How to transform a batch of simple indicators to make up a unique one?[1]

Come trasformare una batteria di indicatori semplici in un solo indicatore?

Fabio Aiello
Dipartimento di Metodi Quantitativi per le Scienze Umane, Università di Palermo,
fabiello@inwind.it

Massimo Attanasio
Dipartimento di Metodi Quantitativi per le Scienze Umane, Università di Palermo,
attana@unipa.it

In questo lavoro viene analizzato il processo di costruzione di un indicatore composto, in funzione di una batteria di indicatori semplici. Tale processo viene scomposto in due parti: la prima consiste nell'individuare delle funzioni che permettano di trasformare i dati grezzi (gli indicatori semplici) in dati omogenei, la seconda nell'individuare una funzione che sulla base dei primi produca una misura dell'indicatore composto. Vengono forniti, da un lato, degli strumenti matematici e statistici delle trasformazioni più usate nelle applicazioni (con particolare riguardo alle trasformazioni lineari) e, da un altro lato, diversi esempi di indicatori composti ottenuti tramite funzioni additive e non.

Keywords: composite indicator, simple indicator, transformation, link function.

1. Introduction

The main purpose of this work is to explore simple mathematical and statistical mechanisms to build and to investigate multiple component (item) *scales* or *composite indicators*. Composite indicators have to measure a complex and underlying concept, usually named *construct*, which is not directly measurable, so it is broken into measurable components, dimensions or items. Multiple component scales are usually named in clinical trials, psychometrics, medicine, etc. while composite indicators are named in social and educational sciences, in environmental setting, in scientometrics, etc.. The literature on this topic is vast and interesting. Investigation has been carried out according to different criteria: the type of scales, the field of application together with the scientific background of the author, the nature and the structure of the data, and the aim of the study (Fayers & Hand, 2002).
Another approach, based on functional analysis and usually named *dimensional analysis*, represents a breakthrough in the theory of scales and indicators (Aczél, 1987; Luce *et al.*, 1990). They provide a list of theorems which give the mathematical conditions to construct different types of scales. Our approach is much simpler: it

[1] The present paper is financially supported by MURST Funded Research (ex. 60%, 1999) "Standard di vita a Palermo", V. Capursi; MURST Funded Research (ex. 60%, 1999) "Costruzione di indicatori composti per la valutazione dei servizi", M Attanasio; Research Project (2000) "Una misura della qualità nella gestione dei servizi", V. Capursi.

consists on the description of some actual paths to combine the items (simple indicators) into a scale (composite indicator), in the attempt to express the final result in a meaningful way.

The paper is divided into two parts to better explain how we shall proceed to pursue our proposal. First, we deal with some issues concerning data transforming, confining our attention to transformations *T's* aiming at the comparability of different data sets. Second, we deal with *the process of reconstruction* of single indicators to composite indicator, through a *link* function *f*. Therefore, the first function *T* allows to obtain *dimensionless* data that, through the second function *f*, can be put together into one thing, which is the measure of the *construct* or of the *latent variable* **X**.

To explicit the purpose of this work we shall attempt to answer some questions: why and where (in what cases) these *T's* are widely used? what properties do statistical transformation must have? (to be friendly); what distinguishes linear transformations from non linear transformations? which statistical properties should be valued most with regard to the aim of the study? what are the most common mathematical functions *f* that *recompose* the transformed data into something relevant in practical usage? what is the relationship between the transformation *T* and the link function *f*? why has the class of additive function *f* been so widely used? when the class of non additive functions *f* appropriate? In other words, the relationship between *T* and *f* can be written:

$$\boldsymbol{X} = f\left[T_1(x_1),\, T_2(x_2),\, \ldots,\, T_k(x_k)\right] \tag{1}$$

Where x_i is the *ith simple indicator* or *item*, T_i is the *ith* transformation and *f* is the *link* function. Section 2 illustrates some general issues about transformations, Section 3 specific transformation issues to compare different batches of data; Section 4 linear transformations; Section 5 non linear transformations and Section 6 is about link functions.

