

Inquiring Country Olympic Profiles through the “Olympic Studies Centre” Data Base

*Ricerca di Profili per i Paesi Partecipanti ai Giochi Olimpici in Base ai
Dati Raccolti nell’Archivio del “Centro Studi Olimpici”*

Luigi Bollani

Dipartimento di Statistica e Matematica Applicata “Diego de Castro”,
Piazza Arbarello n. 8 – Torino, e-mail: bollani@econ.unito.it

Riassunto: Nel confrontare tabelle di sintesi riferite ai giochi olimpici si incontrano sovente difficoltà di comparazione tra varie fonti. Risulta in particolare una criticità legata al dato, che in questo studio si cerca di comprendere e porre in evidenza. A tale scopo l’analisi parte dai dati elementari riferiti alle singole prestazioni sportive e contenuti nel data base del “Centro Studi Olimpici”. L’aggregazione di tali dati comporta scelte di ordine metodologico nel computo dei partecipanti o delle partecipazioni (ogni partecipante può concorrere in più gare): tali grandezze risultano infatti riferibili a diverse fasi di gara, concluse o meno regolarmente dall’atleta. Per le analisi longitudinali si aprono anche problemi relativi alle modificazioni nel tempo dei confini dei Paesi partecipanti e al presentarsi nel tempo di diverse discipline sportive.

Keywords: Statistics in sports, quality of the data, multivariate analysis

1. Introduction

Statistical reports on Olympic Games (presented in different studies) are often difficult to compare. This problem was for instance pointed out by the “United States Olympic Committee” in 2000, when, in a note to a summary report¹, it introduced the following warning: *“the total number of nations, events and athletes in the respective Olympic Games vary greatly in historical publications, but the totals shown above reflect ...”*.

It seems therefore to be evident that a very critical point concerns to the data themselves, and the choices made by different researchers in order to summarize them in statistical reports. For this reason the analysis shown in this paper is carried out starting from the individual participation data, referred to each Olympic competition in all modern Olympic Games (Athens 1896 – Athens 2004 for Summer Games and Athens 1924 – Salt Lake City 2002 for Winter Games²). In order to analyze these data, the “Olympic Studies Centre” data base³ is used. This data base collects by now about 350.000 records (participations in Olympic competitions), referred to 44 editions and 205 Countries. Less than 45.000 records are referred to Winter Games, which include 19 editions and 101 Countries.

As the main methodological aspects in treating Summer and Winter Games are very similar, in the following sections only Winter Games are considered.

¹ It is possible to find this report in the internet following the bibliographic directions.

² At the moment in which this paper is written (December 2005), i.e. before Turin 2006.

³ The “Olympic Studies Centre” is an Organization depending from the “International Olympic Committee”. Its data base is continuously updated. A special thanks to Mrs. Barbara Schenkel, who selected the useful information to be kept during the extraction from the data base.

2. Information drawn from the “Olympic Studies Centre” data base and choices in summarizing data⁴.

The following table 1 presents, as an example, some records containing the main information used in this paper for further analysis.

