Measures of statistical uncertainty summary

Contact:
Louisa Blackwell
pop.info@ons.gsi.gov.uk

Release date:
30 November 2017

Next release:
To be announced

Table of contents

1 . Overview

The measures of statistical uncertainty are research statistics that aim to give users of Office for National Statistics (ONS) local authority mid-year population estimates (MYEs) information about their quality.

We use the cohort component approach to create the local authority MYEs. The cohort component method uses the 2011 Census for the population base and then incorporates natural change (births and deaths), net international migration and net internal migration, and other adjustments (for example, asylum seekers). The census, international and internal migration are the main sources of uncertainty in the MYEs. For now, the uncertainty methodology assumes that there is zero error in the other components such as births and deaths. Since the MYEs combine various data sources and processes to derive each component, we have used different methods to produce 1,000 simulated values for each component. These are then combined using the cohort component formula to derive the uncertainty associated with the local authority MYEs.

The measures of uncertainty for each local authority for the years 2012 to 2016 are summarised in interactive maps for each year (presented in section 2). These show, for each local authority, how the principal measure of uncertainty has changed over time. The principal uncertainty measure is the root mean squared error (RMSE). The RMSE is the variability of the simulated values around the MYE. The higher the value, the wider the spread of values about the MYE and the greater the uncertainty. When the RMSE is calculated as a percentage of the mean simulated composite measures, this becomes the relative root mean squared error (RRMSE). Uncertainty in the MYEs is typically highest in areas with high levels of population turnover, either as a result of internal or international migration. Uncertainty also increases as we move further away from the census, as a result of uncertainty accumulating around the migration components of population change.

In addition to the uncertainty measures, we also show in the Measures of uncertainty and proportional contributions spreadsheet the proportion of the uncertainty that is attributable to each of the 3 components that drive uncertainty: the census, international migration and internal migration.

We have identified 3 methods for deriving 95% confidence intervals for the published MYEs. Our preferred method is the bias-adjusted confidence intervals, but we supply all 3 in the confidence intervals spreadsheet to support your understanding of our methodological approach and of the options available:

1. **Empirical confidence intervals**

   Empirical confidence intervals for each local authority are created by ranking the 1,000 simulated values of the population estimate and taking the 26th and 975th values as the lower and upper bounds respectively. As the observed MYE generally differs from the median of the simulations, this confidence interval is not centred about the MYE and in some extreme cases the MYE is outside the bounds of the empirical 95% confidence interval.

2. **Centred empirical confidence intervals**

   Centred empirical confidence intervals are created by moving the empirical 95% confidence intervals so that they are centred about the observed MYEs. The difference between the median of the simulated values and the observed MYE is subtracted from each of the lower and upper bounds. While the width of the confidence interval remains the same it does not account for the bias component due to the difference between the MYE and the median of the simulation.

3. **Bias-adjusted confidence intervals**

   Bias-adjusted confidence intervals are calculated as the mid-year estimate plus or minus 1.96 multiplied by RMSE. The RMSE is the variability of the simulated values around the MYE. This confidence interval will be symmetric about the MYE and will include a measure of uncertainty due to bias between the MYE and the simulations.

We favour the bias-adjusted confidence interval because it is wider, reflecting the difference between the published MYE and the mean of the simulated composite measures. The discrepancy between the published MYE and the mean of the simulated composite measures may arise for a number of reasons, including:

- the uncertainty methodology only accounts for uncertainty in the census, international and internal migration components
- the assumptions underlying the uncertainty or mid-year estimates methodology
- the uncertainty figures are estimates and so are subject to some uncertainty themselves

We have also specifically identified the undercounting of young males in the internal migration component of the MYEs, to help explain this discrepancy.

We interpret the bias-adjusted confidence intervals in the following way. If the assumptions we have made in estimating uncertainty are correct, we would expect these intervals on average to capture the mid-year population 95% of the time. However, if the bias is relatively large then these confidence intervals will be conservative, i.e., have coverage greater than 95%. Use and interpretation of the confidence intervals will be reviewed as we approach the 2021 census when uncertainty around the MYEs is at its highest level.

2 . Interactive maps

Uncertainty for local authority mid-year population estimates

348 local authorities in England and Wales, 2012 to 2016