2. General Issues of Data Transformations

There are many reasons why we might want to transform (or re-express) data. Actually data transforming is almost necessary whenever we are in presence of statistical data, and the objectives of such operation are usually more than one. Even data transforming is almost always present in any statistical analysis, there are no books, to our knowledge, that are entirely devoted to that topic. Several books assign a special chapter to transforming data, and, on the other hand there is an enormous number of papers about transformations which usually describe special cases, mostly related to reconduct a set of data to the assumptions of the linear model. (Kendall & Stuart, Ch.37, 1983). In applied works, especially, small paragraphs are usually devoted to transformation issues and data are often transformed automatically, just repeating what has already been done in that field. Log transformation is sometimes applied without any explanation.

Actually, from a statistical point of view transformation itself is a twofold concept, concerning both mathematics and statistics. Transformation itself, as a mathematical operation, is usually put aside by statisticians, so, for instance, the *Encyclopaedia of Statistical Sciences*, item *Transformations*, says "The general effect of a transformation depends on the shape of its plotted curve or a graph. It is this curve, rather than the mathematical formula, that has central interest". Here the idea is to design a new model or a new data set that has important aspects of the original ones and satisfies all the assumptions for the new model. Therefore emphasis is most of the times devoted to the effects of transformation rather than to the relationship between transformed data and original data. In fact a central issue in the statistical literature is addressed to determine the correct distributional form to apply a specific statistical method, so the statistical literature addresses the benefits of transforming with regard to statistical modelling, neglecting some relevant aspects.

The book *Understanding Robust and Exploratory Data Analysis* (Hoaglin *et al.*, 1983) has been a breakthrough because it carries out a new point of view. Following a robust approach, they provide a collection of papers dealing with data transforming. So transformations are not anymore confined to the problem of linearizing or of removing heteroscedasticity, either in the ANOVA setting or in the time series analysis, but they deal with several aspects of data analysis. The authors provide *Robust and Exploratory Data Analysis* methods and tools to: enhance interpretability, get symmetry, get stable spread, give a better graphical representation, and, generally speaking, obtain simple data structure.

To compare sets of data consisting of amounts or counts, *T's* ought to have the following characteristics:

(i) *Smoothness.* Actually we do not refer to the usual meaning of smoothness in calculus, (*i.e.* they have derivatives of all orders), but here the meaning is slightly different. The *T* functions ought to: a) be elementary and well known; b) have a widespread usage in practice; c) preserve the order of any batch of data (so percentiles are transformed to percentiles);

(ii) *Computational ease*. Their usage needs just some elementary calculus;

(iii) *Comparability to the original data*. They ought to re-express a set of data in a nearly comparable way to the original data set;

(iv) *Resistance*. It seems appropriate to refer to resistance instead of robustness because transformations do not involve a breakdown of modelling assumptions. An estimator is defined *resistant* if *it is affected to only a limited extent either by a small number of gross errors or by any number of small rounding and grouping errors*, likewise a linear transformation *T* could be defined *resistant* if it is affected by only a limited extent by a small number of outlier observations (Hoaglin *et al.,* Ch. 11, 1983). So according to this definition, the formula *T* can be defined resistant only if it contains resistant parameters. In practice, it is desirable that the operation of transformation does not let that a strong asymmetry or outliers have effect on a big bulk of the new data set.

3. Transformations to construct Composite Indicators

The aim of this paper is to provide a content which allows to compare the most used transformations in practical applications according to their statistical and mathematical properties. Therefore we need to introduce briefly some issues:

- definitions of transformations and mathematical properties;
- characteristics of x_i (direction, units of measure, magnitude);

– other features (geometry, scales of measurements, statistical Properties).

Definitions of Transformations. In general, transforming means to change a set of objects, numbers or geometric entities, into an other set according to some rule or law. There are some transformations which work on algebraic objects by a one-to-one function between two sets. A transformation of the batch x_1, x_2, …, x_k, is a function T that replaces each x_i by new value $T(x_i)$ so that the transformed values of the batch are $T(x_1)$, …, $T(x_k)$. T is usually elementary, strictly increasing, continuous and differentiable.

Sometimes in the statistical literature transformation and standardisation are used as synonymous. Actually standardisation techniques or methods are used to adjust for the effects of some factors as age or sex, when the objective is to compare populations or samples with different factor structures (Inskip, 1998).

