16.1. Appendix I – How statistical model affects confidence interval

The statistical model reflects the sampling frame by which studies were included in the analysis, and it determines how we can generalize from the studies in the analysis.

- In the case of the fixed-effect (singular) model, all studies are based on one population and our results apply only to that one population. There is only one source of sampling error (we sample people within the population), and so the confidence interval tends to be relatively narrow (10 points wide in Figure 131). The interval is narrow because it applies only to the one population in the analysis.

- In the case of the fixed-effects (plural) model, studies are based on a named set of populations and our results apply only to the specific populations in the analysis. There is only one source of sampling error (we sample people within the populations), and so the confidence interval tends to be relatively narrow (10 points wide in Figure 132). The interval is narrow because it applies only to the studies in the analysis.

- In the case of the random-effects model, our studies are based on populations that have been randomly sampled from a universe of populations, and we can generalize our results to that universe. There are two sources of sampling error (we sample populations from the universe of populations, and then we sample people within each population), and so the confidence interval tends to be relatively wide (60 points wide in Figure 133). It is the wider confidence interval that allows us to extrapolate from the studies in the analysis to the wider universe of comparable studies.

Here, I want to show how this works in practice. To that end, I will use three versions of a fictional analysis where we want to identify the mean score on a math test. All three analyses include twenty samples with fifty students in each, for a total of one thousand students in each analysis. In all three cases, we assume that the standard deviation of scores within a school is 75 points, and the standard deviation of school means is 66.14. Yet, the estimate of the mean is relatively precise for the fixed-effect model and the fixed-effects model, and less precise for the random-effects model.

These are the same examples used in the text, but presented in a different sequence.

Fixed effect (singular)

Suppose that we want to estimate the mean score for a specific school, which has a selective admissions policy. In Figure 131 the circle reflects the *true* mean score in this school. We draw 20 random samples of 50 students each from this school. We report that the mean for this school is 600, with a confidence interval of 595 to 605 [A].

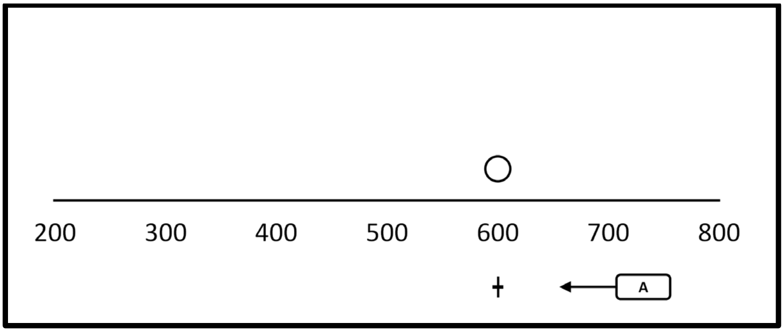

Figure 131 | Fixed effect (singular) | Confidence interval 10 points wide

The mean actually observed in any given sample will invariably fall below or above the circle due to sampling error. However, since all samples are estimating the same parameter, the true effect size for all samples is the same. For that reason, there appears to be only one circle.

The relevant model here is the *fixed-effect* model (where the word *effect* is in the singular). The word *fixed* is generally taken to mean *common*, in the sense that all studies are estimating the identical value. It can also be taken to mean *identified* or *set*, in that our interest is limited to this one school. The word *effect* is in the singular since all samples are estimating the identical effect (the mean for this one school). Here, we limit ourselves to a specific high school, and the results apply only to that high school. The fact that the mean in this school is 600 tells us nothing about the mean for all schools in the city.

In this example we assume that the standard deviation of scores within the school is 75 points. The population variance is given by

$$S^2 = 75^2 = 5625 \tag{10}$$

We draw twenty samples of fifty students each, so the standard error of the mean is

$$SE = \sqrt{\frac{S^2}{N}} = \sqrt{\frac{5625}{1000}} = \sqrt{5.625} = 2.37 \tag{11}$$

where S^2 is the within-study variance and N is the total number of subjects, accumulated across studies. The effect size in the sample will usually fall within two standard error on either side of the mean. In round numbers the confidence interval is 595 to 605, or around 10 points wide (Figure 131).

