

Cumming, G. (2007). Inference by eye: Pictures of confidence intervals and thinking about levels of confidence. *Teaching Statistics*, 29, 89-93.

Journal website: <http://www.blackwell-synergy.com/loi/test> This article may not exactly replicate the final version published in the journal. (It is the version just before copy editing.) It is not the copy of record.

### **Inference by Eye: Pictures of Confidence Intervals and Thinking About Levels of Confidence**

Geoff Cumming

La Trobe University, Melbourne, Victoria, Australia

e-mail: [g.cumming@latrobe.edu.au](mailto:g.cumming@latrobe.edu.au)

**KEYWORDS:** Teaching, Confidence intervals, Displaying data,  $p$  values, Interactive software

Author contact information:

Geoff Cumming,

School of Psychological Science, La Trobe University,

Victoria 3086, Australia

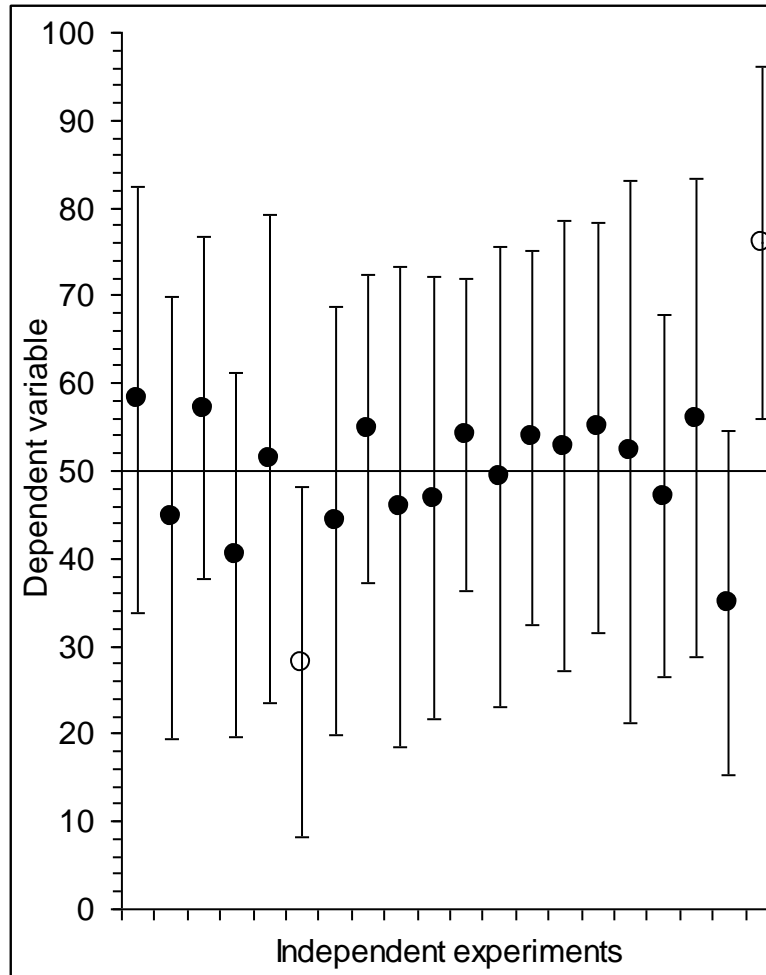
Email: [g.cumming@latrobe.edu.au](mailto:g.cumming@latrobe.edu.au)

A picture of a 95% confidence interval (CI) implicitly contains pictures of CIs of all other levels of confidence, and information about the  $p$  value for testing a null hypothesis. This article discusses pictures, taken from interactive software, that suggest several ways to think about the level of confidence of a CI,  $p$  values, and what conclusions can be drawn from inspecting a CI.

*Inference by eye* is the interpretation of graphically-presented data. A picture of a confidence interval (CI) is a wonderful basis for inference by eye because it gives us both the point estimate—our best bet for what we want to know—and the precision of that estimate. A CI—unlike a test statistic or a  $p$  value—is shown on the measurement scale, and in the units, used by the researcher. It should therefore be easy to give practical meaning to the values signalled by the CI. It is tricky, however, to understand correctly what the level of confidence  $C$  means;  $C$  is expressed as a percentage, and 95% CIs (i.e.,  $C=95$ ) are most common. I offer here several pictures, taken from freely available interactive software, that I hope may help understanding of  $C$ , of levels of confidence, and of what conclusions can be drawn from inspection of a data picture that includes a CI.

