

THE LINEAR INTEGRATION MODEL OF THE SIZE-WEIGHT ILLUSION: ESTIMATING ITS PARAMETERS BY OPTIMAL LINEAR SEPARATION

Michele Vicovaro, Luigi Burigana

University of Padova, Department of General Psychology
michele.vicovaro@studenti.unipd.it luigi.burigana@unipd.it

Abstract

Given a set of two-valued points in the plane (some have value 0, the others value 1), to determine an optimal linear separation of the set means to specify a straight line across the plane so that 0-points and 1-points mostly lie on opposite sides of it. The essentials of a computational procedure to this effect are presented, and the procedure is then applied to the data of an experiment on the size-weight illusion (method of constant stimuli, two stimulus dimensions, 37 cylindrical objects as stimuli, 20 participants) as a means of estimating the parameters of a linear integration model of the illusion.

The size-weight illusion (SWI) takes place when on lifting two objects of equal physical weight but different size we feel the smaller object as being heavier than the larger one. The illusion is clearly due to an influence of size on perceived heaviness. This phenomenon has been known in the psychological literature since the late nineteenth century, thanks to Charpentier (1891; cf. Murray, Ellis, Bandomir, & Ross, 1999).

Several models of heaviness perception have been devised to describe and possibly explain the SWI. On the whole, they can be classified into two main groups: sensory and cognitive models. As examples of the former group we mention the ‘density model’ (Ross & DiLollo, 1970), the ‘information integration model’ (Anderson, 1970, 1981), and a model following the ‘ecological approach’ to perception (Amazeen & Turvey, 1996). As an example of the latter group the ‘expectancy model’ (Nakatani, 1985) may be considered. In this study we are only interested in the sensory models, more specifically in the linear integration model, which we choose as a guide for an experiment on the SWI.

The linear integration model appears to fit experimental data quite well, at least so far as restricted ranges of variation of the involved variables are considered (Anderson, 1970; Masin & Crestoni, 1988; Ellis & Lederman, 1993). The model rests on the assumption that perceived heaviness (in the situation of SWI) is the final result of an additive integration psychological process in which weight information and size information are combined together. The simplest way of expressing this assumption is by the following equation:

$$W^* = \omega W + \sigma S \tag{1}$$

in which W and S are the physical weight and size of the object to be judged (attributes of the stimulus), W^* is the perceived weight of the object (a perceptual or ‘covert’ response by the subject), and ω and σ are (unknown) coefficients characterizing the part played by stimulus variables W and S in conditioning the response variable W^* . The circumstance that W^* depends in the increasing mode on W but in the decreasing mode on S – which typifies the SWI – means that $\omega > 0$ and $\sigma < 0$. In the context defined by (1), a basic scientific task is to obtain plausible estimates of coefficients ω and σ , which are the main parameters of the model. In this paper we present a new way of accomplishing this task, and illustrate it on the data of an experiment on SWI.

A method of combinatorial data analysis: finding optimal linear separators

Our method typically applies to a data-structure $(X,Y)=((W,S),Y)$ formed of a list $X=(x_1,\dots,x_m)=((w_1,s_1),\dots,(w_m,s_m))$ of pairs of real numbers and a corresponding list $Y=(y_1,\dots,y_m)$ of binary numbers (values 0 and 1). A research situation rendering such data may be an experiment in which m different stimuli are used, of which two special attributes W and S are considered (in our experiment, W and S are the physical weight and size of objects), and the participant is asked to react to each stimulus by choosing one of two alternative responses (Y is the list of ‘overt’ responses to the stimuli). Part $X=(W,S)$ (the *point set*) can be represented as a pattern of points in the plane (attribute W on the abscissas, attribute S on the ordinates), and accordingly part Y (the *valuation*) can be represented by depicting those points in two different ‘colors’. For example, the pattern in Figure 1 represents this data-structure:

$$X=((4,8),(12,10),(8,12),(6,4),(8,9),(14,6),(8,6))$$

$$Y=(0,1,0,0,1,1,0).$$

Based on Y , the point set X splits into subset X_0 of 0-points and subset X_1 of 1-points; in the example, $X_0=\{x_1,x_3,x_4,x_7\}$ and $X_1=\{x_2,x_5,x_6\}$.

