



Contents

9.1 Standardized Effect Sizes.....	106
9.2 Meta-analysis.....	107
Appendix.....	108
Standardized Effect Sizes Beyond the Simple Case.....	108
Extended Example of the Meta-analysis.....	109

What You Will Learn in This Chapter

In Part III of this book, we will show that combining data from multiple experiments can provide completely new insights. For example, whereas the statistical output of each experiment itself might make perfect sense, sometimes the combination of data across experiments indicates problems. How likely is it that four experiments with a small effect and a small sample size all lead to significant results? We will show it is often very unlikely. As a simple consequence, if experiments always produce significant results, the data seem too good to be true. We will show how common, but misguided, scientific practice leads to too-good-to-be-true data, how this practice inflates the Type I error rate, and has led to a serious science crisis affecting most fields where statistics plays a key role. In this respect, Part III generalizes the Implications from Chap. 3. At the end, we will discuss potential solutions.

In this chapter, we extend the standardized effects size from Chap. 2 and show how to combine data across experiments to compute meta-statistics.

9.1 Standardized Effect Sizes

As we noted in Part I of this book, much of statistics involves discriminating signal and noise from noise alone. For a standard two sample t -test, the signal to noise ratio is called Cohen's d , which is estimated from data as (see Chap. 3):

$$d = \frac{\bar{x}_1 - \bar{x}_2}{s}$$

Cohen's d tells you how easily you can discriminate different means. The mean difference is in the numerator. A bigger difference is easier to detect than a smaller one, but we also need to take the noise into account. A bigger standard deviation makes it more difficult to detect a difference of means (see Chap. 2). When $n_1 = n_2 = n$, the t -value for a two-sample t -test is just:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{2}{n}}} = \frac{d}{\sqrt{\frac{2}{n}}} = d\sqrt{\frac{n}{2}}$$

So, a t -value simply weights Cohen's d by (a function of) sample size(s). As mentioned in Chap. 3, it is always good to check out the effect size. Unfortunately, many studies report just the p -value, which confuses effect size and sample size. Based on the above equation, we can compute Cohen's d from the reported t -value and the sample sizes:

$$d = t\sqrt{\frac{2}{n}}$$

An important property of Cohen's d is that its magnitude is independent of the sample size, which is evident from d being an estimate of a fixed (unknown) population value.¹

In Chap. 3, we have shown that we can estimate δ by d . However, d is only a good estimator when the sample size is large. For rather small samples, d tends to systematically overestimate the population effect size δ . This overestimation can be corrected by using Hedges' g instead of d :

$$g = \left(1 - \frac{3}{4(2n - 2) - 1}\right)d$$

For nearly all practical purposes Hedges' g can be considered to be the same as Cohen's d . We introduced it here because we will use Hedges' g to compute meta-analyses. The

¹Note that although the sample size n appears in this particular formula, it basically just compensates for t increasing with larger sample size.

Appendix to this chapter includes formulas for when $n_1 \neq n_2$ and for other types of experimental designs.

9.2 Meta-analysis

Suppose we run the same (or very similar) experiments multiple times. It seems that we should be able to pool together the data across experiments to draw even stronger conclusions and reach a higher power. Indeed, such pooling is known as meta-analysis. It turns out that the standardized effect sizes are quite useful for such meta-analyses.

Table 9.1 summarizes statistical values of five studies that concluded that handling money reduces distress over social exclusion. Each study used a two-sample t -test, and the column labeled g provides the value of Hedges' g , which is just an estimate of the effect size.

To pool the effect sizes across studies, it is necessary to take the sample sizes into account. An experiment with 46 subjects in each group counts a bit more than an experiment with 36 subjects in each group. The final column in Table 9.1 shows the weighted effect size, $w \times g$, for each experiment (see the Appendix for the calculation of w). The pooled effect size is computed by summing the weighted effect sizes and dividing by the sum of the weights:

$$g^* = \frac{\sum_{i=1}^5 w_i g_i}{\sum_{i=1}^5 w_i} = 0.632.$$

This meta-analytic effect size is the best estimate of the effect size based on these five experiments. Whether it is appropriate to pool standardized effect sizes in this way largely depends on theoretical interpretations of the effects. If your theoretical perspective suggests that these experiments all measure essentially the same effect, then this kind of pooling is appropriate, and you get a better estimated effect size by doing such pooling. On the other hand, it would not make much sense to pool together radically different experiments that measured different effects.

