Sapienza University of Rome, 22 June 2026

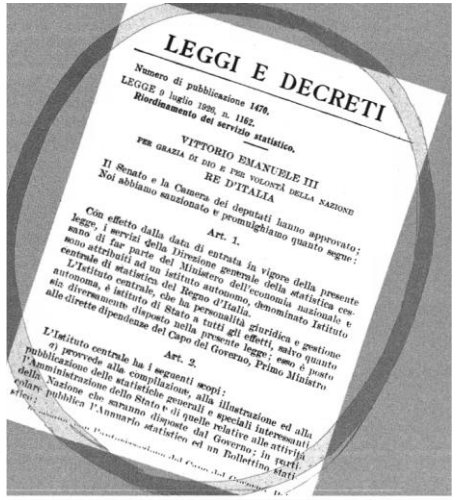
Satellite meeting of the Joint Meeting SIS-FENStatS 2026

100 years of methodological evolution: Official Statistics between scientific rigor and digital innovation

Marco Di Zio
Istat

1926. Istituto Centrale di Statistica

- Coordinate and centralize statistical activities of public administrations
- Regulate and standardize statistical surveys
- Produce, process and disseminate statistical data
- Promote studies on Italy's economic and social conditions


LEGGI E DECRETI

Numero di pubblicazione 1679.

LEGGE 9 luglio 1926, n. 1162.

Riordinamento del servizio statistico.

VITTORIO EMANUELE III
PER GRAZIA DI DIO E PER VOLONTÀ DELLA NAZIONE
RE D'ITALIA

Il Senato e la Camera dei deputati hanno approvato;
Noi abbiamo sanzionato e promulghiamo quanto segue:

Art. 1.

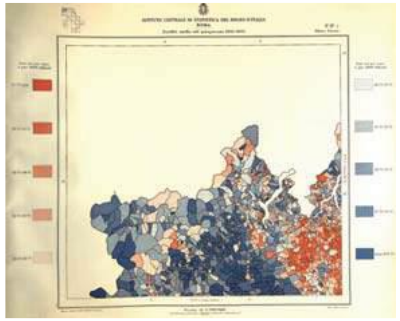
Con effetto dalla data di entrata in vigore della presente legge, i servizi della Direzione generale della statistica cessano di far parte del Ministero dell'economia nazionale e sono attribuiti ad un istituto autonomo, denominato Istituto centrale di statistica del Regno d'Italia.
L'Istituto centrale, che ha personalità giuridica e gestione autonoma, è istituto di Stato a tutti gli effetti, salvo quanto sia diversamente disposto nella presente legge; esso è posto alle dirette dipendenze del Capo del Governo, Primo Ministro

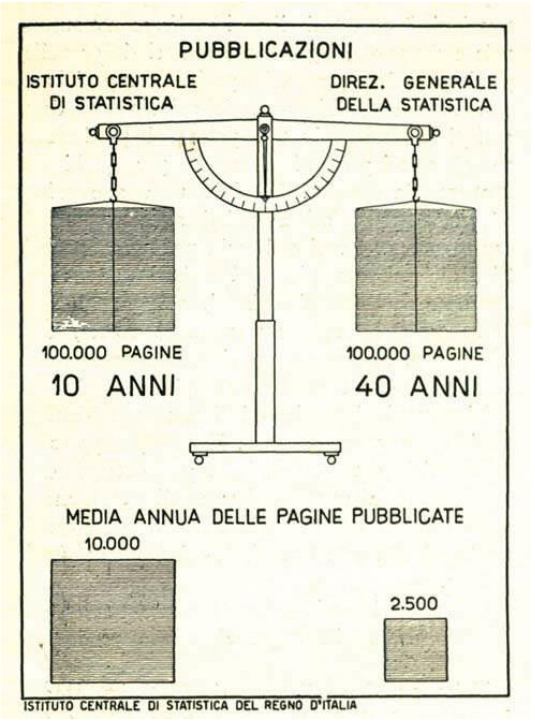
Art. 2.

