



Salanti, Ulm:

Multidimensional isotonic regression and estimation of the threshold value

Sonderforschungsbereich 386, Paper 234 (2001)

Online unter: <http://epub.ub.uni-muenchen.de/>

Projektpartner



Discussion Paper

234

Multidimensional isotonic regression and
estimation of the threshold value

February 2001

Georgia Salanti, Kurt Ulm

1. Introduction

In many medical studies a researcher may strongly believe that the regression function used to describe a relationship between two variables has a particular shape or form which can be characterised by certain order restrictions such as monotonicity; for example exercise and serum cholesterol level, or air pollution and mortality. Given that the proof of such a monotonic relationship is an important criterion needed in establishing causality, several parametric approaches have been proposed. However it is well-known that different parametric models can lead to contradictory results¹. The result may also be highly depended on the quantification of the dose².

In occupational medicine it is of great interest not only to establish a dose response relationship but also to find a threshold concentration dose above which the disease rate increases significantly. In order to answer this important question, i.e. whether a threshold value can be assessed one can apply specific statistical models. However, the quantification of the dose can affect again the results³.

In this paper a non parametric method, isotonic regression, is described as an alternative method of establishing a dose response relationship as well as to assess a threshold value. This method has several advantages compared to parametric models. The only assumption is monotonicity and it does not require any specific form of the relationship. Furthermore, the result is independent of any monotone transformation of the dose. In the following sections the isotonic regression is described together with an algorithm. Subsequently an application is presented.

2. Isotonic regression

2.1 Univariate analysis : the simple order case

Description of the method

This method was first proposed by Barlow et al⁴ in 1972. Starting from the assumption that a monotonic dose-response relationship exists a maximum likelihood estimator under order restrictions of the response is assessed. The pooled advanced violator algorithm (PAVA) provides this estimator .

Consider the case of k dose groups where the dose is in increasing order $d_1 \leq d_2 \leq \dots \leq d_k$ and the outcome of an experiment is Y_{ij} , $i=1..k$, $j=1..n_i$. We do not have to specify the response Y_{ij} which can be binary, poisson, continuous or survival probability. The distribution of Y_{ij} can be written in the following form

$$f(y; \theta, t) = \exp(p_1(\theta) \cdot p_2(\theta) \cdot K(y; t) + S(y; t) + q(\theta, t)) \text{ and } Y_{ij} \sim f(\cdot; \theta(d_i), t)$$

where θ is the parameter of the distribution (for example the mean of the continuous distribution, Poisson parameter, or the positive outcome probability of the binary response) and p_1, p_2, q are some functions that satisfy

$$p_1(\theta) > 0, p_2'(\theta) > 0 \text{ and } q'(\theta, t) = -\theta \cdot p_1'(\theta) \cdot p_2(t) \text{ with respect to } \theta.$$

Then it is well-known that $E[K(Y, t)] = \theta$ and $V[K(Y, t)] = 1/[p_1'(\theta) \cdot p_2(t)]$

The maximum likelihood estimator of the parameter θ is

$$\hat{\theta}_i = \hat{\theta}(d_i) = \frac{1}{n_i} \cdot \sum_{j=1}^{n_i} k(Y_{ij}, t) \quad (1)$$

We wish to have θ_i in non decreasing order, given that $d_i < d_{i+1}$, $i=1\dots k$

That is true only if all the estimates $\hat{\theta}_i$ fulfil the isotonic relationship $\hat{\theta}_i \leq \hat{\theta}_{i+1}$ for each $i \leq k$. Then, denoting $\hat{\theta}_i^*$ the isotonic estimator of $\hat{\theta}_i$, it is trivial that $\hat{\theta}_i^* = \hat{\theta}_i$. If there is somewhere a violator such that $\hat{\theta}_i \geq \hat{\theta}_{i+1}$ for some i , then the isotonic estimator of θ_i is needed to be found.

The max-likelihood estimator for θ_i^* is computed using the pooled advanced violators algorithm (PAVA).

The PAVA starts with $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k)$. If $\hat{\theta}$ is isotonic then $\theta^* = \theta$. Otherwise, both violators $\hat{\theta}_i > \hat{\theta}_{i+1}$ are replaced by their weighted average

$$\hat{\theta}_i^* = \hat{\theta}_{i+1}^* = \frac{w_i \cdot \hat{\theta}_i + w_{i+1} \cdot \hat{\theta}_{i+1}}{w_i + w_{i+1}}, \text{ and } w_i, w_{i+1} \text{ are both replaced by } w_i + w_{i+1}. \text{ This}$$

process is repeated using the new values and weights until an isotonic set of values is obtained.

Note that the weights used in PAVA are determined by $w_i = w(x_i) = n_i \cdot p_2(t_i)$. Consequently, the weights are the frequencies when the response variable is binary, and the inverse of the variance of the mean when the response is continuous.

