2008

U-Scores for Multivariate Data In Sports

Knut M. Wittkowski

Tingting Song

Kent Anderson

John E. Daniels

Journal of Quantitative Analysis in Sports

Volume 4, *Issue* 3 2008 *Article* 7

U-Scores for Multivariate Data in Sports

Knut M. Wittkowski* Tingting Song†

Kent Anderson‡ John E. Daniels**

*The Rockefeller University, kmw@rockefeller.edu

†The Rockefeller University, tsong01@rockefeller.edu

‡University of California at Davis, Clinical and Translational Science Center, kent.anderson@ucdmc.ucdavis.edu

**Central Michigan University, Department of Mathematics, john.daniels@cmich.edu

U-Scores for Multivariate Data in Sports*

Knut M. Wittkowski, Tingting Song, Kent Anderson, and John E. Daniels

Abstract

In many sport competitions athletes, teams, or countries are evaluated based on several variables. The strong assumptions underlying traditional 'linear weight' scoring systems (that the relative importance, interactions and linearizing transformations of the variables are known) can often not be justified on theoretical grounds, and empirical 'validation' of weights, interactions and transformations, is problematic when a 'gold standard' is lacking. With μ-scores (u-scores for multivariate data) one can integrate information even if the variables have different scales and unknown interactions or if the events counted are not directly comparable, as long as the variables have an 'orientation'. Using baseball as an example, we discuss how measures based on μ-scores can complement the existing measures for 'performance' (which may depend on the situation) by providing the first multivariate measures for 'ability' (which should be independent of the situation). Recently, μ-scores have been extended to situations where count variables are graded by importance or relevance, such as medals in the Olympics (Wittkowski 2003) or Tour-de-France jerseys (Cherchye and Vermeulen 2006, 2007). Here, we present extensions to 'censored' variables (life-time achievements of active athletes), penalties (counting a win more than two ties) and hierarchically structured variables (Nordic, alpine, outdoor, and indoor Olympic events). The methods presented are not restricted to sports. Other applications of the method include medicine (adverse events), finance (risk analysis), social choice theory (voting), and economy (long-term profit).

KEYWORDS: statistics, performance, ability, triathlon, baseball, Olympics, soccer, Tour-de-France, multivariate, ranking, voting

---

*This work was partially supported by an NIH/Clinical and Translational Science Award (CTSA) grant UL1 RR024143.

1. INTRODUCTION

Most traditional statistical methods for multivariate data are based on linear weight (lw) scores. A global score is defined as the linear combination (weighted average) of each variable's score. When sufficiently detailed functional models are lacking, as in biomechanics or team interaction (Hughes and Bartlett 2002), the assumptions underlying lw scores can often not be justified on theoretical grounds, so that the definition of scores typically relies merely on computational efficiency, rather than subject matter adequacy.

Non-parametric methods, including u-statistics, are particularly well suited for ordinal data, where a one-unit difference may not carry the same 'meaning' across the range of possible values, For instance, running 10 km in 49 instead of 50 min, can be achieved with moderate training; improving from 39 to 40 minutes requires considerable effort.

μ-Scores (u-scores for multivariate data) cover situations where a one-unit difference may carry a different 'meaning' across variables (Wittkowski *et al.* 2004) and, thus, can integrate information even if the events counted are incomparable or the variables' scales differ, as long as each variable has the same 'orientation'. For instance, when describing a triathlete's fitness, a one minute difference in biking may not be directly comparable to a one minute difference in swimming.

Recently, μ-scores have been extended to situations, where the 'importance' or 'relevance' of a one-unit difference can be ordered (graded) across variables, although the relative importance may be impossible to quantify. Olympic gold medals, for instance, are more valuable than silver or bronze medals (Wittkowski 2003), some Tour de France jerseys are more important than others (Cherchye and Vermeulen 2006, 2007), home runs in baseball are more valuable than triples, doubles, and singles (in that order), and, in soccer, defensive play is penalized by counting a win more than two ties.

Subtle differences between the concepts addressed may have important consequences for the choice of a method. Baseball provides an opportunity to discuss the importance of distinguishing between 'performance', which may depend on the situation and 'ability' ('fitness'), which does not.

How to score batters in baseball has given rise to extensive discussions (James 1982). The diversity of scoring systems (Thorn *et al.* 1985) attests to their subjective nature. 'Batting average' (BA), treating all hits equally, measures a specific aspects of ability. i.e., BA does not depend on the context (position in the batting order, number of runners on bases, etc.). It's usefulness as an overall measure of ability, however, is limited because it ignores that a home run is clearly more indicative of a batter's ability to hit for power than merely reaching first base. Conversely, home runs (HR) measure primarily the ability to hit for power, but less so the ability to run fast. Between these extremes, various sets of relative weights have been proposed ranging from 1:1:1:1 (BA) to 1:2:3:4 ('slugging', SLG). Even

the assumption that these weights should be constant, however, is unrealistic. A difference of a few singles may be more indicative of batting ability among weaker batters than among all-star players. Running speed contributes to ability, except for most home runs. Since the physio- and psychological factors determining a batter's ability are not well understood, content validity of lw scoring systems cannot be established on theoretical grounds.

In baseball, team success is not determined by the profiles of hits, but by the number of batters reaching home (Albert and Bennett 2003). Thus, one might try to justify a particular choice of weights through empirical 'validation', choosing weights that, when applied to a sample, provide a good fit in terms of 'least squares' (l.s.) to this 'gold standard'. However, restricting oneself to linear combinations and l.s. optimality virtually guarantees that the best method is missed (Bassett 1997). Moreover, by fitting against situation dependent performance measures such as runs batted in (RBI) or runs created (RC), the weights also are situation dependent. The extent to which a player's hitting ability contributes to the team's overall performance depends on teammate ability, ballpark design, game situation, etc. Finally, to obtain stable estimates, several years of data may need to be accumulated, so that the results also depend on changes in rules, training practices, and pharmacological interventions. Thus, when situation dependent assessments are sought and a gold standard exists, lw scores may appropriately reflect 'performance', but other approaches may be required to measure ability when gold standards are lacking.

We will draw on extensions to μ-scores for analyzing multivariate ordinal data (Wittkowski 2003), which are based on u-statistics (Hoeffding 1948) (see Section 2.1). The triathlon example demonstrates the distinction between standard μ-scores for ability and lw scores for performance (Section 3.1). In Section 3.2, we apply μ-scores to Major League Baseball data and compare non-parametric ability μ-scores with model-based lw performance scores. In Section 3.3, we introduce hierarchical μ-scores using Olympic medals as an example. This approach is extended in Section 3.4 to censored data. Using soccer as an example, we will then demonstrate how additional knowledge or requirements can be incorporated (Section 3.5). Finally, we will discuss, how μ-scores can be used to derive approximate lw scores that allow us to compare athletes across events (Section 4) and highlight applications to areas other than sports (Section 5).

2. METHODS

2.1. U-Scores for Multivariate Data

μ-Scores do not require any assumptions regarding the variables' functional relationship with 'ability', except that if all other variables are held constant, an increase in this variable is 'good' or 'bad', so that, after changing signs, if necessary, all variables have the same 'orientation'.

Here, k will index m subjects, each characterized by L variables. A partial order among profiles $x_k = (x_{k1}, \ldots, x_{kL})'$ can easily be defined (Wittkowski 1992). Profile k' is higher than profile k if it is higher in some variable(s) and lower in none

$$x_k < x_{k'} \Leftrightarrow \{\forall_{\ell=1,\ldots,L} x_{k\ell} \le x_{k'\ell} \wedge \exists_{\ell=1,\ldots,L} x_{k\ell} < x_{k'\ell}\}. \quad (1)$$

Thus, μ-scores are 'intrinsically valid', i.e., independent of the choice of (non-zero) weights and (monotonous) transformations assigned to the variables.

With (untied) univariate data, all pairwise orderings can be decided. With multivariate data, a pairwise ordering is ambiguous ($x_k \sim x_k'$) if subject k is higher than subject k' in some variable(s), but lower in others. (Variables with missing data in either subject are ignored). The μ-score is the number of subjects being lower minus the number of subjects being higher (w.r.t. the partial ordering (1)). It is undefined if the pairwise order with respect to all other profiles is ambiguous.

2.2. Computational Aspects

Wittkowski et al. (2004) extended Deuchler's (1914) univariate algorithm to depict all pairwise orderings as a symmetric matrix (see several Figures below). Alternatively, one can compute these univariate matrices first and then use matrix operations to combine them (Morales et al. 2008). Recently (Cherchye and Vermeulen 2006) proposed to replace the matrix of pairwise orderings with entries '+1', '0', '?', and '−1', by a computational simpler 'GE' matrix of binary entries $u_{kk'} = \mathrm{I}(x_k \ge x_{k'})$. Combining these two approaches, one can compute univariate GE matrices $u_{kk'\ell} = \mathrm{I}(x_{k\ell} \ge x_{k'\ell})$ first (see function mu.PwO in Figure 1), and then combine them into a multivariate GE matrix (function mu.AND) describing the partial ordering. (R and S-PLUS packages are available from http://cran.r-project.org and http://csan.insightful.com, respectively).

Figure 1: Simplified code for the core functions of the muStat packages

```
mu.PwO <- function(x, y=x)
if (length(y)>1) apply(rbind(x,y),2,mu.PwO,nrow(x))
   else as.numeric(NAtoZer(outer(x[1:y],x[-(1:y)],">=")))
mu.AND <- function(PwO, frml=NULL)
if (is.null(frml)) {
   GE  <- sq.array(PwO); AND <- GE[,,1]^0; nNA <- AND[,1]*0
   for (i in 1:dim(GE)[3]) {
      nNA <- nNA + diag(GEi <- GE[,,i])
      AND <- AND * (GEi + (1-GEi)*(1-t(GEi))) }
      return(as.numeric(AND * ((c(nNA) %o% c(nNA)) > 0))) }
else { # … deal with the formula … (code omitted)
   return(…) }
mu.Sums <- function(PwO) {
   GE   <- sq.matrix(PwO)
   wght <- colSums(GE|t(GE))
   list (score  = (rowSums(GE) - colSums(GE)) *ifelse(wght==0,NA,1),
         weight = wght) }
```

2.3. Censored Data

If additional information about the variables is available, more specific partial orderings can be defined (Wittkowski *et al.* 2004). While valuable in itself, having separate GE matrices available provides the opportunity to compute the univariate GE matrices differently for different variables. In particular, the GE matrix for interval censored variables can be defined by requesting that for two intervals to be ordered, the left limit of the higher interval must be larger than the right limit of the lower interval. In function `mu.PwO` the left and right limit of an interval are entered `x` and `y`, respectively. Thus, the proposed approach allows for scoring several censored (or uncensored) variables comprehensively.

