8 July 2025

USE OF OPEN STREET MAP FOR ACCIDENT INVESTIGATION ON THE ROAD AND MOTORWAYS NETWORKS. Updating year 2023.

1. Road Safety performance Indicators and Big data use

Road Safety Performance Indicators (RSPI) give a multidimensional approach for accident investigation concerning roads, vehicles and persons involved. Combining the use of statistical surveys, administrative geographical information systems (GIS) and Big Data (BD) sources the result gives new elements on planning infrastructure solution, applying policies to reduce deaths and serious injuries, reducing social costs on collectivity and estimating efficiency and effectiveness of safety initiatives.

Preventing road trauma on public roads is a core responsibility for government, its agencies and stakeholder. It requires a common and shared responsibility. The scale of the road safety challenge and the diversity of the effects of road traffic injury underline the importance of exploring synergies among the decision makers of the road network.

Nowadays there is a clear information bias as regards the appropriate reference denominators, used to build statistical indicators linked to road accidents. Resident population is a common proxy for exposed at risk in a specific geographical area, but not always an appropriate solution, especially in the light of the seasonal nature of road accidents and concentration, in some periods of the year, in specific locations. Vehicle fleet (Automobile Club of Italy - ACI) can be another administrative source that gives a more accurate information, but the characteristic of the phenomenon implies a deductible distortion on measures due to the mobility of road users.

The length of the road network (from Open Street Map) gives for sure a consistent first set of information concerning the different territories.

This information is not available from official statistics at national level, although there are archives and detailed road graphs for each municipality, province and region, a harmonized and systematic national road registry has not been established.

The project’s first output, shared in 2019 and updated in 2021 with data from 2016 and 2017, focused on making better use of existing administrative sources, exploring new ones, and analyzing integrated and auxiliary data, in line with Istat’s efforts to modernize its statistical production. The basis of the renewal in the statistical production is to upload any source integration; even any new technique implemented and applied methodology. Every small change that overall effect becomes a process of improvement of the quality of the statistical information provided by Istat.

The aim of project is to calculate road accidents rates, mortality and harmfulness indexes, comparing these measures to the correspondent length in meters of carriageway by road direction from the Open Street Map. Although the product represents a first result, the final purpose of the project is to expand statistical information with the supply of traffic flows (vehicles / km) on the national road network. This would make it possible to calculate the probability of being involved in the accident, taking into account the different exposure to risk.

The project has also been one of the leading innovations for the Institute in recent years, particularly in the practical use of Big Data for statistical purposes. It has even been included among the priority projects of Trusted Smart Statistics (TSS). These projects rely on Big Data, a specific assessment of the robustness of the methodology used, and a potential inclusion—after verifying the necessary requirements—among the Institute's official statistics.

In 2021, the product contained the 2017 data updating, using a new methodology. The update also included revised texts and tables. The main technical innovations introduced in 2021, based on 2017 data, were:

- The addition of traffic points on road segments, extracted from Open Street Map (OSM).
- The development of a new analytical classification for road segments, with updated reference tables.
- The creation of synthetic indicators using "Ranker," a generalized tool designed to define indicator polarity.
- The refinement of indices by removing the "traffic factor" and updating indicators with negative polarity.

This publication presents an update with 2018 data. The preparation of subsequent editions of experimental road accident statistics has required a longer period to maintain the quality standards tested in previous editions and to ensure an appropriate data update.

To reclassify road types proposed in Open Street Map according to the classification used by Istat for road accident data collection, the methodology has incorporated territorial information and Census sections updated to 2021. Istat released this data in 2023 and 2024[1].

As usual, the method involves building overlapping graphical layers by using a dedicated algorithm programmed to perform a "spatial join" operation between the various attributes of the geographical areas considered.

[1] https://www.istat.it/notizia/dati-per-sezioni-di-censimento/

2. The use of GIS systems for a graphical representation

GIS is a geographic system designed to capture, store, manipulate, analyse, manage, and present spatial or geographic data. GIS applications are tools that allow users to analyse spatial information, edit data in maps, and present the results of all these operations. In order to relate information from different sources, GIS uses spatial location as the key index variable (key reference by position). Just as a relational database containing text or numbers can relate many different tables using common key index variables, GIS can relate otherwise unrelated information by using location as the key index variable. This key characteristic of GIS has begun an alternative frontier on producing statistical information. Any variable that can be located spatially using an x, y, and z coordinates representing, longitude, latitude, and elevation, respectively. These GIS coordinates may represent other quantified systems of territories (polygons), road networks (lines) and point of traffic (points).

Join attributes by location is the algorithm that takes an input vector layer and creates a new vector layer that is an extended version of the input one, with additional attributes in its attribute table. The additional attributes and their values come from a second vector layer. A spatial criterion lead the values selection from the second layer that are added to each feature from the first layer in the resulting one (Chart 1).

Chart 1 – Graphical representation of the join attributes by location algorithm

2.1 Istat Census Map localities

The Istat Census Map localities used with the GIS system refer the following administrative units:

• Regions

• Provinces

• Municipalities (yearly updated)

• Localities (Istat Census 2021)

An upgrade of Census Map localities shapes referred to 2021 to municipalities 2023 has been done. It has been built a link table with the aggregation of over 8000 local administrative units territory, at 2021, to the 7899 municipalities at 2023, included in the Italian territory.

The choice of the localities shapes is due to the harmonization need of the complete roads graph to the "road type classification" used by the road accidents survey.

The localities classification includes 4 voices:

1. Urban areas

2. Small inhabited areas

3. Productive areas

4. Wide spread houses

As regards the definitions, the Istat[2] descriptions are below.

Localities - A more or less wide area, usually known by name, on which one or more houses are grouped or scattered; there are two types of localities: inhabited localities and productive locations. The borders of the inhabited localities (center and inhabited area) are the external limit of the buildings placed at the edges of a grouping of at least fifteen buildings. The confines of the localities include gardens and other accessories areas of the considered buildings, non-built neighboring areas not included (such as fields with or without crops). Buildings located at a distance of more than 70 meters, within built-up areas and 40 meters for small-inhabited areas excluded. If the buildings included in the new locality are adjacent or in proximity (to a max distance of 140 meters for urban areas and 60 meters for small inhabited areas) of transport infrastructures or hydrographic limits, the location border will be extended to the middle of these elements.

Urban areas - Aggregate of contiguous or near houses with roads, squares and similar, or however areas characterized by services or public activities (school, public office, pharmacy, shop or similar), detecting a social life and, generally, also a place of collection for the inhabitants of the neighboring areas. The places of tourists meetings, houses, hotels and similar used for the vacation, inhabited seasonally, are considered as temporary inhabited centers too.

