United Nations ECE/CES/GE.41/2014/26

Economic and Social Council

Distr.: General
16 September 2014

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses organized jointly with Eurostat

Sixteenth Meeting
Geneva, 23-26 September 2014
Item 6 of the provisional agenda
Quality management

Internet versus paper mode effects in the 2011 Census of England and Wales: analysis of Census Quality Survey agreement rates

Note by the Office for National Statistics, United Kingdom

Abstract

A sample of households were asked the 2011 census questions again in a Census Quality Survey (CQS). These were compared with their original responses to calculate agreement rates which provide an indication of respondent error in the 2011 Census for England and Wales. One of the aspects that could be tested in this analysis was the agreement rates between census returns completed on the internet versus paper returns, and whether this gave any indication of effects between the two modes. This preliminary analysis shows that most of the differences between the agreement rates were due to the different characteristics of the people who responded by different mode.

I Summary

- Census values (after edit and imputation) were compared to answers independently gathered through CAPI (Computer Assisted Personal Interviewing) process in the CQS
- Questions where the agreement rate between the CQS and census responses provided by internet were significantly better than paper (at the 5 per cent significance level) were:
    - Type of accommodation
    - Tenure
    - Age
    - Marital or civil partnership status
    - Term-time indicator
    - Unpaid care
    - Disability/long-term illness
    - Supervisor
- Questions where the agreement rate between the CQS and census responses provided by paper were significantly better than internet (5 per cent significance level) were:
    - Student in full-time education
    - Ethnicity
    - Religion
    - Address a year ago
    - Employment last week
    - Waiting to start work
- Analysis of these apparently significant differences shows that the differences mostly occurred not because of the design of the questions, but because of the different types of people who answered on paper compared to internet.

II Introduction and background

1. The 2011 Census was the first census in England and Wales to offer respondents the opportunity to complete their return via the internet. Paper forms were made available to all, and each included a unique code to enable the respondents to access their internet form. The percentage of individuals who responded using the internet was 18.8 per cent.

2. A previous UNECE paper (ONS (2012a)) set out the development of the internet data capture system, and introduced expected quality improvements and potential issues due to carrying out a mixed-mode census. A concern in the use of different modes to collect the same data from different members of the population is whether the questions are answered in the same way for each mode. This paper describes some analysis carried out to investigate whether these issues affected the final census data.

3. The Census Quality Survey (ONS (2014)) aimed to estimate response error in the 2011 Census for England and Wales. A sample of households were interviewed and asked the census questions again in a face-to-face interview. Their interview responses were then compared with their original census responses and agreement rates calculated to provide an indication of how accurately the 2011 Census had been completed. The census questions varied widely in their agreement rates and reasons for disagreement. Rates ranged from 55 per cent (year last worked) through to 99.7 per cent (sex). Reasons for the differences include reporting by proxy (where a household member may not have known the exact response for an individual); ambiguity (for instance confusion about which rooms to count);

capture errors (for instance due to scanning handwritten responses). A summary of the agreement rates for all the questions is provided in Annex A. A description of how the agreement rates and confidence intervals were calculated is at Annex B. This analysis concentrates on examining the difference in agreement rates between those who answered by paper and those via the internet.

4. The CQS was an interviewer-based survey, introducing an additional mode. All census questions were asked the same whether or not the respondent used paper or internet census return route, so this analysis effectively ignores any CQS mode effect in order to compare the census modes. However, questions at risk of having an interviewer effect through social desirability bias are highlighted below, where there are different patterns of answers on paper and internet. These were primarily the questions on disability/long-term health problems and marital or civil partnership status questions.

5. The following section discusses possible causes of mode effects in the 2011 Census. Section 4 then shows the results of the analysis, and Section 5 summarises the conclusions.

III Possible causes of mode effects

6. Studies of survey mode effects identify key causes (see ONS (2010)). Table 1 identifies the modal effects that are more relevant for the 2011 Census.

7. The internet questionnaire was not developed at the same time as the paper equivalent, but later. The paper questionnaire format was developed through an extensive programme of consultation and testing. Concerns about modal bias that could be introduced if the internet format was substantially different meant that the two modes were made as similar as possible from a respondent's perspective. This means that some of the advantages that an internet format could have had over paper were deliberately not maximised in the 2011 Census.

