable component ***K***, it seems tautological that following our maximum principle, we have to solve maximization problem:

$$K=argmax_{(x,y)}F_k(H)$$

Positive / Negative

Indifferent

0°

100°

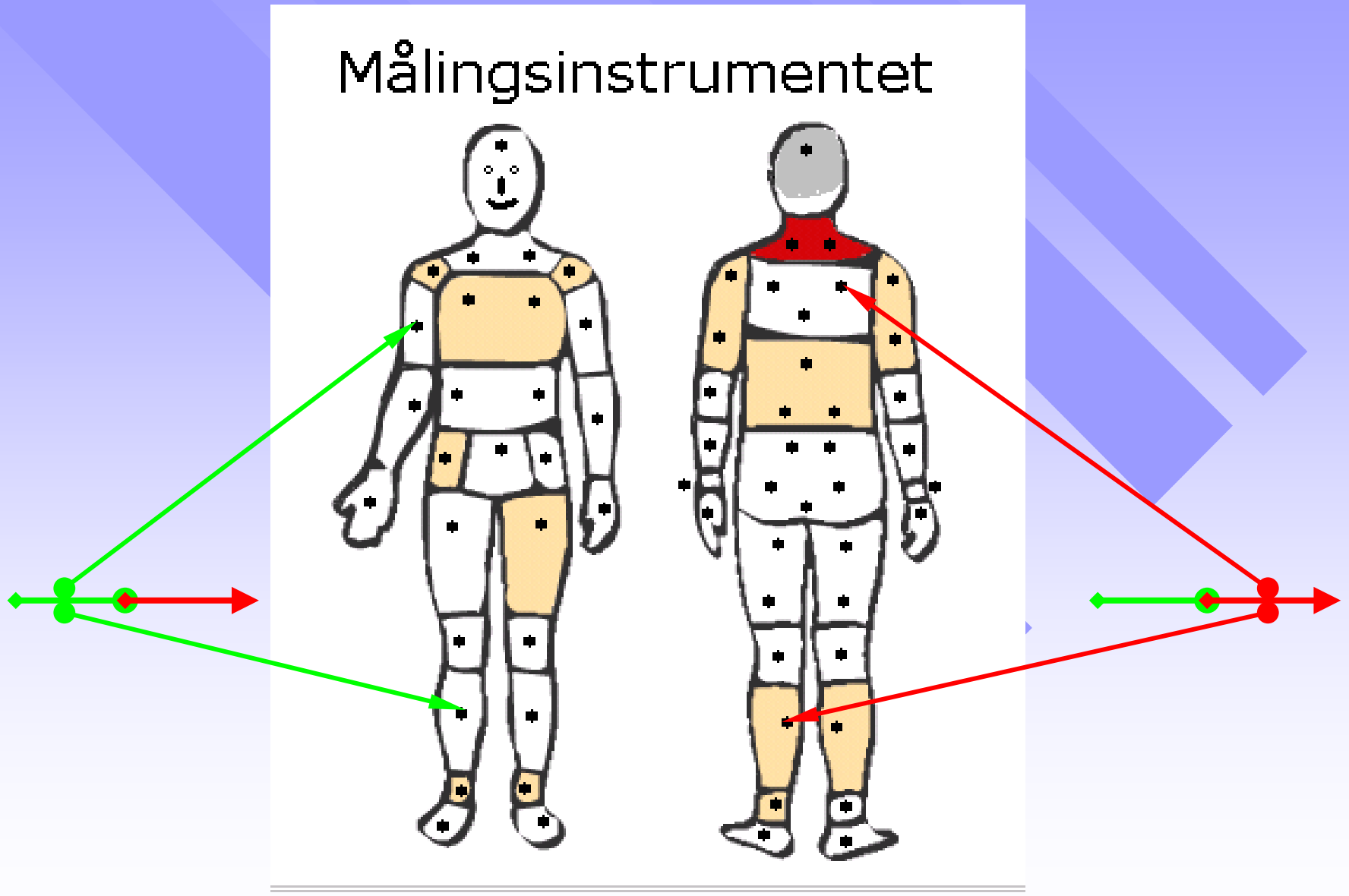
Excellent

Very good

Bad

Very bad

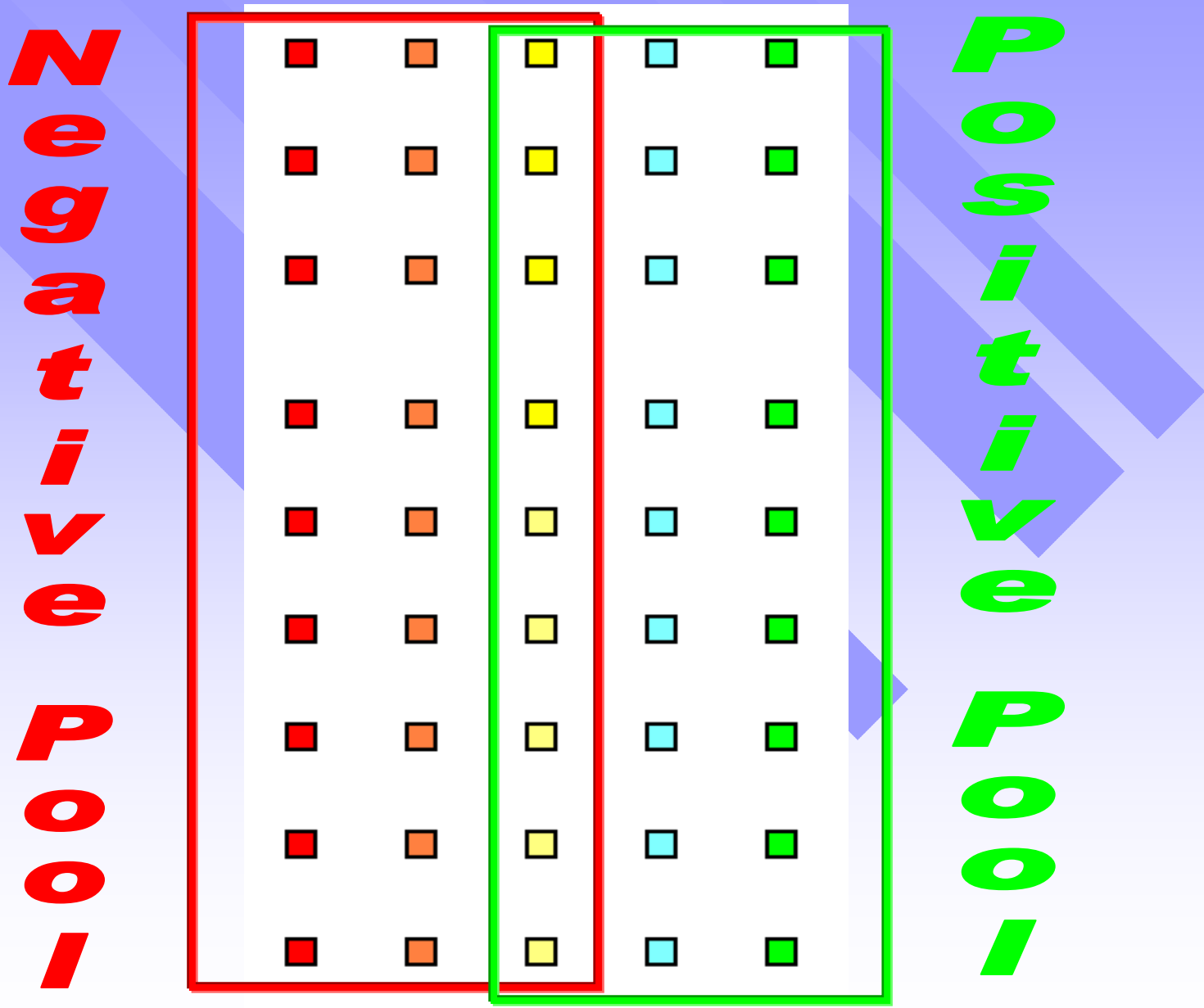
Positive / Negative



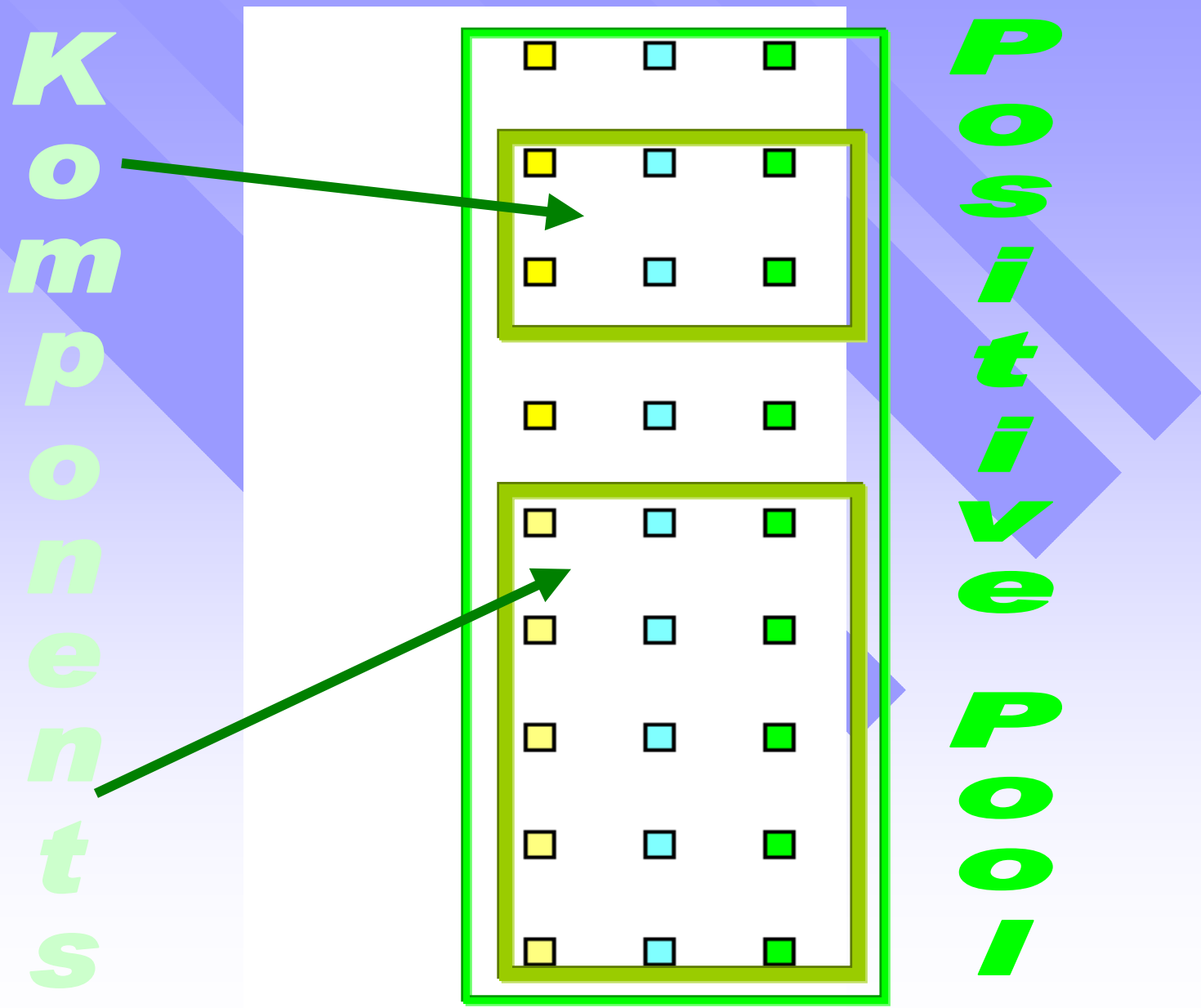
Negative/Positive Scale in the Questionnaire

9.2	Alkohol		Regler for alkohol på arbejdspladsen og i job
9.3	Rygning		Regler for rygning på arbejdspladsen
9.4	Stress		Vejledning i stressforebyggelse og –håndtering
9.5	Rygskole		Kurser i forebyggelse af ryglidelser
9.6	Kost - og slankehold		Vægttabsaktiviteter og kostvejledning
9.7	Motionsplaner og –aktiviteter		Motion og sport sammen med kolleger
9.8	Motionsrum på virksomheden		Fitness på arbejdspladsen
9.9	Adgang til badefaciliteter		Mulighed for bad på virksomheden
9.10	Cykelparkering på arbejdsplads		Mulighed for at parkere cykel på arbejdspladsen

