Statistics 1.

Nikolai Bogduk

Newcastle Bone and Joint Institute

Newcastle, Australia

No conflicts of interest

Confidence Intervals

Much as practitioners abhor statistics, they are a vital tool by which practitioners can protect themselves by being fooled, by others who try to sell them concepts based on data.

This first, brief lesson introduces the concept of confidence intervals of a proportion.

The concept of a confidence interval can be introduced by a hypothetical example in which a speaker announces a 70% success rate in a sample of 10 subjects.

The philosophical and practical question is:

does 7/10 = 70%

One approach to understanding the answer to this question is to appreciate the implications of what the speaker has announced.

Although you might not be a scientist and, therefore, you will not be repeating the experiment, as a practitioner you will adopt the treatment. As a result, each time that you use the treatment you are, in effect, repeating the experiment. The question translates to one more vital to your own professional security.

If the speaker has reported a 70% success rate, does this mean that you will encounter a 70% success rate when you adopt the treatment?

Believing that you will do so implies that each time you and any practitioner adopts the treatment, you will each encounter a success rate of 7/10.

Without indulging any mathematics arguments, the implication should be counterintuitive. Not all patients will be the same. Not all samples of 10 patients will be the same. Therefore, you would expect some degree of variation in outcome, centered on the reported 7/10.

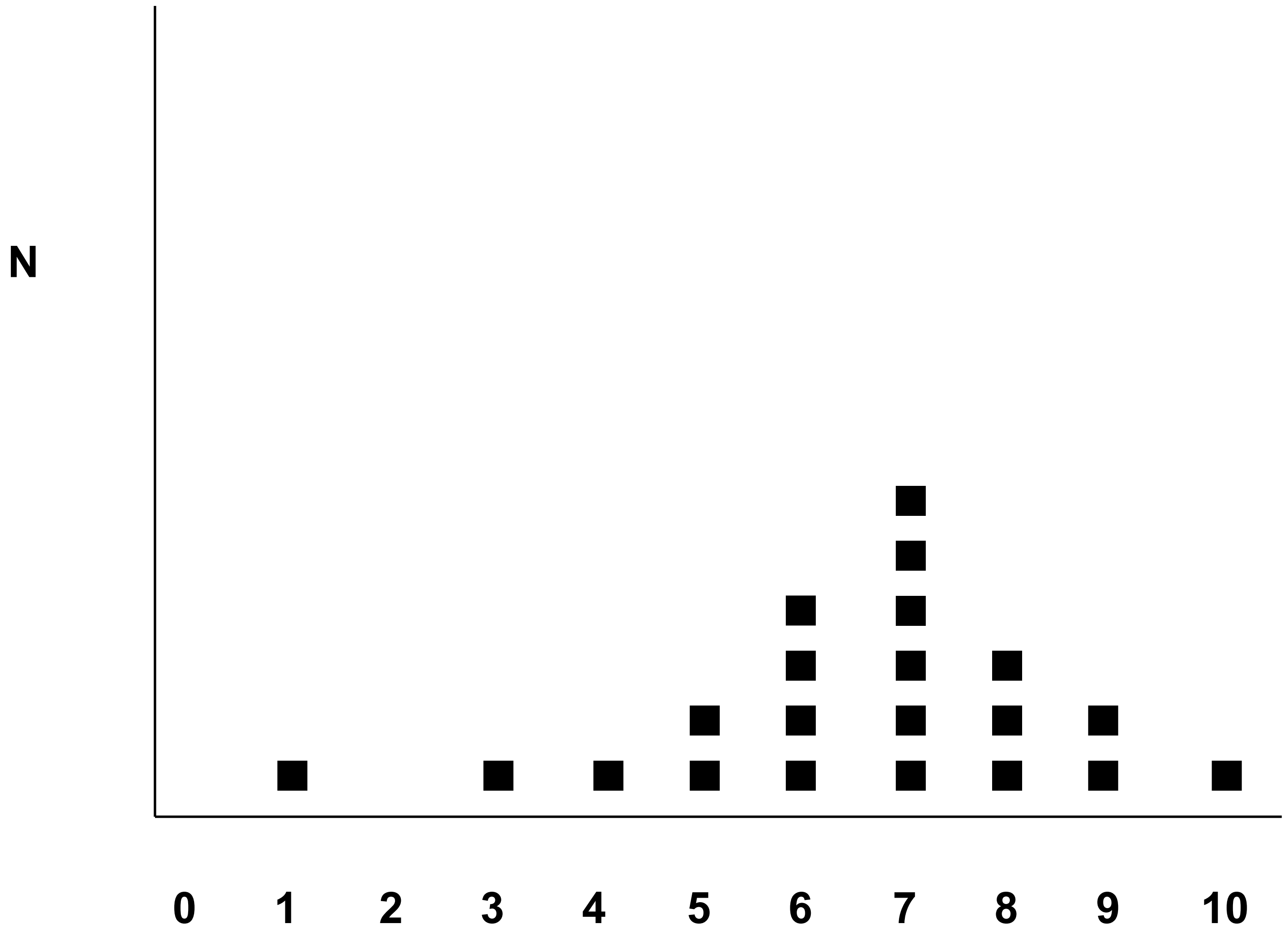
Although initially you might encounter 7/10, on another day you might score a bit less, a bit better, substantially worse, or substantially better.