8. Aspects that did not affect the respondent's perspective (such as auto-filling of names, soft reminders, radio buttons to prevent multiple ticks in single-response questions, question routeing, clarification and instructions on the screen, no need for scanning and not being limited to six household members) were likely to improve the quality of internet responses.

9. On the other hand, some of the longer questions were not always viewable on the internet form without scrolling down. This aspect was tested, and as the 'Next' button was on the bottom of the screen respondents would see all options. Also, respondents could more easily look ahead on the paper form to think about which questions were relevant to them. Internet respondents could navigate through the questionnaire before filling it in, but as this would bring up soft reminders of unanswered questions it was not as straightforward as looking ahead on the paper form, and is was perhaps not clear that they could do this and go back to unanswered questions.

10. Response by proxy was one of the factors behind inconsistent answers between the census and CQS. Responses via the internet were slightly more likely to be answered by proxy than paper responses, so this may be a contributing factor to the difference in agreement rates.

11. Completion of the form via the internet was a self-selected choice in 2011, so the characteristics of the people who completed by internet or on paper were different. As found in other studies of internet collection (see ONS (2010)), the internet options was more likely to be chosen by people with certain characteristics (see Table 1).

Table 1: Causes and effect within different modes

Cause	Details, possible effect
Question routing done automatically on the internet	Where this was not present on paper forms, respondents may have answered questions unnecessarily. However, census processes will have removed unnecessary answers from the final data, so this should not affect final quality.
Soft reminders on the internet	If a question was not completed, a reminder box popped up. The respondent could press 'next' again and ignore it, but it led to very low levels of item non-response for internet completions compared with paper.
Display on page versus on screen	Would a respondent be more likely to select an answer on the top of the screen, or perhaps nearest the 'next' button?
Clarifications/instructions/translations more readily available on the internet	For instance, a respondent could more readily check their work postcode, or translate a phrase into their first language.
Scanning errors from paper questionnaires	This was more prevalent for write-in answers than for tick-boxes.
Characteristics differ for people more likely to answer using the internet compared to paper. (ONS (2012a)).	Significant effects found in the 2011 responses: • young adults • males • English not main language, non-UK born, certain ethnic groups and religions • not disabled (under age 75) • in full-time education • married or in civil partnerships • higher levels of qualifications • people in employment, those working longer hours • larger households (linked to ease of responding for all members on the internet form) • characteristics interacting with household size and also reflecting economic status (household ownership, large number of rooms, number of cars)
The internet questionnaire was not limited to a fixed number of people	This is not analysed here, because the CQS analysis is limited to matched individuals. Separate analysis concluded that the internet completion provided a more uniform coverage pattern, especially of children. See ONS (2012b)
The relationship matrix had a different layout on the internet form	Respondents may have found one layout more intuitive than the other
Some fields on the internet form were pre-filled based on the response to an earlier question	For example, person 1's name would have been added on the address question so person 2 would have said their address a year ago was the same as person 1's, but they did not see that address.
Use of radio buttons on internet	Prevented multiple tick responses.

12. The CQS sample was designed to be representative of the respondent population according to region, hard-to-count, speed of return ('early', before 1 April 2011, or 'late', after 1 April 2011), and also by mode of return (by paper or via the internet). The responses for households were considered to be representative of all respondent households. The sample size was selected to give the agreement rates a margin of error of plus or minus 2 percentage points at the 95 per cent level of confidence for England and Wales. The sample size was not large enough to enable robust comparisons between categories within questions, accounting for the different characteristics of the people who responded by the different modes. However, it was possible to further look at the results within categories to draw out possible reasons for apparently significant results. These are described in the following section.

IV Analysis

13. Table 2 shows a summary of which questions had significantly (at the 5 per cent level) better agreement for one mode or the other, for questions where the sample size was sufficient. Possible reasons for the differences are also suggested although these are suppositions based on knowledge of processing, conversations with those who worked closely with the on-line system and common sense.