### **THE LEVEL OF CONFIDENCE**

Any textbook that offers a correct explanation of  $C$  probably includes a diagram like Figure 1, which shows 95% CIs around the sample means for a set of 20 independent samples, each of size  $n=20$ , taken from a normal population. It is a familiar diagram, but the software provides controls that allow the user to take additional sets of samples, and see the running total of intervals that include  $\mu$ , and also to control many aspects of the display. Playing with the software can illustrate several basic facts about interval estimates, including:



**Fig 1.** The 95% CIs for the population mean  $\mu$ , for a set of 20 independent replications of a study. Each sample has size  $n=20$ . The horizontal line is  $\mu$ . The CIs are based on sample estimates of the population SD and so vary in width from sample to sample. Open circles are the means whose bars do not include  $\mu$ . In the long run, 95% of the CIs are expected to include  $\mu$  (18 do here). Note that the CI varies from sample to sample, but  $\mu$  is fixed and in practice always unknown.

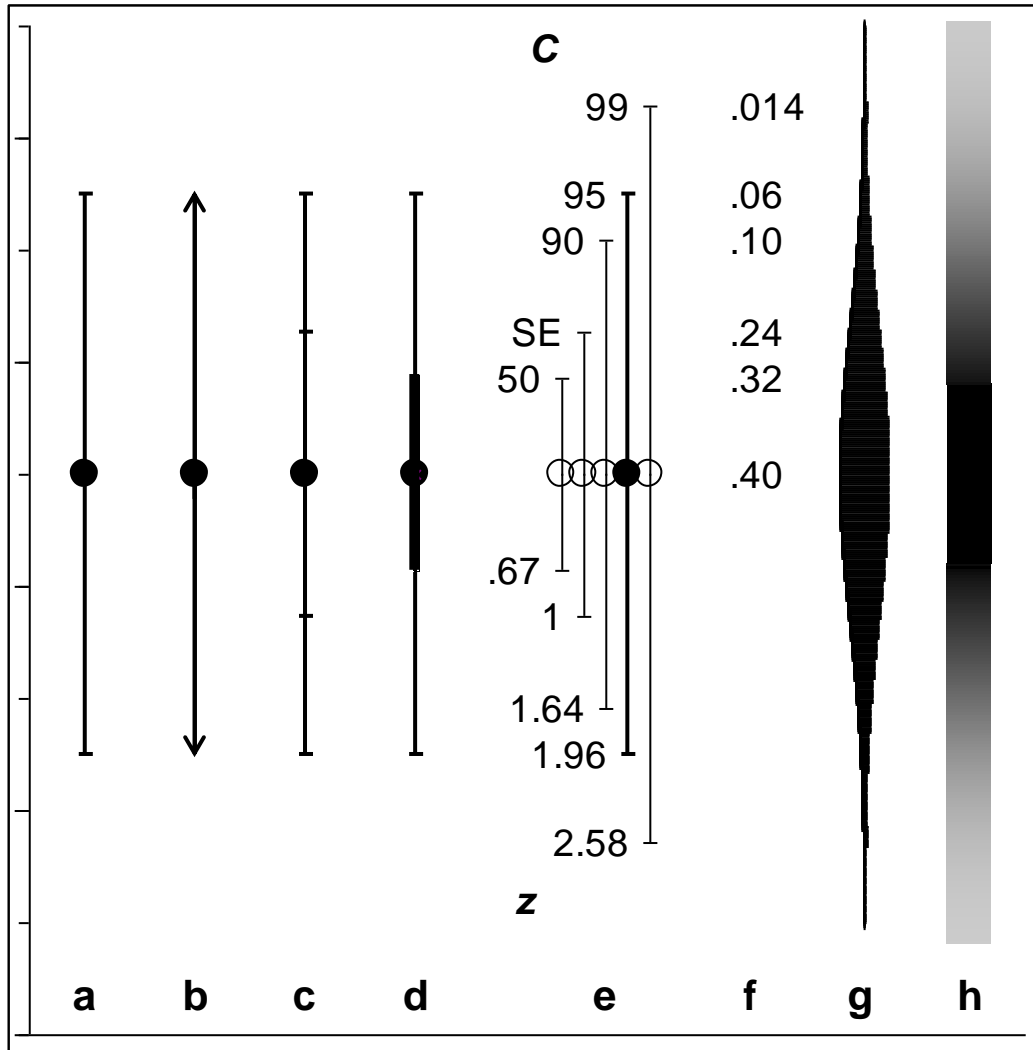
1. The level of confidence  $C$  is the percentage of intervals that, in the long run, include  $\mu$ , the population mean, which is fixed but unknown.
2. If  $C=95$ , sets of 20 samples on average contain one CI that does not include  $\mu$ , but often a set contains none, and occasionally a set contains 3 or even more. If hundreds of sets are taken, however, the overall percentage of intervals containing  $\mu$  will be very close to  $C$ . It is a basic fact of randomness that in the short run results can appear surprisingly 'lumpy', but in the (very) long run expectations are met (very) closely: That's the law of large numbers.
3. It is worth examining the whole set of intervals, but also bearing in mind that a researcher in practice takes only *one* sample, and so has to draw conclusions from just that one

