

Symbols in tables' definitions and help

Introduction

This documentation provides definitions for harmonised symbols to be used in tables, guidance for use, and additional useful information.

It is important to use harmonised symbols as they provide a number of benefits to both users of statistics and statistics providers.

The benefits to data users include:

- Increased understanding – if harmonised symbols are used then no matter what data users' access, the symbols and their meanings will be consistent and familiar.
- Ease of use and comparability – For data users who require multiple datasets, having multiple symbols with multiple meanings could not only be arduous and time consuming, but increase the possibility of mistakes. Harmonised data symbols will allow comparability between multiple datasets.

The benefits to data producers include:

- Helping towards complying with the requirements of the Code of Practice for Official Statistics – Principle 4, practice 6 requires producers of official statistics to 'Promote comparability within the UK and internationally.
- Helping towards complying with the Statistics and Registration Service Act 2007 – point 9 - Definitions etc for official statistics, which is about developing and maintaining definitions, methodologies, classifications and standards for official statistics and promoting their use in relation to official statistics.
- Knowledge management – Use of harmonised symbols reduces vulnerability to staff turnover by sharing knowledge within the organisation, encouraging dissemination of good practice and improving communication with internal and external users.
- Efficient dissemination – Harmonised symbols will facilitate self-help for users of data on the web, thereby increasing efficiency and reducing burden on staff responsible for answering questions.
- Comparability for cross survey analysis and integration – Use of harmonised symbols will increase access to data from different sources and make them comparable across time and source. Provision of consistent data markers will enable the comparison of statistical products and create an environment in which multiple data sources can be integrated.
- Improving the quality of data – Harmonised symbols can be used to reduce avoidable differences in understanding different outputs which can be caused by use of different symbols
- Benefits IM systems –Larger datasets will be accessible in more and more technical ways, through Application Programming Interfaces (APIs) for example, which will enable the more creative and influential users to join up more and more data from different sources. As the push for open data increases and more departments service this demand, it will become

increasingly important that statistical content provided is more joined up and the bringing together of a single set of data markings is just a very small step to helping.

The symbols and definitions

b = break in time series

Break in a series of data occurring when there is a significant change in the definitions or conditions of data collection, meaning that data before the break cannot be directly compared with data after the break. There are many reasons this can occur, including changes in survey scope, changes in sampling methods, changes to questions, changes in mode of data collection, and changes in weighting.

c = confidential

Where presentation of data would disclose confidential information, for example being able to identify details about a single respondent, this data must be suppressed and given this marker to maintain confidentiality clauses. In the previous standard there was a separate category for suppression. Discussion with ONS Statistical Disclosure Control Unit confirmed that reasons for suppression would either be for confidentiality purposes or low reliability, and these are now separately identified in this list of symbols.

† = earliest revision

A period for which the earliest revision was made in a particular time series. No other periods are then marked with a revision. No revisions have been made to earlier periods. An example from Public Sector Employment is the reclassification of public bodies. When colleges were reclassified from the private to the public sector, this increased the series by about 230k from 1993 onwards. A marker was placed next to the 1993 estimate to indicate that this was the point where revisions began. This symbol is only to be used if there is an accompanying Revision Table for users' reference so revisions over the period can be identified. Not to be confused with 'revision' below.

e = estimated

The particular value yielded by an estimator in a given set of circumstances. Tables which are entirely estimates do not require a symbol for each entry provided titles make it clear to users that all figures are estimates.

f = forecast

The statistical value is forecasted. It is not an observed value; it is a calculated future value.

~ = less than half the final digit shown and different from a real zero

For figures that round to zero, so appear as zero in a rounded table, but are not actually zero. For example, international migration statistics published by ONS report to the nearest thousand, so 499 or less would be displayed with the figure '0' and the symbol '~'. Note that "0" should only be used without a symbol when a figure is a true zero.

u = low reliability

This indicates observations/values for which the user should be aware of the low quality, for example where values of statistical significance have been calculated. The use of the 'u' depends on the context, therefore additional information, e.g. explanations about the reliability limits applicable, should be provided in the notes.

z = not applicable

The 'z' symbol is only to be used when an observation is not applicable (e.g. in tables of employment where people under 16 cannot legally be employed).

