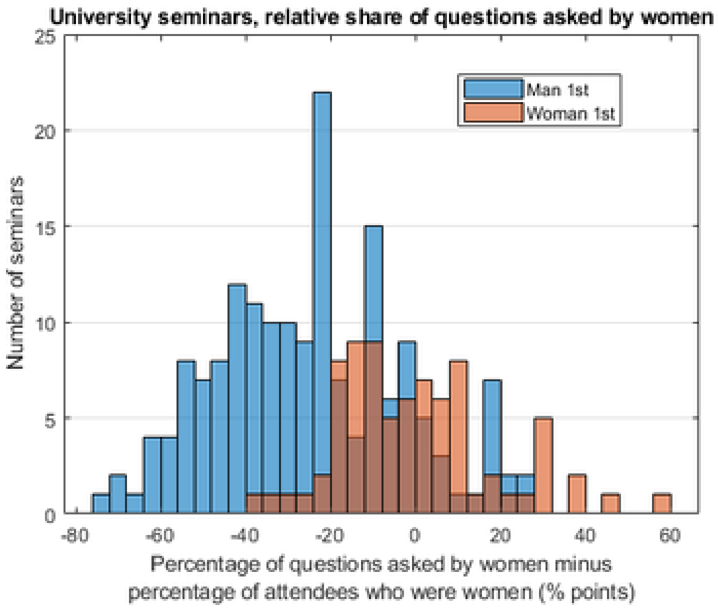
1.

The following article appeared in The Economist on Dec. 7, 2017.

"ONE theory to explain the low share of women in senior academic jobs is thatthey have less self-confidence than men. This hypothesisis supported by data in a new working paper, by a team of researchers from five universities in America and Europe. In this study, observerscounted the attendees, and the questions they asked, at249 departmental talks and seminars in biology, psychology and philosophy that took place at 35 universities in ten countries. On average, half of eachseminar's audience was female. Men, however, were over 2.5 times more likely to pose questions to the speakers—an action that may be viewed (rightlyor wrongly) as a sign of greater competence.

This male skew in question-asking was observable, however, only in those seminarsin which a man asked the first question. When a woman didso, the gender split in question-asking was, on average, proportional to that of the audience. Simply handing the microphone to a woman rather thana man when the floor is opened for questions may make a difference, however small, to one of academia's most intractable problems."



In discussing the paper with a colleague, however, you raise a concern that sorting the data according to who asked the first question might introduce a bias into their metric, particularly if the number of questions per seminar were small.

In order to explore this question of bias, we are going to simulate the sorting procedure used by the authors under the null hypothesis that men and women are equally likely to ask a question regardless of who asks the first question. But first we want some unitary measure of the "effect size" that was actually obtained from the data. In this case, the effect in which we are interested is that of who asked the first question. To start with something simple, let's just look at the difference in the means of the two distributions shown in the figure above. The sign of the difference is important here (Why?), so we'll adopt the convention of subtracting the mean of the "Man 1st" distribution from that of the "Woman 1st" distribution.

What is the mean effect size rounded to the nearest whole number?

2.

Now we want to ask whether the sorting procedure used by the authors might have contributed to the measured effect size. To do this, we will simulate the data collection, sorting and analysis under H0, which is that men and women are equally likely to ask a question regardless of who asked the first one. Assume that there were equal numbers of men and women at each seminar, and further assume that the same number of questions was asked at each of the 249 seminars. For each simulation, calculate the "effect size" as the difference in means between the scores (see x-axis in figure to question #1) just as you did previously.

What is the average effect size you would get if there were only 3 questions asked at each seminar? Round you answer to the nearest whole number.

3.

What if there were 4 questions per seminar?

4.

What if there were 5 questions asked at each seminar?

5.

Do you see a trend here? Write a formula for the average size of the bias effect, using the variable, nQPS, to represent the number of questions per seminar.

6.

We could probably have figured this out without going to the trouble of writing simulation code. So why bother with the simulations? Do they add anything?

7.

For a given number of questions asked, what would happen to our distribution of simulated effect sizes if there were 127 seminars instead of 249?

A. The variance will increase; the mean will stay the same.

B. The variance will decrease; the mean will stay the same.

C. Both the mean and the variance will increase.

D. The variance will increase, but the mean will decrease.

E. Neither the variance nor the mean will change.

8.

What is the smallest number of questions per seminar for which you would be 95% confident that the actual effect size obtained (25) was not purely do to a sorting bias? (Use the actual number of seminars in the study, i.e. 249, and perform a 1-tailed comparison)

9.

How could you eliminate the bias, but still answer the same question about the effect of the sex of the first question asker?

10.

Bonus! (For those who finish early and want to probe further.) For the previous simulations, we assumed that all seminars provided the exact same number of questions. But this is highly unlikely to be the case for the actual survey. Does it matter? Augment your simulation so that you input an *average* number of questions per seminar, but allow it to vary randomly for each seminar using a Poisson distribution. How does the effect size due to the sorting bias compare between the two types of simulations? (i.e. when you have *exactly N questions per seminar* vs. when you have an *average of N questions per seminar*)

A. The sorting bias is **larger** for the more variable case.

B. The sorting bias is **smaller** for the more variable case.

C. The sorting bias is **the same** regardless.

D. There is no sorting bias.