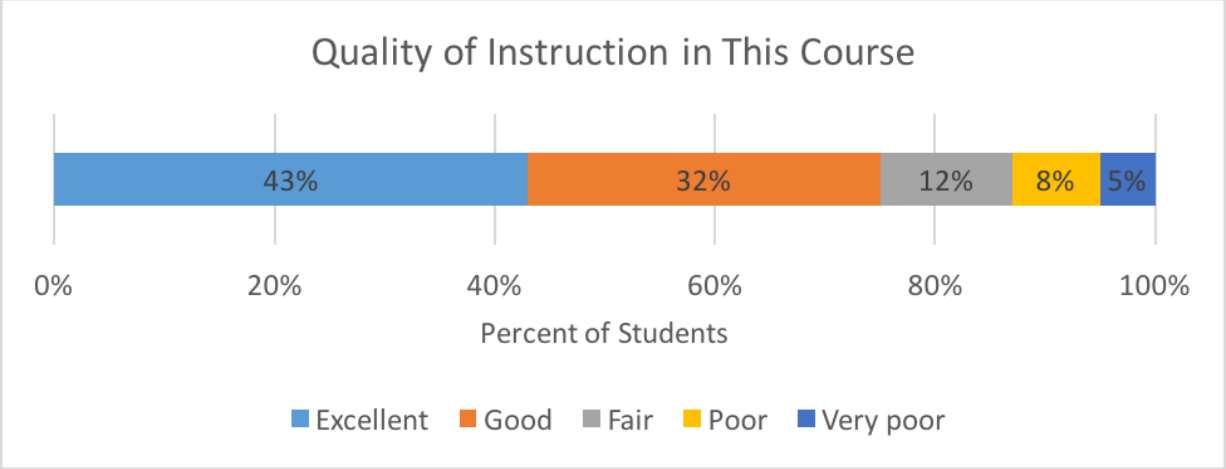
Stanford | Evaluation & Research

Choosing and Interpreting Summary Statistics

Summarizing evaluation data—whether by a mean score or other statistic—can hide important information. Courses with the same mean score can have very different patterns of evaluation responses. Alternatives to the mean score can be considered for summative judgments, while the full distribution of responses is probably the best representation for reflective inquiry.

Let's consider student responses to the question "Overall, how would you describe the quality of the instruction in this course?" Students are asked to select a single rating from a list of five possibilities: Excellent, Good, Fair, Poor, or Very Poor.



From this chart we can derive several summary statistics:

- 43% of students rated the quality of instruction as *Excellent*
- 75% (43% + 32%) rated the quality of instruction *Good* or *Excellent*
- 13% (8% + 5%) rated the quality of instruction *Poor* or *Very Poor*
- If we assign *Excellent* = 5, *Good* = 4, etc., the mean score is 4.0 **Which statistic is best?**

Which of these statistics **best summarizes** the student rating of instruction? That is, if we had to summarize an instructor's performance with a single number, which number would we use? The answer depends on which feature(s) of the ratings we care to focus on. If we care about promoting excellence in teaching, the 43% rating of *Excellent* might be the statistic of choice. If, on the other hand, we want to highlight problematic areas in our courses, the 13% of students rating instruction *Poor* or *Very Poor* might be a statistic of interest.