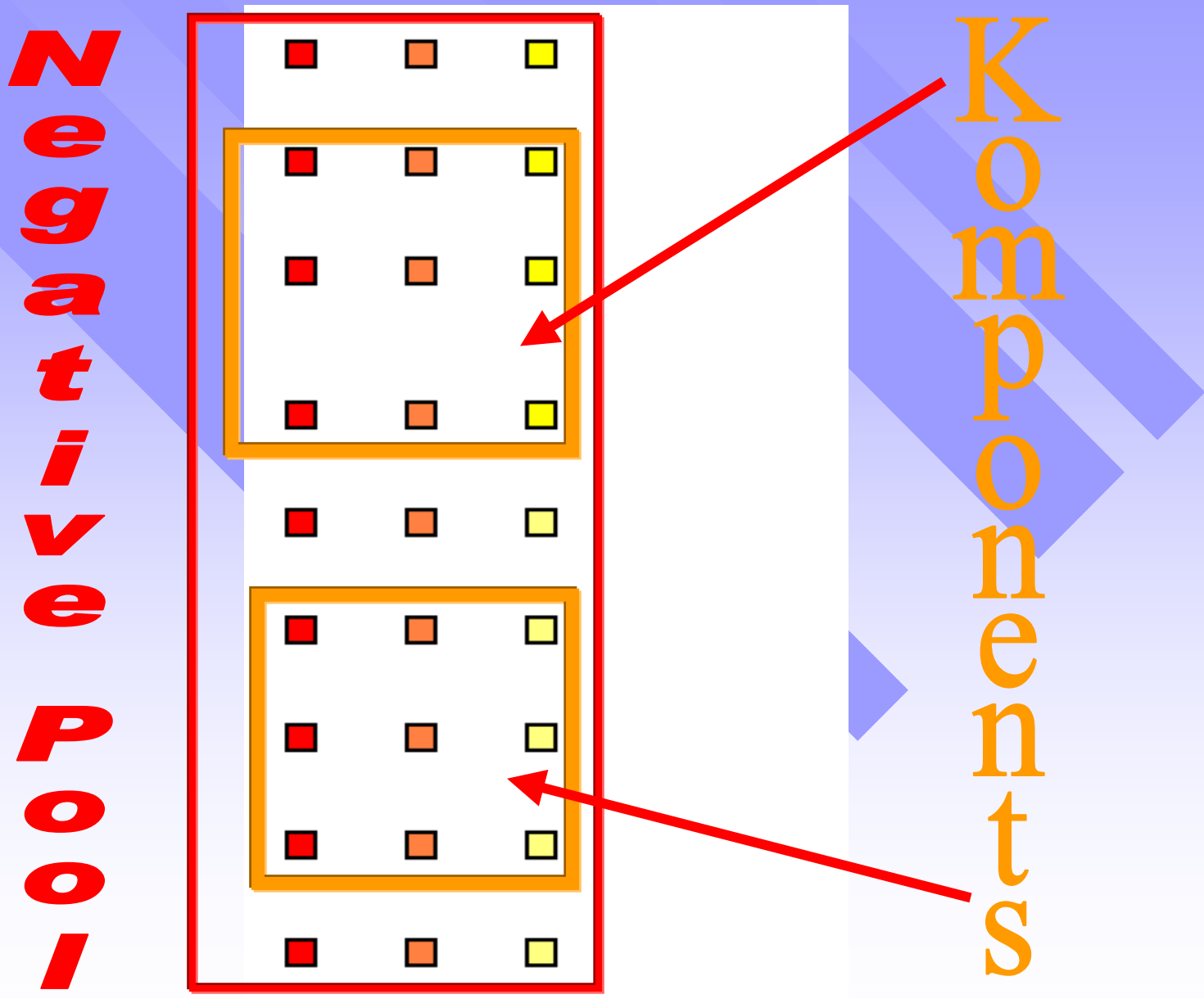
Negative/Positive Scale in the Questionnaire



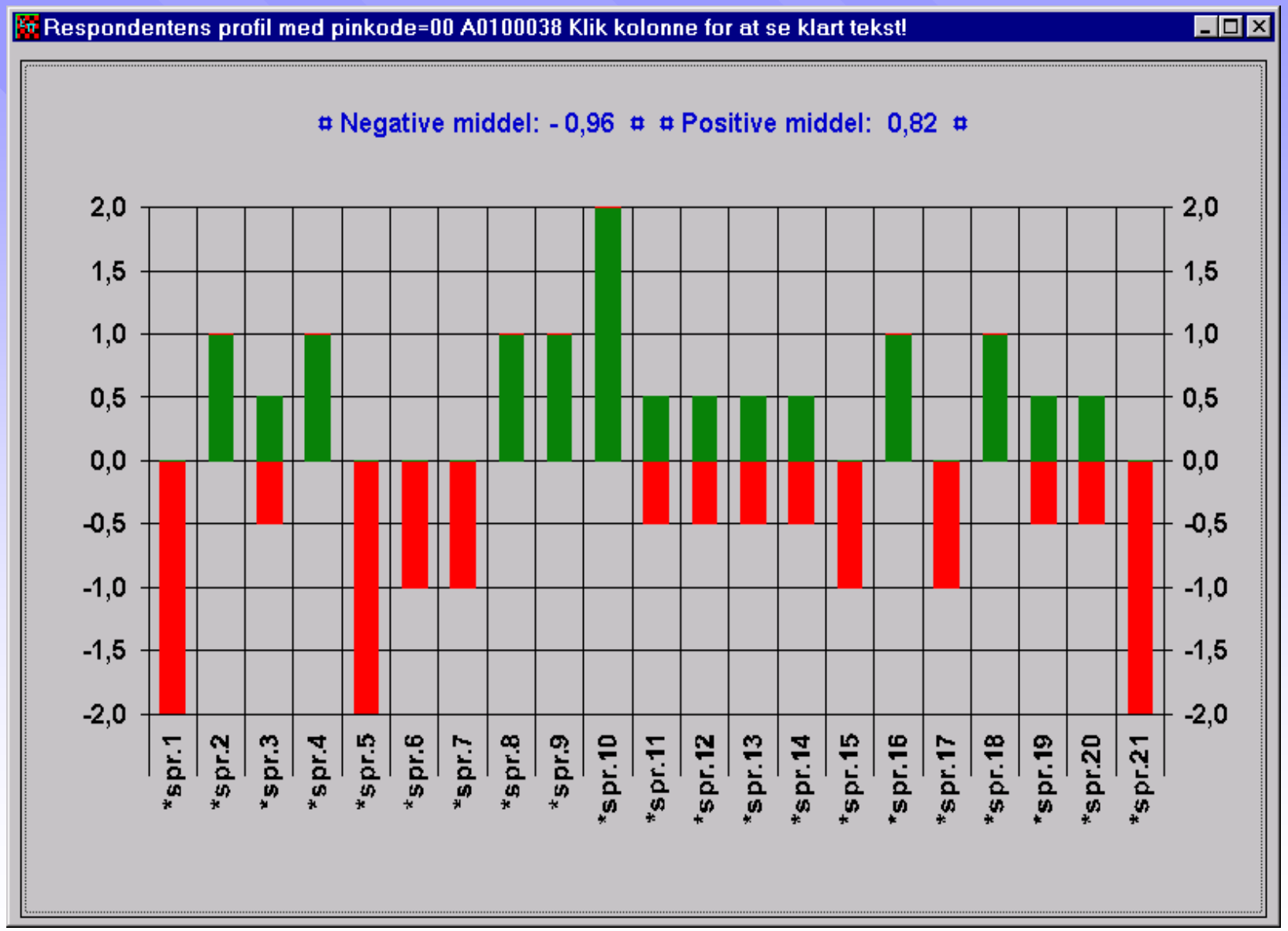
Negative/Positive Scale in the Questionnaire