Table 2: Summary of which questions had significantly better agreement for one mode or the other

Internet significantly higher agreement rate

Question no.	Question	Possible reasons
H7	Type of accommodation	Use of radio button preventing multiple ticks
H12	Tenure	Use of radio button
3	Age	Incorrect capture of handwritten date of birth following scanning
4	Marital/civil partnership status	Group most likely to change answer[1], radio button
8	Term-time indicator	Possibly help information
14	Unpaid care	Group most likely to change answer[1], radio button
23	Disability/long-term illness	Group most likely to change answer[1], radio button
36	Supervisor	Group most likely to change answer[1]

Paper significantly higher agreement rate

Question no.	Question	Possible reasons
7	Student	Group most likely to change answer[1]
16	Ethnicity	Group most likely to change answer[1] Also long question with lots of tick boxes – internet respondent might not have scrolled down far enough
20	Religion	Group most likely to change answer[1]
21	address a year ago indicator	Possibly that Person 2+ respondents couldn't see what Person 1 put as their address a year ago
26	employment last week	Paper format easier to look ahead to decide whether question relevant
29	waiting to start work	Paper format easier to look ahead to decide whether question relevant

[1] Group most likely to change answer: one group most likely to give inconsistent answers in successive surveys was most likely to respond by one mode rather than the other. See following section.

14. Table 3 lists those questions where there was adequate sample size, but showed no significant difference between the agreement rates for the two modes. The remaining census questions did not have sufficient sample size in the CQS sample: often because the question related to a sub-sample of the population, but also sometimes because the level of disagreement warranted a higher sample size to draw robust conclusions.

Table 3: Questions where there was no significant difference between agreement rates by mode, (where sample sizes were large enough)

Question no.	Question
H2	Number of usual residents
H6	Relationship matrix
H8	Self-contained
H10	Number of bedrooms
H11	Central heating
2	Sex
5a	Second address
9	Country of birth
18	Main language
22	Passport
33	Employment status

15. Figures 1 and 2 show the percentage agreement rates and confidence intervals for all household and individual questions. Very wide intervals are indicative of smaller sample sizes.

Figure 1: Agreement rates and confidence intervals between census and CQS, paper v internet: Household questions with sufficent CQS sample