- interval. That interval could be *any* single one of those displayed. (Unfortunately, in real life CIs are not marked as including  $\mu$  or not, as they are in Figure 1!)
4. If  $\sigma$ , the population SD, is assumed known, all intervals are the same width; if  $\sigma$  is not assumed known and  $s$ , the sample SD, is used to calculate each interval, intervals vary in width from sample to sample—as in Figure 1.
  5. Such variation is large for small  $n$ , and is still noticeable even for  $n$  as large as 50.
  6. Points 1 and 2 above—the conclusions about capture of  $\mu$ —apply both for the constant-width intervals based on  $\sigma$  and  $z$ , and for the varying-width intervals based on  $s$  and  $t$ .
  7. The unknown  $\mu$  is more often captured near the centre of an interval than near the lower or upper limit, or endpoint, of an interval.
  8. If 95% CIs are displayed, those few intervals that miss  $\mu$  usually only just miss  $\mu$ , but very occasionally an interval misses  $\mu$  by a considerable distance.
  9. Standard error (SE) bars are about half the width of 95% CIs, if  $\sigma$  is known or  $n$  is roughly at least 10. (SE bars are intervals that extend one standard error above and one standard error below the sample mean  $M$ .)
  10. Increasing  $C$  gives longer intervals, reducing  $C$  gives shorter intervals. Therefore a 99% CI is longer than the corresponding 95% CI, whereas the 90% CI is shorter. Choosing  $C=68$  gives intervals very similar in length to SE bars, if  $\sigma$  is known or  $n$  is roughly at least 10.

### GRAPHICS FOR CONFIDENCE INTERVALS

Figure 2 shows a number of ways that intervals can be displayed. All are calculated using  $z$  and assuming  $\sigma$  is known and, except where marked, all are 95% CIs. I am not suggesting we should use all these representations to report CIs, but I do suggest they may help understand the riches implicit in any CI. Figure 2a shows a conventional graphic for CIs. The discussion above highlights one problem with this graphic: CIs, of any  $C$ , and SE bars all look the same! Authors must be scrupulous always to state what they mean by a pictured interval, and readers must always make certain they know what an interval represents before making any interpretation. Figure 2c is a not very elegant way to solve the problem by showing both SE bars and the 95% CI.

Figure 2b is a suggestion made in the medical literature (Gardner and Altman 1987, p. 746) to emphasise that  $\mu$  may lie beyond the interval. It is true that  $\mu$  may lie outside the interval, but it is probably more important to illustrate in some way Point 7, which notes that a point near the centre of an interval is more likely to fall at  $\mu$  than is a point near either limit. This is an interesting property of intervals that is not made salient by the conventional graphic. The relative chance that various points within and beyond a 95% CI will fall at  $\mu$  is shown in Figure 2g by means of width, and in Figure 2h approximately as the degree of shading. Figure 2d is a simpler possibility that pictures both the 50% and 95% CIs. The values at Figure 2f are the ordinates of a normal distribution and indicate the relative chances that the centre of the CI, and the upper limits of the intervals in Figure 2e, will fall at  $\mu$ . They are proportional to the width of Figure 2g, and approximately to the degree of shading of Figure 2h, at those points. They indicate, for example, that it is about 7 times more likely ( $.40/.06 \approx 7$ ) that a tiny interval at  $M$ , the centre of a CI, will capture  $\mu$  than will a similar tiny interval at the upper (or lower) limit of a 95% CI.