2.4. Graded Variables

Often, a simple transformation of the variables may suffice to reflect additional knowledge. For graded variables, where one unit impacts less in a 'lower grade' than in a 'higher grade' variable (Wittkowski 2003; Cherchye and Vermeulen 2006) one can split each value of variable (ℓ) (sorted by grades) into the value of the lowest grade variable Δ_ℓ and incremental values of the higher grade variables $\Delta_{\ell'=2\ldots\ell}$.

$$
\begin{aligned}
x_{k,(1)} &= x_{k,(1)}\Delta_1 \\
x_{k,(2)} &= x_{k,(2)}\Delta_1 + x_{k,(2)}\Delta_2 \\
&\cdots \\
x_{k,(L)} &= x_{k,(L)}\Delta_1 + x_{k,(L)}\Delta_2 + \ldots + x_{k,(L)}\Delta_L
\end{aligned}
$$

Thus, the profile of counts sorted by grade $\left(x_{k,(1)}, x_{k,(2)}, \ldots, x_{k,(L)}\right)$ can be expressed as the column sums $\left(x_{k,(\geq 1)}, x_{k,(\geq 1)}, \ldots, x_{k,(=L)}\right)$ where $x_{k,(\geq \ell)} = \sum_{\ell'=\ell}^{L} x_{k,(\ell)}$. The partial ordering for graded variables

$$\left(x_{k,(1)},\ldots,x_{k,(L)}\right) < \left(x_{k',(1)},\ldots,x_{k',(L)}\right) \Leftrightarrow \forall_{\ell=1}^{L}\, x_{k,(\geq\ell)} \leq x_{k',(\geq\ell)} \wedge \exists_{\ell=1}^{L}\, x_{k(\geq\ell)} < x_{k'(\geq\ell)} \qquad (2)$$

is equivalent to the regular ordering (1) applied to the cumulative variables $x_{k,(\geq\ell)}$.

Although each profile's outcomes are decomposed into additive components Δ_ℓ, substantially weaker assumptions are made than with lw scores, because the additive components can be unknown and may even differ between pairs, noting that for subjects far apart (subject k is lower than subject k' for each of the variables) or incomparable (some variables higher for subject k and some variables higher for subject k') the weights are irrelevant. Thus, the weights only need to be 'locally similar', rather than 'globally constant'.

2.5. Hierarchically Structured Variables

A downside of μ-scores, in general, is that the number of ambiguous pair-wise orderings increases with the number of variables, unless the variables are highly correlated. Incorporating additional knowledge, can resolve some of these ambiguities. If the variables are related to different 'factors' and the order between subjects A and B is ambiguous with respect to variables related to one factor (e.g., Nordic), unambiguous results with respect to another factor (e.g., downhill) can 'overwrite' this ambiguity. The advantage of creating the matrices reflecting the univariate orderings first (`mu.PwO`) and combining them in a separate step (`mu.AND`) before compute the scores (`mu.Sum`), is that incorporating knowledge about the sub-factor hierarchy through hierarchically combining the matrices can reduce loss of information content (Morales *et al.* 2008).

3. APPLICATIONS TO SPORTS

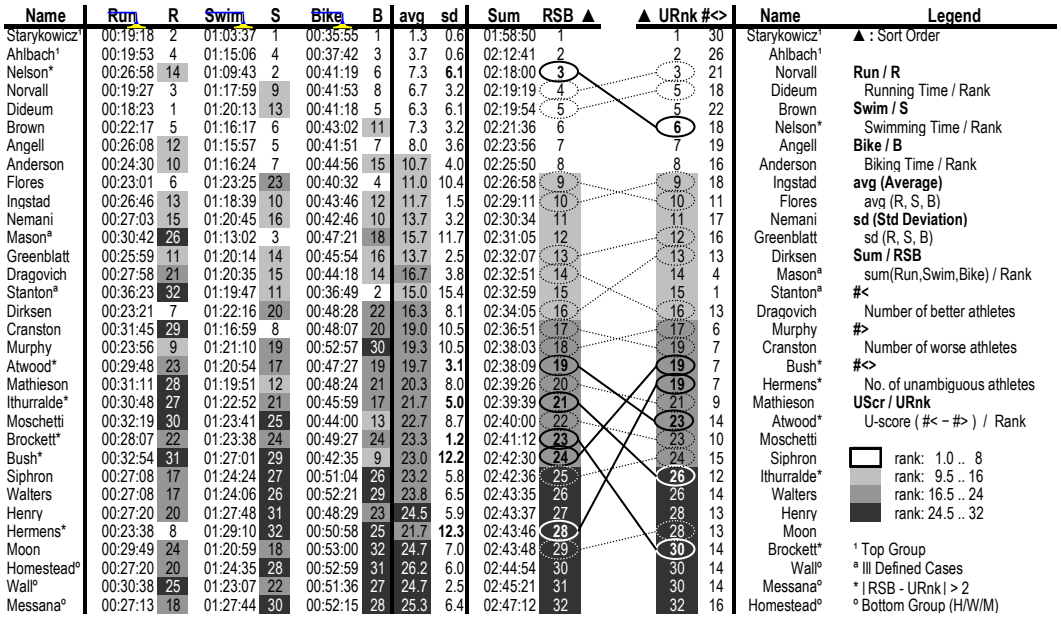
3.1. Scoring Athletes Competing in Several Disciplines (Triathlon)

With bi- or pentathlon, the advantage of a more objective method is obvious. The time or distance of skiing 'equivalent' to missing a ring in shooting is as subjective as is the speed of running to the distance of jumping. Even if all variables have the same scale (time), as in triathlon, total time measures the athlete's relative performance to cover the particular distance (Olympic: 1500 m swimming, 40 km biking, 10 km running) under specific environmental conditions (Paton and Hopkins 2005). A measure for overall ability, however, should depend neither on the 3:80:20 ratio in distance (≈ 1:3:2 in time spent) nor on terrain or weather.

Assuming that increasing speed in either category requires a comparable training effort, overall time favors bikers (≈ 1:15), because they can achieve a one-second difference more easily than swimmers (≈ 0:25). Thus, better performing athletes were found to save energy during swimming (Vleck *et al.* 2006). A mountainous course disfavors swimmers by decreasing the proportion of time spent swimming. Averaging within-category rankings (O'Brien 1984), also known as 'range' voting as a special case of 'utilitarian' voting (Sen 1986), adjusts for differences in scales, but not relevance. By giving the same weight to each category, average ranks overrepresent leg muscle fitness, because running and biking strength are usually correlated. μ-Scores automatically account for biological correlations and depend less on distance and environmental conditions, thereby scoring ability more objectively, unless the character of the competition is changed from endurance to either sprint or ironman.

From Figure 2, μ-score ranks (URnk) are similar to total time ranks (SRB) among the 'best' athletes, except for Nelson, who ranks 3rd (after Starykowicz and

Figure 2: Ranking triathletes. Comparison of ranks based on total time with ranks based on u-scores (top) and computation of u-scores (bottom)



Name	Run	R	Swim	S	Bike	B	avg	sd	Sum	RSB ▲	▲ URnk	#<>
Starykowicz¹	00:19:18	2	01:03:37	1	00:35:55	1	1.3	0.6	01:58:50	1	1	30
Ahlbach¹	00:19:53	4	01:15:06	4	00:37:42	3	3.7	0.6	02:12:41	2	2	26
Nelson*	00:26:58	14	01:09:43	2	00:41:19	6	7.3	**6.1**	02:18:00	3	3	21
Norvall	00:19:27	3	01:17:59	9	00:41:53	8	6.7	3.2	02:19:19	4	5	18
Dideum	00:18:23	1	01:20:13	13	00:41:18	5	6.3	6.1	02:19:54	5	5	22
Brown	00:22:17	5	01:16:17	6	00:43:02	11	7.3	3.2	02:21:36	6	6	18
Angell	00:26:08	12	01:15:57	5	00:41:51	7	8.0	3.6	02:23:56	7	7	19
Anderson	00:24:30	10	01:16:24	7	00:44:56	15	10.7	4.0	02:25:50	8	8	16
Flores	00:23:01	6	01:23:25	23	00:40:32	4	11.0	10.4	02:26:58	9	9	18
Ingstad	00:26:46	13	01:18:39	10	00:43:46	12	11.7	1.5	02:29:11	10	10	11
Nemani	00:27:03	15	01:20:45	16	00:42:46	10	13.7	3.2	02:30:34	11	11	17
Mason[a]	00:30:42	26	01:13:02	3	00:47:21	18	15.7	11.7	02:31:05	12	12	16
Greenblatt	00:25:59	11	01:20:14	14	00:45:54	16	13.7	2.5	02:32:07	13	13	13
Dragovich	00:27:58	21	01:20:35	15	00:44:18	14	16.7	3.8	02:32:51	14	14	4
Stanton[a]	00:36:23	32	01:19:47	11	00:36:49	2	15.0	15.4	02:32:59	15	15	1
Dirksen	00:23:21	7	01:22:16	20	00:48:28	22	16.3	8.1	02:34:05	16	16	13
Cranston	00:31:45	29	01:16:59	8	00:48:07	20	19.0	10.5	02:36:51	17	17	6
Murphy	00:23:56	9	01:21:10	19	00:52:57	30	19.3	10.5	02:38:03	18	19	7
Atwood*	00:29:48	23	01:20:54	17	00:47:27	19	19.7	**3.1**	02:38:09	19	19	7
Mathieson	00:31:11	28	01:19:51	12	00:48:24	21	20.3	8.0	02:39:26	20	19	7
Ithurralde*	00:30:48	27	01:22:52	21	00:45:59	17	21.7	**5.0**	02:39:39	21	21	9
Moschetti	00:32:19	30	01:23:41	25	00:44:00	13	22.7	8.7	02:40:00	22	23	14
Brockett*	00:28:07	22	01:23:38	24	00:49:27	24	23.3	**1.2**	02:41:12	23	23	10
Bush*	00:32:54	31	01:27:01	29	00:42:35	9	23.0	**12.2**	02:42:30	24	24	15
Siphron	00:27:08	17	01:24:24	27	00:51:04	26	23.2	5.8	02:42:36	25	26	12
Walters	00:27:08	17	01:24:06	26	00:52:21	29	23.8	6.5	02:43:35	26	26	14
Henry	00:27:20	20	01:27:48	31	00:48:29	23	24.5	5.9	02:43:37	27	28	13
Hermens*	00:23:38	8	01:29:10	32	00:50:58	25	21.7	**12.3**	02:43:46	28	28	13
Moon	00:29:49	24	01:20:59	18	00:53:00	32	24.7	7.0	02:43:48	29	30	14
Homestead°	00:27:20	20	01:24:35	28	00:52:59	31	26.2	6.0	02:44:54	30	30	14
Wall°	00:30:38	25	01:23:07	22	00:51:36	27	24.7	2.5	02:45:21	31	30	14
Messana°	00:27:13	18	01:27:44	30	00:52:15	28	25.3	6.4	02:47:12	32	32	16

