Statistical significance

A sample-based estimate is more or less close to the ‘unknown’ population value being measured. The size of the deviation d depends on:
- sample size (n)
- percentage observed in the sample (p)
- level of confidence chosen (z)

In this report we will use a level of confidence of 90%. The nomogram on the next page gives the confidence levels for observed percentages and sample sizes.
For instance, in a survey of 1,000 respondents, 20% said ‘yes’ to a certain question.
The entry in the table on the next page, at row n = 1,000 and with column percentage of 20, shows δ to be 2.1%. This implies that there is a nine in ten chance that the true population value lies between 17.9% and 22.1% (20 ± 2.1, at a confidence level of 90%). Hence, that there is a five percent probability that the real value is larger than 22.1% and five percent probability that it is smaller than 17.9%. In another example, say 2% of the sample of 2,000 people had been a victim of a particular crime in the last year. There would be a 90% chance that the true level of victimisation lies between 2.5% and 1.5% (2 ± 0.5).
When there is an average victimisation rate for all countries of 5%, for instance, then a value from an individual survey with a sample of 2,000 of more than 0.8% higher or lower than the average will be statistically significant at the 90% level. Where the overall victimisation rate is 2% say, deviations of 0.5% would be significant. (Thus, in absolute size, the standard error is smaller the less frequently a crime occurs, but proportionately it is much larger.) When the sample is 1,000 (of women only for example), deviations from an overall average of 5% of more than 1.1% will be significant, and with an average of 2%, deviations of 0.7% will be.

The formula which is used for calculating d at a confidence level of 90% is:

$$\delta = 1.65 \times \sqrt{p\frac{(100-p)}{n}}$$

When a research population is finite, the deviation d is smaller because the formula is multiplied by:

$$\frac{N-n}{N-1}$$

in which N is the population size.

Nomogram for a level of confidence of 90%

Sample size	Percentage observed										
	2	5	10	15	20	25	30	35	40	45	50
	98	95	90	85	80	75	70	65	60	55	50
25	4.6	7.2	9.9	11.8	13.2	15.2	15.1	15.7	16.1	16.4	16.5
50	3.3	5.1	7.0	8.3	9.3	10.7	10.7	11.1	11.4	11.6	11.6
100	2.3	3.6	4.9	5.9	6.6	7.6	7.5	7.9	8.1	8.2	8.2
200	1.6	2.5	3.5	4.2	4.7	5.4	5.3	5.6	5.7	5.8	5.8
300	1.3	2.1	2.9	3.4	3.8	4.4	4.4	4.5	4.7	4.7	4.8
400	1.2	1.8	2.5	2.9	3.3	3.8	3.8	3.9	4.0	4.1	4.1
500	1.0	1.6	2.2	2.6	2.9	3.4	3.4	3.5	3.6	3.7	3.7
600	0.9	1.5	2.0	2.4	2.7	3.1	3.1	3.2	3.3	3.3	3.4
700	0.9	1.4	1.9	2.2	2.5	2.9	2.9	3.0	3.0	3.1	3.1
800	0.8	1.3	1.7	2.1	2.3	2.7	2.7	2.8	2.9	2.9	2.9
900	0.8	1.2	1.6	2.0	2.2	2.5	2.5	2.6	2.7	2.7	2.7
1,000	0.7	1.1	1.6	1.9	2.1	2.4	2.4	2.5	2.5	2.6	2.6
1,200	0.7	1.0	1.4	1.7	1.9	2.2	2.2	2.3	2.3	2.4	2.4
1,400	0.6	1.0	1.3	1.6	1.8	2.0	2.0	2.1	2.2	2.2	2.2
1,600	0.6	0.9	1.2	1.5	1.6	1.9	1.9	2.0	2.0	2.0	2.1
1,800	0.5	0.8	1.2	1.4	1.6	1.8	1.8	1.9	1.9	1.9	1.9
2,000	0.5	0.8	1.1	1.3	1.5	1.7	1.7	1.8	1.8	1.8	1.8
3,000	0.4	0.7	0.9	1.1	1.2	1.4	1.4	1.4	1.5	1.5	1.5
4,000	0.4	0.6	0.8	0.9	1.0	1.2	1.2	1.2	1.3	1.3	1.3
6,000	0.3	0.5	0.6	0.8	0.8	1.0	1.0	1.0	1.0	1.1	1.1
8,000	0.3	0.4	0.6	0.7	0.7	0.8	0.8	0.9	0.9	0.9	0.9
10,000	0.2	0.4	0.5	0.6	0.7	0.8	0.8	0.8	0.8	0.8	0.8
20,000	0.2	0.3	0.3	0.4	0.5	0.5	0.5	0.6	0.6	0.6	0.6
30,000	0.1	0.2	0.3	0.3	0.4	0.4	0.4	0.5	0.5	0.5	0.5
40,000	0.1	0.2	0.2	0.3	0.3	0.4	0.4	0.4	0.4	0.4	0.4