**Fig 2.** Some of the many ways to display interval estimates. Graphics **a-d** show a 95% CI. In **c**, SE bars are also shown, and in **d** the 50% CI is also marked. In **e**, 4 CIs are shown, with **C** marked above, and SE bars are also included; the margin of error is shown below, for  $\sigma$  assumed known, as the  $z$  value. The relative chance of capturing  $\mu$  at various points is indicated in **f** by the values, in **g** by width, and in **h** approximately by shading. (The vagaries of printing, and the human visual system, mean that **h** may not give an accurate impression of the relative chance.)

Note the slightly awkward wording in the previous paragraph, for example “a point near the centre of an interval is more likely to fall at  $\mu$  than is a point near either limit”. It might seem more natural to say “ $\mu$  is more likely to be at  $M$ , the centre of an interval, than near either limit”, but saying that may suggest  $\mu$  is a variable, instead of an unknown but fixed value. The wording used above is meant to remind us that it’s the *intervals* that vary, and any probability refers to what happens in the long run in the set of all intervals. Always keep Figure 1 in mind!

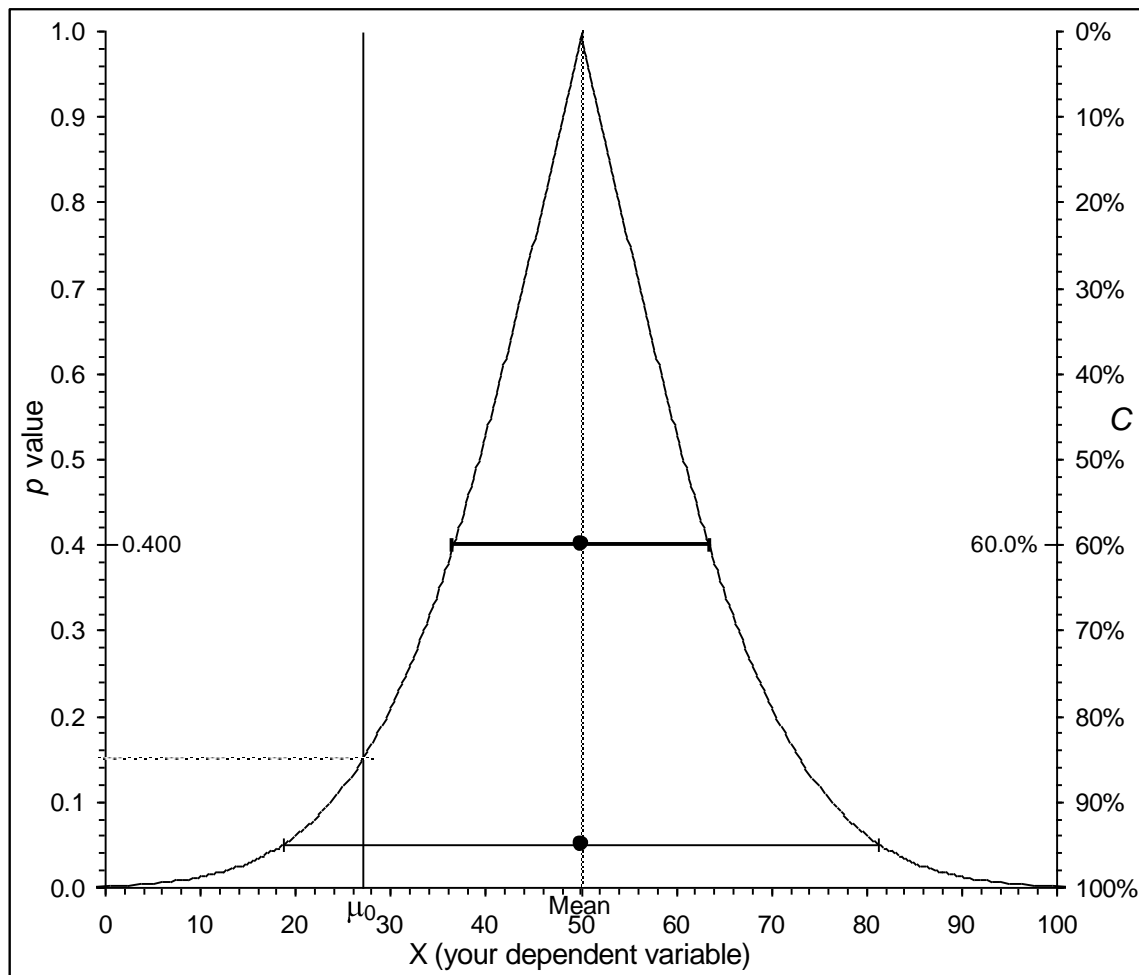
The conventional picture of Figure 2a implicitly contains all the information shown in the other parts of Figure 2, and more.

