

How to arrive at a sample size?

Intuitively, larger the sample size we can be sure about our conclusion. In order to arrive at a population parameter, we sample and study the sample parameters.

Say a new change is rolled out on Google.com and we study the change over 100 samples. Say 100 days. Ideally we need to study comparable samples and given that there might be seasonality across days of week, metric to study would be say number of saves per visit in Google maps.

So across 100 samples, saves per visit is recorded. The samples have a mean as μ_s with standard deviation as S.D. σ_s . Standard error is $\sigma_s/\sqrt{n} = \sigma_s/\sqrt{100}$. Higher the sample size, lower the standard error.

Based on this what we can conclude about population parameter?

The population mean would lie between $\mu_s \pm 1 \frac{\sigma_s}{\sqrt{n}}$ and we can be 65% confident about it. within $\mu_s \pm 2 \frac{\sigma_s}{\sqrt{n}}$ and we can be 95% confident about it. Say after studying 10000 people and we arrived at % favouring Party A is 63%. (Every person was given a scale between 1 & 100% to choose from) and S.D. = 12%. Standard error = $\frac{12\%}{\sqrt{10000}} = 0.12\%$

So we are 65% confident that overall % favouring is $\mu_s \pm 1 SE = 63\% \pm 0.12\%$ between 62.88% and 63.12%. we can be 95% that population mean is between $63\% \pm 2(0.12\%)$ that is 62.76% and 63.24%. If we want to be 95% for smaller margin? Increase sample size.