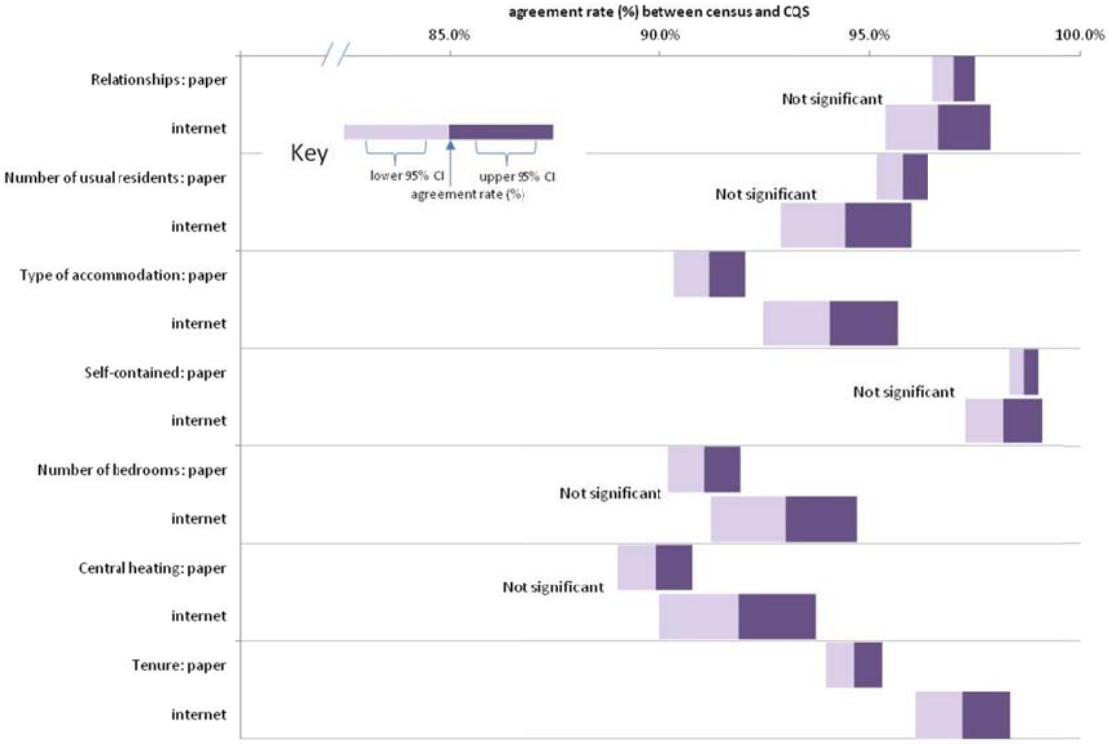
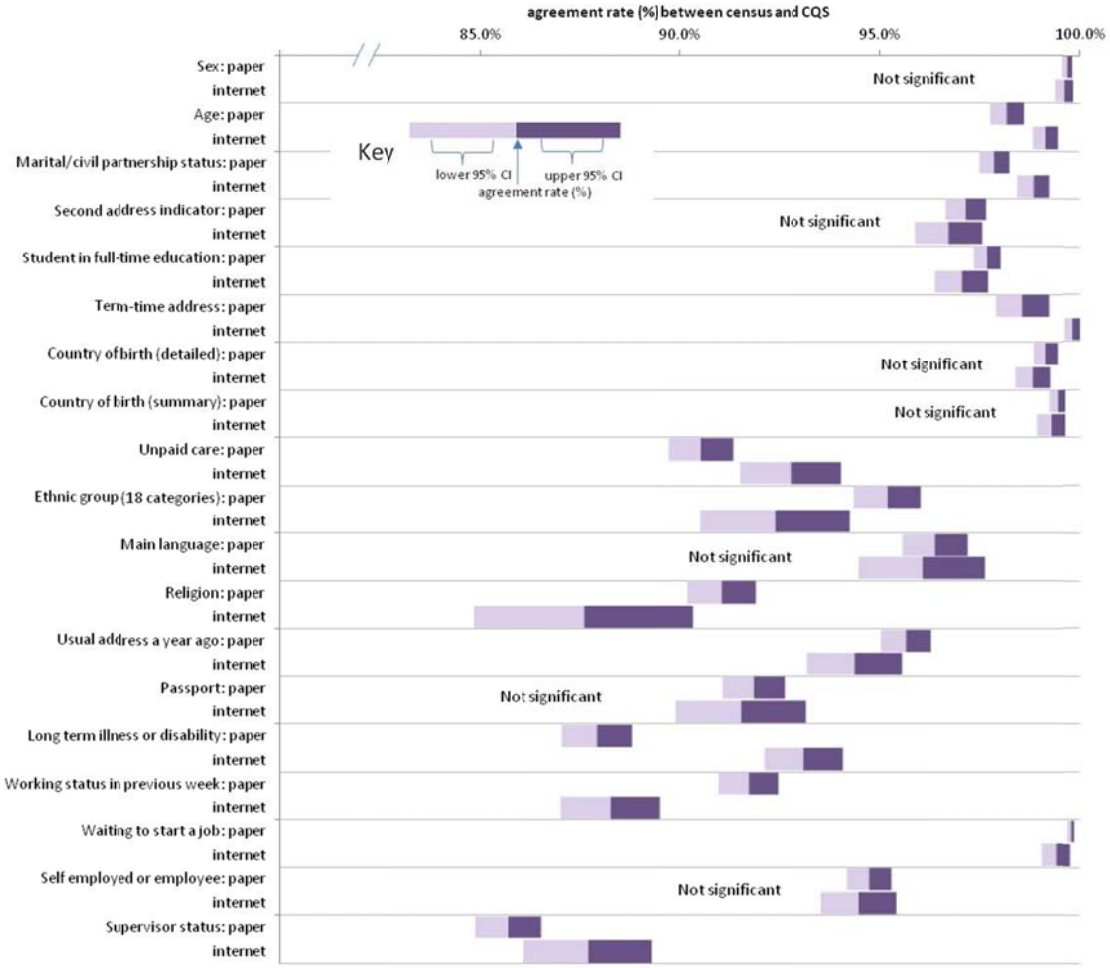


Figure 2: Agreement rates and confidence intervals between census and CQS, paper v internet: Individual questions with sufficient CQS sample



A Groups most likely to change answer focused in one or other mode

16. Marital or civil partnership status: there was higher than average response on paper for people who were separated or divorced. A mode effect may be caused by the CQS interview due to social desirability bias on behalf of the respondent, perhaps not wanting to mention a failed relationship. There is also the possibility that a separated couple may have reunited in the intervening time between the census and CQS interview.

