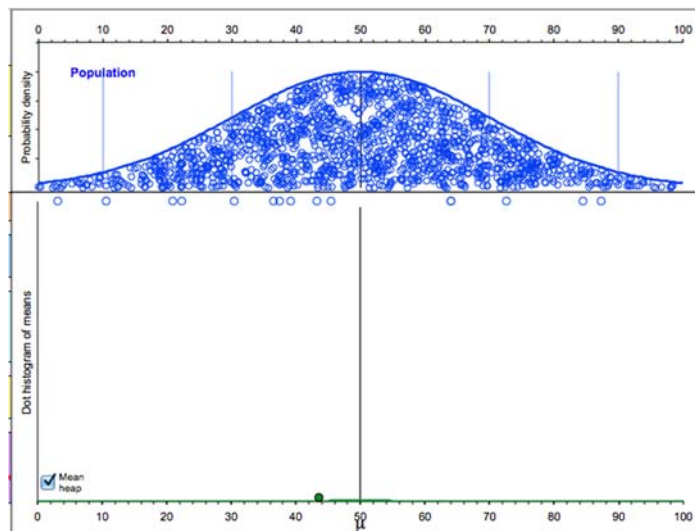


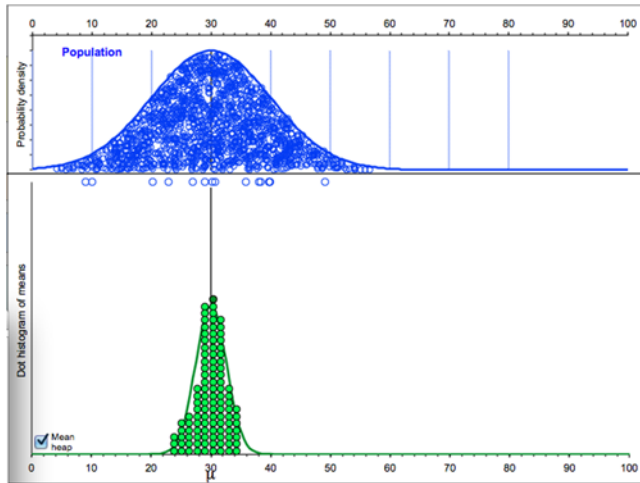
Chapter 4: Changing the Way We Do Statistics: The New Statistics

4.1.1 Application activity for ESCI and confidence intervals

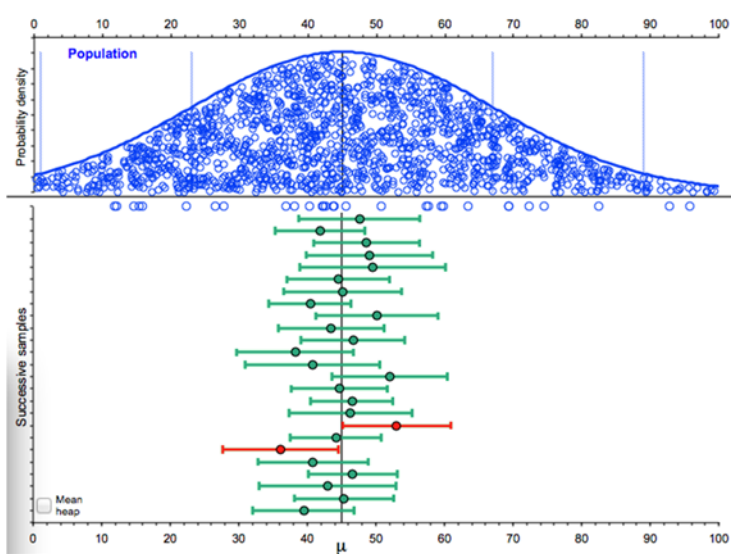
- 1 Here's a screenshot I took after following the directions. It will look different from yours in that the sampled dots will be different, since each sample is randomly drawn from the population.



