

Estimation of Margin of Error Based on Anticipated Sample Size

This teaching note might come in useful for faculty colleagues who wish to discuss the possible Margin of Error which might result from an anticipated sample of n people – the limerick below is intended to make it more fun!

In sampling, such as in political polls, we often need to compute Margin of Error (MoE) or Half-width (h) of a Confidence Interval, so that we know what percentage error we have on each side. For example, if we find that 51 percent of those surveyed approve of a particular policy, when we extrapolate that to the population, we must recognize that we could be wrong by a few percentage points in either direction, say ± 3 percentage points in either direction. This is the MoE, and is expressed as a proportion, say 0.03.

Margin of Error or Half-width of a Confidence Interval is defined as $h = \pm z * s_p$ where $s_p = \sqrt{\frac{[p*(1-p)]}{n}}$

This can easily be done after the study is completed because we know all the terms in the equation. However, we often need to plan our sample size, or the minimum number of people we need to survey, to get a particular MoE. In such cases, the computation is not a trivial one, because we do not know either p or n . Fortunately, the numerator $[p * (1-p)]$ or $[P * (1-P)]$ exhibits certain properties. By using calculus, we can show that, regardless of what the value of p is, the numerator can never be greater than 0.25, and so we can simply use that value in the numerator, and then either compute h for a particular sample size, n , or compute n for a determined h .

The statement in the preceding sentence can be proved in the following way:

$$\frac{\partial [P*(1-P)]}{\partial P} = \frac{\partial [P-P^2]}{\partial P} = 1-2P \quad [\text{Note that this is mathematically the same as: } -(2P-1)]$$

The maximum value of $[P * (1-P)]$ occurs when the first derivative = 0 (assuming that the second derivative is negative, which it is in this case, being $- 2$).

So, putting $(1-2P) = 0$ yields $P = 0.5$

Therefore, the maximum value of $[P * (1-P)]$ occurs when $P = 0.5$; so the maximum value of $[P * (1-P)] = 0.25$

$$\text{Thus, } s_p = \sqrt{\frac{[P*(1-P)]}{n}} = \sqrt{\frac{0.25}{n}} = \sqrt{\frac{1}{4n}} \quad \text{So, } h = \pm z * s_p = \pm z * \sqrt{\frac{1}{4n}}$$

Now, this entire analysis may be expressed as a mathematical limerick:

**Derivative of P times (one minus P),
Where P is a number, you see,
Is *negative* ($2P$ less one).
Which, when equated to none
Yields $P = 0.5$, you see.**

**So, P times (one minus P),
If P is a probability,
Can never be more
Than one over four,
Regardless of the value of P !**

**Thus, maximum MoE,
What Margin of Error might be,
If you interview n
Women or men;
Is [root of (one by $4n$)] times z .**