Small-inhabited areas is an area without the place of collection, characteristic of the urban area. It is based on a group of at least fifteen contiguous and near buildings, with at least fifteen families, with roads, paths, squares, farmyards, small gardens and similar, as long as the distance between the buildings does not exceed thirty meters and it is lower than the distance between

---

[2] Istat - Descrizione dei dati geografici e delle variabili censuarie delle Basi territoriali per i censimenti: anni 1991, 2001, 2011 e 2021 https://www.istat.it/notizia/basi-territoriali-e-variabili-censuarie/

the center and the nearest of the houses clearly scattered.

Productive areas - Extra-urban area not included in the centers or residential areas with more than 10 local units, or with a total number of employees' upper than 200. The local units are contiguous or close, with roads, squares or similar, or anyway in a continuous line, not exceeding 200 meters; the minimum area must be 5 hectares.

Wide spread houses - Houses scattered in the municipal territory at a distance not enough to constitute a built-up area.

2.2 Open Street Map and road arch type

Open Street Map (OSM) [3] is a collaborative project aimed on creating free content maps of the world. The project aims at a collection world of geographical data, with the main purpose of creating maps and cartography. The key feature of the geographic data present in OSM is having a free license, the Open Database License. It is therefore possible use them freely for any purpose with the only constraint of mentioning the source. Everyone can contribute by populating or correcting data. The maps use the data recorded by portable GPS devices, aerial photographs and other free sources. Most of the Android and iOS GPS navigation software on portable devices powered by OSM as WisePilot, Maps.me, NavFree, Scout etc.

The Open Street Map vector layers, used in this work, daily updated and free downloadable data, are the following:

- Road graph;
- Point of traffic (POT);

Added shape:

- Buildings;
- Use of the land;
- Natural; Places;
- POWF (Point of Worship); POIS (Point of interest);
- Railways; Transport;
- Water; Water ways.

Although OSM it is an Open Source tool based on information from a community, the product

[3] OpenStreetMap provides geographic data on thousands of websites, mobile and hardware devices. OpenStreetMap is built by a community of mappers, who contribute, update and monitor data on roads, cafes, railway stations and much more, all over the world - OSM: http://www.openstreetmap.org/about

provides data to be considered reliable and consistent, so much that the major part of GPS Android and iOS navigation software on portable devices are powered by OSM, for example WisePilot, Maps.me, NavFree, Scout, etc.

Table 1 contains the list of different type of road arch by Open Street Map.

Table 1 - OpenStreetMap road arch classification (a)

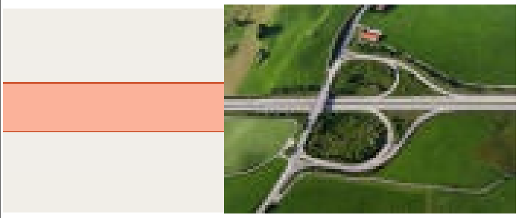
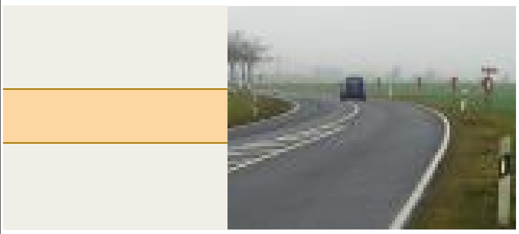
	**Motorway** Free or by toll payment highways, including motorway connections such as the Milan ring road and the Grande Raccordo Anulare of Rome. Equivalent to Freeways, Autobahns, etc ...
	**Motorway Link** The link roads (sliproads/ramps) leading to/from a motorway from/to a motorway or lower class highway. Normally with the same motorway restrictions.
	**Trunk** Roads type between Motorway, Motorway connections and Primary road. The junction section of a motorway-ring road that leads to the city center can also be classified as trunk. Some special primary road called in Italian "Superstrade" with two lanes, could be included in this category. In "trunk" are included extra-urban roads with only one lane in each direction, without crossings, with exits and acceleration and deceleration lanes. They do not include intersections and roundabouts.
	**Trunk Link** The link roads (sliproads/ramps) leading to/from a trunk road from/to a trunk road or lower class highway.
	**Primary** Roads of national and regional importance not classified as motorways, trunks, or their link. They connect the main cities to each other. Usually, they are classified as SS (Main Roads) or SR (Regional Roads), however there are some exceptions such as in small mountain towns where the main road road crosses the village but the primary road is a ring road of modern construction, around the inhabited center. In urban areas, they represent the external ring of the city (e.g. Milan) and are normally classified as Viali.
	Primary_link The link roads (sliproads/ramps) leading to/from a primary road from/to a primary road or lower class highway.

Table 1 - OpenStreetMap road arch classification (continued) (a)

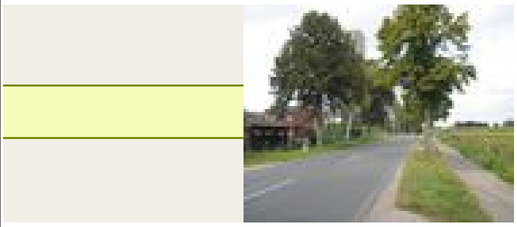
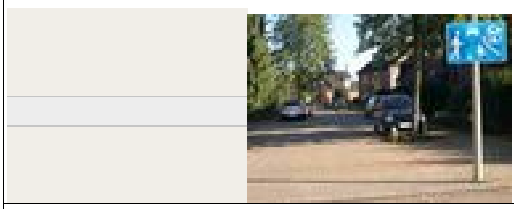
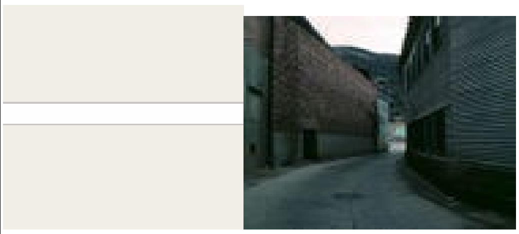
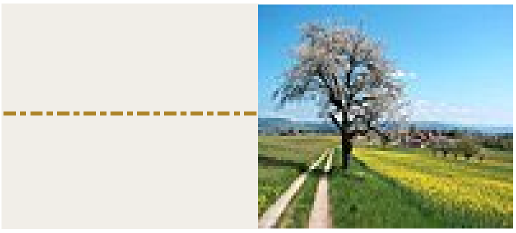
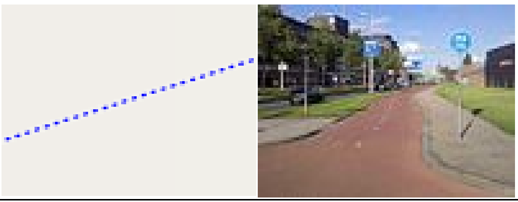
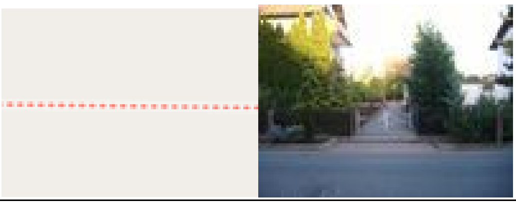
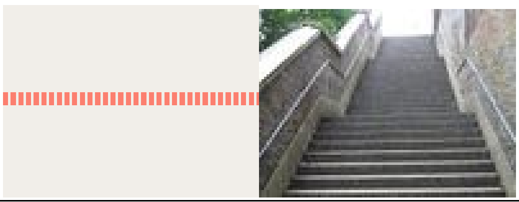
	**Secondary** Roads of regional and provincial importance. They link together the main municipalities of a region. They are, usually, classified as SP (provincial roads) but there are some exceptions. In urban areas, they are normally classified as important streets with two lanes in each direction.
	Secondary link The link roads (sliproads/ramps) leading to/from a secondary road from/to a secondary road or lower class highway.
	**Tertiary** Roads of local rank. They connect smaller municipalities together. In urban areas, they are side roads to primary and secondary roads with a medium flow of traffic.
	Tertiary link The link roads (sliproads/ramps) leading to/from a tertiary road from/to a tertiary road or lower class highway.
	**Unclassified** Classification for some extra-urban road. In urban areas, they are used to reach close destinations and the traffic flow is lower than in higher classification roads. They often connect villages and hamlets.
	**Residential** Roads in a residential area, which serve as an access to housing, without function of connecting settlements. Often lined with housing.
	**Living Street** Residential road where pedestrians have legal priority over cars, speeds are kept very low and where children are allowed to play on the street (category not frequently present in Italy, anyway, if used, the signal of residential area is installed).
	**Pedestrian** Pedestrian areas (roads or squares in urban areas), accessible mainly or exclusively to pedestrians.
	**Service** Access roads or internal service areas, beaches, campings, industrial areas, shopping centers, residences, parking places, landfills, installations, etc.

