Inference for sampling

On Tuesday, the bottles of Arizona Iced Tea filled in a plant were supposed to contain an average of 20 ounces of iced tea. Quality control inspectors sampled 50 bottles at random from the day's production. These bottles contained an average of 19.6 ounces of iced tea.

Population:
    Bottles of Arizona Iced Tea
Parameter of interest:
    Average iced tea per bottle, 20 ounces.
Sample:
    The 50 sampled bottles
Statistic:
    Iced tea per bottle, 19.6 ounces.

The Gallup Poll asked a random sample of 515 US adults whether they believe in ghosts. Of the respondents, 160 said yes.

Population:
    All US adults
Parameter of interest:
    proportion p of US adults who believe in ghosts
Sample:
    the 515 people who were interviewed.
Statistic:
    proportion $\hat{p} = 160/515 \approx 0.31$ of the sample who say they believe in ghosts

A **parameter** is a number that describes a characteristic of the population.

A **statistic** is a number that describes a characteristic of the sample.

Population parameter	Population mean μ	Population standard deviation σ	Population variance σ^2	Population proportion p
Sample Statistic	Sample mean $\bar{x}$	Sample standard deviation s_x	Sample variance s_x^2	Sample proportion $\hat{p}$

pop.sd uses `÷n`, but **sample standard deviation** (sd) of $(x_1, x_2, \ldots, x_n)$ uses `÷(n-1)`

$$\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2} = \sqrt{\frac{1}{n-1}[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2]}$$

Example. Compute sample standard deviation of the dataset (1, 2, 3).

Answer:
sample mean = $\dfrac{1+2+3}{3} = 2$,

standard deviation =

$$\frac{1}{\sqrt{3-1}}\|(1,2,3)-2\| = \frac{1}{\sqrt{2}}\|(-1,0,1)\| = \frac{1}{\sqrt{3-1}}\|(1,2,3)-2\| = \frac{1}{\sqrt{2}}[(-1)^2+0^2+1^2] = \frac{\sqrt{2}}{\sqrt{2}} = 1$$

One large, well-designed random sample is representative of the population:

$$\bar{x} \approx \mu \qquad s_x \approx \sigma \qquad \hat{p} \approx p \qquad \text{when sample size } n \text{ is big}$$

Law of Large Numbers & Central Limit Theorem: Take many SRS of the same moderate size. Then the average of these sample statistics approach the population parameter.

6 - Design of studies

Sample survey is a kind of...

Observational study observes individuals and measures variables of interest, but does not attempt to influence the response.
If well-designed, this produces evidence of **association or no association**.

Experiment deliberately imposes treatments on individuals to measure responses.
If well-designed, this produces evidence of **causation or no causation**.
Individuals here are called **experimental units** or **subjects** if they are human.

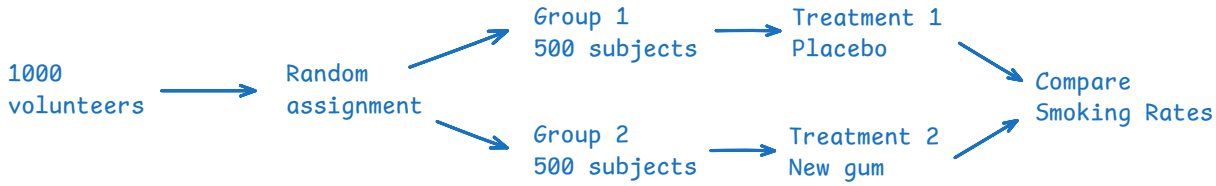
> To perform the study, researchers call 3040 US adults and ask them to rate current US economic conditions and whether the US economy is getting better or worse.

Observational study

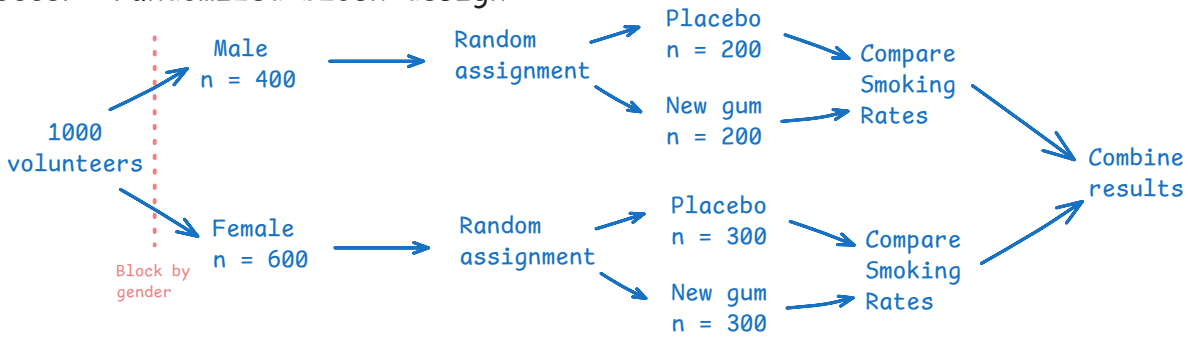
> A company wants to test the effectiveness of a new nicotine gum developed to help people quit smoking. The company identifies ten adults who are heavy smokers. Subjects are given the new gum. After two months, a significant number of subjects have quit smoking.

Experiment. Critique:
- Needs more subjects.
- Add comparison with a control group (dummy treatment).
- Give control group placebo/sugar gum (single-blind experiment).
- And don't let experimenter know which gum is assigned (double-blind experiment).
- Control for gender and other factors to minimize them from being lurking or confounding variables.

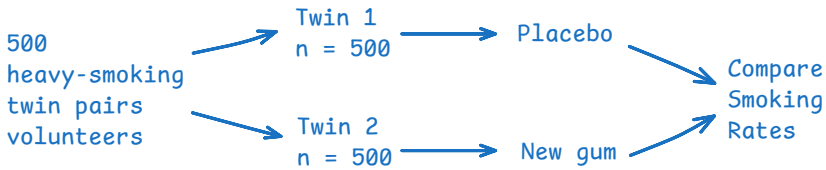
First improved experiment: completely randomized design



Better: randomized block design



Also better: matched pairs



Response differences between the groups in an experiment that are too large to be explained by chance variation in the random assignment, we say the result is **statistically significant**.