Name	#>
Starykowicz¹	
Ahlbach¹	
Norvall	21
Dideum	18
Brown	22
Nelson*	18
Angell	19
Anderson	16
Ingstad	11
Flores	18
Nemani	17
Greenblatt	16
Dirksen	13
Mason[a]	4
Stanton[a]	1
Dragovich	13
Murphy	6
Cranston	7
Bush*	7
Hermens*	7
Mathieson	9
Atwood*	14
Moschetti	
Siphron	15
Ithurralde*	12
Walters	14
Henry	13
Moon	13
Brockett*	14
Wall°	14
Messana°	14
Homestead°	16

Legend

▲ : Sort Order

Run / R
Running Time / Rank

Swim / S
Swimming Time / Rank

Bike / B
Biking Time / Rank

avg (Average)
avg (R, S, B)

sd (Std Deviation)
sd (R, S, B)

Sum / RSB
sum(Run,Swim,Bike) / Rank

#<
Number of better athletes

#>
Number of worse athletes

#<>
No. of unambiguous athletes

UScr / URnk
U-score (#< − #>) / Rank

rank: 1.0 .. 8
rank: 9.5 .. 16
rank: 16.5 .. 24
rank: 24.5 .. 32

¹ Top Group
[a] III Defined Cases
* | RSB - URnk | > 2
° Bottom Group (H/W/M)

Name	R	S	B	avg	RSB	S	A	N	D	B	N	A	A	I	F	N	G	D	M	S	D	M	C	H	B	M	A	M	S	I	W	M	H	B	W	M	H	#<	#>	UScr	URng
Starykowicz	2	1	1	1.3	1	0	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	30	-30	1.0
Ahlbach	4	4	3	3.7	2	1	0	0	0	-1	0	-1	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	25	-24	2.0
Norvall	3	9	8	6.7	4	1	0	0	0	0	0	0	0	-1	0	-1	-1	-1	0	0	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	20	-19	3.0
Dideum	1	13	5	6.3	5	0	0	0	0	0	0	0	0	0	-1	-1	-1	-1	0	0	-1	-1	0	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	18	-18	4.5
Brown	5	6	11	7.3	6	1	1	0	0	0	0	0	-1	-1	0	0	-1	-1	0	0	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	2	20	-18	4.5
Nelson	14	2	6	7.3	3	1	0	0	0	0	0	0	0	0	0	-1	0	0	-1	0	-1	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	17	-16	6.0
Angell	12	5	7	8.0	7	1	1	0	0	0	0	0	0	-1	0	-1	0	0	0	0	-1	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	2	17	-15	7.0
Anderson	10	7	15	10.7	8	1	1	0	0	1	0	0	0	0	0	0	-1	0	0	0	0	0	-1	0	0	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	3	13	-10	8.0
Ingstad	13	10	12	11.7	10	1	1	1	0	1	0	1	0	0	0	0	0	0	0	0	-1	0	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	5	13	-8	9.0
Flores	6	23	4	11.0	9	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	-1	0	0	-1	-1	0	-1	0	-1	-1	0	-1	-1	-1	-1	2	9	-7	10.0
Nemani	15	16	10	13.7	11	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	6	11	-5	11.0
Greenblatt	11	14	16	13.7	13	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	6	10	-4	12.0
Dirksen	7	20	22	16.3	16	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	-1	0	-1	0	-1	-1	-1	-1	-1	5	8	-3	13.0
Mason	26	3	18	15.7	12	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	2	2	0	14.0
Stanton	32	11	2	15.0	15	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	15.0
Dragovich	21	15	14	16.7	14	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	-1	0	-1	0	-1	-1	0	0	0	8	5	3	16.0
Murphy	9	19	30	19.3	18	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	5	1	4	17.0
Cranston	29	8	20	19.0	17	1	1	0	0	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	7	19.0
Hermens	8	32	25	21.7	28	1	1	1	1	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	7	19.0
Bush	31	29	9	23.0	24	1	1	1	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	7	19.0
Mathieson	28	12	21	20.3	20	1	1	1	0	1	1	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	9	21.0
Atwood	23	17	19	19.7	19	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	-1	0	0	12	2	10	22.5
Moschetti	30	25	13	22.7	22	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	10	22.5
Siphron	17	27	26	23.2	25	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	-1	13	2	11	24.0
Ithurralde	27	21	17	21.7	21	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	12	25.5
Walters	17	26	29	23.8	26	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	13	1	12	25.5
Moon	24	18	32	24.7	29	1	1	1	1	1	1	1	1	1	0	1	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	13	0	13	27.5
Henry	20	31	23	24.5	27	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	13	27.5
Brockett	22	24	24	23.3	23	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0	14	30.0
Wall	25	22	27	24.7	31	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	14	0	14	30.0
Messana	18	30	28	25.3	32	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	14	0	14	30.0
Homestead	20	28	31	26.2	30	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	16	0	16	32.0
corr_avg	.68	.81	.82	1.	.97	2	4	3	1	5	14	6	12	10	13	6	15	11	7	26	32	21	9	29	8	31	28	23	30	17	27	17	24	20	22	25	18	20				.97
corr_RSB	.58	.84	.83	.97	1.	1	4	9	13	6	2	5	7	7	10	23	16	14	20	3	11	15	19	8	32	29	12	17	25	27	21	26	17	29	18	31	24	22				.96
corr_URng	.67	.76	.81	.97	.96	3	8	5	11	6	5	7	15	12	4	10	16	16	22	18	2	14	30	20	25	9	21	19	13	26	17	29	26	32	23	24	27	28				1.

Note: For simplicity and generalization to other sports, times for changing clothes between disciplines are excluded from all analyses. As they attribute for < 5 % of the total time, this yields essentially the same ranking as splitting transition times 50/50 among the adjacent disciplines (results not shown).

Ahlbach) due to his 2nd rank in biking (B), although he ranks only 14th and 6th in swimming (S) and running (R), respectively. His URnk is 6th ($\#_{>}:\#_{<} = 1:17$) after-Norvall (20:1), Dideum (18:0), and Brown (20:2). μ-Scores award bronze to Norvall, rather than Nelson (total time) or Dideum (average rank), because Norvall's profile is more 'typical', allowing more pairwise orders to be decided ($\#_{<>}=21$) than Nelson and Dideum (18 each).

Bush and Hermens have better μ-scores, at 19:24 and 19:29, respectively, while Atwood, Ithurralde, and Brockett score worse (22.5:19, 25.5:21, and 30:23, respectively). As μ-score tend to score profiles with low information content (Bush, Hermens) towards the median, an exceptionally fast biker/runner (Nelson) with the shortest total time may still loose to a more balanced athlete (Norvall). As exemplified by Nelson vs. Hermens, μ-scores assign less weight to biking ('Bike' time $\approx 1:15$), so that swimmers ('Swim' time $\approx 0:25$) and runners ('Run' time $\approx 0:45$) are not disadvantaged. Averaging ranks ('avg') also has this feature, yet implies all disciplines being equally important. However, as biking and running times are highly correlated (0.58), but less so with swimming (0.29 and 0.25, respectively), this disadvantages swimmers. μ-Scores do not suffer from this fallacy. Even including a variable twice would not affect the scores.

3.2. Ranking Baseball Batters by Profiles of Hits

When evaluating baseball batters, some important (and endlessly debated) questions are:

1) Who is the better hitter?
2) Who contributed more to the team?
3) Who is more a 'valuable' player?

A formal assessment of the third question, the Most Valuable Player (MVP) award is beside the point, because, according to Baseball Writers Association of America (BBWAA) ballot rules, "there is no clear-cut definition of what Most Valuable means." With the advent of computers, methods such as the Player Game Percentage (Bennett 1993) became available to retrospectively assess how much a player contributed to his team's success (second question), providing managers with information on his best position in the batting order for this team's performance.

To pick a new player, however, a general manager needs to assess how the player's ability qualifies him to play under different conditions. For specialists, HR or BA score ability, but a comparable method to rank versatile players is lacking. Home runs indicate more batting ability than triples (3B), doubles (2B), and singles (1B), yet the amount for each additional base reached is *a-priori* unknown. Still, one can translate a (1B, 2B, 3B, HR) profile into counts of making it at least to first (BA = 1B+2B+3B+HR), second (2B+3B+HR), or third base (3B+HR) and

HR. μ-Scores can then be computed from profiles of these cumulated hits divided by at bats (AB). For simplicity, we exclude lower grad variables such as hit-by-pitch, walks, steals, and sacrifices, thereby replacing the overall performance measures OPS and OPA with batting performance measures BPS and OPB. These measures happen to be ordered with respect to the relative weight given to hits:

Batting Average:	$BA = \left(1.0 \times 1B + 1.0 \times 2B + 1.0 \times 3B + 1.0 \times HR\right) / AB$
BA plus SLG	$BPS = \left(1.0 \times 1B + 1.5 \times 2B + 2.0 \times 3B + 2.5 \times HR\right) / AB$
Offensive Performance	$OPB = \left(1.0 \times 1B + 2.0 \times 2B + 2.5 \times 3B + 3.5 \times HR\right) / AB$
Slugging percentage:	$SLG = \left(1.0 \times 1B + 2.0 \times 2B + 3.0 \times 3B + 4.0 \times HR\right) / AB$

From the cumulative proportions in Table 1, Barry Bonds played on a level all by himself. Hitting home runs in 33.8% of his hits made him the top player in all but the first cumulative categories ($\mathrm{c1B} = \mathrm{BA} = .341$), where Pujols and Helton scored first (.359) and second (.358), respectively. Special skills result in different ranks in lw scores and cumulative categories. A. Rodriguez, Ortiz, and Edmonds, e.g., made it less frequently to first base (.298, .288, and .275), but then often continued, making home runs in $> 24\%$ of their hits (.077, .069, and .087) compared to $\approx 24\%$ for Pujols, Sheffield, and Ramirez. Helton and Mueller hit frequently (.358 and .326, respectively), but made home runs in $< 16\%$ of their hits. Figure 3 displays for each batter, sorted by μ-scores, the above four and the 'LWTS' (Thorn *et al.* 1985) lw score (see Section 4) as separate symbols. μ-Scores are also given as a straight line.

A close correlation of the μ-score ability measure with the performance measures for 'typical' players, is to be expected. 'Specialists' (top and bottom three excluded) is identified based on their normalized standard deviation among the ranks in the cumulative categories

$$sd\left(rank\left(c_{1\mathrm{B}}\right), ..., rank\left(c_{\mathrm{HR}}\right)\right) / hmean^{3/4}\left(rank\left(u\right), n+1-rank\left(u\right)\right) \qquad (3)$$

A total of 32 (top and bottom three excluded) shown as empty (first decentile, bold names) and gray (second decentile) circles. The empirically determined exponent ¾ compromises between a focus on the center (½) and the tails (1).