Fixed effects (plural)

Suppose that we want to estimate the mean score for 20 schools that are under the control of one specific school board. We identify these 20 schools by name, and then draw a random sample of 50 students within each of these schools. In Figure 132 the circles reflect each of the twenty schools that are included in our sample. We report that the mean for this specific set of schools is 400, with a confidence interval of 395 to 405 (A).

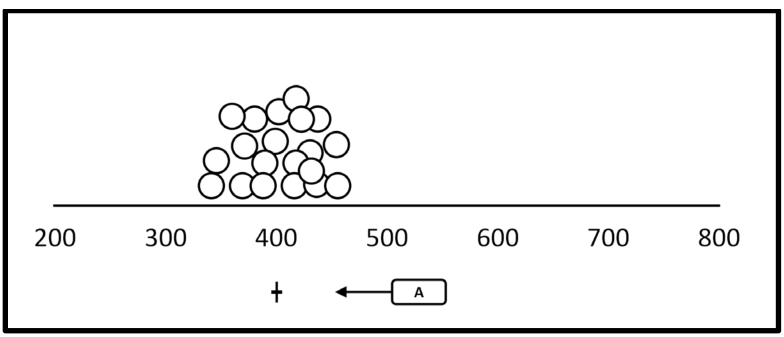

Figure 132 | Fixed effects (plural) | Confidence interval 10 points wide

The relevant model here is the *fixed-effects* model (where the word *effects* is in the plural). The word fixed means *identified* or *set*, in that our interest is limited to this set of schools. The word *effects* is in the plural since each sample is estimating the effect (the mean) in a different school. Here, we limit ourselves to a specific set of schools, and the results apply only to this set. Since these schools are not representative of all schools in the city, the fact that the mean in these 20 schools is 400 tells us nothing about the mean for all schools in the city.

In this example we assume that the standard deviation of scores within any single school is 75 points. The population variance within schools is given by

$$S^2 = 75^2 = 5625\,. \tag{12}$$

We draw twenty samples of fifty students each, so the standard error of the mean is

$$SE = \sqrt{\frac{S^2}{N}} = \sqrt{\frac{5625}{1000}} = \sqrt{5.625} = 2.37 \tag{13}$$

where S^2 is the within-study variance and N is the total number of subjects, accumulated across studies. The effect size in the sample will usually fall within two standard error on either side of the mean. In round numbers the confidence interval is 395 to 405, or around 10 points wide (Figure 132).

Fixed effect (singular) vs. Fixed effects (plural)

For purposes of computing the standard error of the mean effect, the between-study error is assumed to be zero under both the fixed-effect model (11) and the fixed-effects model (13).

Why is the between-study error zero? In both cases we could say that this is because we will *not* be extrapolating to a wider universe. That is, the results apply only to the population(s) actually included in the analysis. In the case of the fixed-effect (singular) model we could also say that the between-study variance for one population is zero, by definition.

Random effects

Suppose that we want to estimate the mean score for all high schools in a large city. We draw a random sample of 20 schools from this universe of schools, and then draw a random sample of 50 students within each of these schools. In Figure 133, the normal curve reflects the distribution of means for all schools in the city, and the circles reflect the true mean for each of the schools included in our sample. We report that the mean for all schools in the city is 500, with a confidence interval of 470 to 530 [B].

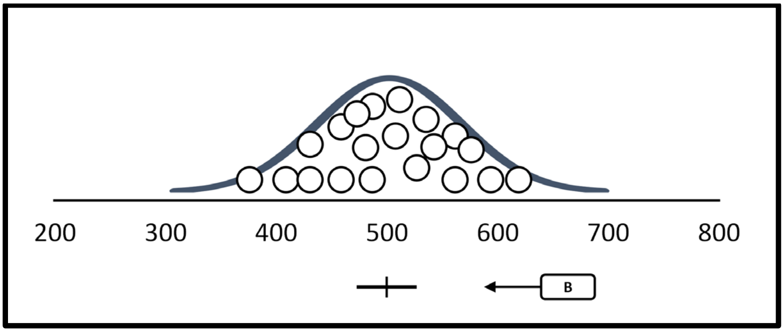

Figure 133 | Random effects| Confidence interval 60 points wide

The relevant model here is the *random-effects* model. The word *random* reflects the fact that the schools in our analysis are a *random* sample from the defined universe. The word *effects* is in the plural, since the effect (the mean score) is assumed to vary from school to school. Here, we obtain a random sample from all public high schools in the city, and the results allow us to generalize to that entire universe of schools.

