There are three kinds of lies: lies, damned lies and statistics.

**Direction’s request**

At the University of Teacher Education Vaud, the quality of teaching is assessed by selecting a sample of modules every semester and inviting all enrolled students to evaluate them by completing an online questionnaire.

Based on these data, the direction asks that an annual result is computed for the whole institution.

What calculation should be done on the data to accurately reflect the voice that students gave through their answers?

**Theory**

In this poster, we take advantage of Cashin’s (1995) distinction between “evaluations”, which refer to the student answers, and “ratings”, which refer to the data to be interpreted.

- In terms of answers, teaching evaluation questionnaires provide information on the quality of the education received by giving the students a voice.
- In terms of data, teaching evaluation questionnaires deliver results that offer valuable guidance for pilots and teachers in education systems.

Viewing student evaluations as data rather than as answers may help to raise the issue of the accuracy of the results on which guidance is based.

**Data set**

- Student evaluations of the modules they attended during the 2016-2017 academic year.
- 64 modules out of 224 were evaluated by 2,108 student ratings, with a response rate of 35%.

**Four calculation methods**

To calculate the requested overall score, we highlight 4 possible methods, each based on a specific modeling approach:

**Calculation 1:** Computing the average of all student ratings
- Level of data analysis: students exclusively
- Correct calculation for data arising from a Simple Random Sampling
- Disregarding the fact that the ratings are nested within the modules, a biased result is obtained

**Calculation 2:** Computing the average of student ratings per module
- Level of data analysis: modules exclusively
- Correct for a One-stage Sampling of modules, which would have been evaluated by all registered students
- Disregarding the fact that variable response rates from one module to another (non-response error) result in unequal sampling fractions, a biased result is obtained

**Calculation 3:** Weighting each module’s score according to the proportion of its respondents
- Level of data analysis: students x modules
- Corresponts to the level of data collection: calculation relies on a Two-stage Modelling that truly reflects the hierarchical structure of the data
- Omitting the fact that both modules and student ratings are samples (sampling error), this result is incomplete

**Calculation 4:** Adding the confidence interval (CI)

- In order for the previous result (sample statistic) to be used as an estimate of the population parameter, its confidence interval must be calculated: CI quantifies the confidence that can be placed in the overall annual result

**Illustration on two questions**

The requested overall annual score is the proportion of positive evaluations.
Each of the 4 methods produces a specific result.

**Proper answer to direction**

Applying the only correct method – Calculation 4 – to respond to the direction’s request, we can communicate the following, taking for example Question 1:

- The result for the 64 surveyed modules (sample result) is 88%
- And we are 95% confident that the proportion of all students that appreciated their teaching is between 81% and 95%

Note that CI per se also act as an indicator of the quality of teaching:

- Small CI indicate a strong consensus, both between student answers and module ratings.
- Large CI denote wide variability, which may be due to either student answers (intra-variability) or differences between modules (inter-variability), or both

**Tips**

- Use the calculations ordered by the sampling plan
- Compensate for the non-response error with the appropriate weighting
- Report confidence intervals

Our warmest thanks to Erika Antal, PhD in statistics and senior researcher at FORS (Swiss Center of Expertise in the Social Sciences, University of Lausanne), for her expert and smart assistance in statistical modeling.