### THE CONFIDENCE INTERVAL FUNCTION

The simple relation between CI width and  $C$  arises, of course, because the margin of error—which is half the total CI width—is simply the SE multiplied by the critical value of  $z$  (or  $t$ ) for the chosen  $C$ . Figure 2e illustrates the relative lengths of CIs with selected values of  $C$ , and of SE bars. The upper numbers are the values of  $C$ . The lower numbers are the corresponding critical values of  $z$ , which apply when  $\sigma$  is known or  $n$  is large, and suggest some simple relations that may be worth bearing in mind. Relative to a 95% CI:

- SE bars are about half as wide.
- The 50% CI is about one-third as wide—as illustrated also in Figure 2d.
- The 90% CI is about one-sixth shorter.
- The 99% CI is about one-third longer.

More accurate values are, respectively: .51, .34, .16, .31.

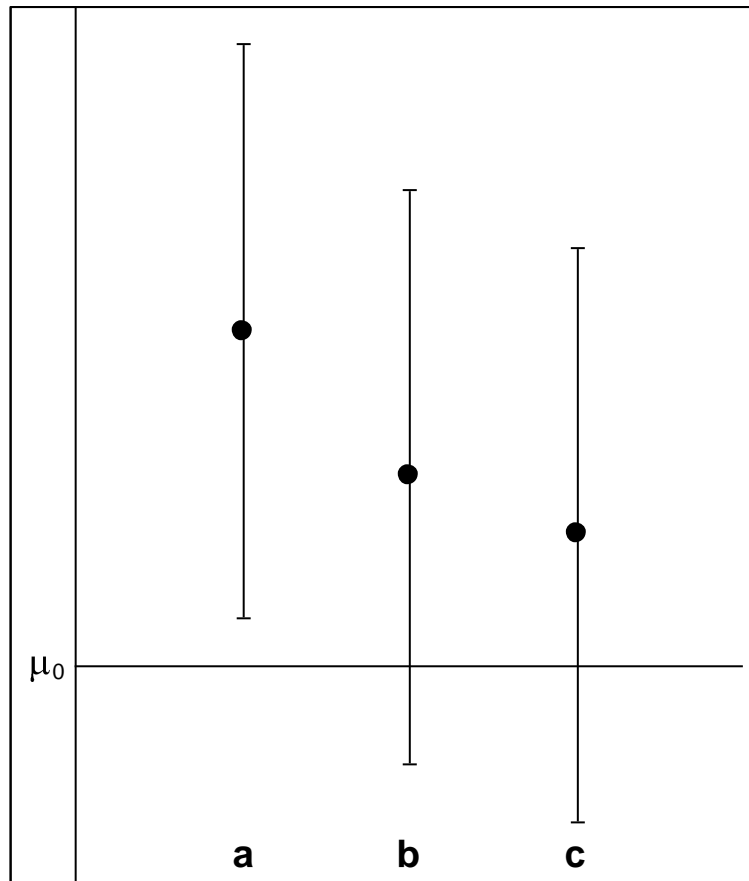


**Fig 3.** The CI function, also called the  $p$  value function. CIs with  $C=60$  and  $C=95$  are shown. For any  $C$ , if a  $C\%$  CI is pictured at a vertical position lined up with  $C$  on the right hand axis, the limits of this CI will lie on the function. The comparison value  $\mu_0 = 27$  is marked with a vertical line, which allows the corresponding two-tail  $p$  value of .151 to be read opposite where the line cuts the function.