Figure 3 also clearly demonstrates the fundamental differences between the existing measures for performance and the novel approach to measure ability. The curvature of the relationship between LWTS scores and μ-scores for 'typical' batters confirms that the relationship is not linear. For both extremes, small differences in the ability score have large effects in terms of performance. Moreover, having a novel measure for a different concept is shown to help with identifying players with special skills sets.

Table 1: Batting statistics and scores for the top 2003 MLB batters sorted by (all player) μ-scores

Player, Team	AB	H	2B	3B	HR	BA		SLG		OPS		BPS		c1B	c2B	c3B	cHR	UScr	
B. Bonds, SF	390	133	22	1	45	0.341	(3)	0.749	(1)	1.278	(1)	1.090	(1)	0.341	0.174	0.118	0.115	162	(1)
A. Pujols, StL	591	212	51	1	43	0.359	(1)	0.667	(2)	1.106	(2)	1.025	(2)	0.359	0.161	0.074	0.073	154	(2)
G. Sheffield, Atl	576	190	37	2	39	0.330	(5)	0.604	(5)	1.023	(4)	0.934	(4)	0.330	0.135	0.071	0.068	145	(3)
T. Helton, Col	583	209	49	5	33	0.358	(2)	0.630	(3)	1.088	(3)	0.988	(3)	0.358	0.149	0.065	0.057	138	(4)
M. Ramirez, Bos	569	185	36	1	37	0.325	(8)	0.587	(9)	1.014	(6)	0.912	(5)	0.325	0.130	0.067	0.065	131	(5)
C. Delgado, Tor	570	172	38	1	42	0.302	(34)	0.593	(7)	1.019	(5)	0.895	(7)	0.302	0.142	0.075	0.074	122	(6)
T. Nixon, Bos	441	135	24	6	28	0.306	(28)	0.578	(10)	0.975	(9)	0.884	(9)	0.306	0.132	0.077	0.063	120	(7)
J. Guillen, Oak/Cin	485	151	28	2	31	0.311	(21)	0.569	(13)	0.928	(18)	0.880	(11)	0.311	0.126	0.068	0.064	119	(8)
A. Rodriguez, Tex	607	181	30	6	47	0.298	(43)	0.600	(6)	0.995	(8)	0.898	(6)	0.298	0.137	0.087	0.077	116	(9)
R. Hidalgo, Hou	514	159	43	4	28	0.309	(24)	0.572	(12)	0.957	(12)	0.881	(10)	0.309	0.146	0.062	0.054	112	(10)
V. Wells, Tor	678	215	49	5	33	0.317	(10)	0.550	(17)	0.909	(30)	0.867	(13)	0.317	0.128	0.056	0.049	108	(11)
A. Huff, TB	636	198	47	3	34	0.311	(22)	0.555	(15)	0.922	(22)	0.866	(15)	0.311	0.132	0.058	0.053	107	(12)
M. Ordonez, CWS	606	192	46	3	29	0.317	(11)	0.546	(19)	0.926	(21)	0.863	(16)	0.317	0.129	0.053	0.048	101	(13)
G. Anderson, Ana	638	201	49	4	29	0.315	(13)	0.541	(20)	0.885	(37)	0.856	(17)	0.315	0.129	0.052	0.045	97	(14)
D. Ortiz, Bos	448	129	39	2	31	0.288	(64)	0.592	(8)	0.961	(10)	0.879	(12)	0.288	0.161	0.074	0.069	93	(15)
D. Young, Det	562	167	34	7	29	0.297	(45)	0.537	(23)	0.909	(30)	0.835	(21)	0.297	0.125	0.064	0.052	88	(16)
G. Jenkins, Mil	487	144	30	2	28	0.296	(49)	0.538	(22)	0.913	(26)	0.834	(22)	0.296	0.123	0.062	0.057	86	(17)
L. Gonzalez, Ari	579	176	46	4	26	0.304	(31)	0.532	(26)	0.934	(17)	0.836	(20)	0.304	0.131	0.052	0.045	85	(18)
H. Blalock, Tex	567	170	33	3	29	0.300	(39)	0.522	(35)	0.872	(43)	0.822	(28)	0.300	0.115	0.056	0.051	84	(19)
B. Boone, Sea	622	183	35	5	35	0.294	(52)	0.535	(25)	0.902	(32)	0.830	(25)	0.294	0.121	0.064	0.056	84	(19)
B. Mueller, Bos	524	171	45	5	19	0.326	(6)	0.540	(21)	0.938	(16)	0.866	(14)	0.326	0.132	0.046	0.036	83	(21)
C. Jones, Atl	555	169	33	2	27	0.305	(30)	0.517	(37)	0.920	(24)	0.822	(29)	0.305	0.112	0.052	0.049	81	(22)
M. Giles, Atl	551	174	49	2	21	0.316	(12)	0.526	(30)	0.917	(25)	0.842	(18)	0.316	0.131	0.042	0.038	79	(23)
N. Garciaparra, Bos	658	198	37	13	28	0.301	(36)	0.524	(32)	0.870	(44)	0.825	(27)	0.301	0.119	0.062	0.043	79	(23)
A. Soriano, NYY	682	198	36	5	38	0.290	(59)	0.525	(31)	0.863	(49)	0.815	(32)	0.290	0.116	0.063	0.056	75	(25)
J. Payton, Col	600	181	32	5	28	0.302	(35)	0.512	(43)	0.865	(47)	0.813	(34)	0.302	0.108	0.055	0.047	73	(26)
S. Sosa, ChC	517	144	22	0	40	0.279	(83)	0.553	(16)	0.911	(27)	0.832	(23)	0.279	0.120	0.077	0.077	72	(27)
J. Edmonds, StL	447	123	32	2	39	0.275	(93)	0.617	(4)	1.002	(7)	0.893	(8)	0.275	0.163	0.092	0.087	71	(28)
P. Wilson, Col	600	169	43	1	36	0.282	(76)	0.537	(24)	0.880	(39)	0.818	(31)	0.282	0.133	0.062	0.060	70	(29)
B. Giles, Pit/SD	492	147	34	6	20	0.299	(41)	0.514	(39)	0.941	(14)	0.813	(35)	0.299	0.122	0.053	0.041	69	(30)
S. Rolen, StL	559	160	49	1	28	0.286	(71)	0.528	(28)	0.910	(29)	0.814	(33)	0.286	0.140	0.052	0.050	68	(31)
L. Berkman, Hou	538	155	35	6	25	0.288	(63)	0.515	(38)	0.927	(19)	0.803	(39)	0.288	0.123	0.058	0.046	65	(32)
C. Beltran, KC	521	160	14	10	26	0.307	(26)	0.522	(34)	0.911	(27)	0.829	(26)	0.307	0.096	0.069	0.050	62	(33)
J. Bagwell, Hou	605	168	28	2	39	0.278	(85)	0.524	(33)	0.897	(34)	0.802	(40)	0.278	0.114	0.068	0.064	62	(33)
J. Kent, Hou	505	150	39	1	22	0.297	(47)	0.509	(45)	0.860	(52)	0.806	(38)	0.297	0.123	0.046	0.044	59	(35)
E. Chavez, Oak	588	166	39	5	29	0.282	(75)	0.514	(40)	0.864	(48)	0.796	(43)	0.282	0.124	0.058	0.049	59	(35)
M. Lowell, Fla	492	136	27	1	32	0.276	(91)	0.530	(27)	0.881	(38)	0.807	(37)	0.276	0.122	0.067	0.065	59	(35)
J. Posada, NYY	481	135	24	0	30	0.281	(78)	0.518	(36)	0.922	(22)	0.798	(41)	0.281	0.112	0.062	0.062	58	(38)

Legend: AB .. HR: batting statistics, Source: http://sports.espn.go.com/mlb/stats/batting?league=mlb, BA, SLG, BPS, OPB: linear weight scores (ranks); shading: (above the 10, 50, and 90 percentile), c1B=BA, c2B=(2B+3B+HR)/AB), c3B=(3B+HR)/AB, and cHR=HR/AB), UScr: u-scores (ranks)

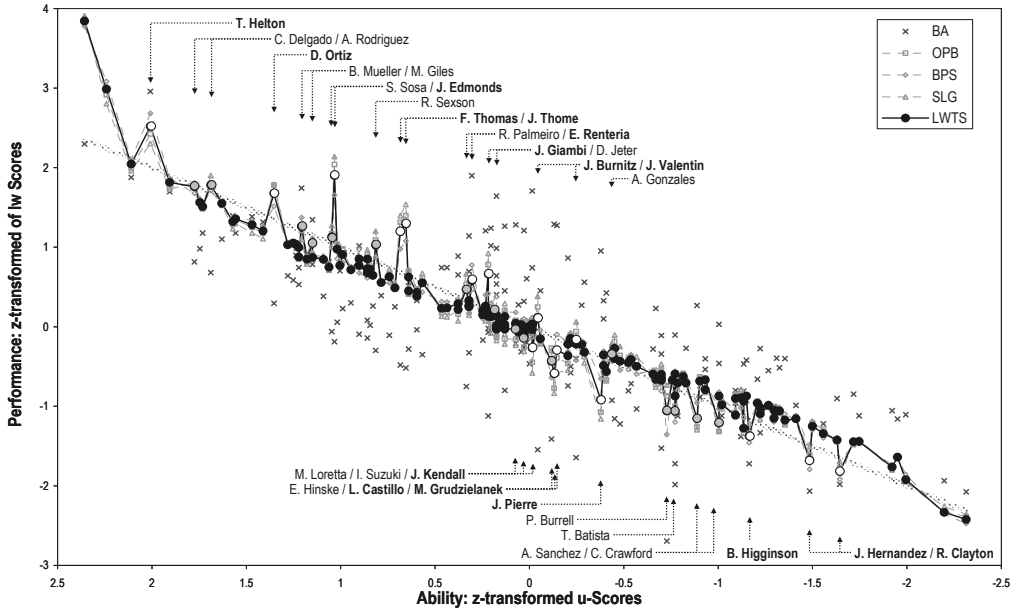


Figure 3: Correlation between μ- and z-transformed lw scores (BA (×), SLG, BPS, OPB, LWTS) among 2003 MLB Batters (see Table 1 for sources and text for details)

3.3. Ranking Countries by Olympic Medal Profiles

At the 2002 Winter Olympics, $m = 25$ countries won at least one medal. Let the number of gold, silver, and bronze medals for country C_k be g_k, s_k, and b_k, respectively. Commonly used weighting schemes assign identical weights to all types of medals, linear weights, e.g., ($b = 1$, $s = 2$, $g = 3$), give gold medals an additional bonus point, or rank countries first by gold medals using other medals only to break ties. Rewriting the latter hierarchical schemes shows it to be a special case of lw schemes where a ceiling $\lceil x \rceil$ is an arbitrary integer larger than x:

Identical:	$IScr = g + s + b$
Linear:	$LScr = 3g + 2s + 1b$
Exponential:	$EScr = 2^2 g + 2^1 s + 2^0 b$
Hierarchical:	$HScr = \left(\lceil \max_k b_k \rceil \lceil \max_k s_k \rceil\right) g + \left(\lceil \max_k b_k \rceil\right) s + b$

In Table 2, the ranks IRg, LRg, ERg, and HRg, based on these weighting schemes agree only for the most extreme cases, Germany, Slovenia, and Belarus. Austria and Finland are ranked 6:12 with IRg or 12:8 with HRg. Countries differ by as much as dRg = 6.5 (Austria) and 7.0 (Sweden) ranks. Comparing the U.S. to Norway highlights a shortcoming of hierarchical weights. To have a single gold medal more than the U.S. is sufficient for Norway to lead. μ-scores agree with the other scores that the U.S., with almost twice as many silver and bronze medals, should score better. For Estonia vs. Sweden, however, one might argue that one silver and three bronze medals do not compensate for the lack of a gold medal, i.e., hierarchical weights are more appropriate than linear scores. Again, μ-scores agree with the 'common sense' assessment. As these examples show, any fixed set of weights to reflect the higher value of gold vs. bronze medals may be difficult to justify, while μ-scores, being more flexible, cover a 'middle ground'.