Table 1 - OpenStreetMap road arch classification (continued) (a)

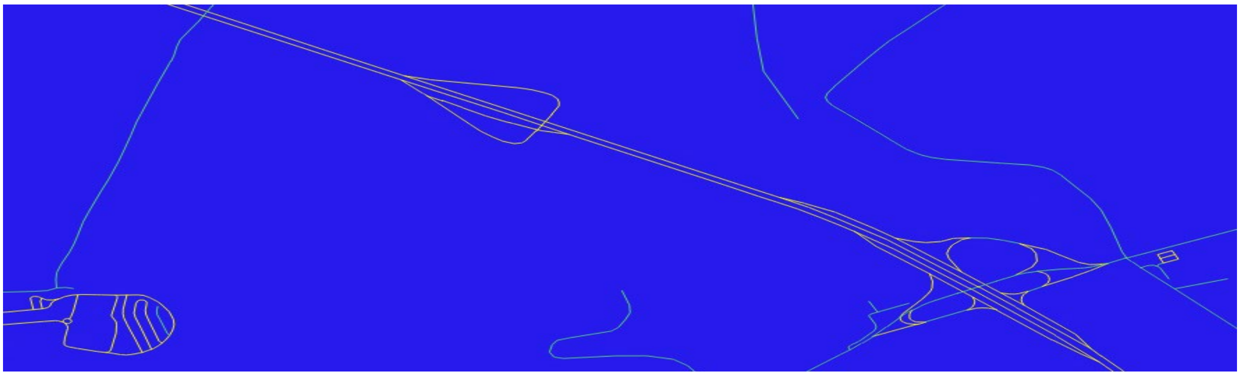
	**Track** Roads for mostly agricultural or forestry uses. To describe the quality of a track, are often rough with unpaved surfaces.
	**Bridleway** A way intended for use by horse riders (primarily) and pedestrians (depending on country-specific regulations). They may be legally designated as rights of way for pedestrians and equestrians (horses), but need not be. Depending on the country, cyclists may be also permitted, though the surface may not be suitable.
	**Cycleway** Cycle paths on dedicated carriageway, mainly or exclusively for cycling tourism.
	**Footway** Paths mainly/exclusively for pedestrians. This includes walking urban tracks, paths in a public park and footpaths also if managed and not maintained at natural status.
	**Path** Paths not structured for a public use
	**Steps** Stairs in steps, exclusively accessible by pedestrians
	**Unknown** Not classified

(a) Road Arches by Open Street Map – update 1/1/2024.

An innovative method of measuring the length in meters of the road graph is given by the information on the number of carriageways of each road arch of OSM.

The snapshot in Chart 2 shows how the yellow arches represent the one-way arches, while the green ones the two-ways. In the future, in order to provide a more detailed information, the researchers used the information of the number of lanes, containing each carriageway.

Chart 2: The OpenStreetMap information on carriageway numbers



2.3 Classification of OSM road arches and road accidents localization

The arch road types selected to calculate the indicators analysed are referred to the motorized vehicles flow: motorway, trunk, primary, secondary, tertiary, unclassified, residential, living street, motorway link, trunk link, primary link, secondary link, tertiary link, service, unknown. Pedestrian, track, track grade, bridleway, cycle way, footway, path, steps are not object of the survey definition.

To build road accidents indicators, with denominator represented by the arches length in Open Street Map, since the last edition, we built a "bridge matrix" between road categories, classified by functional road type, used by OSM, and the categories linked to the roads holder, used by Istat road accidents survey. This methodology is **a systematic classification technique**.

The systematic classification of road arches (Open Street Map), classified in the categories used by Istat, has been modified, in respect of the first release of the experimental statistics. A new **analytical classification** has been adopted, using a more refined technique of attributing single road arches, about three and a half million in total (OSM), to the Istat classification groups See table 2). The operational criterion applied provides the roads classification, through the textual analysis of the Name and Reference attributes, according to the different classes of road arch and spatial attribution of the location type.

Table 2 – Bridge table between roads arches classification by OSM, localities and road type (a)

Road Archs classification by OpenStreetMap	Localities at Census 2011			
	Urban areas + Small inhabited areas		Productive areas + Wide spread houses	
	Road Localisation by Road accidents survey			
	Motorways	Urban Roads	Motorway	s
Motorway	X		X	
Trunk	X		X	
Primary		X		X
Secondary		X		X
Tertiary		X		X
Unclassified		X		X
Residential		X		X
Living Street		X		X
Motorway Link	X		X	
Trunk Link	X		X	
Primary Link		X		X
Secondary Link		X		X
Tettiary Link		X		X
Service		X		X
Unknown		X		X

(a) Istat computing.

The transition from systematic to analytical classification of road graphs follows a road management approach. This change improves the understanding and organization of road networks. The method uses available metadata, such as road name, reference code (a short identifier for the road), the type of area (urban or rural), and the functional road class based on OpenStreetMap (OSM) criteria. Table 1 and the OSM wiki[4] provide more details.