In our example we sampled twenty schools within a relevant universe (all schools in the city) and our goal was to estimate the mean math score for all schools in the city. In Figure 133, the normal curve reflects the distribution of means for all schools in the city, and the circles reflect each of the twenty schools that are included in our sample. In each case the circle corresponds to the true score for the school. The mean actually observed in that school will invariably fall below or above the circle due to sampling error.

In this example we assume that the standard deviation of scores within any single school is 75 points. The population variance within schools is given by

$$S^2 = 75^2 = 5625\,. \tag{14}$$

We also assume that the standard deviation of (true) school means across schools is 66.14. In Figure 133, the normal curve represents the distribution of school means for all schools in the city, and shows that some 95% of school means fall in the range of 366 to 734, which is the mean plus/minus two standard deviations. (The full curve extends from 300 to 700, which captures some 99% of all schools and is the mean plus/minus three standard deviations).

The population variance for school means across all schools in the city is

$$S^2 = 66.14^2 = 4375\,. \tag{15}$$

The standard error for the summary effect size is given by

$$SE = \sqrt{\frac{S^2}{N} + \frac{T^2}{k}} = \sqrt{\frac{5625}{1000} + \frac{4375}{20}} = \sqrt{5.625 + 218.75} = 14.98\,. \tag{16}$$

The mean effect size computed in our meta-analysis will usually fall within two standard error on either side of the city mean. In round numbers the confidence interval is 470 to 530, or around 60 points wide (Figure 133).

Note that the error variance in this case includes two terms. The first term

$$\frac{S^2}{N} = \frac{5625}{1000} = 5.625 \tag{17}$$

is identical to the term in (13), and is the error based on within-study variance. If our goal was to report the mean score for the twenty schools in the analysis, and not extrapolate from these to all schools in the city, then we would be using the fixed-effects (plural) model rather than the random-effects model, and the standard error would be 2.37.

The second term is

$$\frac{T^2}{k} = \frac{4375}{20} = 218.75\,. \tag{18}$$

This corresponds to the between-study error variance. The variance of school means is 4375 and (as always) the error variance in a sample will be equal to the population variance divided by the number of units in the sample. Here, this works out to 4375/20.

Put another way, suppose that after we sampled 20 schools, we did not draw a sample of 50 students within each school but instead tested 100% of the students in those schools so that we knew the mean score for each school with zero error. In that case the sampling error variance for the overall mean in our analysis would be 218.75 and the standard error would be the square root of this number.

Since we did not include all students in each school, but did sample 50 students within each school, we have two distinct sources of sampling error and need to include both of them in the analysis. Since the two terms are independent of each other, we can simply add the two variances, as we did in (16) to get a standard error of roughly fifteen points. The confidence interval (as always) is roughly four standard error, which works out to sixty points.

16.1.1. Random-effects vs. fixed-effects (plural)

The random-effects model and the fixed-effects (plural) model both assume that the effect size differs from study to study. What distinguishes between the two models is the fact that the fixed-effects model is intended to report the mean effect size for the studies in the analysis, while the random-effects model is intended to extrapolate from the studies in the analysis to the universe from which these studies were sampled.

In both models, the error variance includes the term S^2/N, which reflects the fact that the mean in any sample is only an estimate of the true mean in the corresponding population. The random-effects model adds the additional term T^2/k to reflect the fact that the studies in the analysis will be used to extrapolate to a wider universe, while the fixed-effect model does not include this term, since the studies in the analysis will not be used to extrapolate to a wider universe. Put another way, it is this term which allows us to extrapolate to the wider universe.