

'Optimizing the use of Existing Evidence’

Using statistical significance to interpret National Student Survey results

### Dr Alex Buckley (University of Strathclyde)

# Why are confidence intervals useful?

Confidence intervals represent the range around an actual NSS score in which we can be confident that the ‘true’ value falls (i.e. the margin of error). They are useful because they provide a way of taking into account the inherent uncertainty in survey results. They are particularly valuable because they allow us to evaluate the statistical significance of comparisons that we make using NSS scores. A difference is statistically significant if we can be confident that it represents a genuine difference in students’ experiences rather than random variation in the data.

# What confidence interval data is available for the NSS?

On behalf of all four UK nations, the Office for Students makes publicly available on its website 95% confidence intervals for ‘percentage agree’ scores. Confidence intervals are available for all institutions at subject level, using both JACS and CAH subject groupings.

The data can be found at: <https://www.officeforstudents.org.uk/advice-and-guidance/student-information-and-data/national-student-survey-nss/get-the-nss-data/>

# How can confidence intervals be used to assess statistical significance?

When comparing two NSS scores, the statistical significance of the difference between them can be evaluated using the level of overlap in their confidence intervals.

* If the 95% confidence intervals don’t overlap at all, then we can be 99% confident that the difference is statistically significant. 99% is a relatively strict level of confidence
* It can also be appropriate to use a more lenient 95% level of confidence. We can be 95% confident that the difference is statistically significant if the overlap in the 95% confidence intervals is less than half of the average of the confidence intervals.[[1]](#footnote-1)

These examples show the different interpretations of error bar overlap.

*Example 1* illustrates a difference that is significant at the 99% confidence level, as the error bars don’t overlap. *Example 2* illustrates a difference that is significant at the 95% confidence level, as there is only a small amount of overlap. *Example 3* illustrates a difference that is not significant even at the 95% level.

Microsoft Excel allows the addition of error bars to bar charts. Using the ‘custom’ error bar option, the confidence intervals provided by the OfS can be used as error bars.

It is important to bear in mind that the confidence intervals are only one way of representing the uncertainty around the NSS scores. Caution should also be exercised in the use of the binary distinction between ‘significant’ and ‘not significant’: significance testing is just one way of exploring the data and doesn’t replace the need to manually interpret the data, for instance by visually inspecting the confidence intervals.

**Three ways of using the confidence intervals to evaluate the statistical significance of NSS comparisons**

## Evaluating year-on-year changes

Confidence intervals can be used to assess the statistical significance of year-on-year changes in NSS results. The example below shows the year-on-year changes for one (fine-grained) subject at one institution. The size of the confidence intervals means that even quite large changes in raw score can fail to be statistically significant.

Due to the change in the NSS questionnaire in 2017, comparisons can only be made between scores pre- and post-2017 with caution.

## Evaluating comparisons between institutions

Confidence intervals can be used to assess the statistical significance of comparisons between different institutions. The proportion of statistically significant differences can differ markedly by subject and by question. The chart below illustrates the institutional scores, with confidence intervals, for a particular (fine-grained) subject and question, and is provided for illustration.

## Evaluating comparisons with the sector average

As comparisons between individual institutions tend to yield few statistically significant differences, it can be more helpful to compare scores against the subject-level sector average. As sector averages are produced by thousands of student responses, the confidence intervals can be treated as zero. Therefore, assessment of statistical significance does not involve evaluation of the overlap of confidence intervals, but whether the sector average lies above, below or within the confidence interval for the institutional score. Using this method for all of an institutions’ scores yields a ‘heatmap’ which can illustrate in an easily accessible format any particular subjects that might reward further study, and any questions where the institution performs well or badly.

The example below shows an institution with a representative proportion of statistically significant differences from the sector average. Green cells indicate where scores are statistically significantly higher than the average (for that subject and question). Red cells indicate where scores are statistically significantly lower than the sector average (for that subject and question).

Sector averages (at subject level) are not made publicly available, they are currently only available as part of the data that institutions receive directly. Contact your institutional survey team for access to the sector averages.



This document has been produced and published by The University of Strathclyde based on a template provided by the Quality Assurance Agency for Higher Education (QAA). As such, this document may contain content that is not wholly endorsed by QAA.



1. For more information about this pragmatic ‘rule of eye’ see Cumming, G. (2009) ‘Inference by eye: Reading the overlap of independent confidence intervals’, *Statistics in Medicine* 28: 205-220 [↑](#footnote-ref-1)