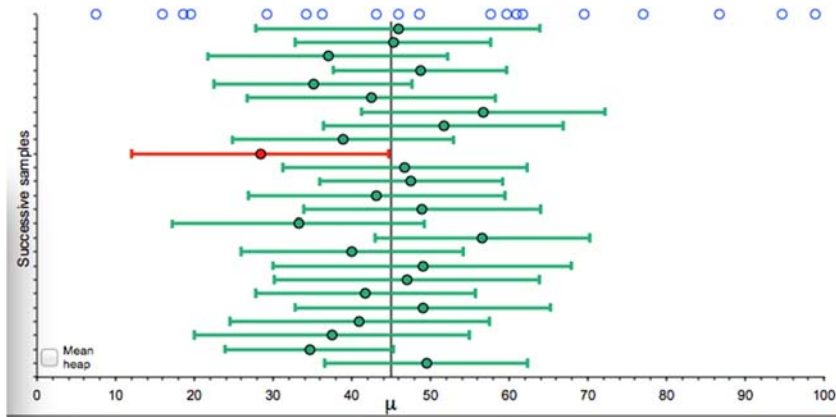
2 Screenshot with ESCI where $\mu = 30$ and $\sigma = 10$, after 100 samples.



3 Screenshot with $\mu = 45$ and $\sigma = 22$. In my figure 25 confidence intervals are visible, and 2 do not contain the population mean.



- 4 Screenshot with $\mu = 45$ and $\sigma = 40$. With a larger standard deviation the confidence intervals are much wider. In my figure 25 confidence intervals are visible, and 1 does not contain the population mean (I did not take a picture of the population curve this time).



- 5 When you make a wider confidence interval (99%), the confidence intervals get longer. You may also notice that if you had one or two CIs that were red with 95% confidence, these have disappeared. Now you only have a 1 in 100 chance of having a CI that does not contain the population mean. When you make a smaller confidence interval (90%), the CIs get shorter, but more red ones creep in. This is because your chance of having a CI that does not contain the population mean goes up.

4.1.3 Application Activity with Confidence Intervals

1 Larson-Hall (2008)

The only test for which there is an effect is the third one for the phonemic discrimination task because it is the only one that doesn't go through zero (although it comes awfully close!). The effect is not very large with any of the tests. Even with the phonemic discrimination task (the only one for which a p -value would be below .05), the middle of the cat's eye view of the confidence interval would show that the most plausible values for the true difference in means

(those closest to M) are only around 5 points out of 96, which is not a very large difference.

Looking at these confidence intervals I would say that the practical effect of having children learn English in a school-type setting, for just one or two hours a week (some parents in Japan spend about \$1000 a year or more to send their children to English conversation schools while young, thinking they learn better while younger) is not worth the difference in scores on any of the linguistic tests.

2 Ellis & Yuan (2004)

There is no effect for any of the groups because all of the confidence intervals go through zero (so none of the comparisons between groups would be statistical at the .05 level). However, we also want to think about the size of the effect, and the confidence intervals let us see that some comparisons have larger effects than others. For example, the comparison between OLP and PTP has a confidence level centered near zero, which means that the most likely values for the true mean are somewhere near zero. For the comparison between PTP and NP, the effect is larger, although if we look at the scale of the numbers along the bottom, we see that they are very small, but that makes sense since the actual number of disfluencies for the groups was quite small. So thinking about the practical effect of different levels of time, it doesn't seem to have been that important in explaining the number of disfluencies.

3 Thought Question

P-values can only tell you whether the test is statistical or not. In both cases, the *p*-value is lower than .05, which means that the correlation is statistical and there is some relationship between age of arrival and scores on a grammaticality judgment test (GJT). But the confidence intervals

also show the size of the effect. Both confidence intervals show very large effect sizes for the correlation between age of arrival and scores on the GJT, even if the true correlation coefficient were found to be at the lower end of the confidence intervals.

4 Poehner & Lantolf (2013)

The 95% confidence interval shows there was definitely an effect of mediation, and that it was quite large, with a plausible range for improvement in scores from 10 to 18 points! Given that there were only 76 possible points on the test, this is quite a large amount of improvement. I would conclude that the mediation was definitely practically important for helping learners to improve their performance on the test.