Meta-analyses can become quite complicated when experiments vary in structure (e.g., published analyses may involve t -tests, ANOVAs, or correlations). Despite these

Table 9.1 Data from five experiments used for a meta-analysis

n	t	g	$w \times g$
36	3.01	0.702	12.15
36	2.08	0.485	8.66
36	2.54	0.592	10.43
46	3.08	0.637	14.17
46	3.49	0.722	15.83

difficulties, meta-analysis can be a convenient way to combine data across experiments and thereby get better estimates of effects.

Take Home Messages

1. Pooling effect sizes across experiments produces better estimates.
2. Combining data across experiments increases power.

Appendix

Standardized Effect Sizes Beyond the Simple Case

When samples sizes are different ($n_1 \neq n_2$), the t -value of a two-sample t -test is:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{\bar{x}_1 - \bar{x}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{d}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = d \sqrt{\frac{n_1 n_2}{n_1 + n_2}}.$$

If a published study does not report the means and standard deviations from the samples, it is possible to compute Cohen's d from the reported t -value and sample sizes:

$$d = t \sqrt{\frac{n_1 + n_2}{n_1 n_2}}$$

For Hedges' g , the calculation with unequal sample sizes is:

$$g = \left(1 - \frac{3}{4(n_1 + n_2 - 2) - 1}\right) d$$

There are similar standardized effect sizes and corrections for other experimental designs. For example, for a one-sample t -test with a null hypothesis of the population mean being equal to the value a , Cohen's d is calculated as

$$d = \frac{\bar{x} - a}{s}$$

which, again, represents signal in the numerator (deviation from the value specified by the null hypothesis) and noise in the denominator (the sample standard deviation). An unbiased version of Cohen's d for the one-sample case is Hedges' g :

$$g = \left(1 - \frac{3}{4(n - 1) - 1}\right) d$$

For repeated measures t -tests, the appropriate standardized effect size depends on how it will be used. Sometimes, a scientist wants an effect size relative to the difference scores that are calculated for each subject. For that use, the one-sample d or g is appropriate. Other times, scientists want to find an effect size that is equivalent to what it would be for a two-sample independent t -test. In that situation it is necessary to compensate for the correlation between scores. When computed from the reported t value of a dependent sample, the formula is:

$$d = \frac{t}{\sqrt{n}} \sqrt{2(1-r)}$$

Unfortunately, most papers do not report the correlation between scores for a dependent sample. For our purposes, the basic idea of a standardized effect size is more important than the specific calculation. However, you should be aware that formulas you may find on the Internet sometimes include unstated assumptions such as equal sample sizes for an independent t -test or $r = 0.5$ for a dependent t -test.

Extended Example of the Meta-analysis

Table 9.2 fills in some intermediate terms that are not present in Table 9.1.

To pool the effect size across studies, we weight each g value by its inverse variance. The calculation of the inverse variance involves multiple steps. For an independent two-sample t -test, the formula for the variance of Cohen's d is

$$v_d = \frac{n_1 + n_2}{n_1 n_2} + \frac{d^2}{2(n_1 + n_2)}$$

and the variance for Hedges' g includes the square of the correction term used earlier:

$$v_g = \left(1 - \frac{3}{4(n_1 + n_2 - 2) - 1}\right)^2 v_d$$

Table 9.2 Detailed meta-analysis, including additional computations, of the data shown in Table 9.1

n_1	n_2	t	g	v_g	w	wg
36	36	3.01	0.702	0.058	17.3	12.15
36	36	2.08	0.485	0.056	17.9	8.66
36	36	2.54	0.592	0.057	17.6	10.43
46	46	3.08	0.637	0.045	22.2	14.17
46	46	3.49	0.722	0.046	21.9	15.83

which is shown in a separate column in Table 9.2. To do the meta-analysis, each standardized effect size is multiplied by its inverse variance:

$$w = \frac{1}{v_g}$$

which is shown in a column in Table 9.2 next to a column listing the product of wg for each experiment. The pooled effect size is computed by summing the products and dividing by the sum of the weights:

$$g^* = \frac{\sum_{i=1}^5 w_i g_i}{\sum_{i=1}^5 w_i} = 0.632.$$

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