Tests

To test the hypothesis H_0 - equality of θ_i - against the isotonic alternative the likelihood ratio test is used. We have the following hypothesis:

$$H_0 : \text{equality of } \theta_i$$

$$H_1 : \theta_i \text{ are isotonic}$$

$$H_2 : \text{places no restrictions on the } \theta$$

In this chapter tests of H_0 versus H_1 (test T_{01}) and of H_1 versus H_2 (test T_{12}) are presented.

Considering again the exponential family the tests are as follows :

$$T_{01} = 2 \cdot \sum_{i=1}^k w_i \cdot \hat{\theta}_i \cdot \left[p_i(\hat{\theta}_i^*) - p_i(\theta_0) \right] + 2 \cdot \sum_{i=1}^k n_i \cdot \left[q(\hat{\theta}_i^*, t_i) - q(\theta_0, t_i) \right]$$

$$T_{12} = 2 \cdot \sum_{i=1}^k w_i \cdot \hat{\theta}_i \cdot \left[p_i(\hat{\theta}_i) - p_i(\hat{\theta}_i^*) \right] + 2 \cdot \sum_{i=1}^k n_i \cdot \left[q(\hat{\theta}_i, t_i) - q(\hat{\theta}_i^*, t_i) \right]$$

where $\hat{\theta}_i^*$ is the isotonic estimator of $\hat{\theta}_i$ assessed using the PAV algorithm and

$$\theta_0 = \frac{\sum_{i=1}^k w_i \hat{\theta}_i}{\sum_{i=1}^k w_i}$$

This test statistic T_{01} follows approximately a weighted Chi-square distribution

$$P(T_{01} \geq c) = \sum_{l=1}^k P(l, k, w') \cdot P[X_{l-1}^2 \geq c]$$

with $P(l, k, w')$ denoting the probabilities that given k subgroups under H_0 the isotonic regression will build l level sets and $P[X_{l-1}^2 \geq c]$ the probability to obtain a value greater than c for a random variable with X^2 distribution with $l-1$ degrees of freedom.

Equivalently the test statistic T_{12} follows the distribution

$$P(T_{12} \geq c) = \sum_{l=1}^k P(l, k, w') \cdot P[X_{k-l}^2 \geq c]$$

This approximation should be used with caution in the case of binary response. Simulations for T_{01} have shown² that when the absolute number of expected events are less than 5 in any of the dose groups, the equivalent exact test gives a critical value higher than the critical value taken from this theoretical distribution.

2.2 Multivariate analysis: Isotonic regression with more than one explanatory variable

Extension of the PAVA for more than one explanatory variables

If more than one covariates is taken into account, the data can be presented in form of an array, where the entries are the outcome (means, proportions ect.) and the margins are the covariates. In order to start with a simple situation consider just two explanatory variables that are categorised in R and C naturally ordered categories respectively. The data are presented in form of a matrix, where in the (i,j) cell is the outcome θ_{ij} of the individual being in the i -th category of the first variable and the j -th category of the second one.

Note that a matrix M is isotonic with respect to the partial order if and only if the elements of M along each row and column are non decreasing : the estimates for the outcome denoted by θ_{ij} fulfil the restriction $\theta_{ij} \leq \theta_{kl}$ for $i \leq k$ and $j \leq l$.

If this requirement does not holds, a row and column smoothing has to be done which can be easily accomplished using the PAVA. The idea is to assess the isotonic rows matrix and subsequently the isotonic columns matrix and to repeat that until convergence. The extension of the PAV algorithm is as follows:

Step 1: Let M^*1 denote the isotonic regression of M over rows. Let $R1=(M^*1-M)$ the first set of row increments.

Step 2: Let M^{**1} denote the isotonic regression over columns of $M+R1$.

Call $C1=M1-(M+R1)$ the first set of column increments.

Step3: At the beginning of the nth cycle $M^*(n)$ is obtained by isotonizing $M+C(n-1)$ over rows. The nth set of row increments is defined by $R(n)=M^*(n)-(M+C(n-1))$.

Next , obtain $M^{**}(n)$ by isotonizing $M+R(n)$ over columns.

Theorem : Both $M^(n)$ and $M^{**}(n)$ converge to the isotonic regression , M^{***} , with respect to the partial order as $n \rightarrow \infty$*

This algorithm as well as the PAV algorithm are programmed in S language^{5,6} and they are presented appendix.

This algorithm can be extended for more than two variables and it is effective when three variables are taken into account. However when the algorithm has to deal with four explanatory variables or more, because of the complexity in computing the estimates, the use of an additive isotonic model^{7,8} is proposed.

Although the additive isotonic model makes no part of the objectives of this paper, it would be interesting at this point just to give an idea about. In an additive isotonic model with three covariates for example, the separate isotonic effect $f_1(x_1)$, $f_2(x_2)$, $f_3(x_3)$ of each covariate in the model is assessed by computing the isotonic functions f_1 , f_2 , f_3 instead to their mixed effect $f(x_1, x_2, x_3)$ that is obtain by the isotonic matrix.

