Inferring Sample Size

Inferring Sample Size
Inferring Sample Size
Inferring Sample Size
Inferring Sample Size
Inferring Sample Size
Inferring Sample Size
Inferring Sample Size

by **Colin Foster**

I was watching some student teachers presenting their school-based inquiry projects, and one of the groups didn't report the sample size for their empirical work. They were quoting results using percentages, so I began to try to infer what the sample size might have been.

"83% of the pupils said..."

Hmmm, that sounded very like 5/6, so perhaps their sample size was a multiple of 6?

"33% of the pupils were..."

Well, that seemed to confirm my conjecture. I started to wonder whether they were deliberately not telling me the sample size because it was small? Perhaps they only asked six pupils?

"25% of the pupils..."

Ah, no – OK, well they must have had at least 12 pupils, then. And so it went on, as each new value contributed to my inferences about their sample size.

This got me thinking about why people use percentages to report their findings. Sometimes it may be easier to grasp the meaning of "50%" than of, say, "238 people out of 476" (Is that about half of them?). It can also feel more permissible to round percentages than to round numbers of people. We wouldn't expect to hear: "Thirty-three-and-one-third percent of participants said..."! Rounding can help to give a clearer sense of the size when the actual numbers could be cumbersome. But there is also a danger with percentages of giving the impression of having far more data than you actually do have. Gorard (2003, p. 50) has commented that "I believe that 'percentage' implies 'in every hundred', so as a rule of thumb I recommend only using percentages for cases numbered in hundreds." I think it can sometimes make sense to use percentages with smaller numbers if you are wanting to make comparisons between groups of different sizes, for instance. But if you have just a single sample, then percentages can suggest that you are implicitly generalizing to the population, and then you need some kind of inferential statistics to accompany any percentage statement.

Getting back to the student presentations, if we suppose that the percentages are all rounded to the nearest 1%, what can we deduce about the possible sample size from a statement such as "84% of participants..."? Let's suppose that we can assume for practical reasons that the sample size $N \leq 30$. What might the sample size have been?

With 84% it turns out that there are only two possibilities. Either there were 21 people out of 25, which is exactly 84%, or there were 16 people out of 19, which rounds to 84%. Which percentages would you expect to have the largest number of possibilities and which would you expect to have the fewest? Would you expect there to be any percentages that are impossible to create? Why?

Table 1 (next page) shows the percentages that are possible for each sample size N from 2 to 30, and Table 2 counts up the number of possible sample sizes for each different percentage. We can see that 50% has the largest number of possibilities and there are *no* ways of making 1%, 2%, 49%, 51%, 98% or 99%. We might ask which *sequences* of percentages (for a fixed sample size) will enable me most swiftly and impressively to deduce the sample size? And what occurs with larger samples, say when $N \leq 100$?

Reference

Gorard, S. 2003 *Quantitative Methods in Social Science*, Continuum, London.

Keywords: Percentages; Rounding; Sample size; Statistics.

Author Colin Foster, School of Education, University of Nottingham, Jubilee Campus, Wollaton Road, Nottingham NG8 1BB.
e-mail: **colin@foster77.co.uk**
website: **www.foster77.co.uk**

Table 1 The possible percentages < 100% (rounded to the nearest 1%) for each sample size N

N	Possible percentages (rounded to the nearest 1%)
2	50
3	33, 67
4	25, 50, 75
5	20, 40, 60, 80
6	17, 33, 50, 67, 83
7	14, 29, 43, 57, 71, 86
8	13, 25, 38, 50, 63, 75, 88
9	11, 22, 33, 44, 56, 67, 78, 89
10	10, 20, 30, 40, 50, 60, 70, 80, 90
11	9, 18, 27, 36, 45, 55, 64, 73, 82, 91
12	8, 17, 25, 33, 42, 50, 58, 67, 75, 83, 92
13	8, 15, 23, 31, 38, 46, 54, 62, 69, 77, 85, 92
14	7, 14, 21, 29, 36, 43, 50, 57, 64, 71, 79, 86, 93
15	7, 13, 20, 27, 33, 40, 47, 53, 60, 67, 73, 80, 87, 93
16	6, 13, 19, 25, 31, 38, 44, 50, 56, 63, 69, 75, 81, 88, 94
17	6, 12, 18, 24, 29, 35, 41, 47, 53, 59, 65, 71, 76, 82, 88, 94
18	6, 11, 17, 22, 28, 33, 39, 44, 50, 56, 61, 67, 72, 78, 83, 89, 94
19	5, 11, 16, 21, 26, 32, 37, 42, 47, 53, 58, 63, 68, 74, 79, 84, 89, 95
20	5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95
21	5, 10, 14, 19, 24, 29, 33, 38, 43, 48, 52, 57, 62, 67, 71, 76, 81, 86, 90, 95
22	5, 9, 14, 18, 23, 27, 32, 36, 41, 45, 50, 55, 59, 64, 68, 73, 77, 82, 86, 91, 95
23	4, 9, 13, 17, 22, 26, 30, 35, 39, 43, 48, 52, 57, 61, 65, 70, 74, 78, 83, 87, 91, 96
24	4, 8, 13, 17, 21, 25, 29, 33, 38, 42, 46, 50, 54, 58, 63, 67, 71, 75, 79, 83, 88, 92, 96
25	4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64, 68, 72, 76, 80, 84, 88, 92, 96
26	4, 8, 12, 15, 19, 23, 27, 31, 35, 38, 42, 46, 50, 54, 58, 62, 65, 69, 73, 77, 81, 85, 88, 92, 96
27	4, 7, 11, 15, 19, 22, 26, 30, 33, 37, 41, 44, 48, 52, 56, 59, 63, 67, 70, 74, 78, 81, 85, 89, 93, 96
28	4, 7, 11, 14, 18, 21, 25, 29, 32, 36, 39, 43, 46, 50, 54, 57, 61, 64, 68, 71, 75, 79, 82, 86, 89, 93, 96
29	3, 7, 10, 14, 17, 21, 24, 28, 31, 34, 38, 41, 45, 48, 52, 55, 59, 62, 66, 69, 72, 76, 79, 83, 86, 90, 93, 97
30	3, 7, 10, 13, 17, 20, 23, 27, 30, 33, 37, 40, 43, 47, 50, 53, 57, 60, 63, 67, 70, 73, 77, 80, 83, 87, 90, 93, 97

Table 2 The number of times each possible percentage (rounded to the nearest 1%) occurs, with a sample size $N \leq 30$ and all percentages < 100%

Possible percentages (rounded to the nearest 1%)	No. of times those percentages occur
1, 2, 49, 51, 98, 99	0
34, 66	1
3, 16, 84, 97	2
6, 9, 12, 26, 28, 37, 39, 61, 72, 74, 87, 91, 94	3
5, 15, 18, 19, 22, 23, 24, 31, 32, 35, 41, 42, 45, 46, 47, 53, 54, 55, 58, 59, 62, 65, 68, 69, 76, 77, 78, 81, 82, 85, 95	4
8, 10, 11, 21, 27, 30, 36, 44, 48, 52, 56, 64, 70, 73, 79, 89, 90, 92	5
4, 7, 13, 14, 20, 29, 40, 43, 57, 60, 63, 71, 80, 86, 88, 93, 96	6
17, 25, 38, 75, 83	7
33, 67	10
50	15