In 2024, researchers analysed over five million road segments in Italy. They identified more than 400,000 possible groupings based on road management characteristics. A segmentation process divided these groups into 23 distinct categories. This approach allows a detailed study of each group's specific features.

The new classification follows the Institute's road localization criteria, used in road accident surveys. This method improves road network management and classification. It also makes accident analysis more effective. In particular, it helps create risk exposure indicators for road accidents. These indicators support standardized planning and monitoring of road infrastructure at the national level.

Analytical classification solves the limits of systematic classification. The previous approach used rigid criteria that did not adapt well to different areas. By using available metadata, the new method improves road network segmentation. It considers both management and functional characteristics of each road segment.

[4] https://wiki.openstreetmap.org/

Reference codes and road names help identify continuity and fragmentation along roads. The urban and rural distinction allows for a better understanding of travel conditions. This is essential for traffic and road safety studies. The OSM classification adds more detail. It helps define the role and function of each road segment more precisely.

This method brings important benefits to traffic management, infrastructure planning, and road risk assessment. Identifying similar road groups makes it easier to plan maintenance, improve safety measures, and optimize road signs. Combining this data with traffic flow and accident records helps create better predictive models and more effective intervention strategies.

Charts 3-6 show cartographic representations of roads by type according to the OpenStreetMap classification.

Chart 3: Motorways. Year 2018
(Istat: OSM classification)

Chart 4: Primary Roads. Year 2018
(Istat: OSM classification)

Source: Istat processing on OpenStreetMap data

Chart 5: Secondary Roads. Year 2018
(Istat: OSM classification)

Chart 6: Residential Roads. Year 2018
(Istat: OSM classification)

Source: Istat processing on OpenStreetMap data

3. The indicators on road accidents

3.1 Road accident indicators in relation to the extended road, resident population, vehicle fleet

The proposal to calculate new road accidents indicators was born to provide, new and more suitable measures, for the risk and probability of being involved in a road accident. Even if Resident Population and Vehicles fleet are both denominators considered, at present, *proxy* for the risk exposure to the road accidents, it is clear that they present some critical points.

As regards the data updating (numerators of indicators) used to build indicators on the road accidents risk, they refers to Istat "road accidents survey", concerning all road accidents resulting in deaths (within the 30th day) or injuries in 2017, involving at least a vehicle circulating on the national road network and documented by a Police authority[5].

As regards denominators used, they are the Resident Population, Vehicle Fleet and the length in meters per carriageway.

The Resident Population[6], not always is an appropriate solution to build road accident indicators, because the seasonality of accidents and the concentration in some specific places during the

[5] Survey on road accidents resulting in death or injuries https://www.istat.it/informazioni-sulla-rilevazione/incidenti-stradali/

[6] Resident Population – 31/12/2023.

year. The Vehicles fleet by registration province (Automobile Club of Italy – Public Register of Vehicles PRA)[7] provides more suitable information than the population, but it does not consider the mobility of users through the country.

The length in meters of carriageway by road direction of road arch (by OSM) [8] provides, instead, a set of more coherent information referred to different areas, because independent by the seasonality and the mobility users influence.

This information is not available at national level, since, although archives and detailed road graphs by single municipality, province or region exist, a national and harmonised road register has been not create yet.

3.2 Road accident indicators: weighting with Traffic Points (PoT) from the Open Street Map

The length in meters of carriageway per direction of the road arch (from Open Street Map), used to calculate the indicators for experimental statistics, provides a first coherent set of information relating to the various territories. Although the product constitutes a first usable result, the project's most ambitious goal is to estimate the real traffic flows (vehicles / Km) on the national road network. This would make it possible to calculate the probability of being involved in a road accident and therefore of real exposure to the risk of accident.

To fill this information gap, other additional information came from the rich source Open Street Map, in particular, data on the traffic points detected on the road arches (PoT Point of Traffic). In fact, a monthly information, downloadable from OSM, is available and the data refers to points over an arch, in which is detected an intensity of traffic. The proposed new road accident indicators, "weighted" with the information on traffic intensity, was built considering, as a discriminating element, the kilometers of carriageway with the presence of a traffic point on the arch.

The authors calculated the length of roads, considering the extension, in meters, of carriageway with arches on which including the presence of traffic points. The researchers processed the relative frequency (f) of the road length with the presence of traffic points on an arch (by province and type of road) and the complementary frequency (1-f) of arches without point of traffic too.

The number of accidents, vehicles involved, deaths and injuries per 100 kilometers of carriageway in the province, excluding the effect on accidents caused by the presence of traffic points on the roads has been processed. The calculation consists in multiplying the value of the indicators for road length and the frequency (1-f) referred to arches without traffic points on the roads of the province.

---

[7] ACI Vehicle fleet (Automobile Club of Italy) all motorized vehicles except the trailers on 31st December 2023.

[8] GIS computing (Census Map + Open Street Map road graph at 1st January 2024) expressed by length in meters per carriageway.

With reference to the notation used for the calculation, the individual items are defined as follows:

LA Tot $_{i,j,k}$ = Total Length of Arches in meters in the province (i), by functional class of the Open Street Map Arch (j) and Type of locality (k);

LA PoT $_{i,j,k}$ = Length of Arches in meters with presence of Points of Traffic (PoT) in the province (i), by functional class of the Open Street Map Arch (j) and Type of locality (k);

LA Tot $_{i,k,l}$ = Total Length of Arches in meters in the province (i), by Type of locality (k) and Type of road by Istat classification (l);

LA PoT $_{i,k,l}$ = Total Length of Arches in meters in the province with presence Points of Traffic (PoT) in the province (i), by Type of locality (k) and Type of road by Istat classification (l);

The total length of the national roads, referred to arches length, with the presence of Points of Traffic (PoT) or in total, is given by the following expressions:

Total Length of Arches with Points of Traffic (PoT):

$$\text{LA PoT}_{\text{tot}} = \sum_{i=1}^{111}\sum_{j=1}^{27}\sum_{k=1}^{4} \text{LA PoT}_{i,j,k} \quad or \quad \sum_{i=1}^{111}\sum_{k=1}^{4}\sum_{l=0}^{9} \text{LA PoT}_{i,k,l}$$

$$\text{LA Total}_{\text{tot}} = \sum_{i=1}^{111}\sum_{j=1}^{27}\sum_{k=1}^{4} \text{LA Totale}_{i,j,k} \quad or \quad \sum_{i=1}^{111}\sum_{k=1}^{4}\sum_{l=0}^{9} \text{LA Total}_{i,k,l}$$

The percentage weight of the Arches Length in meters, with the presence of Points of Traffic (PoT), in the Province (i), by functional class of Open Street Map Arch (j) and Locality type (k) out of the total length is given by the following expression:

$$\text{p LA PoT}_{i,j,k} = \frac{\text{LA PoT}_{i,j,k}}{\text{LA Tot}_{i,j,k}} *100$$

