

Publishing Date: March 1994. © 1994. All rights reserved. Copyright rests with the author. No part of this article may be reproduced without written permission from the author.

Meta Analysis • 2

How to measure effect size

Chuck Chakrapani

From significance to magnitude

In the last article we saw how we can combine different studies in order to reconcile and consolidate their statistical significance. Tests of significance tell us that the obtained differences are 'statistically significant'. All this means is that the differences are greater than 0. In our earlier examples (cited in the February 1994 Imprints), the ratings of your company were significantly higher than those of your competitors. From this we know there is a difference, but not how much. Knowing the size of the difference is particularly important in cases where a particular course of action will be worthwhile only if the difference is a given size. Consequently we need a measure for 'effect size'.

Effect size

Effect size can be defined as the 'degree to which the phenomenon (in our case the differences in ratings) is present in the population. A wide variety of tests are available to measure effect size. In this article we will discuss a few.

ES Index

Let us consider a hypothetical example. A series of studies were conducted to assess the effect of a 'say no to drugs' advertising campaign. (The effect is measured by the stated intention to stay away from drugs) These studies were conducted at different points in time by different researchers using different rating scales. In the previous article we saw how we can consolidate these studies and assess the statistical significance of the results obtained. Here we will see techniques that will help us assess the magnitude of the effect of one variable (advertising campaign) on another (intention to stay away from drugs). The results of four studies that studied the effect of the 'say no to drugs' campaign are given in Exhibit 1.

Exhibit 1
Effect of watching the 'Say no to drugs' campaign

Study	N	Before exposure	After exposure
1*	220	6.6	6.9
2**	250	3.8	3.6
3**	180	3.6	3.9
4*	150	6.7	7.2

* On a 7-point scale

** On a 5-point scale

Standardizing the effects

The first thing we want to do is to convert the difference in each of the studies into standard units. This is done by simply taking the differences found in each study and dividing them by the standard deviation. This measure is called the *d* or the ES Index. [The standard deviation used to standardize the differences can either be the standard deviation of the 'control group' (if a two group design is used) or simply of the pooled (within groups) scores.] Standardization effectively gets rid of the effect of different scales, slightly different questionnaire wording and other confounding variables.

Let us apply this to our data. For the first survey, the means are 6.9 and 6.6. The standard deviation is 0.6.

$$d = (I \times \text{Avg. 1} - \text{xAvg. 2}) / SD$$

Where

xAvg.1 = Mean 1

xAvg.2 = Mean 2

II = Ignore the direction
of the difference eg. $-.5 = +.5$

SD = Standard deviation

$$= (6.9 - 6.6) / 0.6 = .5$$

$$d = 6.9 - 6.6 / 0.6 = .5$$

Calculate the d value for each study in the same fashion. We note that in 3 out of the 4 studies the campaign had a positive impact while in one study (study 2) it did not.

After calculating the d values, we assign positive signs to studies showing a positive campaign impact and negative signs to studies that did not and sum the ds.

$$d \text{ mean} = \text{sum } d / n$$

Where

Sum d = Sum of all differences measured
in standardized units

after assigning signs
to the direction of differences

n = number of studies summed.

$$d \text{ mean} = (0.5 - 0.25 + 0.50 + 0.42) / 4 = 0.29$$

Detailed calculations are shown in Exhibit 2.

Exhibit 2						
Standardizing the effects of the 'Say no to drugs' campaign						
Study	N	Before exposure	After exposure	$I \times \text{xAvg.1} - \text{xAvg.2}$	SD	d
1	220	6.6	6.9	0.3	0.6	.50
2	250	3.8	3.6	0.2	0.8	-.25
3	180	3.6	3.9	0.3	0.6	.50
4	150	6.7	7.2	0.5	1.2	.42
						d / n = 0.29
						Where N = number of studies

Interpreting ES of group differences

What does this mean? It means that the ad campaign increases the 'intention to stay away from drugs' by .29 standard deviation units. As noted earlier, because the scores are standardized, it would not matter if different scales were used in different surveys. Consequently the .29 standard deviation effect applies to the campaign, no matter how the 'effect' is measured.

Measuring the margin of error of effect size

While d mean indicates the effect of one variable on the other, it does not tell us anything about the margin of error that may exist around this figure. Let us work with the 95% confidence interval in the earlier example. We found our effect size to be 0.29. If the margin of error is .10, then we can say that the effect size is not likely to be lower than 0.19 ($0.29 - 0.10$) or higher than 0.39 ($0.29 + 0.10$). But if the margin of error happens to be 0.5,

then the effect size of 0.29 ± 0.5 would include 0. In effect, the 'true effect size' can be anywhere between -0.21 and +0.79. This means that although the effect size is 0.29, the margin of error is so high that it is possible that the ad campaign had no effect (or even a negative effect) on the target audience.

Since all survey research is subject to a margin of error, we need to construct confidence intervals. In our example, we constructed four d scores, one corresponding to each study. We then arrived at a dmean of 0.29. To construct a confidence interval for this figure, we can compute the standard error of these dmean scores. This is shown in Exhibit 3.