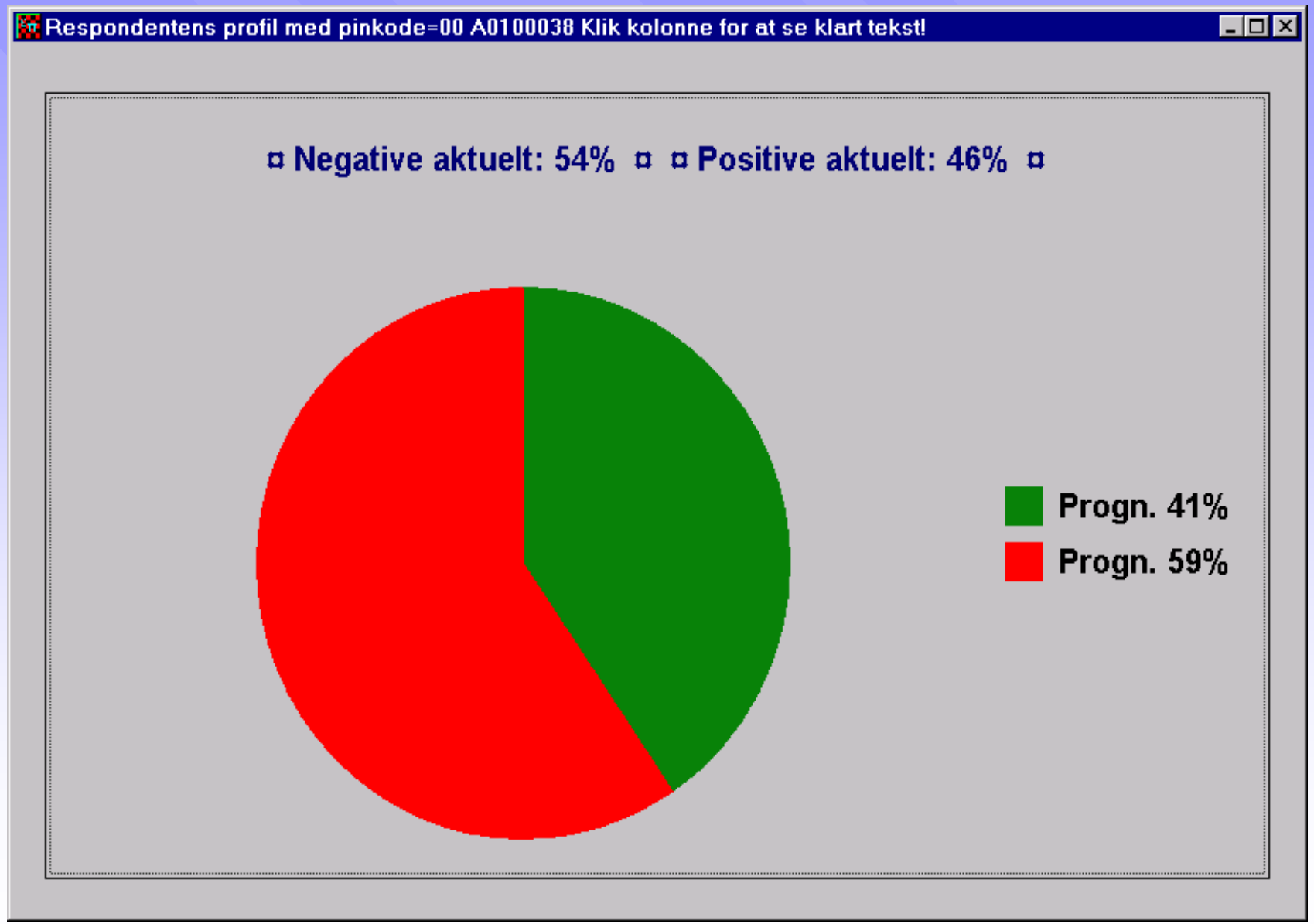
Negative/Positive Scale in the Questionnaire