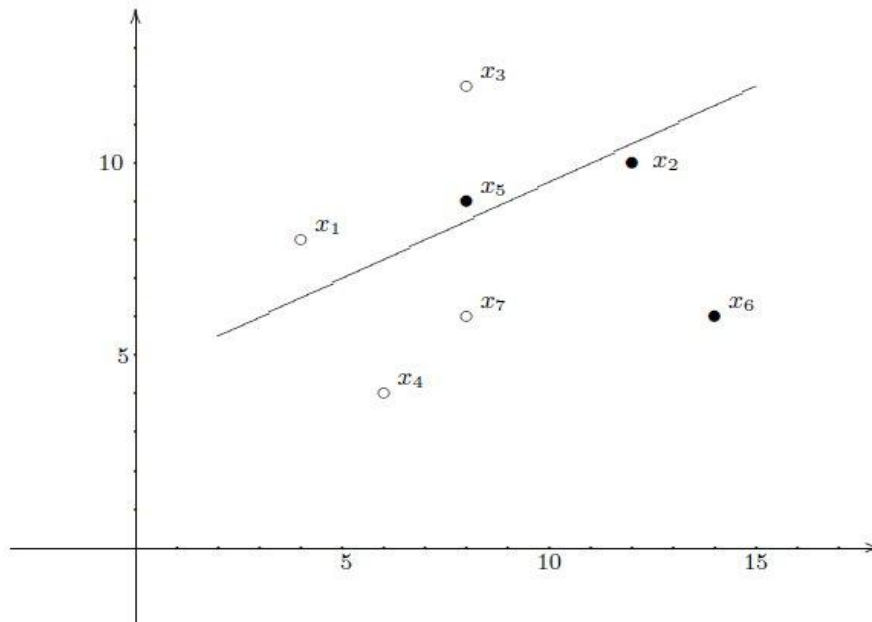


Figure 1. Example of a binary data-structure. Filled and unfilled circles represent 1-points and 0-points, respectively.

A partition $\{P,P'\}$ of point set X (i.e., a pair of non-empty, disjoint, and exhaustive subsets of X) is said to be *linear* if a (straight) line can be drawn across the plane so that points in P and in P' lie on opposite sides of it. Such a line qualifies as a *separator* for the partition. As a line on the (W,S) -plane, a separator is specified by an equation of this form:

$$c_1w+c_2s=c_3 \tag{2}$$

where c_1 , c_2 , and c_3 are suitable coefficients. For example, $\{\{x_1,x_3,x_5\},\{x_2,x_4,x_6,x_7\}\}$ is a linear partition of the point set in Figure 1, and triple $(1,-2,-9)$ specifies a separator of it. On the contrary, $\{\{x_1,x_2,x_3,x_7\},\{x_4,x_5,x_6\}\}$ is not a linear partition of that point set – no linear separator is possible for it.

Any (linear) partition $\{P, P'\}$ of point set X can be compared with the partition $\{X_0, X_1\}$ inherent in the data. More precisely, this quantity can be computed:

$$d(P, P') = \text{abs}(|P \cap X_1|/|P| - |P' \cap X_1|/|P'|) \quad (3)$$

where ‘abs’ means ‘absolute value’, $|P|$ is the cardinality of set P , and similarly for the other sets involved. We refer to (3) as the *discriminating power* of $\{P, P'\}$ (relative to the assigned data-structure). It measures how well $\{P, P'\}$ fits $\{X_0, X_1\}$, that is, how well any linear separator of $\{P, P'\}$ separates 0-points from 1-points in the plane. Possible values of (3) are in the interval $[0, 1]$; maximum 1 is reached when $P=X_1$ or $P'=X_1$, which means that partitions $\{P, P'\}$ and $\{X_0, X_1\}$ are the same (so that 0-points and 1-points in the data are linearly separable); minimum 0 is reached when both partitions are independent of each other. For example, linear partition $\{P, P'\} = \{\{x_1, x_3, x_5\}, \{x_2, x_4, x_6, x_7\}\}$ has discriminating power $d(P, P') = \text{abs}(1/3 - 2/4) = 1/6$ relative to the partition $\{X_0, X_1\} = \{\{x_1, x_3, x_4, x_7\}, \{x_2, x_5, x_6\}\}$ which is inherent in the data represented in Figure 1.