Table 1: *Examples of records used in the analysis.*

Year	N.O.C.	Country	Athlete	Discipline	Gender	Event	Phase	Rank	Medal
2002	Salt Lake City	Italy	DOLFINI, Angelo	Figure skating	M	individual	Final	90000	
2002	Salt Lake City	Italy	DOLFINI, Angelo	Figure skating	M	individual	Short Program	26	
2002	Salt Lake City	Italy	FONTANA, Silvia	Figure skating	W	individual	Final	10	
2002	Salt Lake City	Italy	FONTANA, Silvia	Figure skating	W	individual	Short Program	11	
2002	Salt Lake City	Italy	GIUNCHI, Vanessa	Figure skating	W	individual	Final	20	
2002	Salt Lake City	Italy	GIUNCHI, Vanessa	Figure skating	W	individual	Short Program	21	
2002	Salt Lake City	Italy	COBISI, Michela	Figure skating	X	pairs	Final	19	
2002	Salt Lake City	Italy	DE PRA, Ruben	Figure skating	X	pairs	Final	19	
2002	Salt Lake City	Italy	COBISI, Michela	Figure skating	X	pairs	Short Program	19	
2002	Salt Lake City	Italy	DE PRA, Ruben	Figure skating	X	pairs	Short Program	19	
2002	Salt Lake City	Italy	FUSAR POLI, Barbara	Figure skating	X	ice dancing	Final	3	Bronze
2002	Salt Lake City	Italy	MARGAGLIO, Maurizio	Figure skating	X	ice dancing	Final	3	Bronze
2002	Salt Lake City	Italy	FUSAR POLI, Barbara	Figure skating	X	ice dancing	Short Program	3	Bronze
2002	Salt Lake City	Italy	MARGAGLIO, Maurizio	Figure skating	X	ice dancing	Short Program	3	Bronze
2002	Salt Lake City	Italy	FAIELLA, Federica	Figure skating	X	ice dancing	Final	18	
2002	Salt Lake City	Italy	SCALI, Massimo	Figure skating	X	ice dancing	Final	18	
2002	Salt Lake City	Italy	FAIELLA, Federica	Figure skating	X	ice dancing	Short Program	17	
2002	Salt Lake City	Italy	SCALI, Massimo	Figure skating	X	ice dancing	Short Program	17	

This very short example shows that, depending on the interest in reporting, it is possible to count as follows: 18 total participations (all the records); 9 participations in a final phase; 8 participations in a final phase concluded by the athlete (rank 90000 or more means “did not finish”, “disqualified”, etc.); 5 competitions arrived to a final phase; 9 participants; 8 participants which concluded a final phase; and so on.

So it is necessary to clarify the choices in reporting, specially referring to the count of participations or participants, to the count of only the final phases or also the other ones (like the short programs in the example) and to the count, in pairs or team competitions, of all the team members or of only one unit for the group.

In further analysis the count of participation in final phases, correctly concluded by the athlete, is assumed.

Another problem as to the data arises, in longitudinal studies, because of the different starting time in Country participations and because of the geographical or political changes of some Countries during the modern Olympic period. For instance Germany competed united in the Olympic Games until 1945, then it was present as both the Federal Republic and the Democratic Republic⁵ and at last, starting from 1990, it competed united again.

Because of this kind of behavior, common to many Countries in the whole modern Olympic period, an aggregation for the areas of Germany, Russia, Yugoslavia and Czechoslovakia is carried out in further longitudinal analysis.

⁴ Some remarks and warnings pointed out in this section were already discussed in Bollani (2004) and presented in the Italian *XLI SIEDS Scientific Meeting*. In this paper further observations and analysis related to the different disciplines, their presence during the modern Olympic period and their relationships with the competing Countries are also introduced. Moreover, before the presentation of the results predicted for June 2006, the new data about Torino 2006 Games will be available.

⁵ In this period the few participations resulting for the United Team of Germany depend on particular zones of the Federal Republic controlled by France.

Another warning concerns the disciplines, which have sometimes changed during the modern Olympic period, as is shown in table 3 for Winter Games.

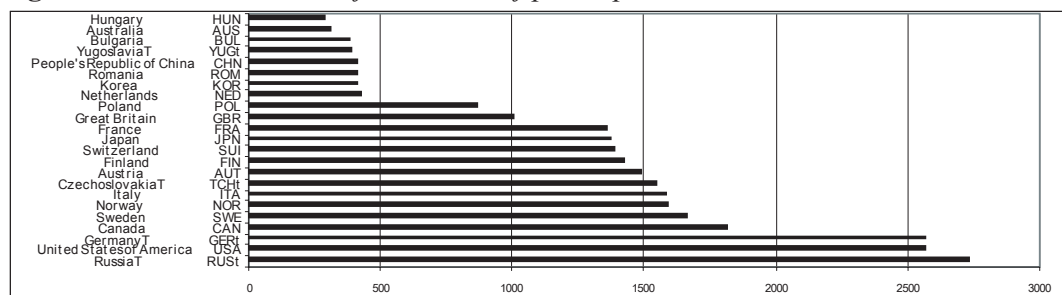
Table 3: *Winter disciplines in time (percentages of athlete participations)*

