CONFERENCE OF EUROPEAN STATISTICIANS
Expert Meeting on Statistical Data Collection Session 4
27-30 September 2021, Online 18 August 2021

Combination Methods in Tourism Statistics

Putra, Amanda Pratama; Munaf, Alfatihah R. M. N. S. P.; Ruslani, Agus; Statistics Indonesia (Statistics Indonesia)
amanda.putra@bps.go.id, alfa@bps.go.id, agus.ruslani@bps.go.id

Abstract

Statistics Indonesia has been using Mobile Positioning Data (MPD) since 2017. At the beginning of the study, the use of MPD went hand in hand with conventional surveys. By improving the algorithm in explaining humanmobility and increasing the accuracy of determining the usual environment to 96.23%, Indonesia is increasingly confident in using MPD.

MPD only produces data from the area of origin and tourist destinations, and we cannot get demographic information or the amount of travel expenditure from MPD. Therefore, MPD is combined with digital surveys to get a complete picture of the required tourism statistics.

By using this combination method, the advantages are that estimates can be carried out in smaller areas (municipalities/cities), shorter data periods (monthly), do not require data collection officers (data collection solutions during a pandemic), can be used for other statistical calculations (commuter, migration): all for half the cost of doing a conventional survey.

Introduction

Human movement is a topic of research that is always renewable. Various policies are taken based on observations on human movement: residential zones, business zones, transportation facilities and routes, education, tourism, and many more. Therefore, up-to-date data is always needed to meet decision-making needs. In addition to always moving, humans also need communication with other human beings. With the support of technology, this communication no longer knows the boundaries of place and time. Furthermore, one of the technology products that is experiencing rapid growth in meeting human communication needs is a cellular phone device. Until 2019, the statistics of the national socio-economic survey show that the estimated household that owns cellular phones in Indonesia has reached 63.53%.

As a result of the widespread use of cellular telephones, administrative records of the use of cellular telephones are available, which were initially only used for billing needs, but in fact, have patterns and can be used to describe human movements. Mobile Positioning Data has become one of the new data sources that are considered to have good sustainability to describe the movement of people over time in a massive, accurate, and real-time.

The use of MPD in Indonesia as official statistics was initiated by the Ministry of Tourism in collaboration with

Statistics Indonesia to record foreign tourist traffic at cross-border posts where there was no immigration office in 2017. In 2018, Statistics Indonesia's journey to study domestic tourists using MPD began.

The Data

In principle, the structure of the MPD is simple. It only stores the mobile phone number (which contains the country code and operator code), the time the transaction occurred, and the unique code from the Base Transceiver Station (BTS) that records the place where the transaction occurred. Each BTS has information in the form of latitude and longitude coordinates. A collection of data from the administrative records of cellular phones has meaning. The record of the most positions in an area in a certain period indicates that person's tendency to stay longer in that area. This may be interpreted as the location of a person's daily activities or known as the concept of the usual environment.

Table 1. Sample data from one of the subscribers

msisdn	datetime	source	bts_lat	bts_lon	prov	kab	kec	desa	trx_date
6281 ...	2018-02-08 13:33:03	LBA_ALL	-4.5469	120.3583	SULAWESI SELATAN	BONE	TANETTE	CELLU	2018-02-08
6281 ...	2018-02-08 13:51:39	CHG_POST	-4.5405	120.3077	SULAWESI SELATAN	BONE	TANETTE	MACANANG	2018-02-08
6281 ...	2018-02-08 14:00:39	CHG_POST	-4.5359	120.3037	SULAWESI SELATAN	BONE	TANETTE	MACANANG	2018-02-08

In modeling the data, BPS uses data from mobile phone operators that have been anonymized so that BPS never knows who is traveling (not against the principle of confidentiality of personal data). In addition, BPS also utilizes MPD data as output in the form of regional aggregates to ensure data confidentiality. However, to evaluate the algorithm's accuracy, BPS recruits volunteers who are willing to submit their cellular data logs to BPS as sample data for later analysis and verification of their journey. This volunteer data has never been published. The accuracy of house location predictions at the Regency/City level continues to increase along with algorithm improvements, with an accuracy value reaching 96.23% in 2018.

The study of the use of mobile positioning data as the data source of official statistics was carried out by BPS together with Telkomsel, a mobile network operator (MNO), which is a subsidiary of Telkom, a State-Owned Enterprise. Telkomsel is also the largest MNO in Indonesia, with a market share of > 60%.

Tourists from MPD's point of view

Based on the usual environment concept, most people’s location at night is determined as the estimated location of their home, and the location of the most people during the day is the estimated location of their daily activities. Locations outside the locations recorded as the usual environment are defined as a movement. A collection of movements over a given period is considered a trip. If the trip is grouped in a certain area with a duration of more than 6 hours, then this area is determined as the destination of the trip. A set of trips is defined as the movement that occurs from the time a person leaves his home location until he returns to his home. BPS uses these assumptions in producing Domestic Tourist Statistics.

However, from the nature of the data itself, MPD can only show the trip's origin, destination, and duration. Meanwhile, tourism statistics need more than that. Tourism statistics require spending on travel, the purpose of travel, mode of transportation, etc. In the conventional method, this can all be obtained by household surveys. However, the household survey has one weakness: the respondent cannot remember well where he has traveled. Household surveys have the potential for memory lapses, which can have an impact on underestimating tourism statistics.

Combine Methods

Indonesia uses big data not because it is the latest technology but because we have no choice. Indonesia has a vast area (1.905 million km²), with a large population (270,203,917). Every year BPS holds more than 200 surveys, and costs are expensive. Nevertheless, the main issue is the lack of resources (± 16,000 employees). The demand for data is increasing each day, and we cannot always rely on surveys. Hence, big data as a new data source needs to be considered to produce official statistics. This is a must, not an option.