$E(Y| x_1, x_2, x_3) = f(x_1, x_2, x_3)$ using multidimensional isotonic regression

$E(Y| x_1, x_2, x_3) = f_1(x_1) + f_2(x_2) + f_3(x_3)$ using additive isotonic model

Tests

The same approach used in the univariate analysis can be extended to take additional possible explanatory variables into account. In the case of two covariates if one is interested in the effect of a single covariate the following test can be used:

Consider a $R \times C$ matrix whose entries are the means from a continuous response. The usual model with interaction, is

$$y_{ijl} = \mu + a_i + b_j + c_{ij} + e_{ijl}$$

$i = 1, 2, \dots, R \quad j = 1, 2, \dots, C \quad l = 1, 2, \dots, n_{ij}, \quad n_{ij}$ the frequency of each cell

where μ is the overall mean, $a_i = y_{i..} - y_{...}$ the residual of the i -th row mean, $b_j = y_{..j} - y_{...}$ the residuals of the j -th column mean, and c_{ij} the difference between the cell's mean and the overall mean. Let $y^*_{i..}$ the isotonic estimation of $y_{i..}$, assessed using the PAV algorithm. Suppose that the column variable has a given isotonic effect to the response.

Then we would like to test the isotonic effect of the row variable. The idea is that if the row variable has no significant increasing trend , then the isotonic estimation of the row margins $y^*_{i..}$ will not differ significantly from the overall mean $y_{...}$ So, we have the following hypothesis:

$$H_0 : a_1 = a_2 = \dots = a_R = 0$$

versus

$$H_1: a \text{ is isotonic with respect to a quasi order on } \{1, 2, \dots, R\}.$$

Denoting $\hat{a}_{H1} = (y^*_{1..} - y_{...}, y^*_{2..} - y_{...}, \dots, y^*_{R..} - y_{...})$ and $\hat{a}_{H0} = (0, 0, \dots, 0)$

then analogously to the test used in the univariable approach we define

$$T_{01} = \frac{C \cdot \sum_{i=1}^R n_{ij} \cdot (y_{i..}^* - y_{...})^2}{C \cdot \sum_{i=1}^R n_{ij} \cdot (y_{i..} - y_{...})^2 + \sum_{ijl} (y_{ijl} - y_{ij.})^2} \cdot \frac{nRC}{nRC}$$

This test statistic should approximately follow the weighted Chi-square distribution described in 2.1 section. However, one can expect that this approximation does not always hold for the same reasons explained in the univariate case.

This test can also be used when the response is binomial, considering that the proportions follows approximately the normal distribution having mean p_i and $p_0 \cdot (1 - p_0) / n_i$ is an estimator of the variance, where $p_0 = \sum p_i \cdot n_i / \sum n_i$

This is a very common approximation but it holds only if the proportions are around 50%. The solution to this problem is again the performance of a permutation test or a simulation procedure. However, if more than two covariates are taken into account, due to large computing time, the use of an additive isotonic model is recommended.

Some computational problems

In a previous section we described the algorithm used to assess the isotonic estimator, when the original data are presented in form of an array. The theorem outlined in that section, that assures convergence, would not necessarily hold if the data array has a lot of zero weighted cells, since in that case the individual row and column isotonic regressions would not be uniquely determined.

In order to illustrate that, consider the following example taken from the study “Chronic Bronchitis” : The matrix which is to isotonize is the event rate as a result of exposure time (in 5-years groups) in total inhalable dust and its concentration (in 0,5- mg/m^3). An extract of the data matrix is presented : (events/number of exposed)

Table 1 : Extract of the original data matrix, example of no convergence.

mg/m ³	years		
	35-40	40-45	45-50
5,5-6	0	0	0
6-6,5	1/1	0	0
6,5-7	0	1	0
7-7,5	0	0	0
7,5-8	0	0	0
8-8,5	0	0	0
8,5-9	0	1/2	2/4
9-9,5	0	0	0
9,5-10	0	0	0

The figure 1 shows the isotonic-columns matrix and the isotonic-rows matrix. It is evident that the algorithm fails to converge. The column smoothing and the row smoothing do not result in a common isotonic matrix.

This problem may be avoided by using the following procedure proposed by Robertson and Write⁹ :(1) remove the zero-weight cells and disregard all orderings which involve these cells (2) carry out the row-column isotonic regression (3) insert any values the zero-weight cells.

However this alternative is not always efficient. Applying this procedure to the matrix there is no way to fill the cells with the appropriate values (table 2).