In these terms and under these conditions, two tasks of combinatorial data analysis may be considered. (i) Relative to a data-structure (X, Y) , find (all) *optimal linear partitions*, that is, partitions of point set X that are linear and maximize the discriminating power, as defined by (3). Such partitions are those which, while complying with the linearity condition, best fit the dichotomy (X_0, X_1) in the data. (ii) For each optimal linear partition, find a representative linear separator (an *optimal linear separator* for the data), that is, find a triple of coefficients (c_1, c_2, c_3) that, through (2), specifies a line separating the parts of the partition. These two tasks are potentially relevant for some problems of psychological research. Specifically, there are research situations in which to be able to perform those tasks means to be able to estimate the parameters of a model linking a binary response (described by Y) to a pair of properties of the stimulus (described by X). A situation of this kind forms the subject matter of our present study, and is detailed in the paragraphs to follow. On the computational side, we have constructed a program in R language (called OSL for ‘Optimal Separating Lines’) for performing the abovementioned tasks. The mathematical background of the program is defined in Burigana (2006).

A way to apply the combinatorial method to the size-weight illusion

A kind of experimental data which may be analyzed using ‘optimal linear separation’ is that obtainable by the method of ‘constant stimuli’ (the CS method), when precisely two attributes of the stimuli are considered. In each single trial of a CS experiment, the subject is presented with a ‘comparison stimulus’ o_i and a ‘standard stimulus’ o and is asked to judge whether or not the former exceeds the latter in a specified perceptual attribute (cf. Torgerson, 1958, § 7.6). For example, in a CS experiment on SWI, stimuli o_i and o would be two objects whose physical weight W and physical size S are considered, and the subject is asked to judge whether or not the former exceeds the latter in apparent weight W^* . If θ is the apparent weight of standard stimulus o , $W^*(o_i)$ is the apparent weight of comparison stimulus o_i , Equation (1) is presumed concerning the dependence of the apparent weight on physical attributes $W(o_i)$ and $S(o_i)$, and response $Y(o_i)$ in comparing o_i with o is coded as 1 or 0 depending on whether or not o_i is judged to exceed o in apparent weight, then this psychophysical relation is implied:

$$Y(o_i) = 0 \text{ or } 1 \text{ depending on whether } \omega W(o_i) + \sigma S(o_i) \leq \text{ or } > \theta. \quad (4)$$

This formula describes how the overt response Y depends on physical attributes W and S of the comparison stimulus (through the mediation of the covert response W^* , which is related to W and S by (1), according to the linear integration model).

The complete schedule for a subject in a CS experiment is a series of m trials in which the subject separately judges m different comparison stimuli o_1, \dots, o_m in relation to standard stimulus o , which is the same across the trials. Thus, the set of data relating to a single subject in the experiment may be organized in this form:

$$(X, Y) = (((w_1, s_1), \dots, (w_m, s_m)), (y_1, \dots, y_m))$$

in which, for $i=1, \dots, m$, component $(w_i, s_i) = (W(o_i), S(o_i))$ is the pair of physical attributes of comparison stimulus o_i (e.g., physical weight and size, in an experiment on SWI), and component $y_i = Y(o_i)$ is the binary answer rendered by the subject when comparing o_i with standard stimulus o . It is directly seen that, if Model (4) holds true (i.e., an ‘ideal subject’ really responds according to rule (4)), then partition $\{X_0, X_1\}$ (i.e., the division of stimulus set X into two parts, depending on the binary answers of such an ideal subject) must be linear, and triple (ω, σ, θ) of parameters in the model specifies through (2) a linear separator for that partition. For this very reason, if $Y = (y_1, \dots, y_m)$ is the sequence of binary responses actually produced by a ‘real subject’ in the experiment, and (c_1, c_2, c_3) is a triple found (by computation) as specifying an optimal linear separator for data-structure (X, Y) , then that triple may be taken as a plausible estimate of parameters (ω, σ, θ) in the model.