L'Istituto centrale ha i seguenti scopi:
a) provvede alle compilazione, alla illustrazione ed alla pubblicazione delle statistiche generali e speciali interessanti l'Amministrazione dello Stato e di quelle relative alle attività dell'Amministrazione che saranno disposte dal Governo; in parti-
colare pubblica l'Annuario statistico ed un Bollettino stati-
stico;

Census and administrative data

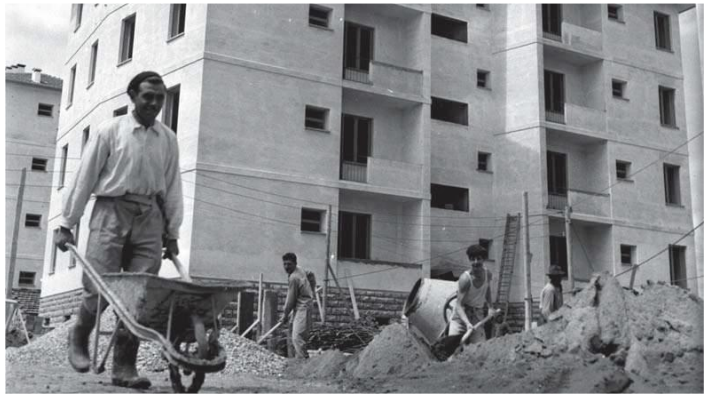
- In the first decades, statistics were based on administrative data and large census operations.
- Face-to-face interviews with paper questionnaires.
- Data processing progressively became mechanized. Punched-card systems.





The post-war period

- Growing demand for statistics to support reconstruction and public policies.
- More detailed demographic, social and economic information became essential.



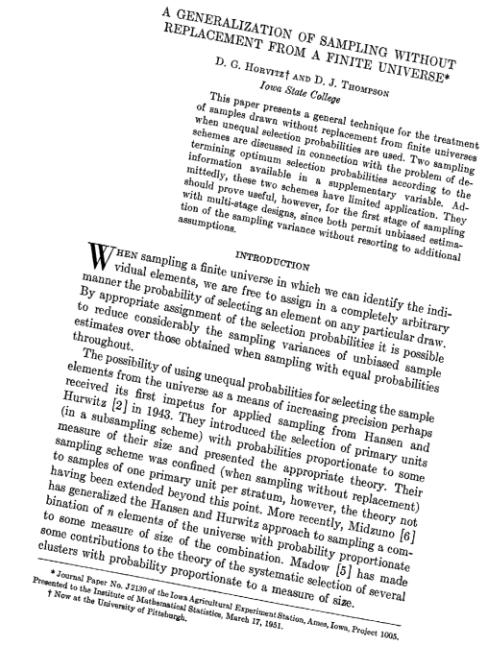
In that period

Society

- *10 September 1952 at 9.00 p.m. ,from the studi Rai di Milano, the first Italian RAI news program (Telegiornale) was broadcast on television*

Statistics

- *Horvitz and Thompson (1952).* A Generalization of Sampling Without Replacement from a Finite Universe *Jour. of the American statistical Association, 47, 663-685.*



A GENERALIZATION OF SAMPLING WITHOUT REPLACEMENT FROM A FINITE UNIVERSE*

D. G. Horvitz† and D. J. Thompson

Iowa State College

This paper presents a general technique for the treatment of samples drawn without replacement from finite universes when unequal selection probabilities are used. Two sampling schemes are discussed in connection with the problem of determining optimum selection probabilities according to the information available in a supplementary variable. Admittedly, these two schemes have limited application. They should prove useful, however, for the first stage of sampling with multi-stage designs, since both permit unbiased estimation of the sampling variance without resorting to additional assumptions.

INTRODUCTION

When sampling a finite universe in which we can identify the individual elements, we are free to assign in a completely arbitrary manner the probability of selecting an element on any particular draw. By appropriate assignment of the selection probabilities it is possible to reduce considerably the sampling variances of unbiased sample estimates over those obtained when sampling with equal probabilities throughout.