The percentage weight of the Arches Length in meters, with the presence of Points of Traffic (PoT), in the Province (i), by Locality type (k) and Type of road by Istat classification (l) out of the total length is given by the following expression:

$$\text{p LA PoT}_{i,k,l} = \frac{\text{LA PoT}_{i,k,l}}{\text{LA Tot}_{i,k,l}} *100$$

The percentage weight of the Arches Length in meters, with the presence of Points of Traffic

(PoT), at national level, out of of the total length, is given, at last, by the following expression:

$$p\ LA\ PoT_{tot} = \frac{LA\ PoT_{tot}}{LA\ Total_{tot}} *100$$

where **i**= 1,…, 103, 108, 109, 110, 111 (Istat Code of province)

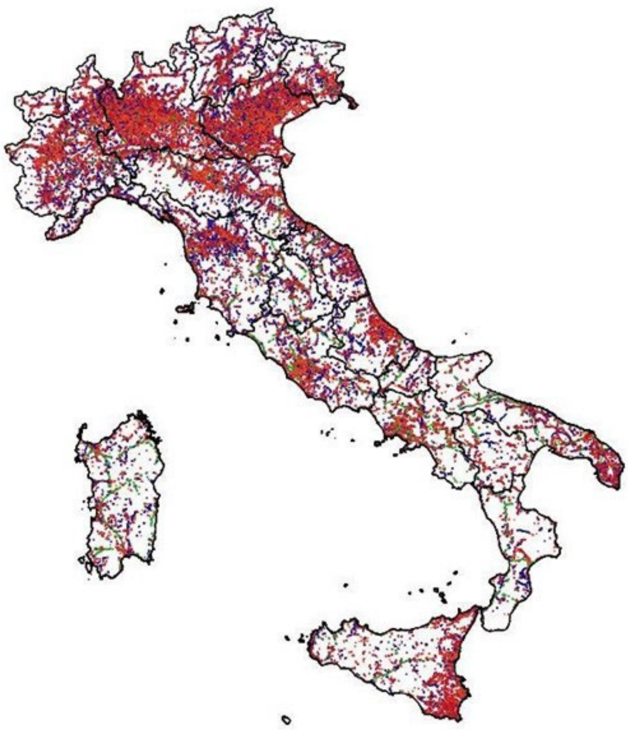
j= 1,.....,27 (Functional class of the Open Street Map Arch)
1=Motorway; 2=Motorway_Link; 3=Trunk; 4=Trunk_Link; 5=Primary; 6=Primary_Link; 7=Secondary; 8=Secondary_Link; 9=Tertiary; 10=Tertiary_Link; 11=Unclassified; 12=Unknown; 13=Residential; 14=Living_Street; 15= Pedestrian; 16=Service; 17=Path; 18=Steps; 19=Track; 20=Track_Grade1; 21=Track_Grade2; 22=Track_Grade3; 23=Track_Grade4; 24=Track_Grade5; 25=Bridleway; 26=Cycleway; 27=Footway

k=1,.......,4 (Type of locality: 1=Urban areas, 2=Small inhabited areas, 3=Productive areas, 4=Wide spread houses)

l= 0,……,9 (Type of road by Istat classification: Regional roads outside urban area (0), Urban roads (1), Rural roads outside urban area (2), Main roads inside urban area (3), Local rural roads (4), Rural roads outside urban area (5), Main roads outside urban area (6), Other roads outside urban area (8), Regional roads outside urban area (9))

In 2023, there are over 390,000 Traffic Points on the road arches of the national road network and almost 75,000 km of carriageway, considering the extension of all the arches that contained at least one traffic point. This extension represents, in Italy, about 3% out of the total network. The types of roads, according to the Istat classification, which have the highest percentages of traffic points, are motorways (14.1%), primary roads (9.0%), secondary roads (4.2 %) and residential roads (3.2%).
With this criterion, in fact, with the same number of accidents and road kilometers extension, for a similar category of road in two different provinces, for example, the presence of PoT modifies the risk of road accidents, with disadvantage of road sections with less traffic flow, thus they result with an higher danger in terms of vehicles / km (Chart 7).

Chart 7: OpenStreetMap Road Arches with Point Of Traffic. Year 2023



Motorways Primary roads Secondary roads Residential roads

Source: Istat processing on OpenStreetMap data

3.3 Road accidents indicators computing: main results

The indicators use 2023 data and calculate ratios between road accidents, fatalities, injuries, and vehicles, and the length of roads by type. The analysis also includes additional information on traffic count points, resident population, and the vehicle fleet.

This **set of road accident indicators per road length** measures the number of accidents, vehicles involved, fatalities, and injuries for every 100 kilometers of carriageway in each province.
Table 2.14 (see Data Tables) reports the "Rates of accidents, vehicles involved, fatalities, and injuries per 100 kilometers of carriageway by Italian province (systematic classification) – Year 2023." Table 2.15 shows the same indicators for 2023 but breaks them down by a more detailed road classification.
The data in Table 2.15, which uses a more precise road classification, highlight the highest risk exposure—based on the number of vehicles and injuries—on motorways and urban roads, especially in large cities. For example, Milan shows the highest rates for accidents and vehicles per 100 kilometers of motorway (145.15 and 323.49, respectively).

For urban roads, Genoa, Milan, Florence, Prato, and Rome show the highest accident rates per 100 kilometers, with values ranging from 68.24 to 50.46. On extra-urban roads, mid-sized provinces show the highest rates. Monza and Brianza, for instance, reports 35.74 accidents and 73.28 vehicles involved per 100 kilometers.
As for fatalities per 100 kilometers of carriageway, the highest values appear on motorways and junctions in Ascoli Piceno (4.48), on urban roads in Rome (0.60), and on extra-urban roads in Venice (0.76).
When looking at injury rates per 100 kilometers, Milan leads both for motorways (223.64), urban roads (82.86), while Monza, and Brianza shows the highest rate on extra-urban roads (51.96).
The synthetic index, calculated using the MZ method on road segments, ranks Milan at the top, followed by Monza and Brianza, Rome, Naples, and Florence, with index values ranging from -3.4104 to -1.7083.