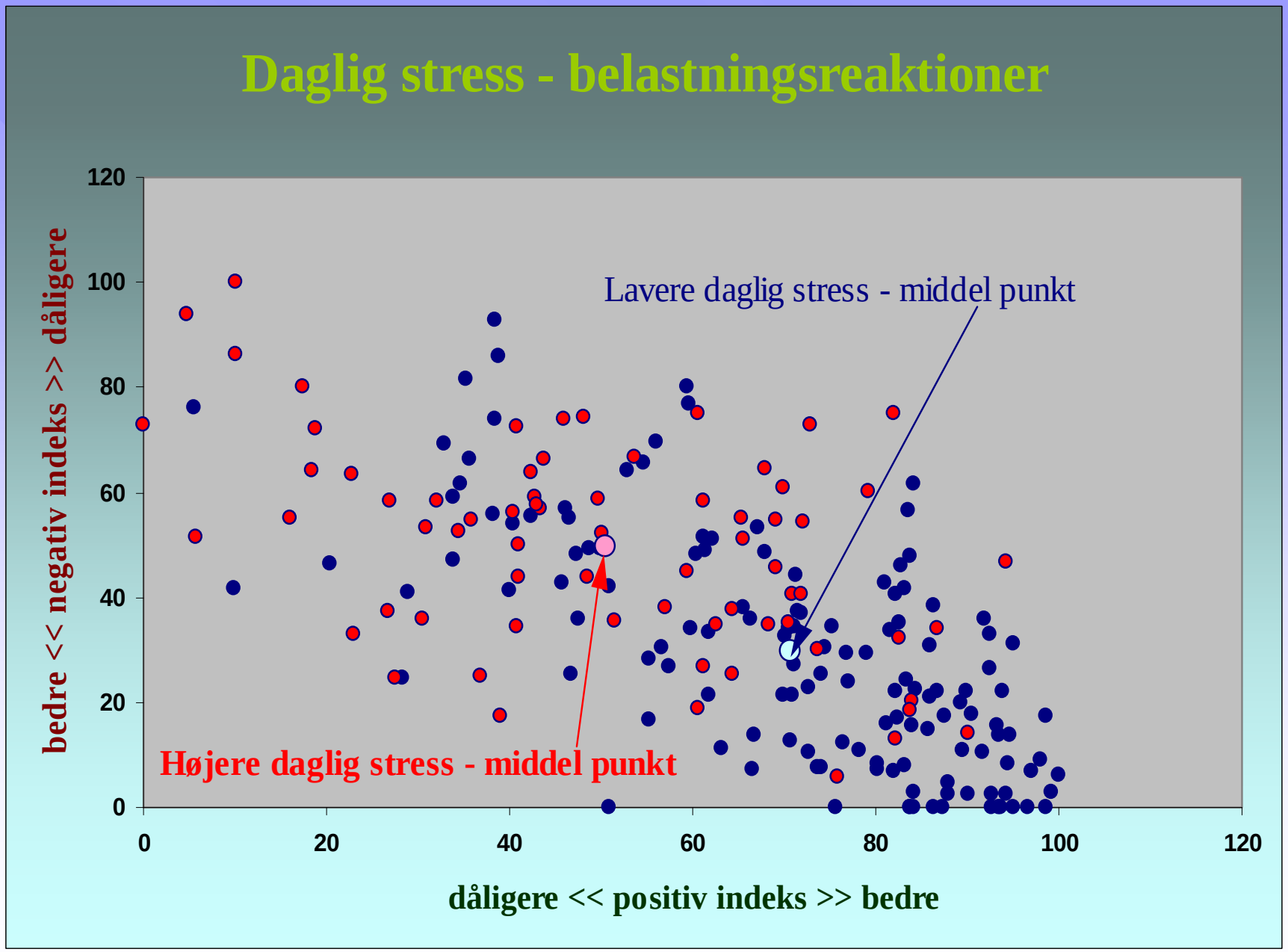
Respondent 00 A0100038 actual



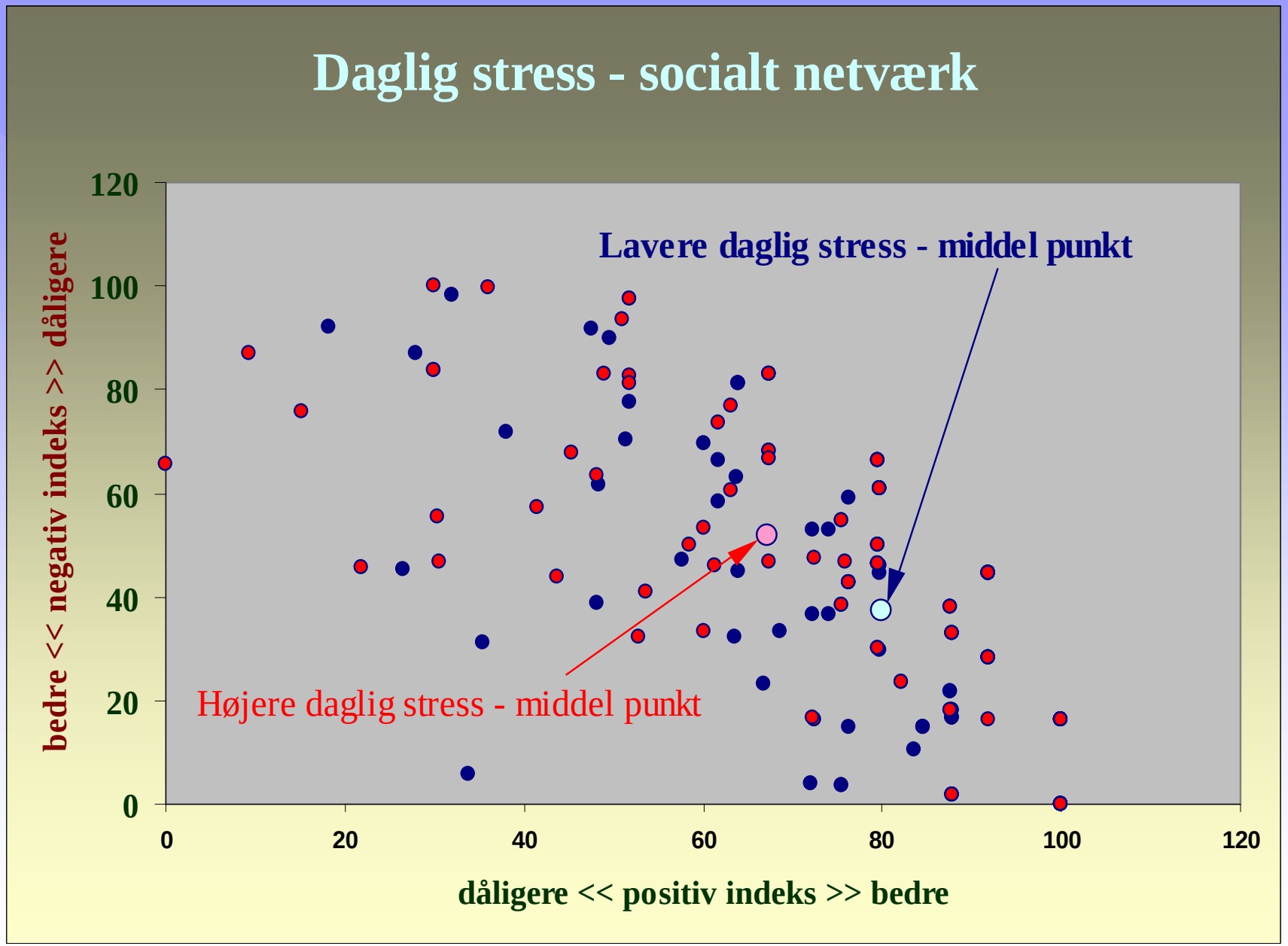
Respondent 00 A0100038 prognoses