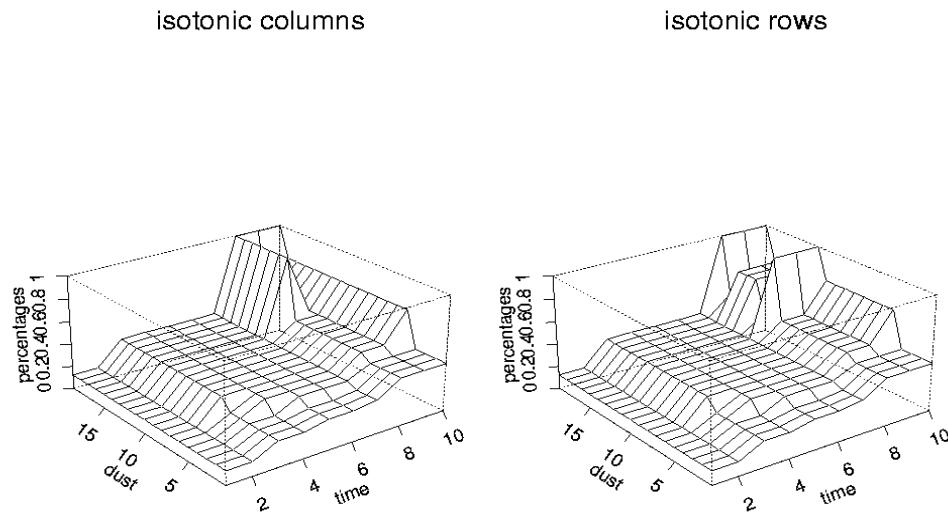


Figure 1: Column isotonic matrix and row isotonic matrix after 100 iterations

Table 2 : Extract of the isotonic matrix

m/cm³ years	35-40	40-45	45- 50
6-6,5	1	NA	NA
6,5-7	NA	0,471	NA
7-7,5	NA	NA	NA
7,5-8	NA	NA	NA
8-8,5	NA	NA	NA
8,5-9	NA	0,5	0,5
9-9,5	NA	NA	NA
9,5-10	NA	NA	NA

One efficient approach in order to avoid this problem, is before starting the isotonic regression procedure to try to eliminate the zero weighted cells by regrouping. Of course, that is not always desired especially when the assessment of a threshold value is of interest.

3. Estimation of the Threshold value

The isotonic regression can be extended to estimate a threshold value. Using a backward elimination algorithm, one can reduce the number of level sets by combining those whose values do not differ greatly. Following the idea of Schell and Singh¹⁰ the threshold value can be found by amalgamating adjacent groups and comparing the likelihood functions.

If the likelihood function does not change a lot, both corresponding groups can be pooled together. If the difference of the likelihood functions exceeds a critical level a , then the dose of the higher group can be assessed as threshold value. Denoting L_i the likelihood function resulting from the data when the i first groups are pooled together, the test statistic

$$D = 2 \cdot (\ln L_i - \ln L_{i-1})$$

can be used to test whether the i th dose level is a threshold value or not, and should follow approximately a χ^2 -distribution with one degree of freedom¹. In order to assess a threshold value, the dose groups are pooled starting with the two lowest groups.

In this paper a significant level of 1% is used, which is justified by a study of Shell and Sing. They showed for the case of normal means that the choice of $a=5\%$ does not yield to a 5% significance level test.

4. Applications to the study “Chronic Bronchitis”

4.1 Description of the Chronic Bronchitis study

The data are taken from the DFG study “Chronic Bronchitis”. The aim of this study has been to investigate the influence of the occupational total inhalable dust exposure besides the well-known factors smoking and age. The disease rate is suspected to increase while the concentration and the duration on the exposure increase. A further aim of this study has been in case of association between dust and disease, whether a threshold value for the dust concentration could be assessed. The endpoint in this study was chronic-bronchitis reaction.

For the analysis the data are grouped into eight dust intervals and eight time groups (less than 10 years exposure, than every 5 years and more than 45 years) in order to have enough observations and events in each category. The data of the cohort of smokers and the cohort of non-smokers are described in tables 3 and 4.

Table 3: Descriptive statistics of the sample of non-smokers and ex-smokers.

	<i>Total : 326</i>	<i>Without CBR :275</i>	<i>With CBR : 51</i>
Median (Min-max)			
Total inhalable dust concentration (in mg/m³)	1,43 (0-15,04)	1,43(0-15,4)	1,45 (0,36-8)
Time since first exposure (in years)	24 (1-66)	23 (1-55)	33 (8-66)

Table 4: Descriptive statistics of the sample of smokers.

	<i>Total : 920</i>	<i>Without CBR :679</i>	<i>With CBR : 241</i>
Median (Min-max)			
Total inhalable dust concentration (in mg/m³)	1,40 (0,20-15)	1,07(0,20-15)	4,62(0,25-12,07)
Time since first exposure (in years)	25 (3-51)	24 (3-51)	28 (6-49)

4.2 One-dimensional isotonic Regression

Starting from the assumption that time has an effect on CBR, one would like to test the effect of the dust concentration and to establish a dose-response relationship. In order to illustrate the univariable isotonic regression and the related tests, we attempt a univariate approach, although the results are expected to be not reliable, because of the influence of the time. The isotonic regressions and the tests for total inhalable dust are as follows:

Table 5: Isotonic regression for total inhalable dust ,non-smokers and ex-smokers

Group	Upper limit of class mg/m ³	Event rate (with CBR)	Isotonic regression	tests	p-value
1	-0,39	0,050	0,050		
2	-0,44	0,111	0,111	T ₀₁ : 6,23	0,048
3	-0,64	0,156	0,556		
4	-1,39	0,271	0,172		
5	-4,03	0,125	0,172		
6	-5,23	0,105	0,172	T ₁₂ : 6,32	0,314
7	-6,26	0,250	0,200		
8	6,26+	0,150	0,200		