Two notes are added on the possible use of linear separation on the data from a CS experiment. (i) In each trial of such an experiment, three (rather than two) alternative answers might be allowed, which are ‘ o_i is more than o ’, ‘ o_i is less than o ’, and ‘ o_i does not differ from o ’. Such answers might be coded by numbers 1, 0, and 0.5, so that the complete list $Y = (y_1, \dots, y_m)$ of responses from a subject would be a sample from $\{0, 0.5, 1\}$ (rather than from $\{0, 1\}$). Our program in R language can also be applied to a data-structure (X, Y) with such a ternary valuation, as the discriminating power of any linear partition $\{P, P'\}$ of X is in fact computed by this formula:

$$d(P, P') = \text{abs}(\sum \{y_i: x_i \in P\} / |P| - \sum \{y_i: x_i \in P'\} / |P'|). \quad (5)$$

It is readily seen that when Y is a binary valuation then (5) is the same as (3). (ii) A CS experiment is ordinarily run on a sample of n subjects, each of which produces a binary (or ternary) valuation $Y_j = (y_{1,j}, \dots, y_{m,j})$ on the same set $X = (x_1, \dots, x_m)$ of stimuli. Based on the individual lists of answers Y_1, \dots, Y_n , a mean valuation $Y = (y_1, \dots, y_m)$ (with values in $[0, 1]$) can be computed, by applying this rule, for all $i=1, \dots, m$:

$$y_i = (y_{i,1} + \dots + y_{i,n}) / n. \quad (6)$$

When associated with the point set X , mean valuation Y amounts to an ‘empirical psychometric function’ on the plane. We point out that our program in R language is also able to deal with a data-structure (X, Y) of this kind, simply because the discriminating powers are computed by (5), which makes sense whenever Y is a list of numbers. The results of the computation may be interpreted in a way similar to that explained above (i.e., as estimates of parameters (ω, σ, θ)), but within the frame of a more complex model, which is expressed by this equation:

$$\text{Prob}(Y=1) = F(\omega W + \sigma S - \theta) \quad (7)$$

where F is a suitable cumulative probability function symmetric around zero (e.g., a normal distribution). This ‘probabilistic’ model derives from Model (4) (which is ‘deterministic’ in character) when a random component Z is presumed to participate in the linear integration process, so that $W^* = \omega W + \sigma S + Z$.

The experiment

Stimuli and procedure

We performed an experiment on the SWI applying the method of constant stimuli. The ‘comparison stimuli’ were 37 aluminum cylinders (painted black) with the same diameter (85 mm) and differing from one another in length (hence in volume) and/or in physical weight. The set of stimuli $X=\{x_1,\dots,x_{37}\}$ may be represented as a pattern of points in the plane, the coordinates of which are the physical weight in grams (the abscissas) and the length in millimeters (the ordinates). The pattern we used is shown in Figure 2. Stimulus $x_{19}=(210\text{ g}, 150\text{ mm})$, which is the center of the pattern, also served as the ‘standard stimulus’.

The participant sat at a table of height 50 cm from the floor. At each trial two cylinders, one ‘comparison’ and one ‘standard’, were placed on the table in front of the participant. A cardboard screen was placed on the table, so that the operations of selecting and positioning the stimuli (by the experimenter) were not visible to the participant. He or she was told first to take the cylinder on his or her right with the favorite hand, lift and move it in any way in order to appreciate the weight, then to lay down the cylinder in the starting position on the table, and then to perform the same operation with the cylinder on the left. In each trial the participant had to judge which of the two objects appeared heavier (the answer of indifference was allowed). The participant was told that if uncertain about the appropriate answer he or she could repeat the weighing of both objects once more. The standard stimulus was the same in all the trials; the 37 comparison stimuli were selected in random order, varying across participants. The left-right position of the standard and comparison stimuli within trials was reversed from the first to the second half of the series of participants.