17. Unpaid care: there was higher than average response on paper for people with caring responsibilities, and most confusion here between the number of hours of care provided rather than whether did or did not look after people. The higher agreement rate by internet is therefore not an effect of the mode, but simply a consequence of the respondents' characteristics.

18. Disability/long-term health problems: respondents who answered 'yes' ('Yes, limited and little' or 'Yes, limited a lot') were more likely to answer by paper, and due to recall error are more likely to change between the two 'yes' answers. So once again this shows the effect of the respondents' characteristics rather than the mode used. However, a

mode effect may be caused by the CQS interview, as those who answered 'yes' can also have higher risk of social desirability bias, in that they may 'play down' their health problems in a response to an interviewer (ONS(2010)).

19. Supervisor: respondents who were no longer working were more likely to respond on paper, and also were possibly more likely to have recall error about whether they used to supervise employees.

20. Student or schoolchild in full-time education: students were more likely to have responded by the internet option. The 'student' question was answered more consistently on the internet by people who answered 'yes' on the census form and was answered more consistently on paper by people who answered 'no'. The largest difference in responses was where the census indicated that the person was a student or schoolchild, but the respondent in the CQS said they were not. Further investigation identified that around half of these were children aged under four. It is possible that this difference is caused by the CQS interviewer being able to clarify that pre-school or nursery were not full-time education. The mode of response for these pre-school children was more likely to be the internet, so accounts for a large part of the lower agreement rate. As above, this is therefore not an effect of the mode, but a consequence of the type of person who chose to respond by the internet.

21. Ethnicity: The group most at risk of changing their ethnicity response in successive surveys (i.e. 'Mixed/Other' ethnicity) had higher than average internet response, with the consequence that their characteristics have led to the internet agreement rate being lower than the paper agreement.

22. Religion: The group most at risk of changing their religion response in successive surveys (i.e. 'No religion'/'Other' categories of religion) had higher than average internet response, leading to lower agreement for internet responses.

B Use of radio buttons

23. Marital or civil partnership status: there were very few cases of internet completions needing editing or imputing for this question, compared to for paper completions (0.1 per cent and 4.3 per cent respectively, in the CQS sample). This is likely to be attributable to the use of radio buttons on the internet form preventing multiple-tick responses. Paper forms with multiple ticks will have been amended through the edit process and the 'true' answer may not have been selected. Removing the records affected by edit and imputation brings the paper agreement rate much closer to the internet agreement.

V Conclusions

24. This preliminary analysis suggests the minimisation of mode effects in the 2011 Census was partially achieved, with significant differences mostly occurring not because of the design, but because of the different types of people who answered on paper compared to internet. There were also possible design issues around viewing Person 1 address and being able to view all questions to see which were relevant. Designing for the internet and promoting internet response as the default option could minimise these issues. Using the CQS data set, ONS are considering undertaking further statistical analysis into the differences between internet versus paper responses by removing the effect of the different population types who responded by each mode. This will shed further light on the impact of mode on the accuracy of the 2011 Census.

References

ONS (2010) The application of alternative modes of data collection on UK Government social surveys (Available from: http://www.ons.gov.uk/ons/guide-method/method-quality/general-methodology/data-collection-methodology/reports-and-publications/alternative-modes-of-data-collection/index.html)

ONS (2012a), 2011 Census for England and Wales: Internet Data Capture, Fourteenth meeting of Group of Experts on Population and Housing Censuses, 24-25 May 2012. UNECE. (Available from: http://www.unece.org/stats/documents/2012.05.census2.html)

ONS (2012b), Mode effects in the 2011 Census within matched households. Unpublished paper. Available on request.

