Algorithms and official statistics

Introduction

1. The Office for National Statistics (ONS) is the UK’s National Statistical Institute, and largest producer of official statistics.

2. ONS aims to provide a firm evidence base to inform decision-making, and support democratic debate. We publish data and analysis about a range of key economic, social and demographic trends, including statistics on the value of the UK economy and the size, geographic distribution and characteristics of the population, as well as indicators of price inflation, employment, earnings, crime and migration.

3. As is set out in further detail below, ONS uses algorithms in both the collection and production of official statistics. Because ONS only collects data for statistical and research purposes – never for operational reasons – we do not use algorithms for decision-making on individuals, households, or businesses.

4. Nonetheless, it is ONS’s clear view that transparency around the application of algorithms in the collection and production of statistics – and, indeed, in the context of data access and sharing more broadly – is key to building and maintaining public trust. There are a range of measures in place at ONS, detailed below, designed to promote transparency while safeguarding private data.

ONS’s use of algorithms in the collection and production of official statistics

5. ONS typically produces official statistics by applying algorithms to detailed ‘microdata’, such as surveys and administrative data (that is, information collected by public bodies in the course of delivering day-to-day services to citizens). Figures 1 to 3, below, provide specific examples of the ways in which algorithms are applied in the collection and production of official statistics.

6. Under the Statistics and Registration Service Act 2007, ONS is only able to collect data for statistical and research purposes, never for operational reasons. Although data may be held at the level of individuals,
Written evidence submitted by the Office for National Statistics (ALG0060)

households or businesses, this information is only ever used to generate and analyse aggregate statistics about the UK’s economy, society and demography. ONS does not use algorithms for decision-making on individuals, households, or businesses.

**Figure 1: Using algorithms to collect data**

One of the ways ONS use algorithms is in the collection of data. For example:

- Sample selection algorithms are used to decide whether (or not) to carry out data collection for a household, and/or to select a person within a household to interview. Similar algorithms are also used for business surveys.
- Algorithms are also used to validate collected data, and identify whether or not a survey response is an outlier. Based on this, ONS may contact the survey respondent to query and potentially correct the value provided.

**Figure 2: Using algorithms to develop geographies for the collection and publication of statistics**

The UK Census is undertaken every 10 years. Its purpose is to collect population and other statistics essential to those who have to plan and allocate resources.

For both the 2001 and 2011 Censuses, ONS worked with colleagues at the University of Southampton in using an optimisation algorithm to design geographies (referred to as Super Output Areas) for the collection and publication of small-area statistics. The algorithm identified and drew together neighboring Census reporting zones, with common demographic characteristics, into areas of approximately equal population size.

The development of these geographies has provided policy-makers with an improved basis for the comparison of small areas across the country, because units are more similar in size than, for example, electoral wards. The geographies are also intended to be stable; enabling the improved comparison and monitoring of policy over time. Further detail on the geographies is available on the ONS website.

**Figure 3: Using algorithms to identify areas with shared characteristics**

Similarly, in both the 2001 and 2011 Censuses, a clustering algorithm was used to create a number of area classifications. The algorithm used Census data to form groups of areas with similar socio-economic characteristics - which could then be described as a particular type (for example Expanding Areas and Established Cities). It was applied to local authorities, as well as to much smaller areas.

The application of this algorithm, which ONS worked on in collaboration with University College London, has resulted in several outputs which are used extensively by a wide range of users. They can be found here.

7. Data published by ONS may also be used by other organisations as inputs to algorithms that are used for decision-making. For example, predictive models may make use of Census and other ONS data, and these models may in turn be used to make decisions about individuals. The robust and detailed metadata
Promoting transparency, while safeguarding privacy

8. It is ONS’s clear view that transparency around the application of algorithms in the collection and production of statistics – and, indeed, in the context of data access and sharing more broadly – is key to building and maintaining public trust. There are a range of measures in place at ONS designed to promote transparency while safeguarding private data.

9. There is a strong legislative framework in place, governing ONS’s use of data. The Data Protection Act 1998, and its accompanying data protection principles, provide a most important safeguard around ONS’s use of data. The Statistics and Registration Service Act 2007 - in addition to limiting the functions of ONS to the production and publication of official statistics - provides a strong criminal penalty for the unlawful disclosure of personal information.

10. While safeguarding data in accordance with this legislative framework, ONS adopt a transparent and “open by default” approach, to enable the greatest possible use of data for wider public benefit.

11. We publish the statistics we produce openly online, available free of charge to all. In line with the Code of Practice for Official Statistics, all of ONS’s statistical publications include details of the methodology used to collect and produce statistics (including details of algorithms applied).

12. ONS also supports researchers to realise the potential public benefit from data, while safeguarding individuals' confidentiality. Specific permissions exist for central and local government, as well as health bodies, so they can do the detailed analysis they need to plan effectively. In addition, other users can apply for limited access to de-identified unpublished data for statistical and research purposes, through an Approved Researchers Scheme. In all cases, it is ONS's policy only to permit access to de-identified unpublished data for statistical research purposes that serve the public good, and only when access does not compromise the guarantee of confidentiality provided to respondents. Once the research is complete we review the final results to ensure there is no risk to the confidentiality of the data subjects, and we require that all research outputs are published to maximise their public benefit.

13. Finally, to ensure that the access, use and sharing of public data for research and statistical purposes is ethical and for the public good, the National Statistician recently established the National Statistician’s Data Ethics Advisory Committee. The Committee considers project and policy proposals, which make use of innovative and novel data from ONS and the wider Government Statistical Service, and advise the National Statistician on the ethical appropriateness of these. Committee members meet quarterly, and play a key role in ensuring that the access, use and sharing of public data is ethical and for the public good.

---

1 For information on the improved Researchers Scheme, see [here](#).
role in ensuring transparency around the access, use and sharing of data for statistical purposes. All meeting papers and minutes are published on the UK Statistics Authority website.\(^2\)

April 2017

\(^2\) Meeting papers and minutes available [here](#).