

Methodological Issues and First Results of a Study to Identify and Evaluate the Error of Some Tourism Statistics

*Problemi Teorici e Primi Risultati di una Indagine per l' Individuazione e la Valutazione dell'Errore Presente in Alcune Statistiche del Turismo.*¹

Anna Maria Parroco, Franco Vaccina
Dipartimento di Metodi Quantitativi per le Scienze Umane
Viale delle Scienze, 90128 Palermo
e-mail: parroco@unipa.it

Riassunto: In questo articolo si analizzano alcune misure del fenomeno turistico, mettendone in evidenza dei difetti legati alla modalità di aggregazione dei dati elementari. In particolare si fa riferimento alle rilevazioni che riguardano gli arrivi, le presenze e la permanenza media dei turisti negli esercizi ricettivi e ai correlati empirici utilizzati per la loro determinazione. Evidenziato il motivo e l'origine del problema, si propongono i primi risultati di una indagine realizzata per valutarne la presenza in un distretto turistico della Sicilia.

Keywords: official tourism statistics, quality of statistic aggregates

1. Introduction

This study examines some quality aspects of official tourism statistics; attention is paid to both contents, that is objects, population, and type of statistical measures, and time aspects.

Tourism Statistics often used refer to tourist arrivals and average stay. It is obvious that both spatial and time aspects are their characteristics. Given the economic importance of tourism in many countries, tourism managers and policy makers need to know tourism distinctive features in order to make proper use, in the development of practical sustainable territorial policies, of the information collected by official institutions in charge of producing and publishing tourism statistics.

According to the recommendations on tourism statistics (World Tourism Organisation, WTO), arrivals are defined as tourists (overnight visitors) registered at lodging establishments in a specific time period (months, years), nights spent by tourists are measured as the number of nights they spend in lodging establishments, and average length of stay is defined by the ratio of nights spent to arrivals.

This way to register arrivals correspond to a substitutive process of measurement, in particular it is noted that the demand dimension (number of arrived persons –tourists- in a specific time period) is measured by the supply side (proxy variable); that is the number of “new occupied bed place” is measured instead of arrived tourists. Therefore

¹ The present paper is financially supported by MIUR, Research found, Prin, 2003 “Metodi e Tecniche per la valutazione statistica del sommerso. Quando la vacanza diviene turismo”.

leading to a measurement bias in tourism statistics, due to a discrepancy between the “real number of tourists” at a destination and the measure of arrivals.

The purpose of this study is to explain the cause of the above mentioned problem and to describe an exploratory study we realized to identify and evaluate the gap pointed out.

2. Methodological Issues

In a recording system in which the unit-time is a day and the space-unit is the single accommodation establishment, arrivals at destination, measured and published in the

official statistics, are calculated as $A = \sum_{j=1}^k \sum_{t=1}^T A_{jt}$ where A_{jt} are the arrivals at accommodation j in the unit-time t .

As already stated this recording system correspond to a substitutive process of measurement: the number of arrived persons in a specific time period is measured from the supply side by the number of “new bed place occupied”. Therefore arrivals (considered like tourists) are oversized and for each tourist the wrong size depends on the number of overnights stay. Particularly, for each tourist the size of the error assumes a maximum value of $(n-1)$ where n equal to overnights stay and when he moves every day from one to another lodge of the same territory.

Average length of stay is obviously undersized by this multiplication of “apparent” tourists.

According to us, the indicator used is perfect only for a time unit and for a single point of observation while it becomes greatly biased when considering a territory where data do not show the same properties of the selected indicator (for variables like, for example, nationality, consumer behaviour, etc.). In fact, available data are not independent and so cannot be aggregated making an addition. The more we collapse territorial units, for example summing up arrivals of each region to obtain arrivals of a larger territory, the higher the probability is of considering each tourist more than once. Moreover, several other factors can influence tourist mobility during their vacations and we will synthetically represent them as “tourism targets”. At the same time, various tourism targets, with their own disposition to mobility, can cause effects on the aggregated data which can be considerable different from each others.

Finally, it is important to point out that a problem of “replications”, that occurs when spatial and time data are aggregated, is pointed out in the “Methodological Notes” of the ISTAT survey on tourism statistics (2004), but it does not refer at all to census data. However, the marked question has a different nature from that we have pointed out².

The World Tourism Organisation (WTO) and the Italian Office for Exchange (UIC), only as minor information, acknowledge the presence of a “replication”, specifying that data refer to the number of arrivals and not to the number of persons. Moreover, data

² Regarding to the quarterly survey on tourism statistics, based on a sample of Italian families so called “Viaggi e Vacanze” (ISTAT), ISTAT points out that some replication could threaten the validity of the total quarterly data for the number of travellers. If travels are distinguished from travellers, it is correct adding the quarterly data relating to the first aspect (the travels), while for the second (the travellers) it is suitable to calculate their arithmetic mean, because it is an estimate of the same aggregate, making use of independent sample, of the same population.

show arrivals by country of origin disregarding the fact that they are not measuring the number of arrived persons. Although the WTO acknowledge the problem, the published data still combine persons' attributes to aggregates that do not represent persons.