Characteristics of x_i's. First, it is important to stress each variable x_k is measured with different *direction*, *magnitude* and *units of measure*, where: ***a.*** *direction* concerns the algebraic sign of the *i-th* variable versus the latent variable X: if high values of x yield high values in x the direction is *concordant*; while, if high values of x yield low values of x the direction is *discordant*; ***b.*** *magnitude* of x is equal to m, if $x = a \cdot 10^m$; ***c.*** *unit of measure* is defined as a special fixed and conventional quantity.

Data comparison must be done taking into account the group structure that a transformation involves and the statistical issues derived by that operation. Therefore our goal is to obtain T's that are not relied to their original direction, magnitude and units of measure, i.e. they have to be dimensionless quantities. A number is dimensionless if it is just a number, not just as a result of same measuring process applied to some type of physical quantity.

Other Features:

(i) *Geometry*. From a geometric point of view, a transformation in which data vectors are transformed in a fixed coordinate system is called *alibi transformation*. In contrast, a transformation in which the coordinate system has changed, leaving vectors in the original coordinate system *fixed* while changing their representation in the new coordinate system is called *alias transformation*. In geometry there are several coordinate plane systems (oblique, Cartesian or rectangular, polar, elliptic cylindrical, and finally, parabolic). The most popular are Cartesian and polar coordinates. The choice of the coordinate system depends on the nature of the data, on the field of application and on the aim of the study. They determine the "best" geometrical representation and in this context, for instance, moving from a coordinate system to another one is a graphical appropriate way of re-expressing data.

In this paper we deal just with first family of transformations. These belong to the *affine transformation family*, which preserve the *collinearity* and the *distance ratios*.

(ii) *Scales of measurement*. Those one-to-one T's have also a very interesting interpretation in terms of group structure and scales of measurement as suggested by Stevens (1946). He reports a *Classification of Scales of Measurements*, in which there is an interesting linkage between: scales, basic empirical operations, and mathematical group structure. Instead, dimensional analysis develops the latter approach to a larger extent (Luce *et al.*, 1990).

(iii) *Statistical Properties*. Therefore the selected T's are just those handy and capable to address practical data analysis problem. We chose a list of mathematical and statistical properties in order to describe T's: ***a.*** units and *scale of measurements*; ***b.***

main *statistical parameters* (mean, variance, range); ***c.*** reduction *of variability* compared to the original data; ***d.*** *resistance*, as defined above; ***e.*** *field of application*.
At first the T's can be classified into two families: *linear T*'s (*LTs*) and *non linear T*'s (*NLTs*). In this paper the concern is mostly given to the *LTs*, even if there will be a brief description of the most popular *NLTs*, for purposes of brevity.
LTs permit to change the origin, the scale and the unit of measurements, but they do not change the shape. *LTs* re-express a value x $\{x: x \in \Re^+\}$ in the form:

$$T(\mathrm{x}) = y = a + bx \quad \Rightarrow \quad a, b \in \Re^+ \tag{2}$$

The most important characteristics of a linear transformation is *proportionality*. This is a very important property because it allows to save the same ratio between observations with a different origin (if $a \neq 0$) and scale ($b \neq 0$). In this paper we chose five simple and widely used *LTs*, labelled *LT1*, *LT2*, *LT3*, *LT4*, and *LT5* (see Tables 1 & 2).
Furthermore, we consider just one *non linear* transformation named *Ranking Scoring* (*RS*).