A partial ordering can be depicted as a lattice, i.e., a directional graph, where nodes (countries) are connected by edges whenever their pairwise ordering can be decided. Unlike trees, lattices can have loops. In Figure 4, one path from Bulgaria to Finland passes Sweden, the other Britain, Estonia, Korea, and Croatia. Switzerland has more gold medals than Austria, so that it would rank higher with hierarchical weights, while Austria has more medals in total, so that it would rank higher with identical weights. As the order depends on the choice of weights, these two countries are not connected by a line.

As with lw scores, countries can be tied (having the same rank) but μ-score ties come in two flavors (Wittkowski 1998). Australia and Spain, as well as Italy and France, have identical rows of signs (exact tie). Great Britain also has an μ-score of 18.0, but its pairwise orderings include Bulgaria and Estonia (inexact tie).

Table 2: Medals won at the 2002 Winter Olympics by country with scores and ranks for u-statistics vs. various linear model weighting schemes (countries sorted by u-scores)

Country	G	s	b	IScr	IRg	LScr	LRg	EScr	ERg	HScr	HRg	MnRg	MxRg	dRg
Germany	12	16	7	35	1.0	75	1.0	87	1.0	121607	1.0	1.0	1.0	.0
Norway	11	7	6	24	3.0	53	3.0	64	3.0	110706	2.0	2.0	3.0	1.0
U.S.A.	10	13	11	34	2.0	67	2.0	77	2.0	101311	3.0	2.0	3.0	1.0
Canada	6	3	8	17	4.0	32	5.0	38	5.0	60308	5.0	4.0	5.0	1.0
Russia	6	6	4	16	5.5	34	4.0	40	4.0	60604	4.0	4.0	5.5	1.5
Italy	4	4	4	12	7.0	24	7.0	28	6.5	40404	7.0	6.5	7.0	.5
France	4	5	2	11	8.5	24	7.0	28	6.5	40502	6.0	6.0	8.5	2.5
Austria	2	4	10	16	5.5	24	7.0	26	8.0	20410	12.0	5.5	12.0	6.5
Switzerland	3	2	6	11	8.5	19	9.5	22	9.5	30206	10.0	8.5	10.0	1.5
Netherlands	3	5	0	8	10.5	19	9.5	22	9.5	30500	9.0	9.0	10.5	2.5
Finland	4	2	1	7	12.0	17	11.0	21	11.0	40201	8.0	8.0	12.0	4.0
China	2	2	4	8	10.5	14	12.0	16	12.0	20204	13.0	10.5	13.0	2.5
Croatia	3	1	0	4	14.5	11	13.0	14	13.0	30100	11.0	11.0	14.5	3.5
Korea	2	2	0	4	14.5	10	14.0	12	14.0	20200	14.0	14.0	14.5	.5
Estonia	1	1	1	3	17.0	6	17.0	7	18.0	10101	17.0	17.0	18.0	1.0
Sweden	0	2	4	6	13.0	8	15.0	8	16.0	204	20.0	13.0	20.0	7.0
Australia	2	0	0	2	21.0	6	17.0	8	16.0	20000	15.5	15.5	21.0	5.5
Spain	2	0	0	2	21.0	6	17.0	8	16.0	20000	15.5	15.5	21.0	5.5
Great Britain	1	0	2	3	17.0	5	19.0	6	19.0	10002	18.0	17.0	19.0	2.0
Bulgaria	0	1	2	3	17.0	4	20.5	4	21.0	102	21.0	17.0	21.0	4.0
Czech Rep.	1	0	1	2	21.0	4	20.5	5	20.0	10001	19.0	19.0	21.0	2.0
Poland	0	1	1	2	21.0	3	22.5	3	22.5	101	22.5	21.0	22.5	1.5
Japan	0	1	1	2	21.0	3	22.5	3	22.5	101	22.5	21.0	22.5	1.5
Slovenia	0	0	1	1	24.5	1	24.5	1	24.5	1	24.5	24.5	24.5	.0
Belarus	0	0	1	1	24.5	1	24.5	1	24.5	1	24.5	24.5	24.5	.0

Legend: g/s/b: Number of gold, silver, and bronze medals, respectively. IScr/IRg: Scores and ranks for identical (1:1:1) weighting. LScr/LRg: Scores and ranks for linear (3:2:1) weighting. EScr/ERg: Scores and ranks for exponential (4:2:1) weighting. HScr/HRg: Scores and ranks for hierarchical (10000:100:1) weighting. MnRg/MxRg: Minimum and maximum among the four ranks. dRg: MxRg – MnRg

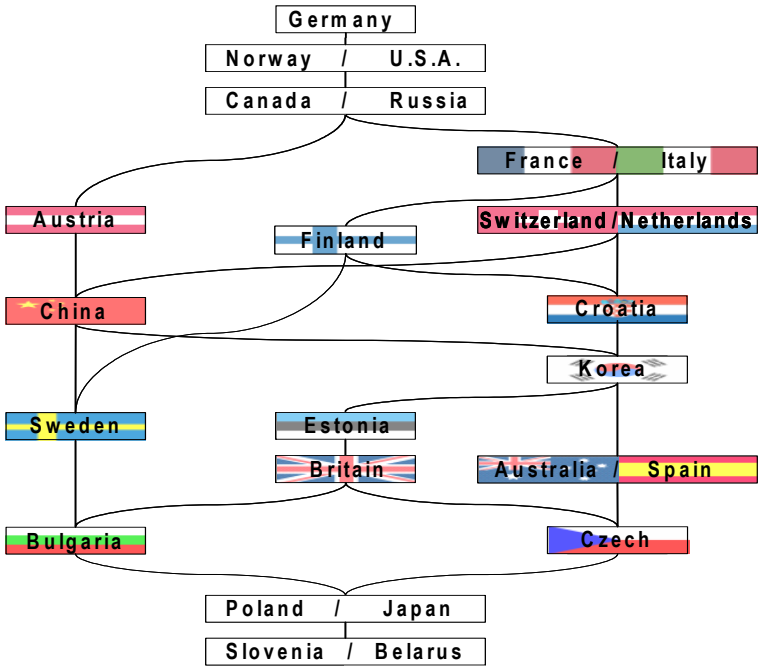


Figure 4: Lattice structure of countries by Salt Lake City medal profiles (see Table 2 for data). Connecting lines indicate the pairwise orderings that are independent of the choice of the lw scoring system. Countries displayed next to each other, e.g., France and Italy, are equivalent with respect to u-scores, i.e., they have the same pairwise orderings with respect to all other countries.

One limitation of μ-scores is that information content tends to decrease as the number of variables increases, because it becomes more likely that at least one pair of variables has orderings with different directions. In the extreme, all pairwise orderings for a given subject could become ambiguous, rendering the resulting score for this subject non-informative. On the other hand, some variables could be related to the same 'factor', i.e., are highly correlated with each other. For instance, several of the twelve cross-country disciplines, the 10 downhill disciplines, and the 10 speed skating disciplines are often won by the same athlete or country. In Table 3, the Netherlands and Korea won all their eight medals in speed skating and short-track skating, respectively, while Croatia won medals in downhill skiing only. Italy and Switzerland, in contrast, won medals in at least three of the main categories (Nordic, alpine, outdoor, and indoor).

Table 3: Number of medals (G: gold, S: silver, B: bronze) by country and hierarchical structure of disciplines. Nordic (N): Cross-Country (CC: 12), Combination (Cb: 4), Biathlon (Bi: 8), Alpine (A: 10): Downhill (DH: 10), Freestyle (FS: 4), Snowboard (SB: 4), Outdoor (O): Bobsleigh (BS: 3), Luge (Lg: 3), Skeleton (Sk: 2), Ski Jumping (SJ: 2), Indoors (I): Curling (Cu: 2), Figure Skating (FS: 4), Ice Hockey (IH: 2), Short Track (SH: 8), Speed Skating (SS: 10)

Country	N-CC			N-Cb			N-Bi			A-DH			A-FS			A-SB			O-BS			O-Lg			O-Sk			O-SJ			I-Cu			I-FS			I-IH			I-SH			I-SS		
	G	S	B	G	S	B	G	S	B	G	S	B	G	S	B	G	S	B	G	S	B	G	S	B	G	S	B	G	S	B	G	S	B	G	S	B	G	S	B	G	S	B	G	S	B
Germany	1	2	1	1	3		3	5	1			1							2	1	1	2	2	1																			3	3	2
Norway	3	4	3				4	2		2	1	1	1																	1													2
U.S.											2			3		2	1	2	1	1	1		1	1	2	1							1		2		2		1	1	1	3	1	4
Canada			1											1	1															1	1		1			2			2	1	3	1		2
Russia	3	3	1				1		2																								2	3			1							
Italy	2	2	1							1	1	1						1				1												1						1					
France								1	1	2	2				1	1	2																1													
Austria		1	1			3			1	2	2	4												1		1																		
Switzerland			1									1				1		1	1	1						1	2		1	1														
Netherlands																																											3	5	
Finland				3	1								1															1	1																
China																																		1						2	2	3			
Croatia										3	1																																	
Korea																																								2	2				
Estonia	1	1	1																																									
Sweden									2		1	1					1																					1							
Australia													1																											1					
Spain	2																																											
Great Britain												1															1			1														
Bulgaria									1																															1	1				
Czech			1										1																																
Poland																												1	1																
Japan															1																												1		
Belarus															1																														
Slovenia						1																																						
	12	13	11	4	4	4	8	8	8	10	10	10	4	4	4	4	4	4	3	3	3	3	3	3	2	2	2	2	2	2	2	2	2	5	3	4	2	2	2	8	8	8	10	10	10