Participants were 20 volunteers aged from 20 to 60 years. They were 14 females and 6 males. All but one were right-handed. None had had prior experience of judging apparent heaviness in psychophysical experiments.

Results and analysis

The data-set resulting from our experiment is a collection (Y_1,\dots,Y_{20}) of $n=20$ ternary valuations, as many as the number of subjects participating in the experiment. Each valuation Y_j (for $j=1,\dots,20$) is a list $(y_{1,j},\dots,y_{37,j})$ of $m=37$ ternary values specifying the answers rendered by the j -th participant to stimuli (x_1,\dots,x_{37}) , these being labeled as shown in Figure 2. For example, the fifth participant produced the following list of answers:

$$Y_5=(0,0,0,0,0,0,0,0.5,0,0,0,0.5,0.5,0,0,1,1, \\ 0.5,0.5,0,1,1,0.5,1,1,1,1,1,1,1,1,1,1,1,1) \quad (8)$$

This is represented in Figure 2 by depicting each stimulus-point as a filled, crossed, or unfilled circle depending on whether the corresponding answer was 1 (comparison appears heavier than standard), 0.5 (comparison and standard do not differ), or 0 (comparison appears lighter than standard). Based on the 20 individual valuations, a mean valuation $Y=(y_1,\dots,y_{37})$ on the same pattern of stimuli can be computed, by applying (6). The mean valuation we obtained is as follows:

$$Y=(0,0,0,0,0,0,0,0.1,0.125,0,0.025,0,0.6,0.275,0.1,0.025,0.85,0.8, \\ 0.5,0.425,0.2,0.95,0.975,0.9,0.85,1,1,1,0.925,0.85,1,1,1,0.975,1,1,0.975) \quad (9)$$

This amounts to an empirical psychometric function on the pattern of 37 points in the plane.

Our OSL (Optimal Separating Lines) computational procedure can be applied both to the collective data-structure (X,Y) in which X is the set of 37 stimulus-points in the plane and Y is the mean valuation, and to each individual data-structure (X,Y_j) in which Y_j is the

valuation yielded by one single participant (possibly transformed from ternary to binary form). By the former application we may come to know the basic features of the linear integration model of SWI, as they are shared within the sample of participants. By the latter we may ascertain specific properties of the same model, presumably peculiar to any single participant.

In practice, by applying OSL to the collective data-structure (X, Y) with Y described by (9), these results are obtained:

$$\begin{aligned}
 MD &= 0.8325 \\
 NP &= 2 \\
 OP &= (13, 17, 18, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37) \\
 &\quad (13, 17, 18, 19, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37) \\
 MS &= (0.947, -0.319, 133.678) \\
 &\quad (0.916, -0.399, 122.783).
 \end{aligned}$$

They mean that the maximum discriminating power (MD) of linear partitions for that structure is 0.8325 (a good result concerning linearity, considering that 1 is the highest possible value of statistic MD); there are two linear partitions of stimulus set X reaching that maximum ($NP=2$); one of these optimal partitions (OP) is formed of set $\{x_{13}, x_{17}, \dots, x_{19}, x_{20}, x_{22}, \dots, x_{37}\}$ and its complement in X , and the other of set $\{x_{13}, x_{17}, \dots, x_{19}, x_{22}, \dots, x_{37}\}$ and its complement in X (thus, both partitions only differ in the membership of stimulus x_{20}); triple $(c_1, c_2, c_3) = (0.947, -0.319, 133.678)$ specifies (through (2)) a median linear separator (MS) for the former of these partitions, and triple $(0.916, -0.399, 122.783)$ specifies a median linear separator for the latter. These two triples are alternative estimates of parameters (ω, σ, θ) in the ‘deterministic’ integration Model (4) (or the ‘probabilistic’ integration Model (7)) for the size-weight illusion, based on the results of our constant stimuli experiment.