The possibility of using unequal probabilities for selecting the sample elements from the universe as a means of increasing precision perhaps received its first impetus for applied sampling from Hansen and Hurwitz [2] in 1943. They introduced the selection of primary units (in a subsampling scheme) with probabilities proportionate to some measure of their size and presented the appropriate theory. Their sampling scheme was confined (when sampling without replacement) to samples of one primary unit per stratum, however, the theory not having been extended beyond this point. More recently, Midzuno [6] has generalized the Hansen and Hurwitz approach to sampling a combination of n elements of the universe with probability proportionate to some measure of size of the combination. Madow [5] has made some contributions to the theory of the systematic selection of several clusters with probability proportionate to a measure of size.

* Journal Paper No. J-2139 of the Iowa Agricultural Experiment Station, Ames, Iowa, Project 1005. Presented to the Institute of Mathematical Statistics, March 17, 1951.

† Now at the University of Pittsburgh.

1952. The age of survey sampling

- Sampling offered a cost-effective yet scientifically rigorous alternative.
- 1952. First probability sample surveys on Labour Force and Agricultural Production.
- Employed in (agriculture, industry, service sector) and unemployed.
- Production of wheat, grapes, and olive trees.
- The introduction of sample surveys: a particularly important milestone in the production of official Italian statistics.

1980 - 2000. Istituto Nazionale di Statistica - Sistan

- 1986. Istat has a role of Public research body (National decrees) dedicated to the production and analysis of statistical data.
- 1989. Istat is named Istituto Nazionale di Statistica
- 1989. The National Statistical System (Sistan) is officially established. An integrated network of public agencies and administrations coordinated by Istat, responsible for collecting and disseminating statistical information


DELLA REPUBBLICA ITALIANA — Serie generale - n. 66

Art. 7.

Comparto del personale delle istituzioni e degli enti di ricerca e sperimentazione

1. Il comparto di contrattazione collettiva del personale delle istituzioni e degli enti di ricerca e sperimentazione comprende il personale dipendente:

dagli enti scientifici di ricerca e sperimentazione di cui al punto 6 della tabella allegata alla legge 20 marzo 1975, n. 70, e successive modificazioni ed integrazioni;

dall'Istituto centrale di statistica (I.S.T.A.T.);

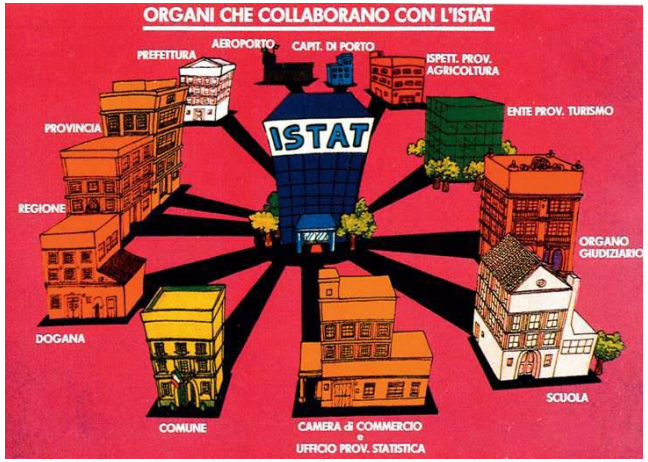
dall'Istituto superiore per la prevenzione e la sicurezza del lavoro (I.S.P.E.S.L.);

dall'Istituto superiore di sanità (I.S.S.);

dall'Istituto italiano di medicina sociale;

dagli Istituti di ricerca e sperimentazione agraria e talassografici;

dalle stazioni sperimentali per l'industria.



1980 - 2000. Administrative data in the production process

- Years characterised by an increasing availability and use of administrative data.
- Use of administrative data for purposes unrelated to the scope for which they were collected.
- Need for the development of methodologically structured processes for data harmonisation, transformation, and validation.
- Administrative data, used as supplementary information in statistical production processes, contribute significantly to improving the efficiency and accuracy of estimates.
- Significant advancements in statistical methodologies for the use of administrative data with a view to integration.