ONS (2014), 2011 Census Quality Survey. (Available from: http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-user-guide/quality-and-methods/quality/quality-measures/assessing-accuracy-of-answers/index.html)

Annex A: Agreement rates between 2011 Census and Census Quality Survey

England and Wales

Question Number	Census Question	Overall Agreement Rates (%)	Paper Agreement Rates (%)	Internet Agreement Rates (%)
H2	Number of usual residents	95.5	95.8	94.4
H6	Relationships	96.9	97.0	96.6
H7	Type of accommodation	91.6	91.2	94.1
H8	Self-contained	98.6	98.7	98.2
H9	Number of rooms	66.5	-	-
H10	Number of bedrooms	91.4	91.1	93.0
H11	Central heating	90.2	89.9	91.9
H12	Tenure	95.0	94.6	97.2
H13	Landlord	87.6	-	-
2	Sex	99.7	99.7	99.6
3	Date of birth	98.4	98.2	99.2
4	Marital and civil partnerships	98.1	97.9	98.9
5	Second address	97.1	97.2	96.7
7	School children/ students	97.6	97.7	97.1
8	Term-time address	98.9	98.6	99.8
9	Country of birth	99.1	99.2	98.9
13	General Health	68.2	-	-
14	Unpaid care	90.9	90.6	92.8
15	National identity	60.4	-	-
16	Ethnic group (18 tick boxes)	94.7	95.2	92.4
18	Main language	96.3	96.4	96.1
20	Religion	90.4	91.1	87.6
21	Usual address one year ago	95.5	95.7	94.4
22	Passports	91.8	91.9	91.6
23	Limiting long term illness	88.9	87.9	93.1
25	Highest qualification	67.6	-	-
26	Working status in previous week	91.2	91.7	88.3
27	Looking for work	96.2	-	-
28	Available for work	86.2	-	-
29	Waiting to start work	99.8	99.8	99.4
30	Reasons for not working	86.4	-	-
31	Ever worked	94.4	-	-
31	Year last worked	55.0	-	-
33	Self employed or employee	94.7	94.8	94.5
34,35	Occupation Code (Major group)	67.5	-	-
36	Supervisor	86.2	86.0	87.8
37,38	Industry Code (Section)	74.2	-	-
40	Workplace address(Postcode Sector)	82.2	-	-
41	Travel to work	85.5	-	-
42	Hours worked	83.9	-	-

- Indicates insufficient sample size in the CQS

Annex B: Calculation of agreement rates, confidence intervals and significance tests

Calculating agreement rates

Agreement rates are the number of agreements between valid census and CQS responses relative to the total number of matched households or individuals. For individuals, this excludes census proxy responses for respondents aged 16 and over.

Calculating confidence intervals

As with any sample, different people would be selected if the sample was randomly drawn again and slightly different estimates would be produced based on this different sample. The spread of these estimates is known as the sampling variability. Confidence intervals are used to indicate the sampling variability.

A 95 per cent confidence interval is a range within which the true population parameter would fall for 95 per cent of all possible samples that could have been selected. It is a standard way of expressing the statistical accuracy of a survey based estimate. If an estimate has a large error level, the corresponding confidence interval will be very wide.

Standard Error of the agreement rate

Households

$$SE = \sqrt{\frac{p(1-p)}{N}}$$

Where:
p = agreement rate = number of responses which agree / total number of valid responses
N = number of valid responses

Individuals
The sample of individuals was clustered within households. This creates a design effect that reduces the effective sample size and increases the variability of the estimates. This design effect was allowed for in the standard errors, which were calculated using SAS proc survey means see:
http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#surveymeans_toc.htm

Confidence Intervals around the agreement rates

$$95\%\ CI = p \pm 1.96 * SE$$

That is the 95 percent confidence interval for the rate is +/- 1.96 standard errors around the estimate calculated from the sample.

Pooled standard error

In order to compare the agreement rates between England and Wales, and Internet and paper responses, a pooled proportion and pooled standard error are calculated:

$$p_p = \frac{(a_1 + a_2)}{(N_1 + N_2)}$$

Where:
a_1, a_2 = weighted number of responses which agree, in dataset 1 (for example Internet) and 2 (for example paper) respectively.

N_1, N_2 = number of valid responses in datasets 1 and 2.

Households pooled standard deviation

$$SE_p = \sqrt{p_p(1 - p_p) * (\frac{1}{N_1} + \frac{1}{N_2})}$$

Individuals pooled standard deviation

$$SE_p = \sqrt{\frac{(N_1 - 1) * {SE_1}^2 + (N_2 - 1) * {SE_2}^2}{(N_1 + N_2 - 2)}}$$

Significance test for unpaired proportions (Z-test)

Two agreement rates (p_1, p_2) are significantly different if:

$$|p_1 - p_2| > 1.96 * SE_p$$

Where:

1.96 is the Z-statistic at the 5 per cent significance level (for a two-sided test).