The intervals of Figure 2e hint at a curve that relates  $C$  to CI limits. This is the CI function, or  $p$  value function (Poole 1987; Rothman 2002, chapter 6), and is displayed in Figure 3. In the interactive version a slider allows the user to adjust  $C$ , and watch the CI move smoothly up or down, while also changing width. The CI function is the locus of the limits of that CI.

It is well-known that whether a null value  $\mu_0$  lies outside or inside a 95% CI is equivalent to rejecting or not rejecting, respectively, the corresponding null hypothesis at  $\alpha=.05$ , two-tail. If  $\mu_0$  falls exactly at the lower or upper limit of the 95% CI, the two-tail  $p$  value is .05. This relation generalises: If  $\mu_0$  falls at a limit of the  $C\%$  CI, the two-tail  $p$  value is  $(100-C)/100$ .

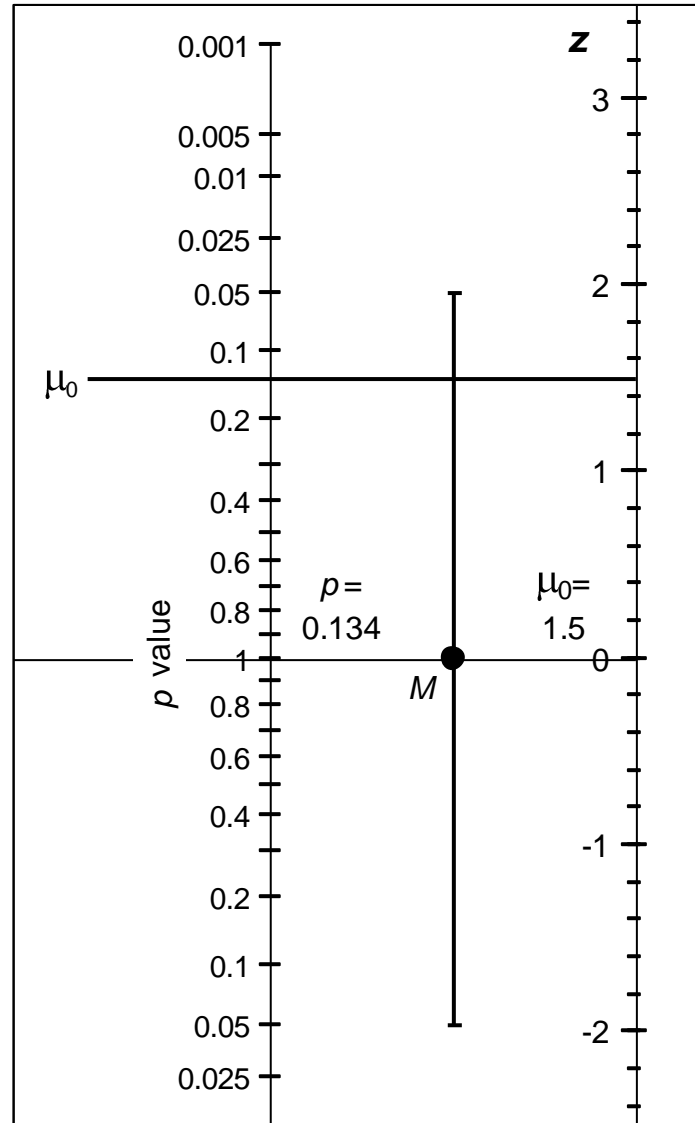
This simple relation between  $C$  and the  $p$  value is shown in Figure 3: For each value of  $C$  on the right hand vertical axis, the corresponding value of  $p$  is shown on the left axis. As Figure 3 illustrates, you can set a  $\mu_0$  of your choice then read off the  $p$  value. With the software, various other features can be explored, for example how much does the CI function change, for a given  $n$ , if  $\sigma$  is assumed known or not known?



**Fig 4.** Three 95% CIs falling in various positions relative to a comparison value  $\mu_0$ . Benchmarks can be used to make eyeball estimates of the two-tail  $p$  value, assuming  $\sigma$  is known or  $n$  is large. Accurate values of  $p$  are .022, .19, and .37, for **a**, **b**, and **c** respectively.