Discipline	Year																			Total
	'24	'28	'32	'36	'48	'52	'56	'60	'64	'68	'72	'76	'80	'84	'88	'92	'94	'98	'02	
Alpine Skiing				1,2	6,5	5,9	6,3	5,5	6,2	6,1	4,7	5,1	4,3	5,5	8,1	10,3	7,1	9,8	7,6	100
Freestyle Skiing																35,6	21,1	15,0	28,3	100
Speed Skiing																100				100
Snowboard																		42,8	57,2	100
Ski Jumping	0,3	2,1	1,9	2,7	2,6	2,4	2,9	2,5	5,9	6,5	6,1	6,1	5,5	6,3	8,9	9,6	9,1	9,9	8,6	100
Bobsleigh	1,3	4,9	4,1	4,1	3,8	3,9	5,3	0,0	4,5	4,9	4,5	5,6	4,2	6,4	7,5	8,9	8,5	8,3	9,6	100
Biathlon	0,6							1,1	1,8	4,0	3,7	4,0	5,5	6,8	7,3	16,1	14,8	14,8	19,8	100
Cross Country S.	0,2	1,2	1,0	2,6	2,3	2,7	4,1	3,3	5,0	5,2	5,0	5,7	4,7	7,2	7,8	10,9	9,8	12,2	9,1	100
Curling	7,7															27,6		33,2	31,5	100
Figure skating	1,3	3,5	2,3	4,9	6,2	3,6	4,5	3,9	4,9	5,5	3,8	5,5	4,7	6,4	6,4	7,3	9,5	8,1	7,6	100
Speed skating	0,7	2,2	0,7	3,2	3,4	3,3	4,1	5,6	6,7	6,5	5,6	5,7	6,8	7,6	7,4	8,1	7,2	7,9	7,5	100
Ice Hockey	1,5	1,5	1,5	1,5	4,5	5,0	5,2	5,2	4,1	4,4	3,6	3,4	3,7	5,3	4,1	8,4	8,5	13,6	15,1	100
Luge									6,6	9,1	9,8	10,6	7,9	8,3	9,1	9,2	10,0	9,1	10,4	100
Nordic Combined	0,7	3,2	3,7	5,2	4,3	2,5	4,0	3,5	3,4	4,6	4,4	3,7	3,3	3,2	8,0	8,5	9,9	10,4	13,6	100
Short Track S.																25,2	40,7	9,8	24,4	100
Skeleton		14,3			16,1														69,6	100

So in an analysis of the discipline profiles it is necessary to pay attention to the ones which were more recently introduced or which were not present for many editions. Then, considering associations between disciplines and Countries, it is important to take into account whether some later participant Countries seem to show stricter relationships than the others with the disciplines more recently introduced, and this because the analysis can be drawn only on the data concerning the last time period.

Otherwise, a facilitation in reports and representations consists in the high concentration of Country participations. Figure 1 shows the first 23 Countries, in order of number of participations, which allow us to lose less than 10% of the total participations. In further analysis and graphical representations only these Countries are shown.