Figure 5 shows how various levels of this hierarchical structure can be represented using functions from the muStat library. With increasing levels of hierarchical order, information content increases from 7.216 to 18.386, indicating the validity of the hierarchical structure. (In fact, the level of information content as-

sociated with hierarchical structures can be used to determine the best 'factor structure', in cases where the prior information does not suffice to determine the relevant 'factors'). Not surprisingly, Italy and Switzerland rank higher in the hierarchical model, while the Netherlands, Korea, and Croatia rank lower.

```
# G <- g, S <- g+s, B <- g+s+b    # gold /silver/bronze

U0 <- mu.score(IR) # equivalent to:
U0 <- mu.score(IR,"(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,...,45)"
U1 <- mu.score(IR,"("%&%
     "(1,2,3),(4,5,6),(7,8,9)," %&%
     "(10,11,12),(13,14,15),(16,17,18)," %&%
     "(19,20,21),(22,23,24),(25,26,27),(28,29,30)," %&%
     "(31,32,33),(34,35,36),(37,38,39),(40,41,42),(43,44,45)"
     %&% ")")
U2 <- mu.score(IR,"("%&%
     "((1,2,3),(4,5,6),(7,8,9))," %&%
     "((10,11,12),(13,14,15),(16,17,18))," %&%
     "((19,20,21),(22,23,24),(25,26,27),(28,29,30))," %&%
     "((31,32,33),(34,35,36),(37,38,39),(40,41,42),(43,44,45))),"
     %&% ")")
```

Figure 5: Wrapper (mu.score) around the functions of Figure 1 to allow hierarchically structured variables to be represented.

3.4. Life-time performance of cyclists

Each summer, the question is raised, who is the best Tour-de-France cyclist of all times and who may take the helm soon. Innumerable rankings, each one more subjective than the others, have been proposed, their large number attesting to the fact that an answer is not easily found. Based on how often a cyclist has carried the yellow jersey along Champs-Elysées to the goal, Lance Armstrong with 7 wins is the best Tour cyclist of all times. Jacques Anquetil, Bernard Hinault, Miguel Indurain, and Eddy Merckx, all won 'only' 5 times. However, other criteria also play a role in determining the ability of a cyclist. Most cyclometricians will include the number of other victories (e.g., green and red polka dot jerseys). Is Armstrong with his 7 final and 22 day victories really better than Merckx with 5 yellow jerseys, 34 day victories, 3 green, and 2 red polka dot jerseys?

Assuming the order: yellow jersey (Y) > second place in the final results (2) > third place in the final results (3) > stage victory (S) > green (G) or red (R) polka dot jersey (the order of the latter unknown), we account for the different importance by defining cumulative variables:

```
cY <- nY
c2 <- nY + n2
c3 <- nY + n2 + n3
cS <- nY + n2 + n3 + nS
cR <- nY + n2 + n3 + nS + nR
cG <- nY + n2 + n3 + nS + nG
```

With this, we can again describe the scores by the formulae used with the muStat function `mu.score(x,frml)`. With censored data, `mxY` is either `Y` (retired cyclist) or `99` (active cyclist):

```
cCurr <- "cY, c2, c3, cS, cG, cR"
cCens <- "[cY:mxY),[c2:mx2),[c3:mx3),[cS:mxS),[cG:mxG),[cR:mxR)"
```

The top part of Figure 6 shows the top 18 cyclists. AI μ-scores rank cyclists by the jerseys already won. The first place is shared between Armstrong and Hinault. Both dominate 524 cyclists, but are dominated by none. Merckx, who ranks third, is also dominated by nobody (not even Armstrong or Hinault), but dominates two fewer cyclists (Zoetemelk, Ullrich). Anquetil ranks fourth (dominating 520 cyclists but being dominated by Armstrong, Hinault and Merckx).

For active cyclists (bottom part of Figure 6), the final counts are not yet known. Still, they can be scored by how many cyclists they are already domineering. By AI μ-scores, Eric Zabel tops the active cyclists by dominating 477 cyclists (active cyclists cannot be dominated). The corresponding scores for 'retired' cyclists appear in the last column under 'H'.

Rnk (adj)	Cyclist (Year of Birth)	Ylw	2nd	3rd	Stg	Grn	Red	"worse" Cyclists	"better" Cyclists	μ-score AI	(R / H)
1	Armstrong, Lance	7	0	0	22	0	0	524 (526/479)	0 (1/ 0)	524	(525/479)
1	Hinault, Bernard	5	2	0	28	1	1	524 (526/479)	0 (1/ 0)	524	(525/479)
3	Merckx, Eddy	5	1	0	34	3	2	522 (524/477)	0 (3/ 0)	522	(521/477)
4	Anquetil, Jacques	5	0	1	16	0	0	520 (522/475)	3 (4/ 3)	517	(518/472)
5	Indurain, Miguel	5	0	0	12	0	0	516 (521/472)	4 (5/ 5)	512	(516/467)
6	Zoetemelk, Joop	1	6	0	10	0	0	513 (520/469)	2 (6/ 3)	511	(514/466)
7	Van Impe, Lucien	1	1	3	9	0	6	511 (519/467)	4 (7/ 5)	507	(512/462)
8	Ullrich, Jan	1	5	1	7	0	0	508 (518/464)	3 (8/ 4)	505	(510/460)
9	Thévenet, Bernard	2	1	0	9	0	0	509 (517/465)	6 (11/ 7)	503	(506/458)
9	Fignon, Laurent	2	1	0	9	0	0	509 (517/465)	6 (11/ 7)	503	(506/458)
11	LeMond, Greg	3	1	1	5	0	0	504 (512/460)	5 (14/ 6)	499	(498/454)
12	Bahamontès, Federico	1	1	1	7	0	6	504 (511/460)	7 (16/ 8)	497	(495/452)
12	Bobet, Louison	3	0	0	7	0	0	503 (511/459)	6 (16/ 7)	497	(495/452)
14	Poulidor, Raymond	0	3	5	7	0	0	496 (509/453)	0 (17/ 2)	496	(492/451)
15	Gaul, Charly	1	0	2	10	0	2	501 (508/457)	7 (18/ 8)	494	(490/449)
16	Janssen, Jan	1	1	0	7	3	0	498 (507/454)	10 (19/11)	488	(488/443)
17	Pantani, Marco	1	0	2	8	0	0	498 (506/454)	11 (21/12)	487	(485/442)
18	Gimondi, Felice	1	1	0	7	0	0	495 (504/451)	14 (22/15)	481	(482/436)
2 (1)	Contador, Alberto (1982)	≥1	≥0	≥0	≥1	≥0	≥0	458 (517/499)	n/a (10/ 2)	458	(507/497)
5 (3)	Boonem, Tom (1980)	≥0	≥0	≥0	≥6	≥1	≥0	442 (503/479)	n/a (18/10)	442	(485/469)
3 (2)	Basso, Ivan (1977)	≥0	≥1	≥1	≥1	≥0	≥0	455 (498/479)	n/a (11/24)	455	(487/455)
3 (5)	Klöden, Andreas (1975)	≥0	≥1	≥1	≥0	≥0	≥0	455 (490/471)	n/a (12/28)	455	(478/443)
1 (4)	Zabel, Erik (1970)	≥0	≥0	≥0	≥12	≥6	≥0	477 (483/477)	n/a (4/42)	477	(479/435)
6 (6)	Evans, Cadel (1977)	≥0	≥1	≥1	≥2	≥2	≥2	435 (488/459)	n/a (22/24)	435	(466/435)
6 (6)	Pereiro, Oscar (1977)	≥0	≥1	≥1	≥2	≥2	≥2	435 (488/459)	n/a (22/24)	435	(466/435)

Legend:

AI (age-independent): top: `mu.score(x,cCurr)` bottom: `mu.score(x,cCens)`
R (rescoring): `mu.score(cbind(mu.score(x,cCens),YOB))`
H (hierarchical): `mu.score(cbind(x,YOB), "(" %&% cCens %&% "), cYOB")`

Figure 6: Top life-time Tour-de-France cyclists 1953–2006 (top), top active cyclists 2007 (bottom) and their age-independent μ-scores (see text for details)

Clearly, age belongs to a different 'factor' than the wins. Thus, adding year-of-birth as another variable in the same category as the various wins may cause loss of information content. The 'R' μ-scores result from the naïve approach of rescoring the performance scores together with age. When age is accounted for, Contador takes the helm from Zabel, because he is 12 years younger. However, this approach has drawbacks, as differences in information content are not accounted for.

Hierarchical 'H' μ-scores adjust for age without causing this fallacy. Again, Contador ranks first among the active cyclists. Still, with only two wins by Contator, these results have to be taken with a grain of salt, yet the young Contador (U.S. team Discovery, current life-time rank: 83 / age-adjusted censored rank: 1) may become the next life-time leader when he returns to the Tour in the 2009 season.

3.5. Scoring Soccer Teams

With additional information available beyond the variables' grading and hierarchy, μ-scores can be easily extended. In soccer, one tries to discourage conservative play by scoring a win higher than two ties. Traditionally, from the 1894 American League of Professional Football to the current FIFA World Rankings, this aim has been achieved by assigning $3:1:0$ weights to wins, ties, and losses. μ-Scores based on cumulating wins w and ties t among n games as $2w/n$ and $(2w+t)/n$ would more closely resemble the concept of a tie counting <50% of a win, rather than specifying a-priori that it should be exactly 33%. Still, when applied to a set consisting of all 15 possible outcomes of four games, μ-scores based on this transformation are equivalent to the $3:1:0$ lw scores, except that some ties are broken, giving a slightly higher advantage to wins (Figure 7).