4. Linear Transformations

LT1 and LT2. *LT1* is very common in any field of application because it is easy to be computed and it has a straightforward application and meaning. To divide by the maximum allows to cancel the physical units of the original quantities and forces the results into a shorter interval.
Modifying *LT1* with *LT2* we get a mapping into the easiest [0,1], something attractive for standardization. *LT2* is very often used in applied economics because it is a good way to compare spatial and/or temporal data with a reduction of range. *LT1* and *LT2* determine a re-scaling of data into a shorter interval. Even if proportionality is maintained, *LT1* and *LT2* are not convenient in presence of strong asymmetry or in presence of outliers, because they comprise transformed data proportionally, so they might be very dense if the extreme values are outliers. Therefore *LT1* and *LT2* are not resistant, according to the definition above given, in fact *LT1* is *dominated* by the maximum while *LT2* is *dominated* by the maximum and the minimum.
LT3, LT4 and LT5. The use of normal scores as conventional numbers was first suggested by R.A. Fisher and F. Yates in the *Introduction to their "Statistical Tables*", first published in 1938. They introduce z-transformation, named *Fisher transformation*, to get a more treatable sampling distribution of the linear correlation coefficient.
Standard scores are just a standard deviate (*LT3*) with mean equal to zero and variance to one. These values make *LT3* very popular because of their interpretative ease and because of the comprising of variability. Moreover when the raw data are distributed normally or approximately normal, the z-transformation becomes the *standard normal deviate*. It tells us how far the single raw x_i lies from its mean, measured in standard deviations, something very useful to compare different data set.

Table 1: *Synoptic table of LT1, LT2, and LT3*

Property	*LT1*	*LT2*	*LT3*
$T(x)$	$T(x)=\frac{x}{Max(x)}$ $a=0;$ $b=\frac{1}{Max(x)}$	$T(x)=\frac{x-Min(x)}{Max(x)-Min(x)}$ $a=-\frac{Min(x)}{Max(x)-Min(x)};$ $b=\frac{1}{Max(x)-Min(x)}$	$T(x)=\frac{x-M(x)}{\sqrt{Var(x)}}$ $a=-\frac{M(x)}{\sqrt{Var(x)}};$ $b=\frac{1}{\sqrt{Var(x)}}$
Units of Measurement	Pure number	Pure number	Pure number
Scale of measurement	Ratio	Ratio	Ratio
Range	$\frac{Min(x)}{Max(x)} \le T(x) \le 1$	$0 \le T(x) \le 1$	$-\infty < T(x) < +\infty$
Mean	$Max(x)^{-1}\, M(x)$	$\frac{M(x)-Min(x)}{Max(x)-Min(x)}$	0
Variance	$\frac{Var(x)}{(Max(x))^2}$	$\frac{Var(x)}{(Max(x)-Min(x))^2}$	1
Variability Reduction*	$1-\left(Max^2(x)\right)^{-1}$	$1-\left(Max(x)-Min(x)\right)^{-2}$	$1-Var(x)^{-1}$
Derivative	$(Max(x))^{-1}$	$(Max(x)-Min(x))^{-1}$	$\left(\sqrt{Var(x)}\right)^{-1}$
Skewness**	$\frac{Asy(x)}{Max(x)}$	$\frac{Asy(x)}{Max(x)-Min(x)}$	$Asy(x) * Var(x)^{-1}$

Another useful transformation (*LT4*), based on *z*-score, can be developed when the aim is to relate scores of a given group to the scores of a normative group, with given mean and given standard deviation. The resulting data shall be re-expressed and measured onto the new normative scale, with mean and variance given by normative group. *LT4* is widely used in psychometric score tests. Both transformations, *LT3* and *LT4*, are not very resistant because their computation involves the mean and the standard deviation. These parameters can be sometimes affected by the presence of outliers in the original data set.

Finally, *LT5* is similar to *LT4*, but in this case it uses the median as the location parameter and the *MAD* (*median absolute deviation*) as the scale parameter. The median and the *MAD* overcome the presence of outliers so *LT5* is very resistant.

Table 2: *Synoptic table of LT4, LT5, and RS.*

Property	*LT4*	*LT5*	*RS*
$T(x)$	$y \sim [M(y), Var(y)]$ $T(x) = M(y) + \frac{Var(y) * [x - M(x)]}{Var(x)}$	$T(x) = \frac{x - Med(x)}{MAD(x)}$ $a = -\frac{Med(x)}{MAD(x)}$; $\mathrm{b} = \frac{1}{\mathrm{MAD(x)}}$	*Ranking Scoring*
Units of Measurement	Units of *Y*	*Med*, *MAD* score	No
Scale of measurement	Ratio	Ratio	Metric
Range	Domain of *Y*	$-\infty < T(x) < +\infty$	$1 \leq T(x) \leq N$
Mean	$M(y)$	$\frac{M(x) - Med(x)}{MAD(x)}$	$\frac{(N+1)}{2}$
Variance	$Var(y)$	$Var(x)(MAD(x))^{-2}$	$\frac{N^2 - 1}{12}$
Variability Reduction*	$1 - (Var(y) * Var(x)^{-1}$	$1 - \left(MAD^2(x)\right)^{-1}$	$1 - \frac{Var(x)(N^2 - 1)}{12}$
Derivative	$Var(y) * \sqrt{Var(x)}^{-1}$	$(MAD(x))^{-1}$	Piecewise constant
Skewness**	$\left(Var(y) * \sqrt{Var(x)}^{-1}\right) * Asy(x)$	$Asy(x) * MAD(x)^{-1}$	0