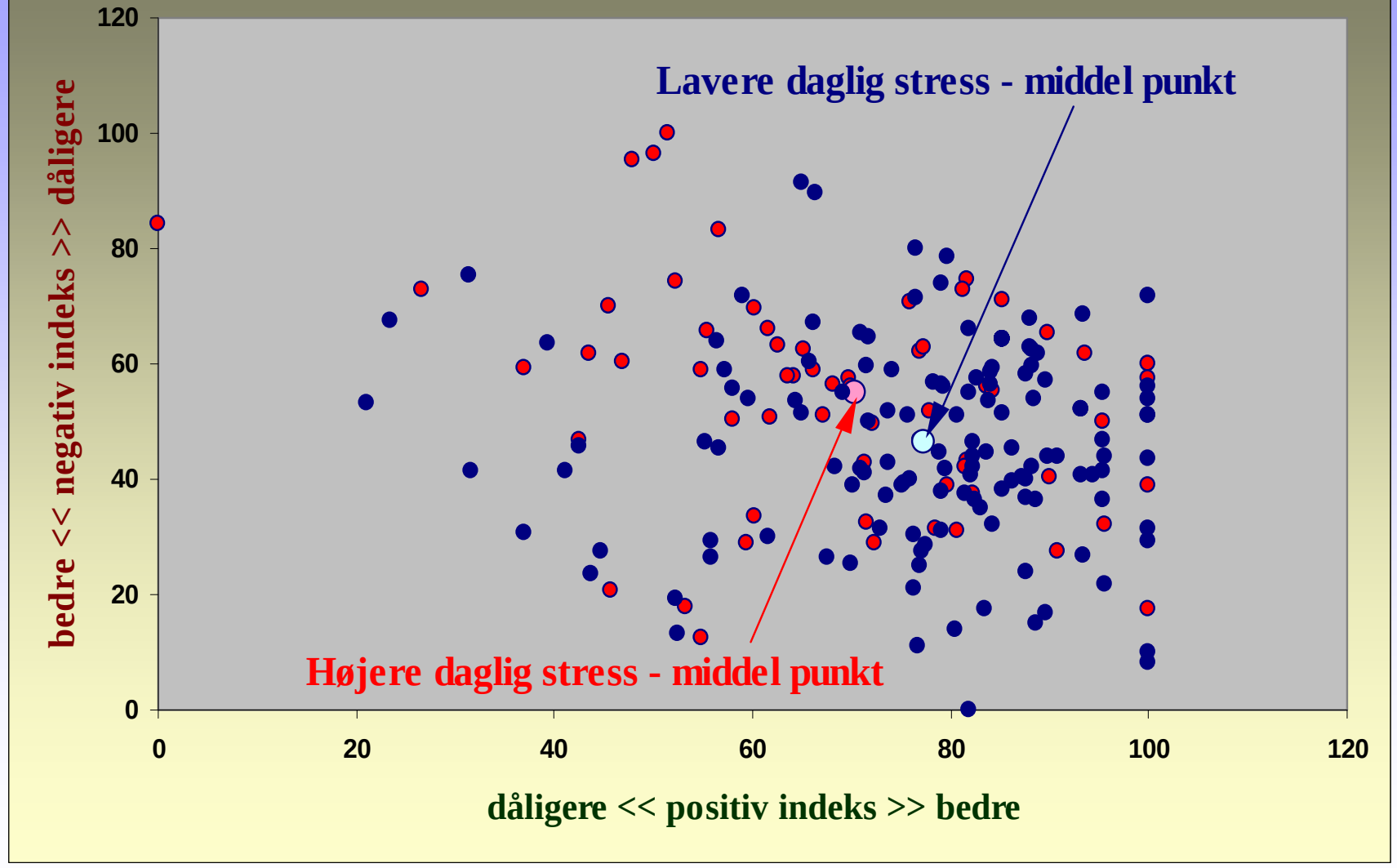
Daglig stress - belastningsreaktioner



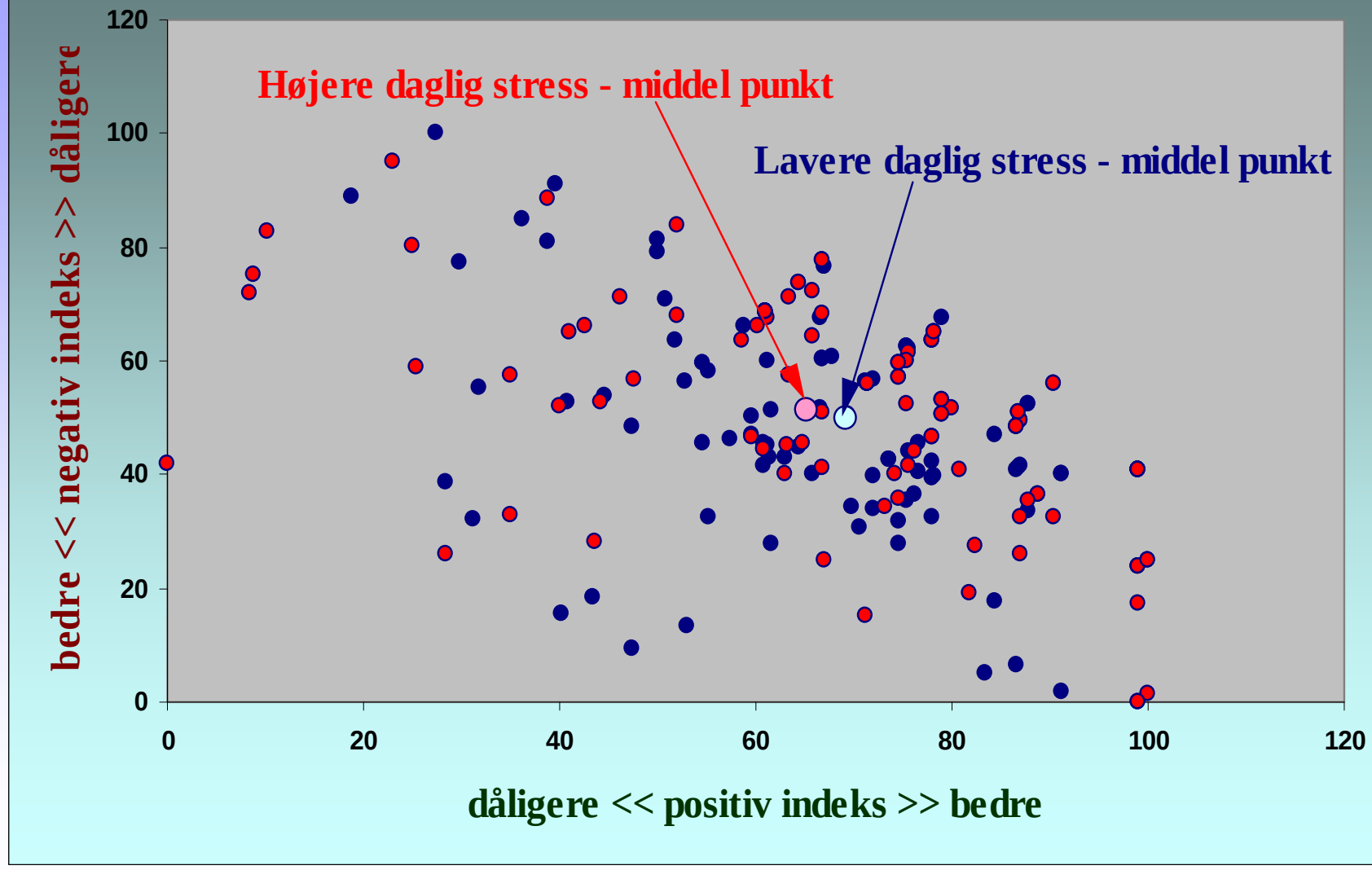
Daglig stress - socialt netværk