To illustrate the use of our method on a binary data-structure, let us consider the following valuation, which is derived from (8) (the response pattern of the fifth subject in the sample) by replacing 0.5 by 0:

$$Y^{\circ}_5 = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1).$$

Then, by applying OSL to data-structure (X, Y°_5) , these results are obtained:

$$\begin{aligned}
 MD &= 0.944 \\
 NP &= 1 \\
 OP &= (17, 18, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37) \\
 MS &= (0.958, -0.284, 163.511).
 \end{aligned}$$

On the presumed data (which are from a single subject) the maximum discriminating power is quite high ($MD=0.944$) and there is only one linear partition attaining it, that formed of set $\{x_{17}, x_{18}, x_{22}, \dots, x_{37}\}$ and its complement in X . Triple $(c_1, c_2, c_3) = (0.958, -0.284, 163.511)$ specifies a linear separator for that partition, which is represented in Figure 2 (the oblique line).

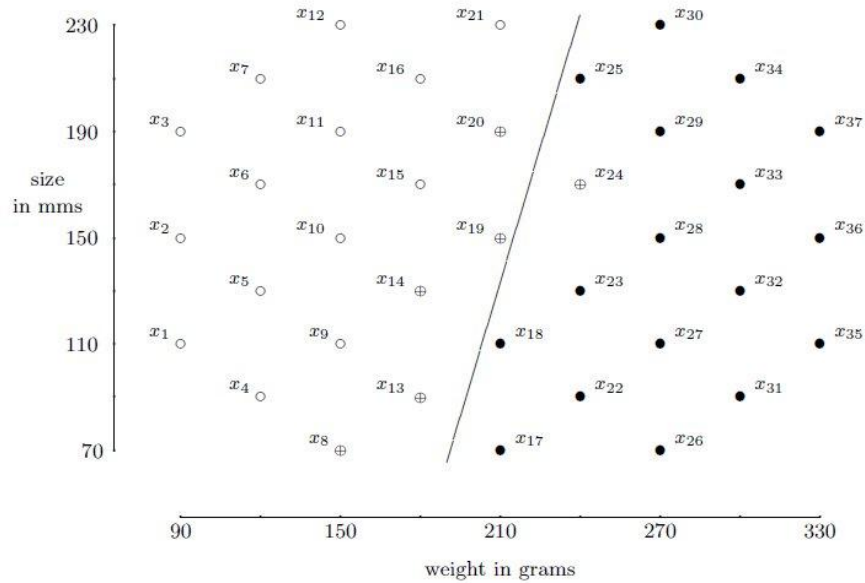


Figure 2. Pattern of responses of a participant in the CS experiment on SWI. The horizontal axis shows the physical weight of cylinders (in grams) and the vertical shows their height (in millimeters). Filled, crossed, and unfilled circles stand for responses in which the comparison cylinder is judged to be heavier than, indifferent from, or lighter than the standard cylinder. The oblique line is the median separator found by OSL (on the binary valuation obtained by replacing 0.5 by 0).

Discussion and conclusion

Two aspects of the above results especially deserve mention. One is the value of statistic MD (maximum discriminating power), which is high when computed on the collective data-structure (0.8325) and on most of the individual data-structures (the range of possible values of MD is $[0,1]$). This aspect signifies that *linear* partitions – the best of them – fit quite well the pattern of responses in our experiment, and in turn this is evidence in favor of the *linear* integration model when applied to the SWI in the frame of a CS experiment (Equations (1), (4), and (7)). The other aspect is the opposite signs of coefficients c_1 and c_2 , which estimate parameters ω and σ in the linear model: c_1 is positive and c_2 negative. This precisely match the substance of the SWI, which lies in the fact that the apparent weight of an object jointly depends on the physical weight and size, in the *increasing* way on the former but the *decreasing* way on the latter.