1980 - 2000. Methodological innovations: calibration estimators

- Based on the adjustment of sampling weights to obtain estimates consistent with known population totals, often derived from administrative sources.
- Improving efficiency of estimates and their consistency with external benchmarks, thereby helping to enhance the overall reliability of the statistical information produced.
- 1997. Labour force survey one of the first application.

- *Deville and Särndal (1992). Calibration estimators in survey sampling. JASA.*

1980 - 2000. Methodological innovations: small area estimates

- A favourable context for the first experimentation of SAE methods.
- SAE integrates survey data with auxiliary information to produce estimates of parameters on small size sub-populations that would be unreliable with traditional methods.
- First application is on the estimates of unemployed and employed for “local labour systems” updated to 2001

- *Fay R. E. and Herriot R. A.(1979) Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data. JASA*
- *Battese G.E. , Harter R.M., and Fuller W.A. (1988).* An error components model for prediction of county crop area using survey and satellite data. JASA
- *Rao J.N.K. (2003). Small Area Estimation. John Wiley & sons*

1980 - 2000. Methodological innovations: Data Integration

Statistical methods for data integration: probabilistic record linkage and statistical matching.

- Probabilistic record linkage identifies records belonging to the same entity across different datasets, assigning weights based on the discriminatory power of the fields .
- Statistical matching integrates data sources that do not share common units but have common variables.
- First record linkage application at the end of ’90 in the construction of Archivio Statistico delle Imprese Attive (ASIA).
- Statistical matching: integration of time use survey and labour force.
- *Fellegi, I.P. and Sunter, A.B. (1969). A theory for record linkage. JASA.*
- *Jaro, M.A. (1989). Advances in record-linkage methodology as applied to matching the 1985 Census of Tampa, Florida. JASA*
- *D'orazio, M., Di Zio, M., & Scanu, M. (2006). Statistical matching: Theory and practice. John Wiley & Sons.*

Electronic questionnaires and computer-assisted techniques

- 1990s marked the transformation of personal computers from niche office equipment to mainstream household staples.
- Technological advancements directly influenced data collection techniques, promoting the use of computer-assisted interviews.
- Paper questionnaire were gradually replaced by software interfaces used by an interviewer through a telephone interview (CATI) or a computer (CAPI).
- Advantages: quality, possibility of real-time validation checks, timeliness and standardization.

2000-2015 European harmonisation and Web

- Regulations, codes, and manuals of best practices developed by Eurostat and NSIs strongly guided the design and revision of statistical production processes.
- 1998. Revision of the Labour Force Survey: conducted every week of the year, replacing the quarterly survey. A mixed-methods data collection was introduced to reduce costs and statistical burden.
- 2005. The European Statistics Code of Practice was adopted. It establishes fundamental principles to ensure the quality, independence, impartiality, and credibility of official statistics produced by Eurostat and NSIs.

2000-2015 Web

- Internet access grows
- CAWI (Computer Assisted Web Interview).
- The first electronic questionnaires submitted via the web were those for surveys on businesses and institutions (1998, 1999)
- The first survey to use a CAWI component as a supplement to a telephone survey was the 2009–2010 survey on the professional integration of Ph.D. graduates

2016 –NSIs Modernization

Official statistics over the past few decades have been characterized by

- declining response rates
- the need to reduce production costs and the burden on respondents,
- need to produce statistical information that more timely and more detailed.

2016 – Data deluge and Machine learning - AI

At the same time, enormous opportunities

Data

- The digitilization and datafication of phenomena have expanded the possibilities for gathering information
- This opens new perspectives for understanding social and economic phenomena, improving the statistics produced, and broadening the scope of information thanks to their timeliness and temporal and spatial granularity.

Methods

- Rapid development of machine learning methods and, more generally, artificial intelligence (AI)
- They make it possible to process large volumes of data, identify complex relationships, and leverage unstructured sources, such as texts.