Figure 7: Computation of soccer u-scores (2 ties < 1 win) based on the cumulative counts 2W (2 × #wins) and 2W+T (… plus ties) and comparison with soccer lw scores (tie = 1/3 win) based on the sum 3W=T

Team	Wins	Ties	Loss	3W+T	u(3W+T)	2W	2W+T	(0,0,4)	(0,1,3)	(0,2,2)	(0,3,1)	(1,0,3)	(0,4,0)	(1,1,2)	(1,2,1)	(1,3,0)	(2,0,2)	(2,1,1)	(2,2,0)	(3,0,1)	(3,1,0)	(4,0,0)	#<	#>	U
(0,0,4)	0	0	4	0.00	-14	0.00	0.00	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	14	-14
(0,1,3)	0	1	3	0.25	-12	0.00	0.25	1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	13	-12
(0,2,2)	0	2	2	0.50	-10	0.00	0.50	1	1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	2	12	-10
(0,3,1)	0	3	1	0.75	-7	0.00	0.75	1	1	1	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	3	10	-7
(1,0,3)	1	0	3	0.75	-7	0.50	0.50	1	1	1	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	3	9	-6
(0,4,0)	0	4	0	1.00	-3	0.00	1.00	1	1	1	1	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	4	8	-4
(1,1,2)	1	1	2	1.00	-3	0.50	0.75	1	1	1	1	1	0	0	-1	-1	-1	-1	-1	-1	-1	-1	5	8	-3
(1,2,1)	1	2	1	1.25	0	0.50	1.00	1	1	1	1	1	1	1	0	-1	-1	-1	-1	-1	-1	-1	7	7	0
(1,3,0)	1	3	0	1.50	3	0.50	1.25	1	1	1	1	1	1	1	1	0	0	-1	-1	-1	-1	-1	8	5	3
(2,0,2)	2	0	2	1.50	3	1.00	1.00	1	1	1	1	1	1	1	1	0	0	-1	-1	-1	-1	-1	8	5	3
(2,1,1)	2	1	1	1.75	6	1.00	1.25	1	1	1	1	1	1	1	1	1	0	-1	-1	-1	-1	-1	10	4	6
(2,2,0)	2	2	0	2.00	8	1.00	1.50	1	1	1	1	1	1	1	1	1	1	0	-1	-1	-1	-1	11	3	8
(3,0,1)	3	0	1	2.25	10	1.50	1.50	1	1	1	1	1	1	1	1	1	1	1	0	-1	-1	-1	12	2	10
(3,1,0)	3	1	0	2.50	12	1.50	1.75	1	1	1	1	1	1	1	1	1	1	1	1	0	-1	-1	13	1	12
(4,0,0)	4	0	0	3.00	14	2.00	2.00	1	1	1	1	1	1	1	1	1	1	1	1	1	0	-1	14	0	14

Thus, as in the case of LWTS weights in baseball, μ-scores provide a formal justification for lw scores based on intuition and experience, yet have significant advantages. In practice, however, one might want to take advantage of μ-scores being adaptive to the particular configuration of observed outcomes.

4. DISCUSSION

For Olympic medals (Section 3.3), Morton (2002) suggested a parametric solution to account for population size P and gross domestic product D. Fitting u-scores based on his model $\mathrm{u}\left(\left(\gamma G+\sigma S+B\right)\Big/\left(P^{\pi}D^{\delta}\right)\right)$ against $\mu\left(G,S,B,-P,-D/P\right)$ for the Summer Olympics 2000 yields $\gamma:\sigma:1\approx 5:2:1$, and $\pi\approx\delta\approx 1/2$. Ignoring the countries' characteristics ($\pi=\delta=0$) yields $4:2:1$ (results not shown), the (exponential) weights suggested in Table 2. The counterintuitive result (Morton 2002) that gold medals carry no more weight than silver medals ($2:2:1$, $1:1:1$ before removing the two 'outliers' Cuba and India) shows how even small deviations from model assumptions affect parameter estimates when the "goodness-of-fit surface [is] flat" (*ibid*).

As mentioned in Section 3.1, μ-scores could also be used in biathlon to determine whether the current penalties of 1 min or 150 m for each ring missed should be adjusted. The triathlon example shows how μ-scores fit into the current discussion regarding voting methods. As with utilitarian voting, dropping the requirement of "independence of irrelevant alternatives" (IIA) avoids the consequences of Arrow's (1950) improbability theorem. In addition, μ-scores avoid the need for the subjective choice of constant weights to be assigned to the variables.

Baseball (Section 3.2) is perhaps unique for the effort spent on analyzing the contribution of each type of individual success (hit) to the performance of a team. Lindsey (1963) analyzed 373 games during 1959/60. His tables yield estimates for the expected $1\mathrm{B}:2\mathrm{B}:3\mathrm{B}:\mathrm{HR}$ run value ratio of $1.00:1.80:2.35:3.13$. Thorn and Palmer (1985) then developed the LWTS model, which included how many runs above average each player contributed indirectly. Using 1901–1977 data, the ratio $1.000:1.739:2:217:3.043$, the so-called LWTS model (Albert and Bennett 2003), became a gold standard for measuring performance.

A general manager looking to fill the leadoff or number three/cleanup position in the batting order, however, would still base his decision on the specialty ability measures BA or HR, respectively. Similarly when comparing different candidate players he would be less interested in their performance with the previous team than in their ability to contribute to the new team. Thus, he would need a measure of overall ability when looking for a hitter to be placed later in the batting order, where the loading of the bases is less predictable. Developing such a multivariate measure for ability, however is fundamentally different from developing performance measures such as the LWTS, because no gold standard (RC, RBI) is available against which the model could be fitted empirically. μ-Scores are the first intrinsically valid scoring system to generalize univariate measures of ability (BA, HR) to multivariate measures of ability, thereby measuring a fundamentally different concept than the existing performance measures.

Still, as ability contributes substantially to performance, ability μ-scores and performance lw scores should correlate well, except for players with unusual skills. Removing the 32 players with specific skills (white and gray circles in Figure 3) results in a nearly perfect rank correlation. The general advantage of nonparametric approaches in populations whose heterogeneity is not accounted for in the model explains, in part, why μ-scores are at least comparable to LWTS scores for typical subjects, even though the only information used is the grading of the variables and the current year's data. The reason for the μ-scores to be more robust is that profiles are only compared to profiles that are, in fact, comparable. With μ-scores, a profile with many singles, but few home runs has an ambiguous pairwise ordering to a profile with few singles, but many home runs. With lw scores, in contrast, all profiles are forced to have a projection onto a one-dimensional scale before the comparisons are made. The price paid for having all pairwise comparisons contribute is that the comparisons between lw scores assigned to incomparable profiles depend heavily on the choice of the weights used.

Due to the high correlation between μ- and LWTS scores of 0.986, an evaluation of non-parametric ability scores can draw upon the model based performance scores. Among better batters (Figure 3, left), players with specific skills (Helton, A. Rodriguez, Ortiz, etc.) contribute more to team success (performance) than their overall ability would suggest, because they are strategically placed in the batting order. Weak batters with specific skills (Castillo and Pierre with their >.300 BA, J. Hernandez and Clayton with their >.023 HR, Figure 3, right), contribute below their overall ability, because they often end up in a position where a more balanced skill set is needed.

To ensure the significance of our empirical validation, we computed the 'special skills' index (3) for 2003 and 2004. The 113 players who played in both seasons, had a correlation of 0.45 (p = .0002). As the outliers identified are at least in part due to specific skills exhibited over several years, the deviation between LWTS and μ-scores could be used to identify players with exceptional skills early in their career and place them accordingly in the batting order.

The Tour-de-France data provides an opportunity to discuss μ-scores in the context of the Impossibility Theorem (Arrow 1950). By definition, μ-scores are democratic and universal (all variables and all data contribute), surjective (all rankings are possible), deterministic, and monotonic (improving any of the outcomes cannot worsen this subject's score). The reason for μ-scores to be possible is that the requirement of 'independent of irrelevant alternatives' does not apply. Consider, for instance, Armstrong and Hinault, who are tied for the first place in Figure 6. Their cumulative profiles are

Name	Ylw,	2nd,	3rd,	Stg,	Grn,	Red
Armstrong	7,	7,	7,	29,	29,	29
Hinault	5,	7,	7,	35,	36,	36

A hypothetical cyclist with 33 stage wins would be dominated by Hinault only and, thus, 'break the tie' in his favor. On the other hand, a hypothetical cyclist with 6 yellow tricots would break the tie in favor of Armstrong. Would either of these cyclists be 'irrelevant'? As have others before (Harsanyi 1953; Callander and Wilson 2006) we would argue that an additional cyclist with 6 yellow tricots or 33 stage wins would 'devalue' the 5 yellow tricots of Hinault or the 29 (implied) stage wins of Armstrong, respectively, because the more cyclists achieve a a large number of a particular type of wins, the less exceptional this achievement becomes. Thus, the ability of additional cyclists to affect the previous order of cyclists is, in fact, a desirable feature.

Of course, no statistical approach is a panacea. As there can be no universally optimal scoring system, one needs to match scoring systems to the particular objectives, e.g., use univariate scores (BA, HR) for specialists, LWTS scores for overall performance, and μ-scores for overall ability. Moreover, being based on pairwise orderings within a sample, μ-scores do not provide an absolute measure of ability that could be used across competitions. Still, as we have demonstrated in the Olympics and soccer examples, μ-scores can be used as a 'gold standard' to justify a particular lw score, which then could be used as an absolute measure of ability across populations (years, regions, …).

Another seeming weakness is that μ-scores based on many poorly correlated variables have little information content. This problem can be ameliorated if the variables can be grouped hierarchically into several categories and sub-categories of better correlated variables. In fact, information content could be used to determine which of several hypothetical partial orderings is most adequate (Morales *et al.* 2008; Diana *et al.* submitted).

Still, the more assumptions about the underlying model are being made, the more the resulting scores depend on the accurateness of these assumptions. μ-Scores, however, at least have the advantage over (more-or-less) arbitrary weights that the assumptions being made are easily explained in terms of the underlying model. On the other hand, lw scores can have advantages over μ-scores if the choice of linear weights can be substantiated, as in the LTWS scores for baseball performance. Thus, the choice between traditional model-based vs. the new non-parametric scores will ultimately rely on the confidence in making strong model assumptions.

5. CONCLUSIONS

Multivariate ordinal data (multiple ordered categorical ratings) are frequently observed to assess semi-quantitative characteristics. Both traditional approaches for combining different measures into a utility function and 'utilitarian voting' require that a relative weight is assigned to each measure. However, when the choice of such weights is not easily justified, any analysis based on such a utility function

may be misleading. In decision tree based methods, objects are separated by the most significant criterion first, and each subset is then separated by a subset-specific variable next in the hierarchy. This approach, of which CART (Breiman 1984) is an example, has the advantage of resulting in easily communicated decision strategies. As we have demonstrated, however, these strategies are merely special cases of linear weight functions with extreme weights.

With μ-scores, no additional assumptions need to be made and validated, as long as each variable increases (or decreases) with the unobservable 'latent' factor. By accumulation over various subsets of variables, μ-scores can be generalized to 'graded' count variables. The proposed scores are valid by construction and, thus, no empirical evaluation is needed. In particular, no assumptions need to be made regarding the relative importance of or the correlation among the variables. Relative importance and correlation do not even need to be constant, but may vary with the level of the underlying, often unknown, latent factor. Finally, adding a highly correlated variable is unlikely to affect any of the existing pairwise orderings and, thus, has little or no effect on the scores.

If additional knowledge about the nature of the variables is available, this knowledge can be easily incorporated, either by modifying the transformation, by choosing a different rule to determine pairwise orderings, or be reflecting a hierarchical factor structure. Of course, if the functional relationship between some variables should be known, latent variable models could be used to reduce the dimensionality prior to computing u-scores (Bartholomew and Knott 1999).