7 / 10 = 70%

$\frac{7}{10}$ $\frac{7}{10}$ $\frac{7}{10}$ $\frac{7}{10}$ $\frac{7}{10}$ $\frac{7}{10}$ $\longrightarrow$

$\frac{7}{10}$ $\frac{6}{10}$ $\frac{8}{10}$ $\frac{3}{10}$ $\frac{5}{10}$ $\frac{9}{10}$ $\frac{10}{10}$ $\longrightarrow$

Were you to accumulate the outcomes of many, samples of 10 cases, you would find that those outcomes center around the announced 70% but also vary around it. The outcomes will assume a distribution.

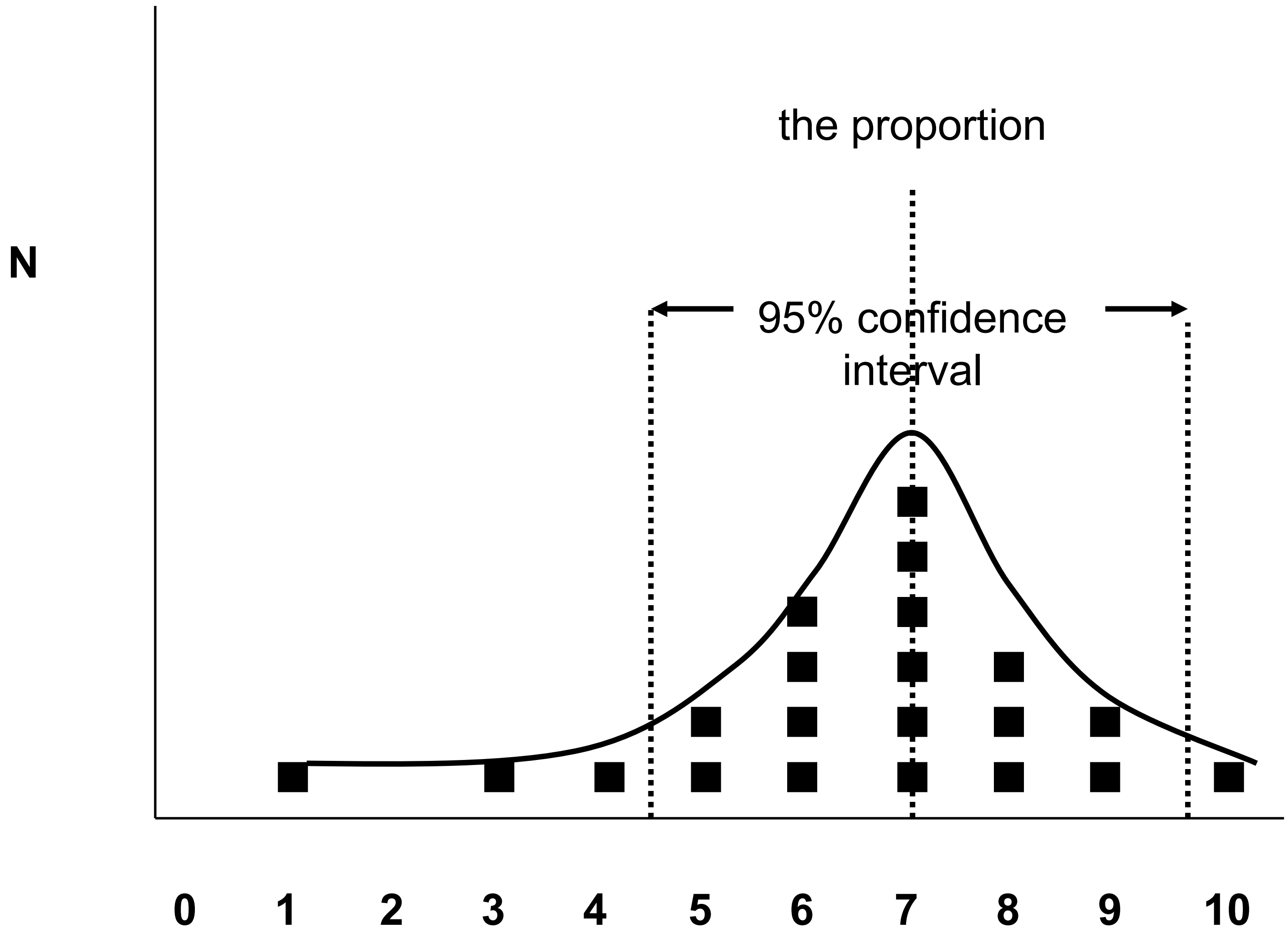

N
0
1
2
3
4
5
6
7
8
9
10

All being well, there will be a range of values within which 95% of repeat observations will fall.

This becomes the 95% confidence interval of the observed proportion.

It means that although some practitioners might encounter the reported proportion, others might encounter lesser or greater values. However, although their individual values might differ, they share in common the range in which those various values fall.

“The range within which 95% of repeat observations should fall”



The 95% confidence intervals of a proportion can be calculated, using mathematics theory.

The formula states that the possible range of values of a proportion is predicated by the magnitude of the observed proportion and the sample size on which it is based.

$$p^* = p \pm 1.96\sqrt{\frac{p(1-p)}{n}}$$

p* = the possible proportion

p = the observed proportion

n = sample size

So, in the case of 7/10, the 95% confidence interval is 42-98%.

7/10 does not equal 70%.

Someone who adopted the treatment could get outcomes as low as 42% or as high as 98%.

95% of people who adopted the treatment would get outcomes somewhere within the range: 42% - 98%.

With respect to advertising, “70%” is not really representative of something that can be as low as 42% or as high as 98%.

This arises because the sample size on which the observation is based is too small.

$7 / 10 \neq 70\%$

$$p^{*} = p \pm 1.96\sqrt{\frac{0.7(1-0.7)}{10}}$$

$p^* = 0.70 \pm 0.28$

$p^* = 0.42 - 0.98 \quad = \quad 42\% - 98\%$

If the sample size is increased to 100, the 95% confidence interval is reduced: to 65% - 75%.

In that case, “70%” is more representative of the possible range of values.