3. The Empirical Research

The technical problem, connected to the bias we have pointed out, is to evaluate the repeated data. Indeed, lodging establishments do not differentiate tourist arriving from other accommodations creating repeated measurements.

Clearly, time, space and tourist habits are the variables of interest to obtain a solution of the problem. It is important to underline that the longer is the time period and/or the greater is the given territory (and more space there is between different accommodations), the more "replications" can occur.

The source of data for this analysis is an exploratory research, conducted with the collaboration of the Sicilian hotel manager association, in the Cefalù district, from mid-July to mid-September, 2005. The survey aimed to measure the difference between tourists and arrivals, (even if this difference can be considered as the minimum value of the error), and make possible the valuation of consequences of the error.

The questionnaire used as survey instrument was provided to all the hotel managers of the Sicilian association (80% of the total in the same district); it included few questions concerning the number of registration and the length of stay of tourists in Sicily (before and after their registration at Cefalù hotels), few demographics variables (such as: nationality) and the type of tourist group they belong. Due to contingent reasons we have not considered the arrivals registered at non-hotel lodging establishments, which, as known, belong to official tourism statistics too.

About 70% of hotel managers has contributed to the questionnaires collection and has given back the filled survey instrument. We have collected 664 questionnaires of which 60% regarding foreign tourists. Totally we have valid data for 13600 arrivals, which are about 1/3 of the arrivals in the district we have considered (during the same period).

Before illustrating the preliminary results of our survey we underline that the size of the error we have calculated is the minimum value of the same one, and this is for a lot of reasons concerning: a) the kind of lodging establishments (only hotels) we have included; b) the summer period for collection of data, which coincides with the arrivals of seaside tourists, typically with a low grade of mobility; c) the space (just a single destination, Cefalù) we have assumed as point of reference to measure both arrivals and tourists.

We observe (Table1) that there is an absolute difference between arrivals and tourists of 1877 units that corresponds to a ratio of arrival to tourists equal to 1.16. This is only for the summer season and within the limit of hotel establishments in the Cefalù district. Besides, it emerges that domestic tourism and foreign tourism have a different ratio and that it is also different in function of the dimension of the group the tourist belongs³.

Without going far with the comments on the size of replications, we can underline that the survey we have realized could be used as guidance to carry out adjustments to the

³ When the number of registrations at lodging accommodations of a destination is unknown it is possible to calculate the effective number of tourists also by the inverse of the index used to evaluate the mean permanence and to estimate, comparing the two values of A and T, the replication effect.

system adopted to gathered official data on tourism. These adjustments could lead the official tourism statistics to agree with the term “tourist” and with the time and place of reference.

Table 1: *Tourists and Arrivals (Cefalù, July- September 2005)*

N° of registrations in Sicily	Arrivals	Turists	Replications
1	10732	10732	0
2	836	418	418
3	1250	417	833
4	539	135	404
5	185	37	148
6	60	10	50
7	28	4	24
Total	13630	11753	1877

Source: our survey

4. Final Remarks

In this paper we have focused our attention to a typical tourism statistics’ problem: the substitution of an elementary units of observation (the tourist) with an indicator (new bed place). Therefore, the necessity to consider the quality of aggregates in addition to the quality of individual measurements is emphasised. As we have pointed out, in the case of tourism arrivals, the properties of the indicator at an elementary level are not the properties of the statistic aggregate. So a lot of analysis procedures, and even data of official statistics, are heavily biased. This bias, together with other sources of inaccuracy (see, for example, unregistered arrivals), leads to a lack of data quality and they all are the cause of erroneous interpretation of the tourism phenomena.

The cited survey is only a way to underline the existence of the gap between aggregate and individual measurements and suggests to introduce some adjustment to the system adapted to gathered data on tourism, collecting and combining properly survey and census data.

References

- ISTAT (2004) Statistiche del turismo.
ISTAT (2004) I viaggi in Italia e all’estero nel 2002.
Parroco A.M., Vaccina F. (2005), Referring to space and time while using territorial data: the case of touristic arrivals. *Proceeding ISI, 55th Session*, Sydney, Australia.
Polfeldt T. (1997) Quality aspects of official environment statistics, *Environmetrics*, 8, 375-385.
Ufficio Italiano Cambi (2005) Turismo internazionale dell’Italia.
[www. uic.it/it/statistiche/pubblicazioni/turismo/glossario.htm](http://www.uic.it/it/statistiche/pubblicazioni/turismo/glossario.htm)
www.world-tourism.org/facts/methodological/method.htm