An additional benefit of the proposed method is its computational simplicity. From (1), it is clear that the computational effort increases only linearly with the number of variables. Table 2 illustrates that an additional subject increases the matrix of pairwise comparisons by one row and one column. Thus, the computational effort increases only with the square of the number of subjects. Having such a highly efficient algorithm available allows some analyses to be conducted in environments better suited for interactive inspection of the data and intermediate results, providing profound insight into the nature of the algorithm and, thus, into the understanding of the results. Spreadsheet programs for small data sets and teaching are available from http://mustat.rockefeller.edu. For larger data sets and more complex partial orderings, packages for R and S-PLUS are available from http://csan.insightful.com, and http://cran.r-project.org, respectively. When various combinations of variables and hierarchies need to be explored, e.g., to find the factor structure resulting in the highest information content (Diana *et al.* submitted), even these tools may not suffice. Our Web site http://mustat.rockefeller.edu provides access to a grid, where jobs uploaded will be parallelized and executed using optimized libraries (Wittkowski *et al.* 2006; Song *et al.* 2007)

μ-Scores, in general, have provided new insights in the medical fields of toxicology (King *et al.* 2003), addiction (Spangler *et al.* 2004), cancer (Paczesny *et al.*

2004), HIV infection (Arrode *et al.* 2005), immunology (Gottlieb *et al.* 2005), hearing (Schick *et al.* 2006), behavior (Shelley *et al.* 2007), diagnostics (Quaia *et al.* 2007), pharmacogenomics (Haider *et al.* 2008), and fertility (Ramamoorthi *et al.* 2008), but also to machine learning (Sapir *et al.* 2005). This paper discusses the use of μ-scores to score athletes or teams in sports and provides several extensions, including the use of μ-scores for graded variables.

μ-Scores for graded variables can also grade outcomes in fields other than sports. When investigating the relative contribution of immune system components (Oliver 2000), side effects could be graded either as grave > severe > (relatively) benign, or I < II < III < IV < V (death). Similarly, various events indicating the risk of terrorist attacks could be graded as imminent, clear, and weak, or as green/blue, orange, yellow, and red (Wittkowski 2003). Finally, when determining personal traits affecting management decisions (Becker *et al.* 2005), which may themselves be assessed by (ungraded) μ-scores, μ-scores can grade earlier profits higher, so that the relative benefit of traits would not depend on the assumption of a particular interest rate.

References

Albert, J., Bennett J.M. (2003). Curve ball: baseball, statistics, and the role of chance in the game. Revised ed. New York, Springer. 410 p.

Arrode, G., Finke J.S., Zebroski H., Siegal F.P., Steinman R.M. (2005). CD8+ T cells from most HIV-1-infected patients, even when challenged with mature dendritic cells, lack functional recall memory to HIV gag but not other viruses. European Journal of Epidemiology 35(1): 159-170.

Arrow, K.J. (1950). A Difficulty in the Concept of Social Welfare. Journal of Political Economy 58(4): 328–346.

Bartholomew, D.J., Knott M. (1999). Latent Variable Models and Factor Analysis. London, Arnold.

Bassett, G.W. (1997). Robust sports ratings based on least absolute errors. The American Statistician 51(2): 99-105.

Becker, O., Feit T., Hofer V., Leopold-Wildburger U., Schütze J., Selten R. (2005). Das Marketingspiel SINTO und seine Vorzüge als Unternehmensplanspiel. Zeitschrift für Systemdenken und Entscheidungsfindung im Management 4(1): 3-18.

Bennett, J.M. (1993). Did Shoeless Joe Jackson Throw the 1919 World Series? The American Statistician 47: 241-250.

Breiman, L. (1984). Classification and regression trees. Belmont, CA, Wadsworth.

Callander, S., Wilson C.H. (2006). Context-dependent Voting. Quarterly Journal of Political Science 1: 227–254.

Cherchye, L., Vermeulen F. (2006). Robust Rankings of Multidimensional Performances: An Application to Tour de France Racing Cyclists. Journal of Sports Economics 7(4): 359-373.

Cherchye, L., Vermeulen F. (2007). Acknowledgement of Priority. Journal of Sports Economics 8(5): 557-.

Deuchler, G. (1914). Über die Methoden der Korrelationsrechnung in der Pädagogik und Psychologie. Z. pädagog. Psychol. 15: 114-131, 145-159, 229-242.

Diana, M., Song T., Wittkowski K.M. (submitted). Studying travel-related individual assessments and desires by combining hierarchically structured ordinal variables. Transportation.

Gottlieb, A.B., Chamian F., Masud S., Cardinale I., Abello M.V., Lowes M.A., Chen F., Magliocco M., Krueger J.G. (2005). TNF inhibition rapidly down-regulates multiple proinflammatory pathways in psoriasis plaques. Journal of Immunology 175(4): 2721-2729.

Haider, A.S., Lowes M.A., Suarez-Farinas M., Zaba L.C., Cardinale I., Khatcherian A., Novitskaya I., Wittkowski K.M., Krueger J.G. (2008). Identification of cellular pathways of "Type 1," Th17 T cells, and TNF- and inducible nitric oxide synthase-producing dendritic cells in autoimmune inflammation through pharmacogenomic study of cyclosporine a in psoriasis. Journal of Immunology 180(3): 1913-1920.

Harsanyi, J.C. (1953). Cardinal Utility in Welfare Economics and in the Theory of Risk-Taking. Journal of Political Economy 61(5): 434-5.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. Annals of Mathematical Statistics 19: 293-325.

Hughes, M.D., Bartlett R.M. (2002). The use of performance indicators in performance analysis. Journal of Sports Sciences 20(10): 739-754.

James, B. (1982). The Bill James Baseball Abstract. New York, NY, Ballantine.

King, T.P., Jim S.Y., Wittkowski K.M. (2003). Inflammatory role of two venom components of yellow jackets (Vespula vulgaris): a mast cell degranulating peptide mastoparan and phospholipase A1. International Archives of Allergy and Immunology 131: 25-32.

Lindsey, G.R. (1963). An Investigation of Strategies in Baseball. Operations Research 11(4): 477-501.

Morales, J.F., Song T., Auerbach A.D., Wittkowski K.M. (2008). Phenotyping genetic diseases using an extension of μ-scores for multivariate data. Statistical Applications in Genetics and Molecular Biology 7(1): 19.

Morton, R.H. (2002). Who won the Sydney 2000 Olympics?: an allometric approach. Journal of the Royal Statistical Society Series D-the Statistician 51: 147-155.

O'Brien, P.C. (1984). Procedures for comparing samples with multiple endpoints. Biometrics 40(4): 1079-87.

Oliver, S.J. (2000). The Th1/Th2 paradigm in the pathogenesis of scleroderma, and its modulation by thalidomide. Current Rheumatology Reports 2: 486-491.

Paczesny, S., Banchereau J., Wittkowski K.M., Saracino G., Fay J., Palucka A.K. (2004). Expansion of Melanoma-specific Cytolytic CD8+ T Cell Precursors in Patients with Metastatic Melanoma Vaccinated with CD34+ Progenitor-derived Dendritic Cells. Journal of Experimental Medicine 199(11): 1503-1511.

Paton, C.D., Hopkins W.G. (2005). Competitive Performance of Elite Olympic-Distance Triathletes: Reliability and Smallest Worthwhile Enhancement. Sportscience 9: 1-5.

Quaia, E., D'Onofrio M., Cabassa P., Vecchiato F., Caffarri S., Pittiani F., Wittkowski K.M., Cova M.A. (2007). Diagnostic value of hepatocellular nodule vascularity after microbubble injection for characterizing malignancy in patients with cirrhosis. AJR Am J Roentgenol 189(6): 1474-83.

Ramamoorthi, R.V., Rossano M.G., Paneth N., Gardiner J.C., Diamond M.P., Puscheck E., Daly D.C., Potter R.C., Wirth J.J. (2008). An application of multivariate ranks to assess effects from combining factors: Metal exposures and semen analysis outcomes. Statistics in Medicine: online prepublication.

Sapir, M., Verbel D., Kotsianti A., Saidi O. (2005). Live logic (TM): Method for approximate knowledge discovery and decision making. In: Slezak D., Wang G., Szczuka M., Düntsch I., Yao Y. ed. Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Part 1, Proceedings, Springer. Pp. 532-540.

Schick, B., Starlinger V., Haberle L., Eigenthaler M., Walter U., Knipper M. (2006). Delayed formation of actin filaments in the outer pillar head plate of VASP-/- mice. Cells Tissues Organs 184(2): 88-95.

Sen, A. (1986). Social choice theory. In: Arrow K.J., Intrilligator M.D. ed. Handbook of Mathematical Economics. Amsterdam, North-Holland. Pp. 1073-1181.

Shelley, D.N., Dwyer E., Johnson C., Wittkowski K.M., Pfaff D.W. (2007). Interactions between estrogen effects and hunger effects in ovariectomized female mice. I. Measures of arousal. Hormones and Behavior 52(4): 546-553.

Song, T., Coffran C., Wittkowski K.M. (2007). Screening for gene expression profiles and epistasis between diplotypes with S-Plus on a grid. Statistical Computing and Graphics 18(2): 20-25.

Spangler, R., Wittkowski K.M., Goddard N.L., Avena N.M., Hoebel B.G., Leibowitz S.F. (2004). Opiate-like Effects of Sugar on Gene Expression in Reward Areas of the Rat Brain. Molecular Brain Research 124(2): 134-142.

Thorn, J., Palmer P., Reuther D. (1985). The hidden game of baseball. Garden City, NY, Doubleday.

Vleck, V.E., Bürgi A., Bentley D.J. (2006). The Consequences of Swim, Cycle, and Run Performance on Overall Result in Elite Olympic Distance Triathlon. International Journal of Sports Medicine(1): 43-48.

Wittkowski, K.M. (1992). An extension to Wittkowski. Journal of the American Statistical Association 87: 258.

Wittkowski, K.M. (1998). Versions of the sign test in the presence of ties. Biometrics 54: 789–791.

Wittkowski, K.M. (2003). Novel Methods for Multivariate Ordinal Data applied to Genetic Diplotypes, Genomic Pathways, Risk Profiles, and Pattern Similarity. Computing Science and Statistics 35: 626-646.

Wittkowski, K.M., Lee E., Nussbaum R., Chamian F.N., Krueger J.G. (2004). Combining several ordinal measures in clinical studies. Statistics in Medicine 23(10): 1579-1592.

Wittkowski, K.M., Haider A., Sehayek E., Suarez-Farinas M., Pellegrino M., Peshansky A., Coffran C., Coker S. (2006). Bioinformatics tools enabling u-statistics for microarrays. Conf Proc IEEE Eng Med Biol Soc 1: 3464-9.