Exhibit 3						
Calculating the margin of error of d						
Study	N	Before exposure	After exposure	d	d - dmean	(d - dmean) ²
1	220			50	.21	.04
		6.6	6.9			
2	250			- .25	-.54	.29
		3.8	3.6			
3	180			.50	.21	.04
		3.6	3.9			
4	150			42	.13	.02
		6.7	7.2			
				dmean Sum d / n =		
				0.29		.39
Calculating the margin or errof for dmean						
1. Standard deviation of the dmean = $[(d - dmean)^2 / (n-1)] = (.39/ 3) = 0.36$						
2. Standard error of the dmean = Standard deviation of the dmean / $n = .36/2 = .18$						
3. 95% confidence interval = $1.96 \times \text{standard error} = 1.96 \times .18 = \pm 0.35$						
0.29 ± 0.35						

The obtained standard deviation of the d scores for our example is 0.36. The standard error is 0.18. For the 95% confidence interval we multiply this number, 1.96, to obtain 0.35. In other words the true effect size is no lower than -0.06 ($0.29 - 0.35$) and no higher than 0.64 ($0.29 + 0.35$) standard deviation units. Note that the range includes 0, meaning that it is possible that the campaign had no effect on the target audience.

Here is what this analysis tell us. Although three out of four studies indicated that the 'Say no to drugs' campaign influenced the target audience in the intended direction, when we consider the consolidated evidence provided by all four studies, we cannot conclusively state that the campaign had any effect at all on the target audience.

Other approaches

Not every researcher prefers to calculate the standard error of the d values to set the margin of error. In some disciplines, it is not uncommon to compare dmean of a particular meta-analysis with the distribution reported in the literature. Some researchers simply use rules-of-thumb to interpret the results obtained. One such guideline is provided by Cohen (1977). See Exhibit 4 for details.

Exhibit 4

Rules of thumb for interpreting d

- 0.2 = Small effect
- 0.5 = Medium effect
- 0.8 = Large effect

If we apply the Cohen criterion to our data, our obtained dmean of 0.29 would be interpreted to mean that the

advertising campaign had a small effect on the target audience as opposed to 'probably no effect' on the basis of the computed standard error.

References

Cohen (1977) *Statistical Power Analysis for the Behavioral Sciences*. (Revised Edition). New York, Academic Sciences.

BOOK REVIEW

13th GEN

Abort, Retry, Ignore, Fail?

by Neil Howe and Bill Strauss
Published by Vintage Books
New York, 1993

"Madonna. Michael Jackson. Spike Lee. Cartoonists Matt Groening, Bill Watterson, and Berke Breathed. Ice T. The creators of Def Jam rap, of *Max Headroom*, of the *Teenage Mutant Ninja Turtles*, of *Ren & Stimpy*, of *Batman Returns* and of Fox TV entertainment programming.

What do all these people have in common?"

They are important forces shaping the culture and reputation of America's 13th Generation or 13th Gen for short. The 13th Gen were born between 1961 and 1981. There are about 80 million of them. They are the largest demographic group, larger than the baby boom generation.

The images shaping the 13ers are not created by the 13ers but by the boomers who preceded them. The images produced by the icons mentioned earlier encompass dark comedy, frenzy, physicality, alienation and low self-esteem.

How different are 13ers from boomers? Here is a comparison provided by a University of Michigan survey of two groups of high school seniors conducted 12 years apart.

Is it 'extremely' or 'quite' important for you to:

- have a lot of money? 15% in 1976; 31% in 1988.
- have clothes in the latest style? 13% in 1976; 24% in 1988.
- be able to find steady work? 63% in 1976; 63% in 1988;
- have at least 2 cars? 5% in 1976; 12% in 1988.
- find purpose and meaning in life? 64% in 1976; 57% in 1988.

13ers are very different from the boomers. If you want to know why you should read this engaging book. In fact they are so different, Strauss and Howe predict the following:

1. Over the next 15 years, the festering quarrel between 13ers and boomers will grow into the next generation gap.
2. 13ers will never outgrow their bad image.
3. 13ers will be one of the most important immigrant generations in history.
4. Early in life, the most successful 13ers will be risk takers who exploit opportunities overlooked by established businesses.
5. 13ers will be the first generation born in this century to suffer a decline in living standards.
6. 13ers will restore the family.

7. 13ers will reach their 50s in collective exhaustion and slow the pace of change.
8. 13ers will be politically conservative.
9. 13ers will produce national leaders who will excel at cunning, flexibility and deft timing.
10. Before the year 2030, events will call on 13ers to make aging boomers get real.
11. 13ers will neither ask for nor receive much assistance from government.
12. 13ers will favour investment over consumption and the needs of the very young over the needs of the very old.
13. 13ers will make caustic, independent yet self effacing elders.

Although this book is written from the American perspective, most of it can easily be related to Canada.

This book is fun to read. There is a lot in this book for us to think about both as marketers/researchers as well as individuals trying to make sense of the radical restructuring that currently seems to engulf us.

Dr Chuck Chakrapani is President of Standard Research Systems Inc. He is the author of several books and is the Editor-in-Chief of the Canadian Journal of Marketing Research.

© 1994. All rights reserved. Copyright rests with the author. No part of this article may be reproduced without written permission from the author.