4.1.5 Application activity with CIs and precision

1 Larson-Hall (2008)

MOE = 1/2 of the CI, so:

- a. Aptitude test CI width = $1.57 + .99 = 2.56$; $2.56/2 = 1.28$
- b. GJT CI width = $5.62 + .76 = 6.38$; $6.38/2 = 3.19$
- c. Phonemic discrimination test = $10.8 - .001 = 10.799$; $10.799/2 = 5.4$

For the aptitude test, the MOE is 1.28 out of a possible 37 points—this seems a fairly precise measurement. For the GJT, MOE is 3.19 points out of a possible 200 points, so again this seems to be fairly precise. For the phonemic discrimination test, MOE is 5.4 out of 96 points, which is not as precise as the others but not terribly imprecise either. This is certainly the test with the least precision of any of the 3.

2 Ellis & Yuan (2004)

The MOE for the comparison between the OLP and PTP groups is about .04 (just looking at the right-hand arm of this confidence interval, which starts near 0 and ends near .04). The other group comparisons seem to have similar MOE. This does seem like a fairly precise measurement and I feel satisfied that I understand that groups differences don't play a large role in explaining the number of disfluencies that were made in the writing task.

3 DeKeyser (2000) and Flege, Yeni-Komshian, and Liu (1999)

The MOE for the DeKeyser (2000) study is $.76 - .42 = .34$; $.34/2 = .17$

The MOE for the Flege, Yeni-Komshian, and Liu (1999) study is $.92 - .87 = .05$; $.05/2 = .025$

The Flege, Yeni-Komshian, and Liu (1999) study has a more precise CI and MOE. This is due to the larger number of participants in this study ($n = 264$) versus the DeKeyser study ($n = 57$).

4 Poehner & Lantolf (2013)

The MOE is half of the CI, $[10.4, 18.3]$. It is thus $18.3 - 10.4 = 7.9$; $7.9/2 = 3.95$. Given that the test had a possible 76 points, this seems fairly precise, and the CI looks fairly small in Figure 4.6.

While a little more precision might be desirable, I feel fairly satisfied that I understand the real effect of the test.

4.3.7 Application Activity with Power Calculation

- 1 Open the G*Power program. Choose TESTS > MEANS > TWO INDEPENDENT GROUPS. Set Tails to "Two," Effect size d to .4, leave α err prob at .05, change Power to .80, and leave the Allocation ratio at 1. Press the "Calculate" button and it says you need 100

participants in each group for a total of 200 participants. If you change the Power to .95, you'll need 164 participants in each group for a total of 328 participants.

- 2 Open the G*Power program. Choose TESTS > CORRELATION AND REGRESSION > BIVARIATE NORMAL MODEL. Set Tails to "Two," Correlation (effect size r) to .4, leave α err prob at .05, change Power to .80, and leave the correlation for the null hypothesis at 0. Press the "Calculate" button and it says you need 67 participants in each group. It appears that 10 participants was much too small in the original study to find an effect size.
- 3 For this question the answers will vary depending on what parameters you choose, but to start, in the G*Power program you should choose TESTS > MEANS > TWO DEPENDENT GROUPS (MATCHED PAIRS).

4.4.1 Application Activity with Precision Calculation

- 1 Open the "Precision two" tab. In the light blue band near red 5, put in 2 for the "Target MOE" and 5 for the "Population σ ." You would need to have two groups, each with 194 participants, to achieve 50% precision for the MOE. For 99% assurance you would need 225 participants in each group.
- 2 Open the "Precision two" tab. In the green band near red 6, put in 1 for the " δ " and .3 for the "Target ave. MOE." You would need to have two groups, each with 97 participants, to achieve this precision for the confidence interval around an effect size of $d = 1$. For 99% assurance you would need 106 participants in each group.
- 3 Open the "Precision paired" tab. Answers will vary here based on what values students pick for the MOE and for the standard deviation.