



Basic principles:

- Every time you see a p-value, someone tested a hypothesis
- A statistic is a number (i.e. a proportion, an average or a disease rate) you calculate from your data sample.
- The statistic is an attempt to describe the truth (a parameter) about a larger population using your sample.
- Hypothesis testing allows a researcher to decide whether a difference between numbers is more likely due to random variation or a true difference

The CLT is very powerful, but it has two limitations:

- 1) it depends on a large sample size
- 2) to use it, we need to know the standard deviation of the population.

IMPORTANT we usually don't know the standard deviation of the population so we use the standard deviation of our sample

t-distribution

- It is a little **shorter** and **wider** than a normal distribution
- The shape of t depends on the **size of the sample**...the larger the sample size, the more confident we can be that 's' is near sigma, and the closer t gets to Z

IMPORTANT n-1 (degrees of freedom).

Because it is not normal, the t-distribution does not follow the 68-95-99 rule, but we can use **t-tables** or **computer programs** to estimate the area-under-the-curve (probability)





To test a sample of normal continuous data, we need:

- 1. An **expected value** = the **population or true mean**
- 2. An **observed mean** = the **average of your sample**
- 3. A measure of spread: standard error
- 4. Degrees of freedom (df)

$$T = \frac{(\bar{X} - \mu)}{\frac{S}{\sqrt{n}}}$$

$t = \frac{(\text{observed mean} - \text{expected mean})}{SE}$