2016 – Istat Modernization plan

Multi-source production system

- A new statistical production framework based on an infrastructure built around the integrated system of registers, which incorporates sample surveys and information from “non-traditional data sources” such as big data.
- The transition to this production system—which is still evolving—has entailed significant organizational, methodological, and technological changes.

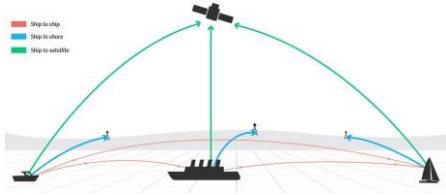
Significant representative case: Permanent Census of Population

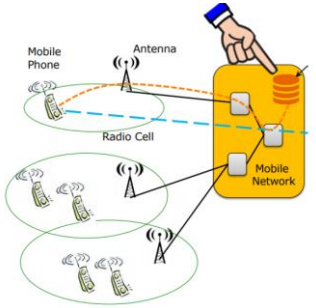
From decennial to annual statistics through the use of

- Administrative data for the population and some core variables (Integrated System of Registries)
- Use of survey sample
- Use of big data (smart meters)

2016 – Work in progress with non-traditional data sources

- Remote sensing: land cover, urban vegetated areas, updating territorial basis
- Web-scraping: Online Job Advertisemet (OJA),..
- Signals: AIS data for maritime transport statistics
- MNO: for tourism, population movements…
- …..





2016 – Work in progress on machine learning and AI

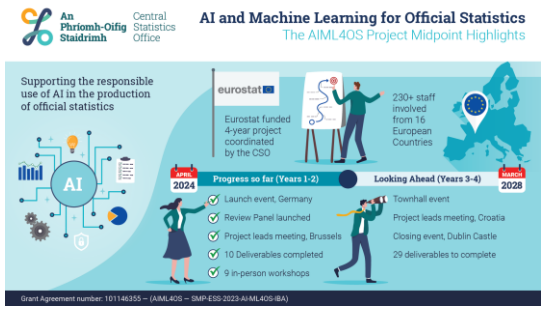
- Smart surveys: use of respondent's smart devices, combining data from web questionnaire with data from sensors
- Machine learning for dealing with missing data
- LLM for classification: ATECO, …
- AI for data validation

2016 – Working in an international network

INTL projects

- 2016-2018. Big data I (led by Istat). 22 countries.
- 2018-2021. Big data II. 28 Partners
- Smart Surveys (2020-2022)
- 2023-2025 Smart Surveys Implementaion. 7 countries – 3 Univ.
- 2021-2025. WIN project. Case study: Online Job Advertisements – OJA and Online Based Enterprise Characteristics – OBEC. 14 countries.
- 2023-2025. MNO Minds. Methodologies for the use of mobile data (led by Istat). 10 countries.
- 2024-2028. AIML for Official Statistics. 16 countries.

Co-funded by the European Union



1926 | 2026 Istat

2016 – Methodological questions

- Integration of sample surveys and big data poses significant methodological challenges because they differ in nature and quality.
- Probabilistic samples ensure representativeness and rigorous inference.
- Big data, extremely rich in information, may not be representative of the population of interest, lacks a controlled statistical design, and often lacks information regarding its quality.
- Challenges are in harmonizing information, managing data quality, and applying classical/new inference techniques.

2016 – Quality and transparency

- Quality indicators must cover all dimensions included in official statistics, and, in the case of integration with sample surveys, the assessment must take into account the different inferential contexts.
- Transparency is closely linked to the interpretability of the models used; in fact, it is essential that the methodologies adopted be understandable and that it be possible to clearly reconstruct the process by which certain results are reached.
- This requirement poses a significant challenge in the field of advanced machine learning and AI techniques, which, despite their high predictive capabilities, are often characterized by limited interpretability.

2016 – next years

The main challenge will be

- the full utilization of new information, technological, and methodological resources to produce more timely, detailed, and relevant results,
- but always in accordance with high standards of quality, transparency, and reliability, to preserve the role of official statistics as a reliable source of information.

Thank you

Marco Di Zio | marco.dizio@istat.it

1926 | 2026

Istat | Istituto Nazionale di Statistica