The **set of road accident indicators based on the vehicle fleet** measures the number of accidents, vehicles involved, fatalities, and injuries per 100,000 registered vehicles in each province. Accident and injury rates by road type, adjusted for the vehicle fleet, reach their highest levels on motorways in Genoa, Bologna, and Savona (51.39, 44.66, and 40.64 accidents per 100,000 vehicles, respectively), and for injuries in Savona, Bologna, and Genoa (81.99, 81.40, and 80.43). For urban roads, Genoa, Prato, and Milan show the highest accident rates per 100,000 circulating vehicles (579.37, 422.92, and 422.65). On extra-urban roads, Grosseto, Ravenna, and Latina lead in accident rates (150.38, 139.46, and 131.53). Injury rates are highest on urban roads in Genoa (671.05) and on extra-urban roads in Grosseto (232.06).
When measuring the ratio between vehicles involved in accidents and the vehicle fleet (per 100,000 circulating vehicles), the highest values occur in Bologna for motorways and junctions (101.37), in Genoa for urban roads (954.61), and in Ravenna for extra-urban roads (252.66).
Fatality rates per 100,000 registered vehicles are highest in Ascoli Piceno on motorways (3.55), in Nuoro on urban roads (5.58), and in Southern Sardinia on extra-urban roads (9.22).
It is important to note that, for motorways, provinces with major infrastructure hubs and seasonal factors may show distorted indicator values. The number of vehicle registrations in these areas often does not match actual traffic flows. In urban areas, the indicators do not reflect commuting patterns, as the vehicle fleet distribution does not capture cross-boundary traffic. On extra-urban roads, the indicators also fail to account for the dense network of consular roads and actual traffic volumes (see Data Tables – Table 2.16).
Looking at the synthetic index calculated using the MZ method based on the vehicle fleet; Bologna ranks first, followed by Genoa, Savona, Piacenza, and Ravenna, with index values ranging from -1.6023 to -1.0257.

The **set of road accident indicators based on the resident population** measures the number of accidents, vehicles involved, fatalities, and injuries per one million residents in each province.

When adjusting accident, vehicle, and injury rates by road type to the resident population, the highest values appear in Genoa for motorways (443.69 accidents per 100,000 residents) and in Savona for vehicles involved in motorway accidents (866.39 per 100,000 residents).
For urban roads, Genoa shows the highest rates per 100,000 residents: 5,002.30 for accidents, 8,242.12 for vehicles, and 5,793.82 for injuries. For extra-urban roads, Grosseto leads with 1,500.19 for accidents, 2,504.94 for vehicles, and 2,315.10 for injuries.
Fatality rates per 100,000 residents are highest in Ascoli Piceno on motorways (34.82), in Nuoro on urban roads (55.50), and in Rieti on extra-urban roads (86.46).
The synthetic index based on resident population and calculated using the MZ method ranks Savona first, followed by Bologna, Ravenna, Florence, and Piacenza, with index values ranging from -1.4749 to -0.9909.
It is important to note that resident population figures do not reflect the actual number of people exposed to road risk in provinces that host key motorway junctions. This often leads to distorted indicator values. Port areas, transit zones, and industrial districts in urban areas also expose a much larger number of individuals to risk than the number of official residents. Similarly, for extra-urban roads, the indicators do not capture the true scale of road use. As a result, these rates may appear exaggerated or misleading (see Data Tables – Table 2.17).

The **new set of road accident indicators per road length, weighted by traffic point presence**, measures the number of accidents, vehicles involved, fatalities, and injuries per 100 kilometers of carriageway in each province, while adjusting for the impact of traffic concentration along the roads.
To calculate these values, the model multiplies the unadjusted indicators per road length by the complement percentage of road segments without traffic points in the province (see Data Tables – Table 2.15 T).
This adjustment reveals notable changes in rankings compared to the unweighted indicators. For example, on motorways, Prato and Bologna switch positions for fatality rates, with Bologna ranking lower after adjusting for traffic presence. Similarly, for extra-urban roads, Rome shows a higher adjusted risk than Venice, highlighting the real road danger beyond traffic volume.
The synthetic index calculated using the MZ method, adjusted for road segments with traffic points, places Milan at the top, followed by Monza and Brianza, Rome, Naples, and Florence, with index values ranging from -3.3299 to -1.7196.
The highest traffic-adjusted accident and injury rates per 100 kilometers of carriageway occur on motorways and junctions in Milan: 139.69 for accidents, 215.06 for injuries, and 315.75 for vehicles involved. For urban roads, Genoa records the highest accident rate (73.25), while Milan has the highest injury rate (90.96). On extra-urban roads, Monza and Brianza leads with the highest values for accidents (39.98), vehicles (80.65), and injuries (60.27).

Fatality rates, adjusted for traffic concentration, peak in Genoa on motorways (13.96 per 100 kilometers of carriageway), in Prato on urban roads (0.72), and in Lecco on extra-urban roads (0.66) (see Data Tables – Table 2.15 T).
The team calculated the index values using the MZ method for road segments, vehicle fleet, and population, based on the analytic classification (see data tables, Table 2.19).
The Ranker software, used to calculate the composite indicators, requires users to set the “polarity” of each indicator to rank values and assign meaning through a + or - sign. For the 2018 data, the team also included the additional element of the "traffic point factor". They used weights based on the complementary percentage (1-p) of road sections with traffic points (p) to calculate the indicators per road length. This method aims to remove the effect of traffic volume and better highlight the impact of driver behavior and infrastructure damage. This approach offers a more intuitive interpretation by assigning a negative polarity to indicators, which reflects the level of road danger in each province (see data tables, Table 2.19).

4. The computing of synthex indicators

To realize a ranking and a classification of indicators, a generalized tool, developed by Istat, was used. Two generalized tools are available for the analysis and benchmarking of results produced by different composite indicators: RankerTool is a desktop and i.Ranker web application (Appendix A).

Both, with few differences, allow the user to:

- acquire, in standard format (csv or .xls), the values of the indicators available for each entity, already calculated and normalized;
- calculate, for each entity, one or more methods among those implemented;
- display the values and rank results for each method, both in tables and graphic way;
- compare the rankings using the different methods.

The Ranker tool used give the possibility to choose different methods.

The methods proposed are: the Average method of standardized values (MZ), the MR method - relative indices (IR), the MPI method - Mazziotta-Pareto Index (De Muro et al. 2010), the Graduations Method (MG) and the Wroclaw Taxonomic Method (MTW). The last two options were excluded a priori, as they were not considered suitable for ranking indicators on road accidents and representing the phenomenon under study. The ranking method, in fact, does not take into account the outliers, while the Wroclaw taxonomic method is based on the distance from an ideal unit in the Euclidean space.

The steps on the computing process is divided in two phases:

- **standardisation** of elementary indicators, the standardization aims to make the indicators comparable as they are often expressed in different units of measurement and may have different polarities. Therefore, it is necessary to bring the indicators to the same standard, reversing the polarity, where necessary, and turning them into pure, dimensionless numbers;

- **aggregation of standardized indicators**: the combination of all the components to form the synthetic index (mathematical function), after the standardisation by Ranker tool[9].

For the selection of the method to be used for synthetic indexes computing, a **robustness test** was carried out and an **influence analysis** was done by applying the COMIC[10] software (COMposite Indices Creator), through synthesis methods and the evaluation of their consistency.

Through the COMIC software, it was possible to make a comparison between the degrees of robustness for the main applicable methods; in particular, table 6 shows the data on the mean and standard deviations of the *shifts* for different methods.