### CONFIDENCE INTERVALS AND $p$ VALUES

Two relationships were noted above: between CI width and  $C$ , and between  $C$  and  $p$  value. Together these allow estimation of two-tail  $p$  value by noting where the 95% CI falls, relative to any  $\mu_0$  value of interest. If the CI falls so that  $\mu_0$  is very close to a CI limit, the  $p$  value is of course close to .05; if the centre of the CI is close to  $\mu_0$ , two-tail  $p$  is close to 1. If  $\mu_0$  is about half way between  $M$  (the centre of the CI) and a limit, it is about at the limit of SE bars so  $C$  is about 68 and the  $p$  value is about .32. Again, it may be worth bearing in mind some simple values corresponding to the four suggested earlier. (Recall that the margin of error is the distance from  $M$  to a limit of the CI; it is the length of one 'arm' of the CI.) If the CI falls so that  $\mu_0$  is:



**Fig 5.** A 95% CI that may be moved to any chosen position relative to the comparison value  $\mu_0$ , and two vertical axes that move with the CI. The right axis shows the  $z$  score for  $\mu_0$ , in relation to  $M$ , the sample mean at the centre of the CI, assuming  $\sigma$  is known. The left axis shows the two-tail  $p$  value for the CI in relation to the comparison value.

- about one third of the way from  $M$  to the limit of the 95% CI, the two-tail  $p$  value is about .50;
- about half the way from  $M$  to the limit, the  $p$  value is about .32;
- about 5/6 of the way from  $M$  to the limit, the  $p$  value is about .10; and
- about 1/3 of the margin of error beyond the limit, the  $p$  value is about .01.

Using those values as benchmarks, it is not hard to estimate the  $p$  values for the configurations shown in Figure 4. For example, in Figure 4a  $\mu_0$  is beyond the limit, but not by as much as 1/3 of the margin of error. It is roughly half way between these benchmarks, so we estimate .03; the accurate value is .022.

Figure 5 illustrates the relationship between two-tail  $p$  value and the position of the 95% CI in relation to  $\mu_0$ . The interactive version allows the  $p$  value to be read off for any chosen position of the CI in relation to  $\mu_0$ . The software also illustrates the relationships for  $\sigma$  assumed known, and for  $\sigma$  not assumed known.

### CONCLUSION

In other articles we have reported studies of how researchers understand CIs, and evidence that various misconceptions are widespread (Belia, Fidler, Williams and Cumming 2005; Cumming, Williams and Fidler, 2004). Cumming and Finch (2005) offered suggestions for inference by eye when two CIs are displayed.

Here I have discussed inference by eye for a single CI. I hope that playing around with the pictures and ideas in this article, and with the software, can help students achieve fuller and more accurate appreciation of CIs, and can also ground their understanding of  $p$  values and statistical significance testing more firmly in their understanding of CIs. After all, pictures that include CIs can be a wonderful basis for inference by eye!

### THE INTERACTIVE SOFTWARE

Figures 1, 3 and 5 come from *ESCI IBI levels of confidence*, a part of *ESCI* (“ess-key”, *Exploratory Software for Confidence Intervals*), which runs under Microsoft Excel. See [www.latrobe.edu.au/psy/esci](http://www.latrobe.edu.au/psy/esci)

**Acknowledgement** The author thanks Fiona Fidler for helpful discussions.

### References

- Belia, S., Fidler, F., Williams, J. and Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, 10(4).
- Cumming, G. and Finch, S. (2005). Inference by eye. Confidence intervals and how to read pictures of data. *American Psychologist*, 60, 170-180.
- Cumming, G., Williams, J. and Fidler, F. (2004). Replication, and researchers’ understanding of confidence intervals and standard error bars. *Understanding Statistics*, 3, 299-311.
- Gardner, M. J. and Altman, D. G. (1987). Using confidence intervals. *The Lancet*, 1(1) for 1987, 746.
- Poole, C. (1987). Beyond the confidence interval. *American Journal of Public Health*, 77, 195-199.
- Rothman, K. J. (2002). *Epidemiology. An introduction*. New York: Oxford University Press.