$* \frac{Var(x) - Var[T(x)]}{Var(x)}$; $\quad$ ** $Asy(\mathrm{x}) = M(x) - Med(x)$

5. Non Linear Transformations

There are many reasons why we might want to make a *NLT*. An immediate reason is to linearize data through the logarithm transformation, in this way original data change their shape. As already said, *NLTs* help to obtain either standard statistical assumptions in the linear models or several other issues. For instance, if our goal is to change the units of measure, but also to change the basic scale of measurement, we need to modify the original distributional shape. Power transformations are a solution to alter the shape of the original structure. (Hoaglin *et al.*, Ch. 4 & 8, 1983).
Among non linear transformations there is also the *rank* transformation and the *ranking scoring* transformation (*RST*). The rank order of a set of *N* observations is the order in which they come when arranged according to the characteristic under study. The individual *rank* denotes the position of each one object onto the constituted ranking. A special case of the rank transformation is given by the *percentile*. In practice, the two of them have the same statistical meaning. The latter is easier to be interpretable because it varies between 1 and 100 and does not depend on *N*. The rank is usually treated as a class of monotone score functions that maps metric data to ordinal data.

Ranking scoring is an operation that assigns scores to levels of ordinal variables. It does not treat scores as scaling of ordinal variables, but as values of interval variables. The most frequent application of ranking scoring is to assign scores to several items of a questionnaire. This is almost always done in customer satisfaction surveys. For instance, with four multiple response ordinal items as *not at all*, *a little*, *quite a bit*, *very much*, 1 is assigned to the first category, 2 to the second, 3 to the third and, finally, 4 to the last category. The higher the score the higher is the level of the degree of accordance with the item. Thus, the scores given to each response category are not treated anymore as ordinal but as metric numbers. In this way it is possible to add up them and derive the overall score of each responder over all items and/or the score of each item over all responders. Sometimes these scores are weighted to construct a weighted mean (Prieto, 1996) but variability measures are usually avoided because their interpretation is somewhat difficult.

6. The Link Function

In this section there are several examples coming from statistical *everyday* practice of constructing composite indicators.
<u>*Example 1*</u>. An additive *LF* is utilised by the financial newspaper *Il Sole 24ore* in the survey *Qualità della Vita* on the 103 Italian Provinces. Assuming equal weights and independence between simple indicators, they sum over 36 indicators x_i, using two types of *T*'s, the first is a *LT* while the second is a *NLT*:

$$\boldsymbol{X} = f\left[T_1(x_1), \ldots, T_1(x_{21}), T_2(x_{22}), \ldots, T_2(x_{36})\right]$$

$$T_1(x_i) = \frac{x_i}{Max\{x_i\}}1000 \qquad \text{when } x_i \text{ concordant to } X \qquad (i = 1, \ldots, 21)$$

$$T_2(x_j) = \frac{Min\{x_j\}}{x_j}1000 \qquad \text{when } x_j \text{ discordant to } X \qquad (j = 22, \ldots, 36)$$

so X takes the following form: $X = \sum_{i=1}^{21} T_1(x_i) + \sum_{j=22}^{36} T_2(x_j)$.

Therefore *LF* is given by summing up 21 directly proportional quantities plus 11 inversely proportional quantities. This operation is not appropriate from a mathematical point of view because it produces a result whose mathematical relationship to the original x_i's is not definable. An easy solution to overcome this problem is modifying the *NLT* into a *LT* as:

$$T_2^{'}(x_j) = -\frac{x_j}{Max\{x_i\}}1000 \quad \text{for} \quad j = 22, \ldots, 36.$$

In this way *LF* is an additive function which sums over 36 *LT3's* (Attanasio and Capursi, 1997).

