- As sample size increase, the variability of the sample statistic tends to decrease and the sample statistics tend to be closer to the true value of the population parameter.
    - The larger the sample size, the less the variability in the statistics, and the less uncertainty in the estimates. Larger samples supply more information.

Interval estimation

- **Interval estimate** - Gives a range of plausible values for a population parameter.
    - A common form for an interval estimate is **statistic +/- margin of error**.
        - Where the margin of error reflects the precision of the sample statistic as a point estimate for the parameter.
    - e.g. A sample of n=755 mobile phone users found an average number of 41.5 text sent and received per day with a margin of error of 12.2. The interval estimate is 29.3 - 53.7.

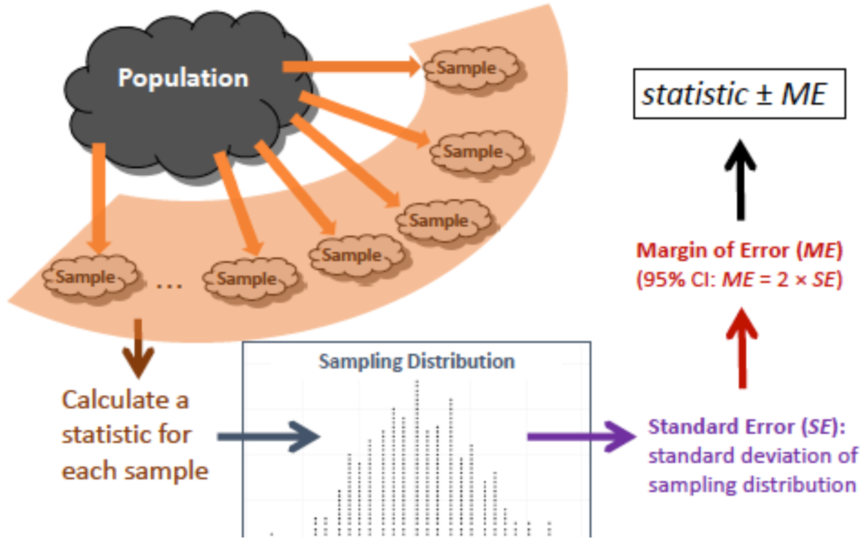
- The margin of error can be determined by the spread of the sampling distribution; the standard error (SE).
    - The more spread of the sampling distribution, the higher the margin of error.

- **Confidence interval** - The interval computed for a parameter from sample data by a method that will capture the parameter for a specified proportion of all samples.
    - **Confidence level** - The success rate (proportion of all samples who intervals contain the parameter).
        - A 95% confidence interval will contain the true parameter for 95% of all samples.
- If you had access to the sampling distribution, how would you find a margin of error to ensure that intervals of the form captured the parameter for 95% of all samples?
    - **Statistic +/- margin of error**.
- If the sampling distribution is relatively symmetric and bell-shaped, a 95% confidence interval can be estimated using **statistic +/- 2 x SE**.
    - In other words, **margin of error = 2 x standard error**.

> A survey of 1,502 Americans finds that 86% consider the economy a "top priority". The standard error for this statistic is 0.01.
>
> • What is the 95% confidence interval for the true proportion of all Americans that considered the economy a "top priority"?
>
> statistic $\pm$ 2 $\times$ SE = 0.86 $\pm$ 2 $\times$ 0.01 = 0.86 $\pm$ 0.02
>
> Hence, the 95% confidence interval for the true population proportion is (0.84, 0.88)

- The confidence interval yields a range of plausible values for the true parameter.
    - 95% of all samples yield intervals that contain the true parameter, i.e. roughly 1 in 20 samples will yield an interval that doesn't contain the true parameter.
    - "*We are 95% confident that the true proportion of all Americans that considered the economy a 'top priority' is between 0.84 and 0.88*."

How to get a confidence interval



To create a 95% confidence interval for a parameter:

1. Take many random samples from the population, compute the sample statistic for each sample.
2. Build up the sampling distribution using all the sample statistics.
3. Compute the SE as the standard deviation of the sampling distribution.
4. Use the estimated statistic +/- 2 x SE (= ME)

Bootstrapping

- Bootstrapping determines how much our sample statistic varied if we only have one sample.
- To create a sampling distribution, we think of the population as just many copies of our sample, and we can just take repeated random samples from this artificial population.
    - As long as the original sample is random, the artificial sample will be representative of the population.
    - In practice, we don’t actually make infinite copies of the sample, but we do essentially the same thing by sampling with replacement from the one sample we do have.
        - Each sample we create is referred to as a **bootstrap sample**.

- **Bootstrap sample** - A random sample taken with replacement from the original sample, and of the same size as the original sample.
- **Bootstrap statistic** - A statistic computed on a bootstrap sample.
- **Bootstrap distribution** - The distribution of many bootstrap statistics.