70 / 100

$$p^* = p \pm 1.96\sqrt{\frac{0.7(1-0.7)}{100}}$$

p* = 0.70 ± 0.05

p* = 0.65 - 0.75 = 65% - 75%

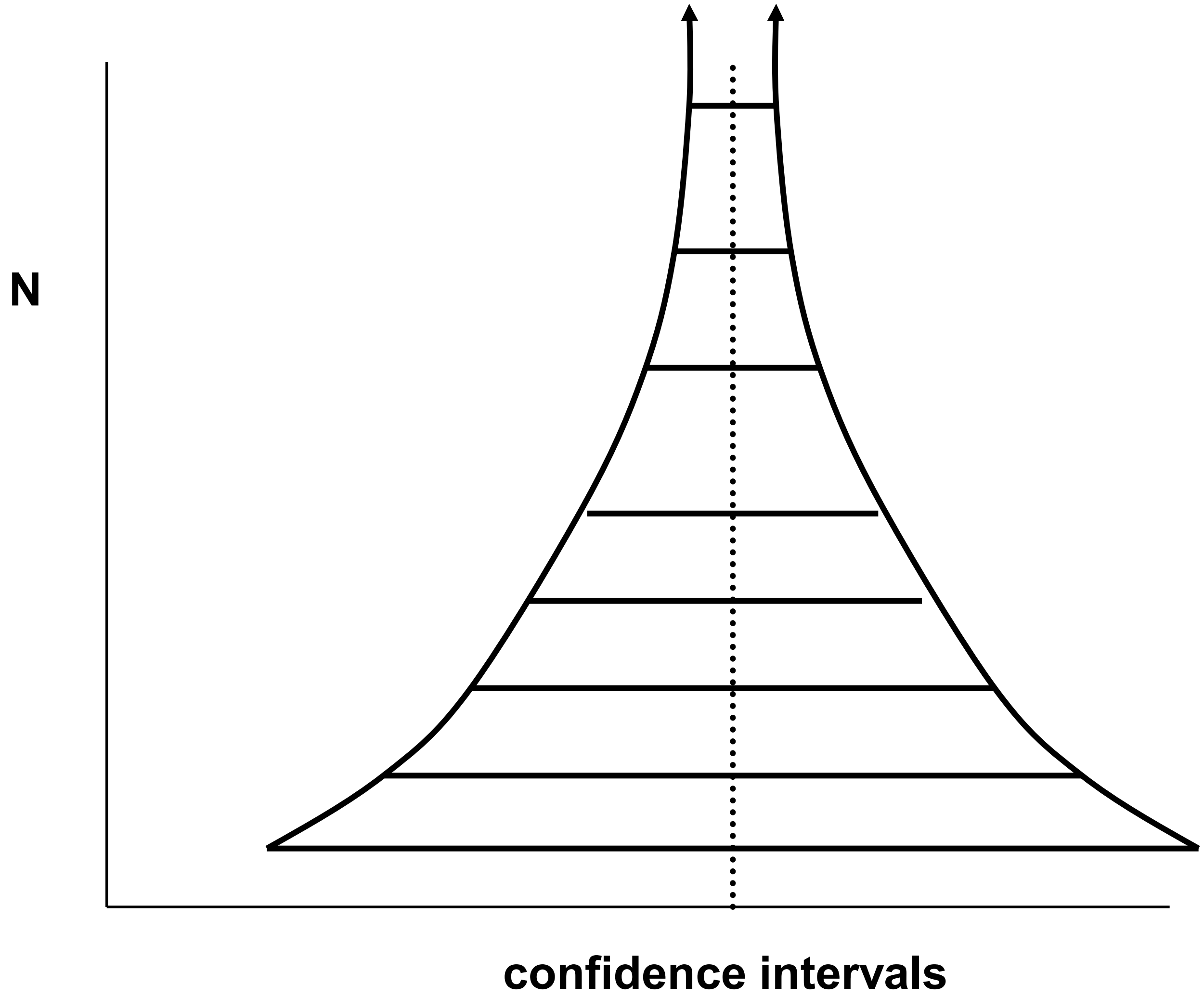
The 95% confidence interval is severely affected by sample size.