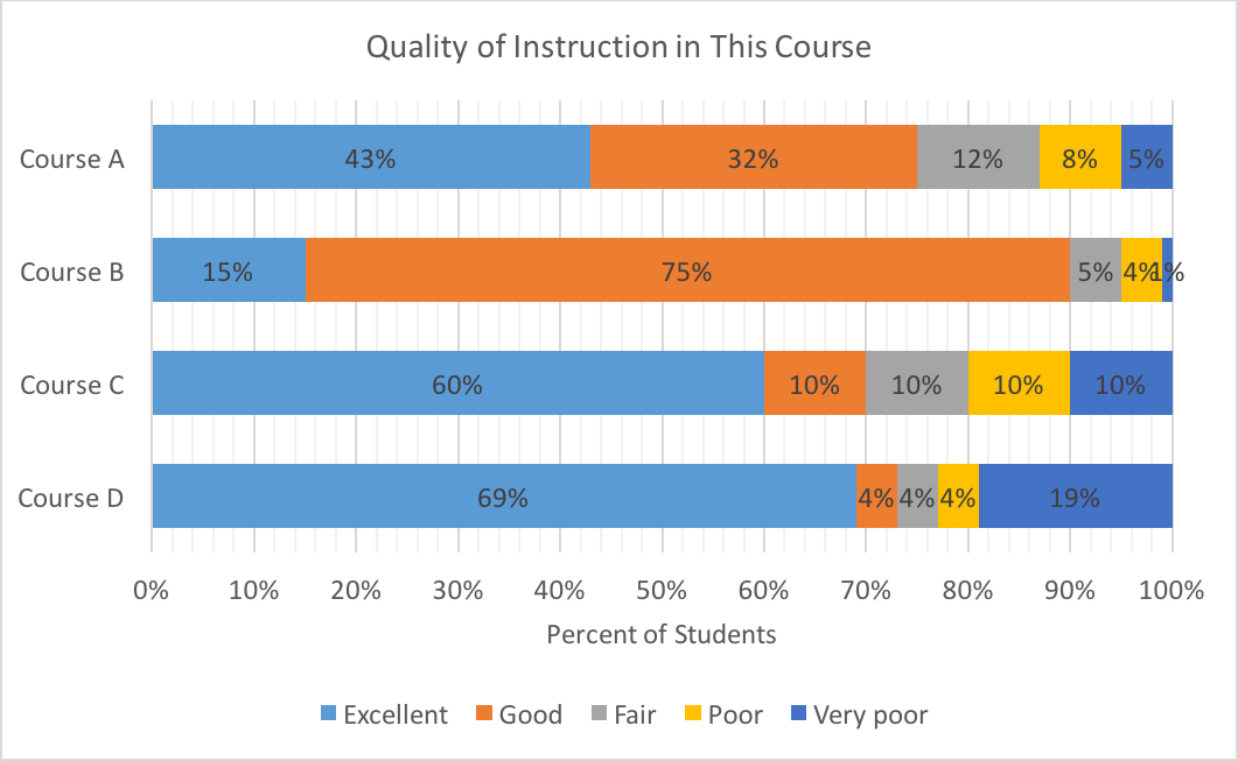
Typically, evaluation reports summarize a course rating using the formula shown in the last bullet point: assign values of 1–5 to the ratings[1], and compute the mean score. This highlights what statisticians call the "central tendency," or what we might more colloquially call the "average rating." There is also some controversy over whether it is appropriate to compute the mean of so-called "ordinal" data in the first place.

Consider the full distribution

Regardless of the statistic chosen, however, we are left with the same problem: no single statistic can capture all of the interesting and meaningful patterns in the data, even in a measure as simple as a 5-point rating scale. Consider the

[1] We note there is no *a priori* reason to assign an equal interval scale such as 1, 2, 3, 4, and 5 to the rating levels. We might decide, for example, that *Excellent* ratings deserve special emphasis, and choose numeric score values of 1, 2, 3, 4, and ***7*** to represent the ratings.

chart below, where each of four courses have been rated for quality of instruction. **As it turns out, all four courses result in a mean score of 4.0.** Course A is the example shown above. In Course B we see that 75% of students—the vast majority—rated the instruction as *Good*, and 90% of students rated the instruction *Good* or *Excellent*. In terms of the percent of students rating *Good* or *Excellent*, Course B's instructional ratings are much higher than Course A's.



Courses C and D show what can happen when there is a split opinion within a class. In course C, a majority (60%) rated instruction as *Excellent*, but 30% of the class rated the instruction *Fair* or lower. Course D exhibits what statisticians call a "bimodal distribution", where we observe two distinct clusters of responses (*Excellent* and *Very Poor*). While Course D results in a mean score of 4.0, there is considerable disagreement among the students regarding the quality of instruction.

We also note that in courses with small enrollments, 10% or more of a score's distribution may reflect a single student's response. Please see the guide *Stability of Course Evaluation Statistics* for discussion of the random variation present in these summary measures.

Recommendation

- Faculty members and other stakeholders should pay attention to the complete distribution of course ratings when making critical decisions.

For more information, please see our website at evals.stanford.edu

Contact Us:
518 Memorial Way, Stanford, CA. 94305 | course-evaluations@stanford.edu