We know that MPD has limitations. It cannot get complete data for statistical tourism needs. However, we also know that every trip can be observed using MPD. Therefore, the use of MPD is combined with the advantages of surveys to obtain more complete data. A digital survey was conducted to complete the MPD to get other essential variables such as demographic, travel purpose, transportation mode, accommodation, type of expenditures, and the number of expenditures.

What is the mechanism for conducting the survey?

1. Obtained a number of X which according to the algorithm determined by BPS was identified as travelers, this X will be designated as a sampling frame.
2. It has been determined that the sample that must be achieved is Y, where Y must be in line with the respondent's travel pattern based on the MPD.
3. X is divided according to the origin of the district/city, then during the survey period, SMS blasting is carried out with the survey link content until the respondent's target for the district/city is achieved.
4. Respondents are given a gimmick if they complete the survey.

Table 2. Comparison of Response Rates Based on the Type of Gimmick in the Digital Survey 2018

No	Gimmick	SMS Blast	Response	Response Rate
(1)	*(2)*	*(3)*	*(4)*	*(5)*
1	Voice on Net 40‘	121.120	302	0,25%
2	Data 50 MB	1.334.375	3.514	0,26%
3	TCash	7.755.842	46.184	0,60%

In the first digital survey in 2018, several possible gimmicks were tried to see which one had the highest response rate (Table 2). The table shows that MNO's customers in Indonesia prefer gimmicks in electronic money (tcash is Telkomsel's electronic money) over telephone packages or data packages.

Table 3. Comparison of Response rate and Gimmick Value of Digital Survey, 2018-2020

No	Year	Gimmick Value	Survey Periode	*Response rate*
(1)	*(2)*	*(3)*	*(4)*	*(5)*
1	2018	4.500	Sep - Dec	0,60%
2	2019	4.500	Sep - Dec	0,47%
3	2020	5.000	Aug - Dec	1,20%

Response rate is influenced by many factors, both internal and external. The implementation of activities in 2019 was two weeks behind compared to 2018, so the possibility of affecting the response rate was lower than before. In 2020 the gimmick increased to 5,000, and the implementation was one month earlier than in 2018;

the response rate also increased to 1.20%. In 2020 there was also a shift in the survey platform; if in 2018 and 2019, it used SurveyGizmo, in 2020, it used the Integrated Collection System (ICS). The survey platform with the domain survey.bps.go.id has become more official so that the trust of the Indonesian people increases. Besides that, BPS has also just completed an online population census that also uses ICS.

Efficiency using MPD

Previous data collection was held by conducting a conventional survey which has limitations in the number of samples, estimation level, data period, the use case that could be covered, and enormous cost. The number of samples was detected have data mobility restricted to only around 120.000 people. It was linear with the cost provided means that an increase in the number of samples will increase the cost needed. The estimation level for data analysis also correlated with the cost of the survey. The more detail the level has to achieve, the more it will cost. It has become common that conventional surveys mostly provide yearly data periods, especially when the survey coverage is quite broad.

By using MPD as an alternative to data collection, the cost can be reduced, but the quality of the data can still be maintained. BPS has saved more than 50% of the survey cost by implementing MPD for the data collection method. The cost of the data procurement with MNO may vary in every country but potentially to get data mobility of a massive number of MNO's subscribers. However, since the data collection by MPD does not need enumerators, a considerable portion of the survey's cost can be eliminated. Reducing conventional data collection can promote eradicating the workload of BPS's regional branches.

Table 4. The Differences between MPD Approach and Household Approach (Conventional)

No	*MPD Approach*	*Household Approach (Conventional)*
(1)	(2)	(3)
1	Estimates for smaller areas: districts/cities	Estimates up to the provincial level
2	Monthly data period	Yearly data period
3	No surveyors needed	The number of surveyors needed depends on the number of samples
4	With a more detailed level of estimation, the required budget is only around IDR 15 billion	To conduct a conventional domestic tourist survey, a budget of around IDR 30-40 billion is needed (based on the 2018 and 2019 conventional domestic tourists survey budget).
5	It has the potential to be able to generate statistics for domestic tourists monthly	Can only generate statistics for domestic tourists on annualy
6	The solution for collecting data for domestic tourists during the covid-19 pandemic, no door-to-door interviews needed	The Covid-19 pandemic condition limits conventional data collection activities
7	MPD can be used for other statistical calculations	One activity is only used for one type of statistical calculation

The cooperation scheme between BPS and MNO is to provide a media data connection, in which the MNO provides MPD data which first goes through a quality assurance process. BPS then applies the developed algorithm for each use case.

By using the usual environment concept, in addition to tourism statistics, BPS uses MPD to calculate commuter statistics. Another advantage of using MPD is that it can generate other use cases with one type of data. In addition to domestic tourists and commuters, BPS also uses MPD for non-permanent mobility and the Metropolitan Statistical Area.

BPS also collaborates with other Ministries regarding the use of MPD, such as with the Ministry of Tourism for Super Priority Tourist Destinations and with the Ministry of National Development Planning of Indonesia for Metropolitan Statistical Areas. Other use cases are also being studied so that the data utilization can be more optimal such as transportation statistics, international trade in services statistics, and accommodation statistics. The other use cases utilize MPD as a new method of data collection is potential to be invented. The main point of this project is how NSO elaborates what they need and what they have by refocusing on another data source that they never thought of before, similar to cellular data before MPD was invented.