Small samples carry large confidence intervals.

That means that conclusions based on small samples are not dependable.

Confidence intervals become progressively less, as the sample size is increased.

Eventually, however, the behaviour of the mathematics is asymptotic. Even though you increase the sample size greatly, the confidence intervals reduce little.


N
confidence intervals

Three practical applications of confidence intervals illustrate their utility in everyday professional life.

The first relates to a claim – say from a controlled trial – that a 70% success rate is better than a 30% success rate.

That conclusion may or more not be true, depending on the sample sizes on which it is based.

For a sample size of 10, the confidence intervals of 70% and of 30% are large, and overlap. Therefore, statistically, 70% and 30% are actually not different.

Even for a sample size of 20, the two proportions are not different.

Only when the sample size reaches 30 do the two proportions become significantly different.

So, ask for the sample size before believing a proportion.

APPLICATIONS. 1 (of three)

70% is better than 30% ?

n = 10

0.70

0.42 —— 0.98

0.02 —— 0.58

0.30

NO

n = 20

0.70

0.50 —— 0.90

0.10 —— 0.50

0.30

NO

n = 30

0.70

0.54 —— 0.86

0.14 —— 0.46

0.30

YES

The second application relates to small numbers. Small numbers can arise in practice in the context, for example, of infection rates, or other complication rates.

Evaluation the proposition that 0.2% is better than 0.5%. This might come from an agent who claims that you can “more than halve your infection rate” if you buy their product.

You protection is to ask: has this been proven? Furthermore, since you know about confidence intervals, you could ask: was the study large enough?

It transpires that a difference between 0.2% and 0.5% cannot be proven with a sample size of 100. Nor can it be proven with a sample of 1,000. The confidence intervals of the small proportions overlap and, so, are not different.

It is not until the sample size is 6,000 that significant difference arises.

So, the short response that you can offer is: do you have a study with a smaple size of 6,000 in each arm that proves that 0.2% is different from 0.5%?

APPLICATIONS.

A complication rate of 0.2% is better than 0.5%

For	n = 100	0.2% =	0.00%	-	1.00%
		0.5% =	0.00%	-	1.90%
For	n = 1,000	0.2% =	0.00%	-	0.48%
		0.5% =	0.10%	-	0.90%
For	n = 4,000	0.2% =	0.00%	-	0.48%
		0.5% =	0.20%	-	0.70%
For	n = 5,000	0.2% =	0.08%	-	0.32%
		0.5% =	0.30%	-	0.70%
For	n = 6,000	0.2% =	0.09%	-	0.31%
		0.5% =	0.31%	-	0.69%

The third application arises when someone reports a zero event. For example they might claim that they had no complications. The question that arises is how confident can you be that if you adopt this intervention that you, too, will encounter zero complications. Reciprocally, the issue is what are the chances that you will encounter a complication, despite the expectation of having no complications.

When dealing with zeros, the formula for confidence intervals can be used to calculate what is the largest proportion whose 95% confidence interval includes zero. That proportion becomes your risk.

It transpires that if safety was determined on a sample of 100, a reported complication rate of 0% is statistically indistinguishable from a rate as high as 3.6%. So, although the reported zero complications is reassuring, the reality is that you should be prepared for a rate up to 3.6%.

For a sample of 200, the expected rate drops to 1.9%.

Only if the sample was 500 could be be confident that your risk is less than 1%.

APPLICATIONS. 3 (of three)

The complication rate is 0%

For	n = 100	0% =	0.0%	-	3.6%
For	n = 200	0% =	0.0%	-	1.9%
For	n = 500	0% =	0.0%	-	0.8%

All proportions come with 95% confidence intervals

$$p^{*} = p \pm 1.96\sqrt{\frac{p(1-p)}{n}}$$