The results of the influence analysis and robustness tests data show that the best methods to be used, among the different alternatives, are the MZ method - arithmetic mean (z-scores) and MPI - Mazziotta -Pareto Index.

The **MZ method - arithmetic mean (z-scores)**, well known method that allow an easy interpretation of results for all common users, has been selected for the presentation of the data contained in this research.

Table 7 – Summary of means and standard deviation of the *"shifts"* (a)

Method for Synthex Indicators	Mean of the mean of the shift	Standard deviation of the mean of the shift	Mean of the Standard deviation of the shift	Standard deviation of the Standard deviation of the shift
Index Mean 0-1	2,092	0,828	2,152	0,857
MZ aritmetic mean z-scores	2,224	0,604	2,356	0,631
Jevons static	2,671	1,416	2,752	1,221
MPI	2,183	0,667	2,440	0,938
MPI correct	2,136	0,862	2,288	1,090
IMG	2,006	1,098	2,530	2,057

(a) Results of the robustness test, using the COMIC software - Synthex indicators for road accidents rates by road arch, vehicles fleet and resident population. Year 2016

The application of different weighting criteria leads to very divergent results. The analysis according to the road infrastructures allows purifying a component of mobility of the phenomenon. The seasonal factor due to a more objective measurement also improves the concept of exposure to the risk of being involved in a traffic accident.

Table 8 shows the covariance values, referred to the variation of each variable contained in the matrix in respect of all others.

[9] The methodological note and the use guide are available in Appendix A of this document and in the links:
http://www.istat.it/en/files/2014/03/RANKER-manuale.pdf
https://i.ranker.istat.it/wr_guida.htm
https://i.ranker.istat.it/wr_guida_notametodologica.htm

[10] COMIC (COMposite Indices Creator) https://www.istat.it/it/metodi-e-strumenti/metodi-e-strumenti-it/analisi/strumenti-di-analisi/comic

Table 8– Covariance Matrix between computing indicators: results obtained by Ranker application. 2023 data

Ranks	Road Arch	Resident Population	Vehicles Fleet
Road Arch	1.0000	0.3794	0.4830
Resident Population	0.3794	1.0000	0.8843
Vehicles Fleet	0.4830	0.8843	1.0000

The values included in table 8 show, in fact, that the risk to be involved in a road accident, within the province of residence (0.3794) or within the vehicle registration province (0.4830) is significantly lower if compared to the indicator out of the length road arch, where the accident occurred.
The road accidents indicators, referred to the road length by province, therefore, allows to obtain a result closer to the road accident risk measures, using traffic flows data. The last data would represent, in fact, the best and correct denominator for road accidents rates.

Table 9 shows the values of the covariance, referred to the variations of the different classification criteria of the road arches.

Table 9 - Covariance matrix for the classification criteria of road arches on 2018 data

Ranks	Systematic classification	Analytical classification	Classification with PoT
Systematic classification	1.0000	0.9927	0.9931
Analytical classification	0.9927	1.0000	0.9995
Classification with PoT	0.9931	0.9995	1.0000

The table shows as the analytical classification criteria for road arches leads to a qualitative improvement in the attribution of the events, according to the Istat road accidents survey variable (0.9931), without significantly altering the two provincial distributions. The inclusion of the traffic factor allows reading the road accidents components net of other factors, such as the infrastructural and behavioral. It leads, moreover, to a more objective evaluation criterion of the phenomenon (0.9995). This measurement makes it more evident the identification of actions to improve road safety by the stakeholders and in the prevention actions by the law enforcement agencies in charge of road control.

APPENDIX A

Ranker tool for a composite index computing

A1. Computation and evaluation of composite indices [11]

A *composite index* is a mathematical combination (or aggregation as it is termed) of a set of indicators [1] that represent the different dimensions of a phenomenon to be measured.

Constructing a composite index is a complex task. Its phases involve several alternatives and possibilities that affect the quality and reliability of the results. The main problems, in this approach, concern the choice of theoretical framework, the availability of the data, the selection of the more representative indicators and their treatment in order to compare and aggregate them.

In particular, we can summarize the procedure in the following main steps:

1. *Defining the phenomenon to be measured*. The definition of the concept should give a clear sense of what is being measured by the composite index. It should refer to a theoretical framework, linking various sub-groups and underlying indicators. Also the *model of measurement* must be defined, in order to specify the relationship between the phenomenon to be measured (concept) and its measures (individual indicators). If causality is from the concept to the indicators we have a *reflective model* – indicators are interchangeable and correlations between indicators are explained by the model; if causality is from the indicators to the concept we have a *formative model* – indicators are not interchangeable and correlations between indicators are not explained by the model.
2. *Selecting a group of individual indicators*. The selection is generally based on theory, empirical analysis, pragmatism or intuitive appeal. Ideally, indicators should be selected according to their relevance, analytical soundness, timeliness, accessibility and so on. The selection step is the result of a trade-off between possible redundancies caused by overlapping information and the risk of losing information. However, the selection process also depends on the measurement model used: in a reflective model, all the individual indicators must be intercorrelated; whereas in a formative model they can show negative or zero correlations.
3. *Normalizing the individual indicators*. This step aims to make the indicators comparable. Normalization is required before any data aggregation as the indicators in a data set often have different measurement units. Therefore, it is necessary to bring the indicators to the same standard, by transforming them into pure, dimensionless, numbers. Another motivation for the normalization is the fact that some indicators may be positively correlated with the phenomenon to be measured (positive *polarity*) [2], whereas others may be negatively

[11] RankerTool desktop software (http://www.istat.it/en/tools/methods-and-it-tools/analysis-tools/ranker)
i.Ranker web application (https://i.ranker.istat.it)

correlated with it (negative *polarity*). We want to normalize the indicators so that an increase in the normalized indicators corresponds to increase in the composite index. There are various methods of normalization, such as *re-scaling* (or Min-Max), *standardization* (or *z*-scores) and 'distance' from a reference (or *index numbers*).

4. *Aggregating the normalized indicators*. It is the combination of all the components to form one or more composite indices (mathematical functions). This step requires the definition of the importance of each individual indicator (weighting system) and the identification of the technique (*compensatory* or *non-compensatory*) for summarizing the individual indicator values into a single number [3]. Different aggregation methods can be used, such as *additive methods* (compensatory approach) or *multiplicative methods* and *unbalance-adjusted functions* (non-compensatory or partially compensatory approach) [4].
5. *Validating the composite index*. Validation step aims to assess the robustness of the composite index, in terms of capacity to produce correct and stable measure, and its discriminant capacity (*Influence Analysis* and *Robustness Analysis*).