Table 6: Isotonic regression for total inhalable dust : smokers

Group	Upper limit of class mg/m ³	Event rate (with CBR)	Isotonic regression	tests	p-value
1	-0,39	0,242	0,204		
2	-0,44	0,217	0,204	$T_{01}: 27,56$	<0,001
3	-0,64	0,190	0,204		
4	-1,39	0,212	0,204		
5	-4,03	0,154	0,204		
6	-5,23	0,319	0,319	$T_{12}: 3,13$	0,69
7	-6,26	0,363	0,363		
8	6,26+	0,393	0,393		

In the sample of non smokers the increase of dust concentration seems to have a low influence to the disease risk since the test T_{01} results to a rather small value that is slightly above the critical value. However, for the sample of smokers one can conclude that a strong monotonic dose response relationship truly exists. The value of the T_{01} test is large resulting in a p-value less than 0,001. Furthermore the T_{12} test (adequacy of the isotonic hypothesis against to no-restrictions hypothesis) results in a non significant value: that means evidence of the increasing dose-response relationship against any other possible regression shape (umbrella or U-shape for example).

The results from isotonic regression for the hole sample are presented in table 7 .

Table 7: Isotonic regression for total inhalable dust :total sample

Group	Upper limit of class mg/m ³	Event rate (with CBR)	Isotonic regression	tests	p-value
1	-0,39	0,195	0,189		
2	-0,44	0,197	0,189	$T_{01}: 25,42$	<0,001
3	-0,64	0,181	0,189		
4	-1,39	0,231	0,189		
5	-4,03	0,146	0,189		
6	-5,23	0,266	0,266	$T_{12}: 3,88$	0,59
7	-6,26	0,333	0,332		
8	6,26+	0,331	0,332		

This analysis gives information about the unadjusted influence of the dust concentration and is just for presenting the results of the univariate approach.

4.3 Two-dimensional isotonic regression

In order to take into account both factors dust concentration and time since first exposure, the two-dimensional isotonic algorithm is used separately for non-smokers and smokers. The results are presented in figures 2 and 3.

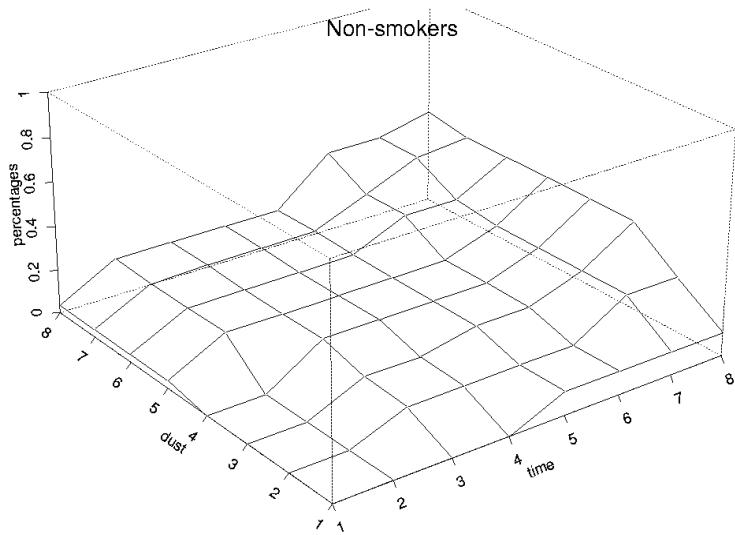


Figure 2: Two-dimensional isotonic regression (Non-smokers)
Likelihood function: $-2\ln L=243,20$

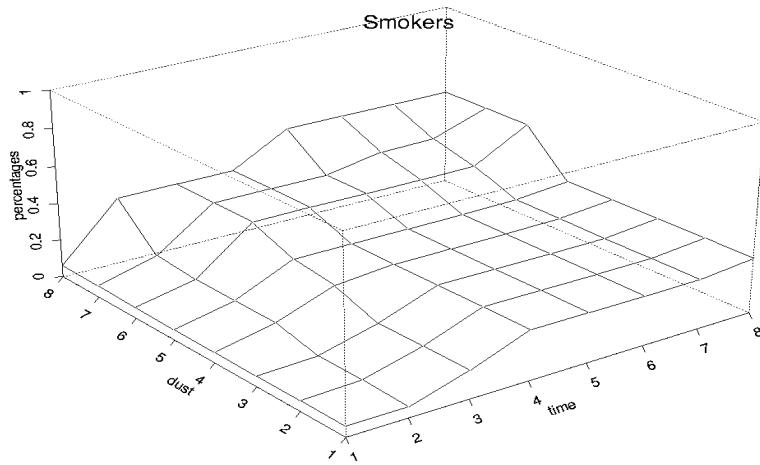


Figure 3: Two-dimensional isotonic regression (Smokers)
Likelihood function: $-2\ln L=969,38$