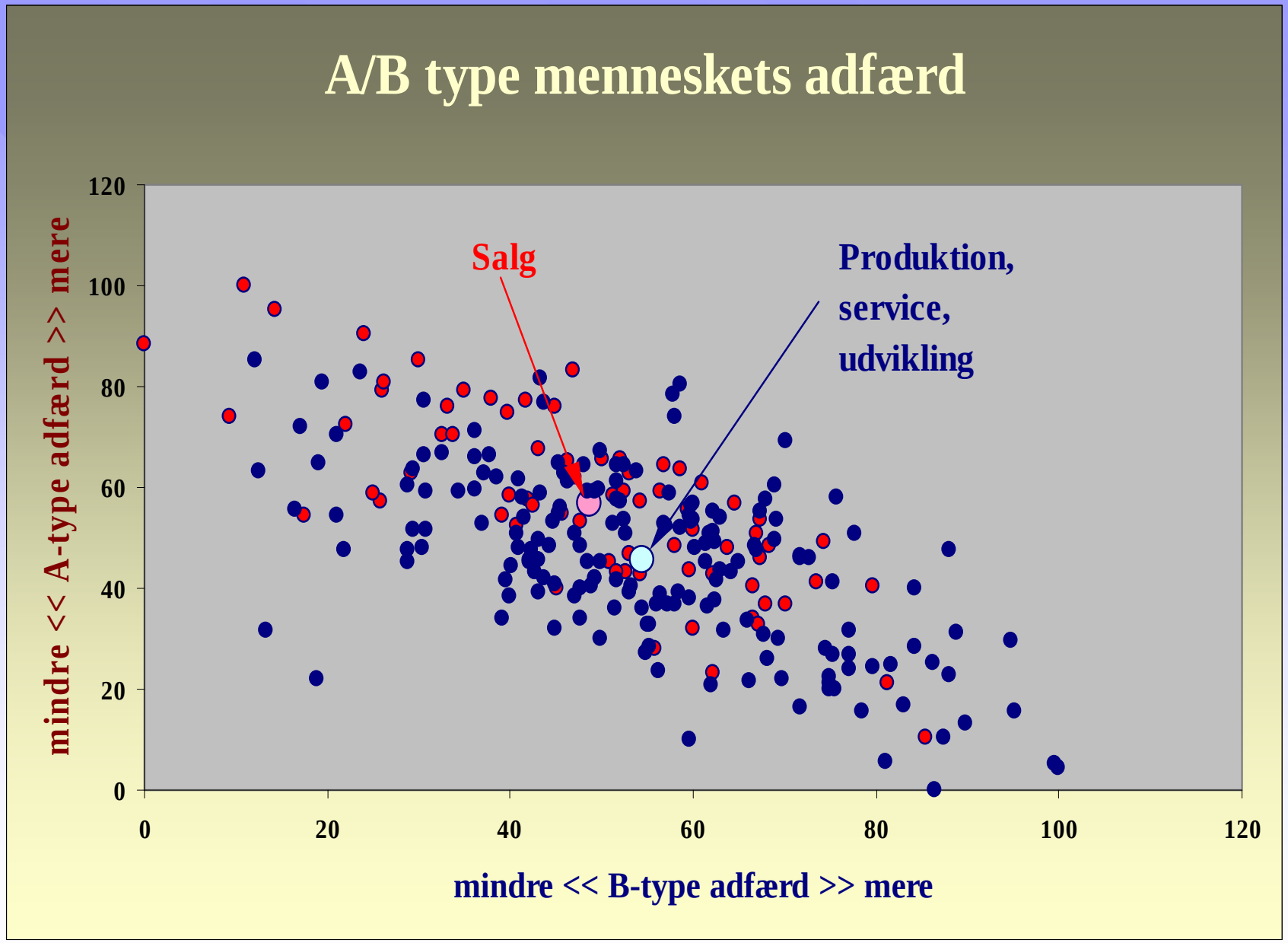
Daglig stress - psykisk arbejdsmiljø



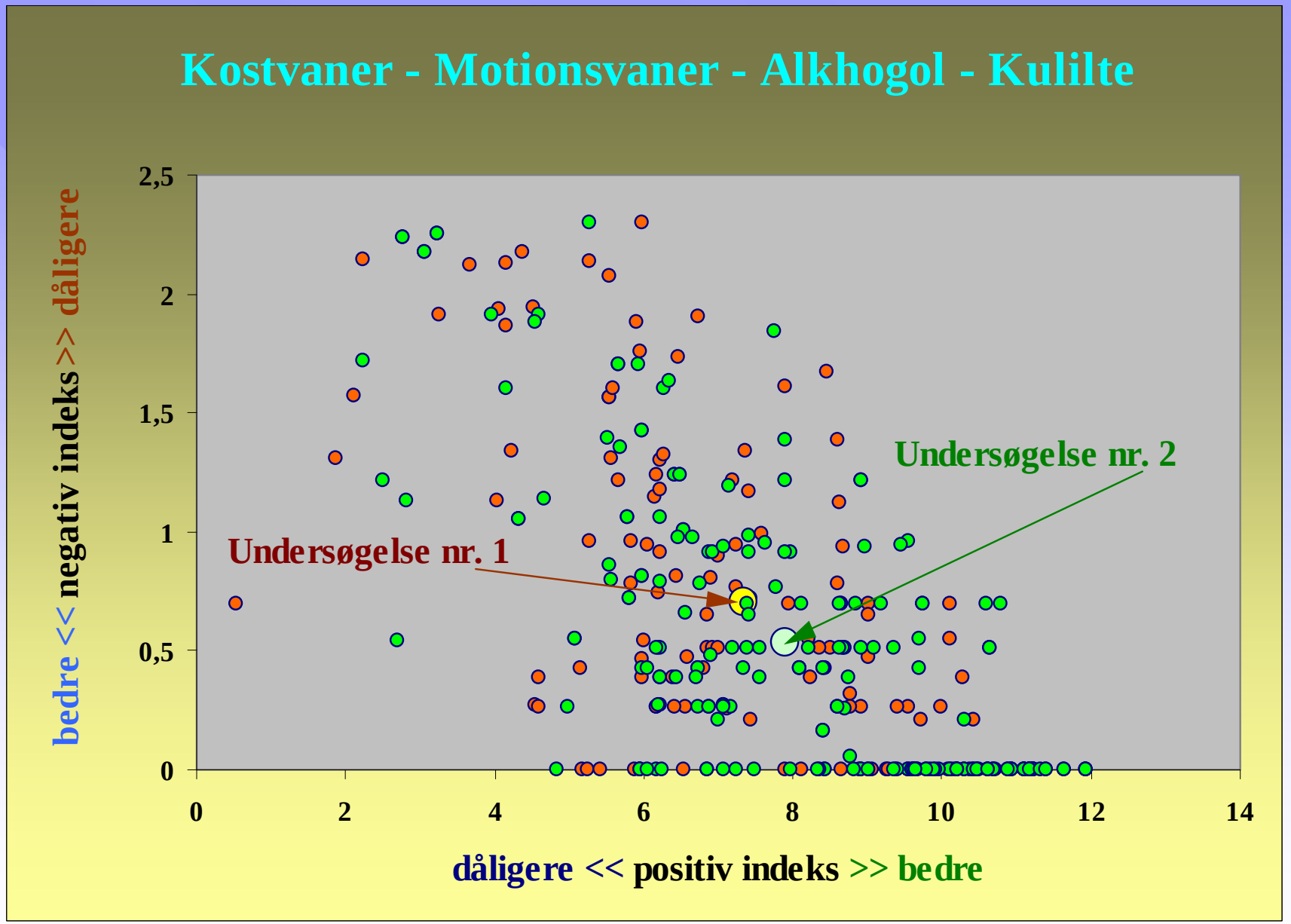
Daglig stress - fysisk arbejdsmiljø



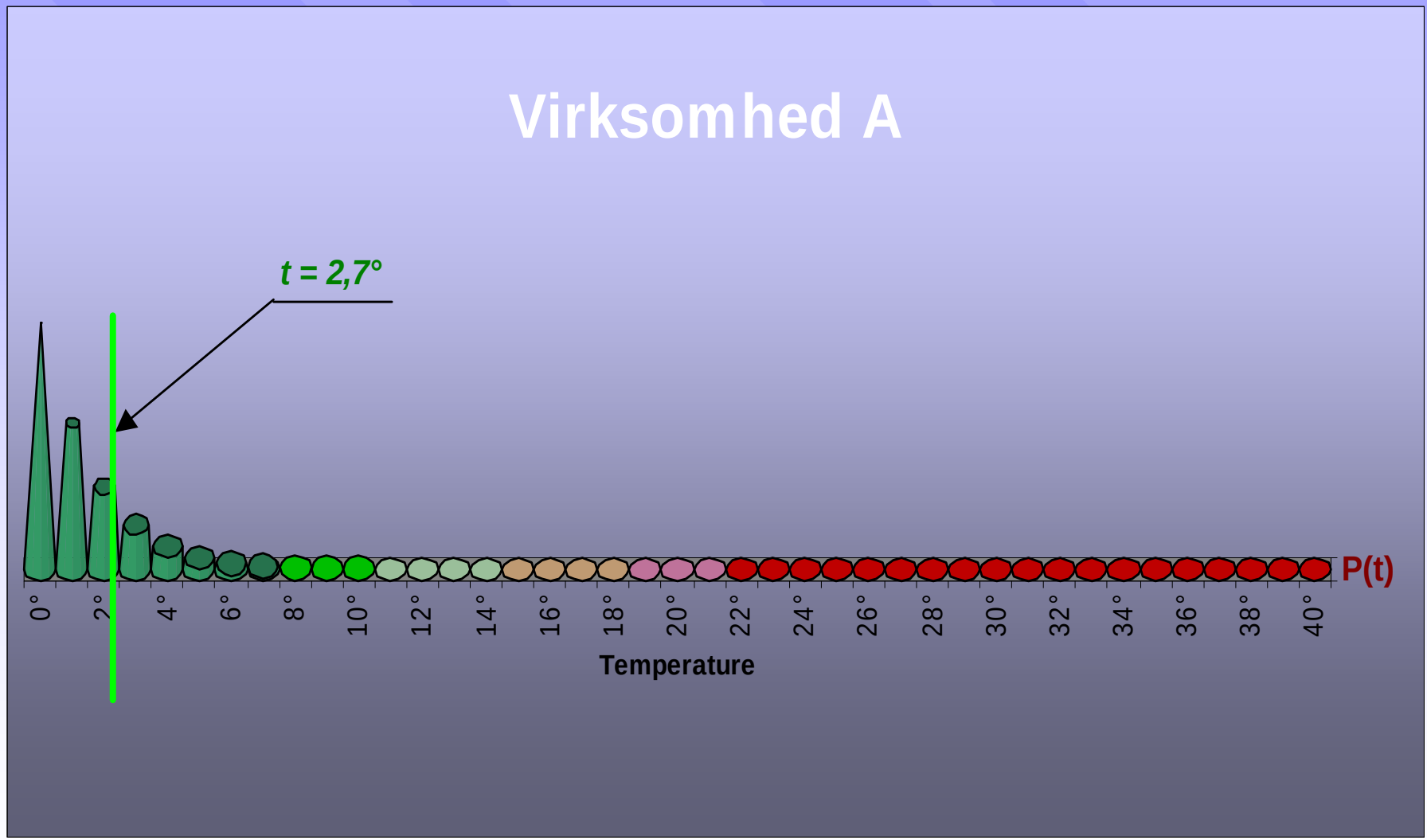
A/B type menneskets adfærd



Kostvaner - Motionsvaner - Alkhogol - Kulilte



Sundheds-feber for virksomhed A



Sundheds-feber for virksomhed B