<u>*Example 2*</u>. Instead, several American studies on <u>*Quality of Life*</u> use a procedure that "converts all variables to the same unit of measure and it allows neighbourhood scores to be added to derive an overall or composite score based on multiple variables. Some of the variables used in the analysis were inverse measures of the quality of life, i.e., a high value indicated a low quality of life condition. The signs of the *Z* scores for these variables were reversed before summing scores for several variables to derive an overall or cumulative score for the quality of life" (www.charmeck.org/…/2002 +Quality+of+Life+Study.pdf). In this case *X* takes the following form:

$$X = f(T_1(x_i), T_2(x_j)) \qquad \text{where } T_1(x_i) = \frac{x_i - M(x_i)}{\sigma(x_i)};$$

$$T_2(x_j) = -\frac{x_j - M(x_j)}{\sigma(x_j)}$$

T_1 is used if x_i is concordant to X, while T_2 if x_j is discordant to X.

$$X = \sum_{i=1}^{I} T_1(x_i) + \sum_{j=1}^{J} T_2(x_j).$$

<u>*Example 3*</u>. *DI* (*Discomfort Index*) is an empirical tool used in physics to measure the indoor (dis)comfort combining the air temperature (x_1) and the humidity (x_2). Here the (1) takes the form:

$$X = f(T_1(x_1), T_1(x_2)) \qquad \text{and} \qquad T(x_i) = x_i \qquad i = 1, 2.$$

$$X = x_1 - (0.55*(1 - 0.01x_2)*(x_1 - 14.5)).$$

LF is a polynomial of degree 2 obtained by means of empirical analysis.

<u>*Example 4*</u>. *ROC* (*Receiver Characteristic Curve*) is an empirical tool used in clinical epidemiology to measure the relationship between *sensitivity* (i.e. true positive rate) and *specificity* (i.e. 1 – false positive rate) of a screening test, measured at different levels. By construction, sensitivity and specificity are discordant, because each of them can only be increased at the expense of the other (Fletcher *et al.*, 1982). The (1) takes the following form:

$$X = f(T_1(x_1), T_2(x_2)) \qquad x_1 = \text{true positive rate}; \qquad x_2 = \text{true negative rate}$$

$$T_1(x_1) = x_1 = \text{sensitivity} \qquad T_2(x_2) = 1 - x_2 = \text{specificity}$$

$$AUC\ (Area\ Under\ the\ Curve) = X = \int_0^{100} f(z)dz \qquad f(z) = x_1 = 1 - x_2$$

The *ROC* gives equal weight to sensibility and specificity, even if it is rare to find situations where false positive cases and false negative cases can be valued equally.
Example 5. *Body Mass Index* (*BMI*) is an empirical tool for indicating weight status. It is a medical diagnostic tool: as BMI index increases, the risk for same disease increases. So:

$X = f(T_1(x_1), T_2(x_2))$ where x_1 = *weight in Kg*; x_2 = *height in cm*

$T_1(x_1) = x_1 \quad T_2(x_2) = x_2^2 \qquad X = \frac{x_1}{x_2^2}$

In this case T_1 is linear, T_2 is non linear and *LF* is multiplicative.
Example 6. *Customer Satisfaction* (*CS*) questionnaires usually content questions/items with different number of categories assuming every item has equal importance. For simplicity, they can be ranked in this way:

Item 1:	□ *Yes*	□ *No*		
Rank:	2	1		
Item 2:	□ *very much*	□ *quite a bit*	□ *a little*	□ *not at all*
Rank:	4	3	2	1

If *RST* is utilised then a measure of CS is given by summing over the scores of the two items for each respondent. As usual (1) takes the form:

$X = f(T_1(x_1), T_2(x_2))$ x_1 = *score item 1*; x_2 = *score item 2*

And to eliminate their different magnitudes, it is possible to transform as:

$$T_i(x_i) = \frac{x_i}{Max\{x_i\}} \quad \forall\, i \qquad \text{and} \qquad X = \sum_{i=1}^{2} T_i(x_i)$$

Moreover to obtain total scores in the interval [0, 1] in presence of *k* items, *LF* can be written as an arithmetic mean:

$$X = \frac{\sum_{i=1}^{k} T_i(x_i)}{k}$$

Example 7. Another attractive application of the scale obtained by means of *RST* is the *Formula One World Championship,* where scores (points) are assigned according to the arrival placement for each Grand Prix. The points assignment rule (*PAR*) is rather peculiar, in fact it is neither proportional to the time race, nor to the usual one-to-one step ranking. In addition, *PAR* was changed in the 2003 season to let the championship be more challenging and attractive till the last races (Table 3).