In order to test the effect of dust concentration and the effect of time a permutation test is performed¹¹. Based on the observed number of patients per time and dust group and the total number of events, 10000 permutations are analysed. Each worker is characterised by three elements (t_i, d_i, s_i) , $i=1 \dots 326$ (non smokers), or $i=1 \dots 920$ (smokers) with t_i denoting the time-group, d_i the dust-group and s_i the disease status (0 absence, 1 occurrence). For the permutation test the pair (t_i, d_i) is considered independent from the status so that the events occur in random allocation. Within each permutation H_0 is considered. The isotonic estimation of each permutations matrix is obtained using the two-dimensional extension of the PAV algorithm and the difference between the likelihood function obtain from each permutations matrix (L_0) and his isotonic estimation (L_{isot})

$$\text{Test } D = -2 (L_0 - L_{\text{isot}})$$

is used as test statistic. Afterwards the test value resulting from the observed data is compared with the 95th percentile of the permutation test's distribution. If the T_{obs} exceeds the 95th percentile, then the observed allocation of the events can not be explained by the hazard and H_0 is to be rejected. The exact p-value is estimated as the probability that the result of a permutation is equal or greater to the observed.

In the sample of non smokers the observed difference $D_{\text{obs}} = -2 (L_0 - L_{\text{isot}}) = 39,59$ and the exact p-value taken from the permutation test is less than 0,001. As to the sample of smokers $D_{\text{obs}} = 88,79$ and the exact p-value is again less than 0,001. In the following table the changes of the Likelihood functions are presented, for both smoker's and non-smoker's samples when the covariates time and dust are taken into account.

Table 8: Likelihood functions

	Non-smokers ,ex-smokers	Smokers
	$- 2 \cdot \ln L$	$- 2 \cdot \ln L$
No covariate	282,79	1058,17
time	261,71	1007,18
dust	276,55	1030,61
Time and dust	243,20	969,38

The method described in section 3 for assessing the threshold value is used and the results are presented in table 9.

In the first column, in the i-th cell is presented the Likelihood function corresponding to data regrouped until the i-th changepoint as well as the p-value corresponding to the change of the likelihood function when the i-th group is pooled.

Table 9 :Estimation of the threshold value

Amalgamating dust groups	Non and ex-smokers		Smokers	
	- 2 · ln L	P value	- 2 · ln L	P value
-0,39 mg/m³	243,20		969,38	
-0,44	245,31	0,073	969,49	0,373
-0,64	247,74	0,060	969,45	0,500
-1,39	257,85	0,001	972,66	0,037
-4,03	260,80	0,043	973,24	0,224
-5,23	260,36	0,500	979,78	0,006
-6,26	261,52	0,141	989,75	0,001
6,26+	262,95	0,116	1008,82	<0,001

In the sample of non-smokers the test statistic D outlined above is too small to indicate a significant increase of risk when the three first groups are pooled. If the next dose group (0,65 – 1,39 mg/m³) is aggregated the change of the likelihood function is large enough to indicate a significant increase in risk from that dose and beyond. A threshold value in the interval 0,64 – 1,39 mg/m³ is obtained.

As to the smoker's sample, applying the same idea a total inhalable dust concentration somewhere above 4,03 mg/m³ lead to a significant increase in the likelihood function. The threshold value can be assessed to be in the interval 4,03-5,23 mg/m³.

4.4 Three-dimensional isotonic regression

It seems logical to assume a priori that the event rate is higher for the smokers than for the non-smokers for each dust and time group. Under this assumption a three-dimensional isotonic estimation of the disease proportions is attempted using an extension of the algorithm presented in section 2.2 by adding a covariate (non-smokers , smokers). For each combination of time and dust the risk for a smoker is equal or greater than the corresponding risk for an ex- or non-smoker.

At each point, the surface of the smoker's event rate is upper the surface of non-smokers. The result of the isotonic regression depicted in figure 4.

If the estimation of a global threshold value (for non smokers and smokers) is of interest, then a logical solution is to obtain the lowest one, which is 0,64 mg/m³. However, since the increasing trend is borderline significant in the non-smokers sample we can not attach a lot of reliability to the threshold value obtained from this cohort.

The idea of assessing the threshold value presented in section 3 can be used for the three dimensional isotonic regression. Adjacent dust groups will be pooled starting from the lowest expositions levels for both the smokers and non-smokers sample. In each pooling the obtained value of the likelihood function is compared with the value assessed from the previous pooling . Using the D likelihood ratio test, we obtain a

threshold value in the interval 0,64 -1,39 mg/m³. In table 10 are presented the changes of the likelihood function for the whole sample.