A2. The Synthesis Methods

The synthesis methods available on i.ranker are based on the assumption of a formative model:

- Mean of standardised values (MZ);
- Meean of relative index (MR);
- Mazziotta-Pareto index (MPI+/MPI-);

Mean of the standardised values (MZ)

Given the matrix **X**={x_{ij}} of *n* rows (units) and *m* columns (indicators), transformation matrix **Z**={z_{ij}} is built, with:

$$z_{ij} = \begin{cases} \dfrac{(x_{ij} - M_{x_j})}{S_{x_j}} & \text{if the indicator j has positive polarity} \\ -\dfrac{(x_{ij} - M_{x_j})}{S_{x_j}} & \text{if the indicator j has negative polarity} \end{cases}$$

where M_{x_j} e S_{x_j} are, respectively, mean and standard deviation of the indicator *j*.

The synthex index, for the unit *i*, is given by the formula:

$$MZ_i = \frac{\sum_{j=1}^{m} z_{ij}}{m}$$

Relative Indices Synthesis (MR)

Given the matrix **X**={x_{ij}} of *n* rows (units) and *m* column (indicators), the transformation matrix **R**={r_{ij}} is built, with:

$$r_{ij} = \begin{cases} \dfrac{(x_{ij} - \mathrm{Min}_{x_j})}{(\mathrm{Max}_{x_j} - \mathrm{Min}_{x_j})} & \text{if the indicator j has positive polarity;} \\[2ex] \dfrac{(\mathrm{Max}_{x_j} - x_{ij})}{(\mathrm{Max}_{x_j} - \mathrm{Min}_{x_j})} & \text{if the indicator j has negative polarity;} \end{cases}$$

where Min_{x_j} and Max_{x_j} are, respectively minimum and maximum of the indicators *j*.

The synthex index, for the unit *i*, is given by the formula:

$$\mathrm{MR}_i = \frac{\sum_{j=1}^{m} r_{ij}}{m} .$$

Mazziotta-Pareto index (MPI $^{+}$/MPI $^{-}$)

Given the matrix **X**={x_{ij}} of n rows (units) and *m* columns (indicators), transformation matrix **Z**={z_{ij}}, is built with:

$$z_{ij} = \begin{cases} 100 + \dfrac{(x_{ij} - \mathrm{M}_{x_j})}{\mathrm{S}_{x_j}} 10 & \text{if the indicator j has positive polarity;} \\[2ex] 100 - \dfrac{(x_{ij} - \mathrm{M}_{x_j})}{\mathrm{S}_{x_j}} 10 & \text{if the indicator j has negative polarity;} \end{cases}$$

where M_{x_j} e S_{x_j} are, respectively, mean and standard deviation of the indicator *j*.

The synthex index, for the unit *i*, is given by the formula [5] :

$$\mathrm{MPI}_i^{+/-} = \mathrm{M}_{z_i} \pm \mathrm{S}_{z_i} \mathrm{cv}_i$$

dove $\mathrm{M}_{z_i} = \dfrac{\sum_{j=1}^{m} z_{ij}}{m}$; $\mathrm{S}_{z_i} = \sqrt{\dfrac{\sum_{j=1}^{m} (z_{ij} - \mathrm{M}_{z_i})^2}{m}}$; $\mathrm{cv}_i = \dfrac{\mathrm{S}_{z_i}}{\mathrm{M}_{z_i}}$.

A3. The synthesis method

Table A1 contains a list of the main features of available methods, to select a suitable synthex index to the phenomenon studied.

The advantages of a synthex index can be summarized in:

(a) one-dimensional measurement of a complex phenomenon, (b) easy interpretation in respect to a set of elementary indicators ("dashboard"), (c) simplification of data analysis (in particular, ordering of geographical units). Table A1 shows the main features of available method:

Table A1. Methods of indices synthesis features

Method of synthesis	Main features
Mean of standardised values (MZ)	Compensatory method It is based on the arithmetic mean of the z-scores. Mean 0 range between -3 and +3. Assumption: the indicators have the same variability.
Mean of relative indices (MR)	Compensatory method It is based on the arithmetic mean of the relative values. Range between 0 and 1. Does not exist a fixed mean.
Mazziotta-Pareto index (MPI^+/MPI^-)	Non-compensatory method It is based on an arithmetic mean penalized on the basis of the imbalance of values. Mean equal to 100, range between 70 and 130 Assumption: the indicators have the same variability. It is applicable to both positive events (MPI-) and negative (MPI +). It can be divided into two parts: "medium" (compensatory) effect and "penalty" effect (imbalance).

Bibliography

Aureli Cutillo E. (1996), *Lezioni di statistica sociale. Parte seconda, sintesi e graduatorie*, CISU, Roma.

Delvecchio F. (1995), *Scale di misura e indicatori sociali*, Cacucci editore, Bari.

Harbison F. H., Maruhnic J., Resnick J. R. (1970). *Quantitative Analyses of. Modernization and Development*, Princeton University Press, New Jersey.

Maggino F. (2008), *La misurazione dei fenomeni sociali attraverso indicatori statistici. Aspetti metodologici*, Università degli Studi di Firenze – Working paper, Firenze.

Mazziotta M., Pareto A. (2011), Un indice sintetico non compensativo per la misura della dotazione infrastrutturale: un'applicazione in ambito sanitario, *Rivista di Statistica Ufficiale*, n. 1/2011, pp. 63-79.

Mazziotta M., Pareto A. (2015). On a Generalized Non-compensatory Composite Index for Measuring Socio-economic Phenomena, *Social Indicators Research*, DOI 10.1007/s11205-015-0998-2.

Mazziotta M., Pareto A. (2013). Methods for constructing composite indices: one for all or all for one?, Rivista Italiana di Economia Demografia e Statistica Volume LXVII n. 2 Aprile-Giugno 2013

OECD (2008), *Handbook on Constructing Composite Indicators. Methodology and user guide*, OECD Publications, Paris.

Silvio-Pomenta J. F. (1973), *Typological study using the Wroclaw Taxonomic Method (A study of regional disparities in Venezuela)*, SHC/WS/316, UNESCO.

---

[1] An simple indicator is computed data by means the ratio of a “raw” to a reference base (for example, “per capita income”).

[2] The polarity of an indicator is the sign of the relationship between the indicator and the phenomenon to be measured

[3] A simple indicator is considered “substitutable” if a deficit of an indicator can be compensated by a surplus in other.

[4] A simple indicator is considered “not substitutable” if a deficit of an indicator cannot be compensated by a surplus in other.

[5] If the composite index is ‘increasing’ or ‘positive’, i.e., increasing values of the index correspond to positive variations of the phenomenon (e.g., the socio-economic development), then MPI- is used. Vice versa, if the composite index is ‘decreasing’ or ‘negative’, i.e., increasing values of the index correspond to negative variations of the phenomenon (e.g., the poverty), then MPI+ is used.

Edited by: Marco Broccoli (broccoli@istat.it) e Silvia Bruzzone (bruzzone@istat.it)