Table 3. *Points by place. Formula One World Champ.*

Place	Points before 2003	Points 2003
1	10	10
2	6	8
3	4	6
4	3	5
5	2	4
6	1	3
7		2
8		1

So, following the usual formula, we get the 2003 total seasonal score, over 16 races:
$X = f(T(x_1), \ldots, T(x_{16}))$ x_i = place *i-th* race; $T(x_i) \Rightarrow$ *assigned points* (Table 3)

$$X = \sum_{i=1}^{16} T(x_i)$$

In this way *PAR*'s might be seen as a weighted *RST*, whose weighting system is rather empirical (or arbitrary). Instead the *LF* is simple additive.

7. Conclusions

Did we give reasonable answers to the questions stated in the title and in the introduction of this paper? The answer is probably yes and no, because actually we made an attempt to formalize the process of constructing a composite indicator from a batch of single indicators, by means of two steps: *transforming non homogeneous data & gathering data transformed*. Answers to the first step are in the *counselling table* which provides some pros and some cons to the most widely used transformations. This task does not seem easy. The answer to the second step comes from a pragmatic approach: additive functions are much more used than the non additive even if there are reported several examples of practical applications which can be considered hints for further extensions. Actually there is an evident lack of theoretical and general bases, and the linkage between first step first and the second one has not been explored in depth. In this direction dimensional analysis might provide interesting clues.

References

Aczél J. (1987). *A Short Course on Functional Equations*. D. Reidel Publishing Company, Dordrecht.

Atkinson A.C. and Cox D.R. (1982). Transformations, in: *Encyclopaedia of Statistical Sciences*. Kotz S. & Johnson N.L. (Eds.). Wiley. New York.

Attanasio M., Capursi V. (1997). Graduatorie sulla qualità della vita: prime analisi di sensibilità delle tecniche adottate. *Atti XXXV Riunione Scientifica SIEDS*, Alghero.

Fayers P.M., Hand D.J. (2002). Casual Variables, Indicator Variables and Measurement Scales: an example from quality of life. *JRRS*, A, 165, 233 – 261.
Fletcher R.H., Fletcher S.W., Wagner E.H. (1982). *Clinical Epidemiology – the essentials*, Williams & Wilkins. Baltimore.
Hoaglin D.C., Mosteller F., Tukey J.W. (1983). *Understanding Robust and Exploratory Data Analysis*. Wiley, New York.
Inskip H. (1998). Standardized Methods, in: *Encyclopaedia of Biostatistics*, Armitage P. & Colton T. (Eds.), Wiley, 6, 4237 – 4250.
Kendall M., Stuart A., Ord J.K. (1983). *The Advanced Theory of Statistics*. Charles Griffin and C. 3, 97.
Luce R.D. *et al.* (1990). *Foundations of Measurements*. Academic Press, New York.
Prieto L. *et al.* (1996). Scaling the Spanish Version of the Nottingham Health Profile: Evidence of Limited Value of Item Weights. *J. Clin. Epi.*, 49, 31 – 38. Elsevier Science.
Streiner D.L., Norman G.R. (1999).(eds.) *Health Measurement Scales. A practical guide to their development and use*. 2nd Ed. Oxford University Press, New York.
Stevens S. S. (1974). Measurement, in: *Scaling: a sourcebook for behavioural scientists*, Maranell M. (ed.), Aldine Publishing Company. Chicago.
UNC Charlotte Dept. Of Geography and Earth Sciences, UNC at Charlotte (2002). Charlotte Neighborhood Quality of Life Study. http://www.charmeck.org/NR/rdonlyres/ …/2002+Quality+of+Life+Study.pdf