Figure 1: *First 23 Countries for number of participations*



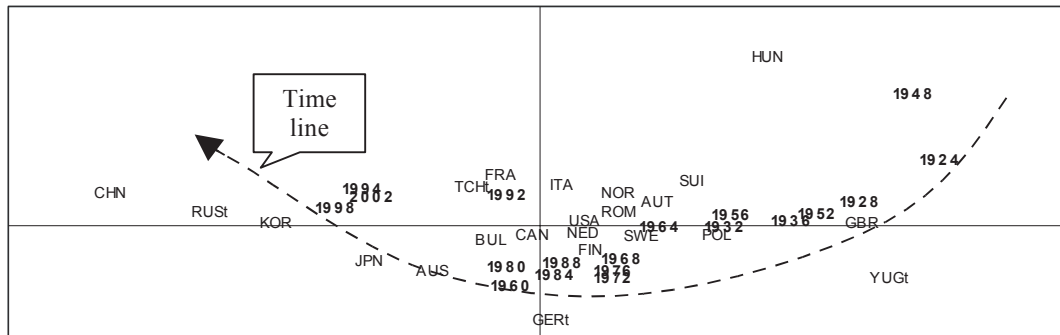
3. Examples of Country profile graphical reports for Winter Games and some warnings in their reading.

In order to find an association structure, referred to Olympic athlete participation, between years and Countries, a symmetric correspondence analysis was drawn.

The first factorial plain, presented in figure 2, shows how the individual Countries had a different participation rate in time, which roughly follows (for groups of years) the sketched line. So the Countries on the right (like Hungary, Great Britain and Yugoslavian area) have a larger participation rate in older games; on the contrary the Countries on the left (like Korea, Russian area and China), have a larger rate in the more

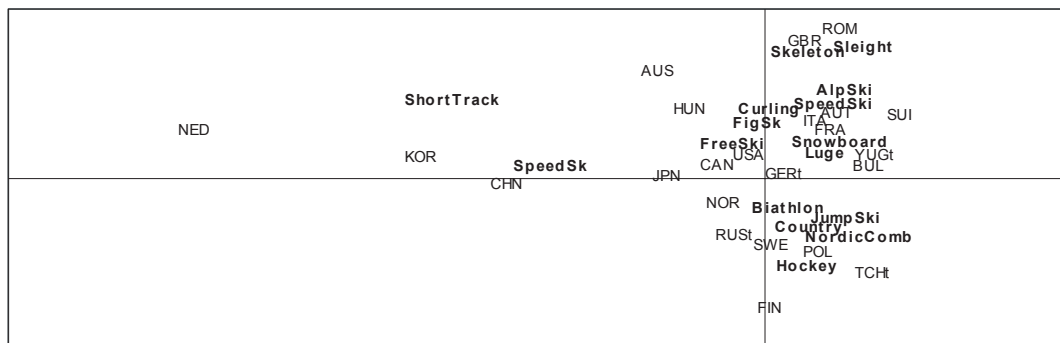
recent games. For a correct interpretation of the sports phenomenon, it is however important to take into account that some Countries (like China and - a little less- Korea) started their participation more recently than the other Countries and so their profile is not comparable with those of the other ones⁶.

Figure 2: *Athlete participation by years and Countries*⁷



In figure 3 the association structure between Countries and disciplines is shown. Also in this case it is important to take into account that Korea and China, which compare on the same side of some recently introduced disciplines (like short tracking and speed skiing), have a not comparable profile with the other ones.

Figure 3: *Athlete participation by years and Countries*⁸



References

- Bollani L. (2004) Giochi olimpici e profili dei Paesi partecipanti, *Rivista Italiana di Economia Demografia e Statistica*, Volume LVIII NN. 1/2, 247-256.
- Mussino A. (2002) *Compass 2002 Progress report*, on the internet site of “Compass”, University of Rome “La Sapienza”, url: <http://w3.uniroma1.it/compass/report/Italy.doc>
- United States Olympic Committee (2000) *Chronology of the Olympic Games*, url: www.olympic-usa.org/education/general_olympic_education/olympic_games_summary.pdf

⁶ If the analysis interest were focused on these Countries, a better comparison would be done considering also the other Countries in the same time period in which the first ones competed.

⁷ The first factorial axis explains 42% of the total inertia and the second one 16% (the sum is 58%). Looking at the square cosines diagnostics it is important to pay attention to the positions of Bulgaria, the Netherlands, Canada and Australia, which do not have a good planar representation.

⁸ The first factorial axis explains 34% of the total inertia and the second one 26% (the sum is 60%). Some warnings in the positions of Free Skiing and Curling are to be formulated.