Our method computes not only the signs of estimates for parameters ω and σ (and threshold θ), but also specific numeric values of them (e.g., ‘ $Y=1$ iff $0.947 \times W - 0.319 \times S > 133.678$ ’ is a numeric form taken on by Model (4) based on the data from the experiment). Of course, those numeric values depend on the coordinates of the stimulus-points in the plane, and in turn such coordinates depend on the units used in measuring the relevant stimulus attributes (gram for weight and millimeter for size, in our case). It is easily seen that, if the units of measurement were changed, then the parameter estimates would also change in their values. There are, however, some valuable invariance properties of the proposed method: if measurements are on interval or ratio scales, then the maximum discriminating power and optimal linear partitions do not change when changing the scales, and there is a simple algebraic relationship between coefficients of linear separators computed (on a fixed set of stimuli) using different scales.

We had two aims with this contribution: to present the essentials of a method for optimally dividing a set of points in the plane by a straight line, and to illustrate its use as a method of data analysis by applying it to the data from an experiment on the size-weight illusion. There are standard methods of statistics that are basically directed towards other goals than linear separation but may nevertheless be adapted to the linear separation task. We mention, in particular, ‘canonical discriminant analysis’ (Klecka, 1981) and ‘multiple logistic regression’ (Hilbe, 2009). Our method, however, has some special merits. We conclude by mentioning three of them. (i) It yields a simple and easily interpretable measure of the consistency of the data with the linear separability condition. The measure is statistic *MD* (maximum discriminating power). (ii) It is an exploratory and exhaustive searching procedure which renders all optimal linear partitions and corresponding linear separators. (iii) It acts in simple geometrical and combinatorial terms. It does not involve any assumption concerning probability distributions, and in this regard it qualifies as a completely ‘distribution-free’ method of combinatorial data analysis.

References

- Amazeen, E.L., & Turvey, M.T. (1996). Weight perception and the haptic size-weight illusion are functions of the inertia tensor. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 213-232.
- Anderson, N.H. (1970). Averaging model applied to the size-weight illusion. *Perception & Psychophysics*, 8, 1-4.
- Anderson, N.H. (1981). *Foundations of information integration theory*. New York: Academic Press.
- Burigana, L. (2006). Optimal linear bipartitions of two-colour sets of points in low dimensional Euclidean spaces. *TPM – Testing, Psychometrics, Methodology in Applied Psychology*, 13, 317-342.
- Charpentier, A. (1891). Analyse expérimentale de quelques éléments de la sensation de poids. *Archives de Physiologie Normale et Pathologique*, 3, 122-135.
- Cross, D.V., & Rotkin, L. (1975). The relation between size and apparent heaviness. *Perception & Psychophysics*, 18, 79-87.
- Ellis, R.R., & Lederman, S.J. (1993). The role of haptic versus visual volume cues in the size-weight illusion. *Perception & Psychophysics*, 53, 315-324.
- Hilbe, J.W. (2009). *Logistic regression models*. Boca Raton, FL: CRC Press.
- Klecka, W.R. (1980). *Discriminant analysis*. Beverly Hills, CA: Sage Publications.
- Masin, S.C., & Crestoni, L. (1988). Experimental demonstration of the sensory basis of the size-weight illusion. *Perception & Psychophysics*, 44, 309-312.
- Murray, D.J., Ellis, R.R., Bandomir, C.A., & Ross, H.E. (1999). Charpentier (1891) on the size-weight illusion. *Perception & Psychophysics*, 61, 1681-1685.
- Nakatani, K. (1985). Application of the method of fixed sets to the size-weight illusion. *Psychological Research*, 47, 223-233.
- Ross, J., & DiLollo, V. (1970). Differences in heaviness in relation to density and weight. *Perception & Psychophysics*, 7, 161-162.
- Torgerson, W.S. (1958). *Theory and methods of scaling*. New York: Wiley.