: = not available

For use when data is unavailable for reasons other than those described in this list, for instance, data not collected in a region.

n = not significant

The figure is not statistically significant. This category is an example of where the Eurostat standard differs with common practices used for UK statistics. For the majority of UK statistics, certainly within ONS, it's more useful to know the level of significance, which is why the '*' are included (see below). To provide clarity for users, and for ease of use for providers, supporting notes can make it clear that if '*' is used then anything without these markings is not significant. Data providers are still able to use 'n' should they wish to highlight statistics that are not significant.

p = provisional

For use when the data are yet to be finalised, or are expected (or almost certainly expected) to be revised. For example, the Annual Survey of Hours and Earnings always publish provisional figures followed by a revised final set of figures, so a time series would include the latest provisional figures alongside former final figures.

r = revised

The figure stated has been revised since it was first published. Not to be confused with 'earliest revision' above.

Statistically Significant

A statistically significant result is one that we are confident is not simply the result of random chance. The fact that a result is statistically significant does not mean that it is necessarily of practical importance. For example, if sample sizes are large, small differences can be statistically significant. Only by considering context can we determine whether a difference is important enough to require action.

*** = significant at 0.05 level**

When we report on statistical significance, we provide an assessment of how likely it is that we would see results as unusual as these if chance alone was operating. The phrase "statistically significant at the 0.05 (or 5%) level" indicates that, if chance alone was operating, a result like this would occur less than 5 times in 100, or less than 5% of the time.

**** = significant at 0.01 level**

When we report on statistical significance, we provide an assessment of how likely it is that we would see results as unusual as these if chance alone was operating. The phrase "statistically significant at the 0.01 (or 1%) level" indicates that, if chance alone was operating, a result like this would occur less than 1 times in 100, or less than 1% of the time.

***** = significant at 0.001 level**

When we report on statistical significance, we provide an assessment of how likely it is that we would see results as unusual as these if chance alone was operating. The phrase "statistically significant at the 0.001 (or 0.1%) level" indicates that, if chance alone was operating, a result like this would occur less than 0.1 times in 100, or less than 0.1% of the time.

In Practice:

In terms of layout, it is recommended that symbols have their own column since this enables them to be easy to read and legible to users:

Figure	Symbol
0	~
4	r
6	p
0	

Below are practical examples.

ONS Migration Statistics published in February 2014.

Main reason for migration	European Union ¹							
	European Union ¹		European Union EU15		European Union EU8		European Union EU2	
	Estimate	+/-CI	Estimate	+/-CI	Estimate	+/-CI	Estimate	+/-CI
YE Dec 10	7	4	3	2	4	3	0	0
YE Mar 11	7	4	2	2	4	3	0	0
YE Jun 11	9	5	3	2	6	4	0	0
YE Sep 11	11	5	5	3	6	4	0	0
YE Dec 11	10	4	5	3	5	3	0	0
YE Mar 12	9	4	5	3	4	2	0	0
YE Jun 12	9	3	5	3	3	2	0	0

ONS Vital Statistics Outputs Branch

ICD-10 code	Underlying cause ¹	Death rates per million population								
			All ages	Under 1	1-4	5-14	15-24	25-34	35-44	45-54
			Europe ²							
A00-R99, U000Y89	All causes, all ages	M	6,191	4,621	188	97	401	663	1,402	2,991
		F	4,510	3,687	146	82	178	340	830	2,030
A00-B99	I Certain infectious and parasitic diseases	M	61	75	14	4 u	3 u	8	22	50
		F	51	68	10 u	4 u	4 u	7	16	23
A00-A09	Intestinal infectious diseases	M	14	c	c	c	c	c	1 u	1 u
		F	16	c	c	c	c	0	2 u	2 u
A15-A19	Tuberculosis	M	5	0	0	0	1 u	c	1 u	6
		F	2	0	c	0	c	2 u	1 u	2 u
A15-A16	Respiratory tuberculosis	M	3	0	0	0	c	c	1 u	4 u
		F	1	0	0	0	0	1 u	1 u	0

Age of mother at birth	Maternities	Births					
		Live			Still		
		Total	Male	Female	Total	Male	Female
All maternities with multiple births							
All ages	11,441	22,845	11,417	11,428	254	138	116
Under 20	226	443	211	232	10	7	3
20-24	1,239	2,479	1,196	1,283	c	c	c
25-29	2,632	5,239	2,611	2,628	c	c	c
30-34	3,760	7,504	3,720	3,784	c	c	c
35-39	2,625	5,250	2,688	2,562	c	c	c
40-44	753	1,513	790	723	c	c	c
45 and over	206	417	201	216	c	c	c