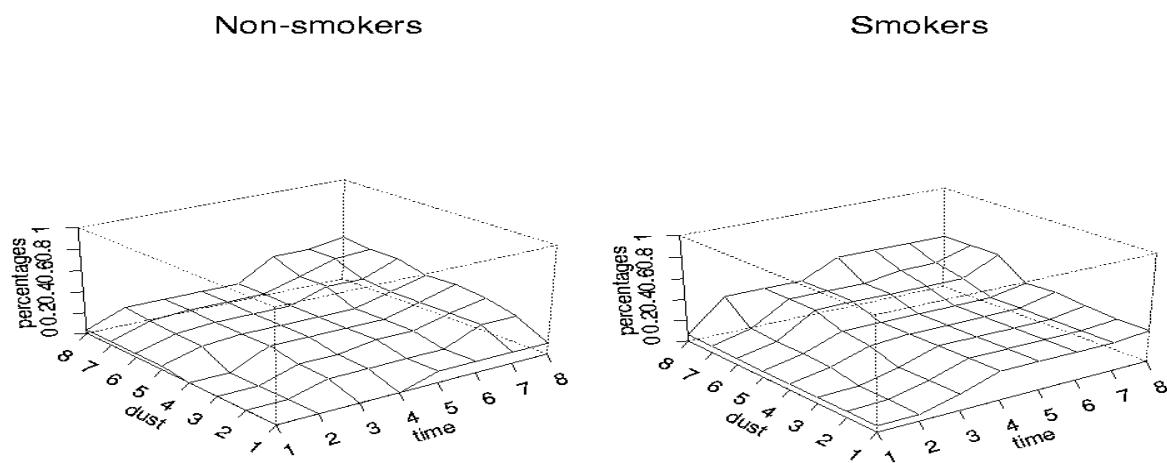


Figure 4 : Three-dimensional isotonic regression

Table 10: Estimation of the global threshold value (total sample)

Amalgamating dust groups	- 2 · ln L	P value
-0,39 mg/m ³	1216,27	
-0,44	1218,46	0,070
-0,64	1220,05	0,104
-1,39	1232,19	<0,001
-4,03	1235,45	0,036
-5,23	1240,85	0,010
-6,26	1251,47	<0,001
6,26+	1271,87	<0,001

5. Discussion

Our primary objective in this paper has been to describe a methodological approach and a computational proposal to the non-parametric regression under order restrictions when the selection of covariates and the assessment of a threshold value are of interest. This approach is more flexible than parametric models and more reliable than other non parametric alternatives. Exact methods are used in order to check the significant influence of the covariates. The idea of pooling blocks, as long as each such pool does not increase a criterion by more than the specified amount can be used to assess a threshold value.

More work remains to be done, because of three main problems related to isotonic regression and to the reduced isotonic regression. The first one concerns the appropriate test statistic T_{01} when one wants to test the equality of proportions against the isotonic alternative. In this paper the large sample approximation described by Robertson et al⁹ is used to assess the corresponding p-value in the univariate approach. However, the large sample approximation does not always hold. Simulations have shown that if the expected events under H_0 are less than 5 in any of the dose groups the large sample approximation fails². In this case a permutation test, as it is illustrated in section 4.3 for two covariates, is proposed as alternative.

The other problem concerns the test of the effect of one covariate when the effect of a second is given. The test presented in section 2 requires normal distribution of the proportions, and that is not always fulfilled. The solution is again the performance of a permutation test.

Isotonic regression was used also to assess the threshold value. The use of a $X_{1,a}^2$ test statistic seems to be reasonable as the test concerns the changes of the Likelihood function, but the main problem is related to the appropriate number of degrees of freedom. Furthermore, the reduced isotonic regression estimator obtained by pooling groups depends perhaps on the choice of a -significance level. Permutations tests performed by Schell and Singh¹⁰ who applied the reduced monotonic regression on the normal means case have shown that using a in the elimination procedure does not yield an a -level test : the appropriate a for an alpha level of 0,05 is certainly less than 0,05.

However given its many attractive features and simplicity, the isotonic regression and the reduced isotonic regression methods could prove to be an important method to fit a changepoint model.

Appendix

A proposal algorithm for computing the isotonic regression in S-language.

Two-dimensional isotonic regression

```
d2.isotonic.fun<-function(data,w,Likelihood,plot=T)
#This algorithm computes the two-dimensional isotonic regression with respect to the partial order of
the data matrix.
```

#Required arguments:

data: The data matrix having rows and columns in increasing dose :The event rate in each group

w: This is a matrix (same dim as data matrix):The exposed workers in each group

#If there are zero-weighted cells they are replaced by 0,00001

#Optional arguments:

Plot (default value plot=T)

Likelihood argument assess the likelihood functions

#The iterative algorithm (T.Robertson "Order Restricted Statistical Inference") is used

```
m <- dim(data)[1]      #nr of rows of the data matrix
n <- dim(data)[2] #nr of columns of the data matrix
C <- matrix(0, m, n)
it <- 0
difference <- matrix(1, m, n)
w<-replace(w, w==0,0.00001)
```

THE ALGORITHM

```
while(any(difference > 0.001) && it < 100) {
  #isotonizing over rows
  G.row <- data + C
  for(i in 1:m) {
    G.row[i, ] <- isotone.simple.order.fun((data + C)[i, ], w[i, ])$isotonic}
  R <- G.row - data - C      # rows increment

  # isotonizing over columns
  G.col <- data + R
  for(j in 1:n) {
    G.col[, j] <- isotone.simple.order.fun((data + R)[, j], w[, j])$isotonic}
  C <- G.col - data - R      # columns increment

  difference <- abs(G.col - G.row)
  it <- it + 1  }
  isotonic.data <- G.col
  dimnames(isotonic.data) <- dimnames(data)

  #*LIKELIHOOD FUNCTIONS*
  if(!missing(Likelihood)) {
    #Computing the Likelihood Functions
    events <- replace(events, events == 0, 1e-07)
```

-2ln L0

```
  events <- data * w
  overall.rate <- sum(events)/sum(w)
  minus2.lnL0 <- (-2) * (sum(events) * log(overall.rate) + (sum(w) - sum(events)) *
  log(1 - overall.rate))
  # -2ln Lcol
  events.col <- apply(events, c(2), sum)
  ni.col <- apply(w, c(2), sum)
  isot.col <- isotone.simple.order.fun(events.col/ni.col, ni.col)$isotonic
  minus2.lnL.col <- (-2) * (sum(events.col * log(isot.col) + (ni.col - events.col) *
  log(1 - isot.col)))
```

```

# -2ln Lrow
  events.row <- apply(events, c(1), sum)
  ni.row <- apply(w, c(1), sum)
  isot.row <- isotone.simple.order.fun(events.row/ni.row, ni.row)$isotonic
  minus2.lnL.row <- (-2) * (sum(events.row * log(isot.row)) + (ni.row - events.row) *
  log(1 - isot.row)))
}

# -2ln Ltotal
  isotonic.data2 <- replace(isotonic.data, isotonic.data == 0, 1e-05)
  minus2.lnL.total <- (-2) * (sum(events * log(isotonic.data2)) + (w - events) * log(1 -
  isotonic.data2)))
  Likelihood.functions <- c(minus2.lnL0, minus2.lnL.col,
  minus2.lnL.row, minus2.lnL.total)

  names(Likelihood.functions) <- c("-2lnL0", "-2lnL.col", "-2lnL.row", "-2lnL.total")
}

else (Likelihood.functions <- c("You could ask for Likelihood functions-
argument:Likelihood"))

#*PLOT*
  if(plot) {par(ask = T)
  persp(t(isotonic.data), zlab = "percentages", zlim = c(0, 1), xlab = "columns
variable", ylab = "rows variable")
  title("Isotonic Regression")
  par(ask = F)}

#*RESULTS*:
#1.Number of iterations for convergence
#2.The isotonic Matrix
#3.The Likelihood Functions
#4.Two-dimentional plot
  return(list(iterations = it, isotonic.data = isotonic.data, Likelihood.functions =
Likelihood.functions))
}

*THE PAVA ALGORITHM*
isotone.simple.order.fun_function(y,w)
{
# The classical PAVA algorithm
#--- y: is a vector (means, proportions etc.) and we are interested to find the isotonic estimator of y
with respect to the simple order
#----w: are the weights. If there are zeros, they are replaced by 0.00001
  w<-rep(w,w==0,0,0001)
  k <- 1:length(y)
  diff <- y - c(0, y[ - length(y)])
  j <- 1:length(y)
  while(any(diff < 0)) {
    viol <- min(j[diff < 0])
    k[k == k[viol - 1]] <- k[viol]
    y[k == k[viol]] <- (y[viol] * w[viol] + y[(viol - 1)] * w[(viol - 1)])/(w[viol] + w[(viol - 1)])
    w[k == k[viol]] <- w[viol] + w[(viol - 1)]
    diff <- y - c(0, y[ - length(y)])
  }
#RESULTS:
#1.the level sets
#2.the isotonic regression of y
#3.the new weights
  return(level.sets = k, isotonic = y, isotonic.weights = w)}

```

REFERENCES

¹Ulm, K., Nonparametric Analysis of Dose-Response Relationships ,Annals of the New York academy of Sciences 895, 223-231 (1999)

²Ulm, K., Dannegger, F., Becker, U., Tests for trends in binary response, Discussion paper 115, SFB 386 (1998)

³Küchenhoff, H., Ulm, K., Comparison of Statistical Methods for assessing Threshold Limiting Values in Occupational Epidemiology, Computational Statistics, 12:249-264 (1997)

⁴Barlow, R.E., Bartholomew, D.J., Bremner, J.M., Brunk, H.D., Satatistical inference under Order Restrictions, Wiley, London, U.K., (1972)

⁵Venables, W., Ripley, B., Modern Applied Statistics with S-Plus, Springer-Verlag, (1994)

⁶Chambers, J., Programming with data, USA ,(1998)

⁷Bacchetti P., Additive isotonic models, Journal of the American Statistical Association,84:289-294 (1989)

⁸Morton-Jones, T., Diggle , P., Parker, L., Additive isotonic regression models in epidemiology ,19:849-859 (2000)

⁹Robertson, T., Wright, F. T. and Dykstra, R. L. Order Restricted Statistical Inference, Wiley, New York, USA,(1988)

¹⁰Schell, M.J., Singh, B., The reduced Monotonic Regression Method, J.Am.Stat.Assoc.,92:128-135 (1997)

¹¹Edgington, E., Randomization tests, Dekker, New York,USA,(1980)