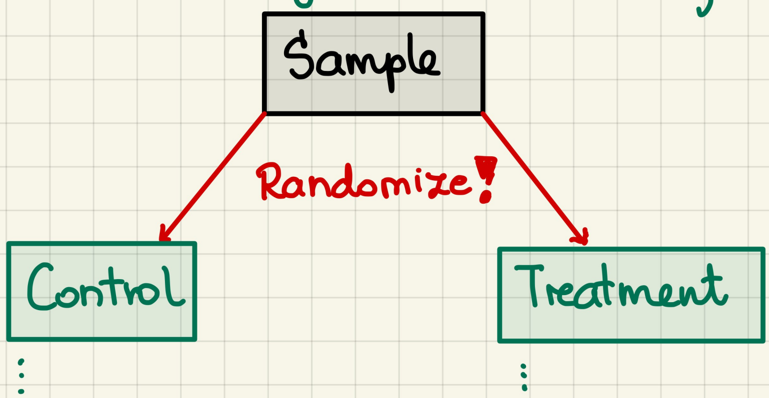
M358K: November 29th, 2021.

Statistical Inference for Two Means.

Inspiration.

Consider an experiment for testing whether a new drug works better than an existing drug.

"old drug" vs. "new drug"



μ_1 ... mean for the (sub)population #1

μ_2 ... mean for the (sub)population #2

Our focus is on $\boxed{\mu_1 - \mu_2}$.

Goals:
- confidence intervals
- hypothesis testing

We should @ the statistic: $\boxed{\bar{X}_1 - \bar{X}_2}$

Assumptions:
- both population dist'ns are normal (or the samples are sufficiently large)
- the two samples are independent

For $i = 1, 2$, we have

$$\bar{X}_i \sim \text{Normal}\left(\text{mean} = \mu_i, \ \text{sd} = \frac{\sigma_i}{\sqrt{n_i}}\right) \text{ w/ } n_i \text{ ... sample size}$$

$$\Rightarrow \boxed{\bar{X}_1 - \bar{X}_2 \sim \text{Normal}\left(\text{mean} = \mu_1 - \mu_2, \ \text{sd} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)}$$

independent samples

$$\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Caveat: We don't know σ_1 and σ_2.

We use:

$$\boxed{\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}} \sim t(df = \underline{\min(n_1,n_2)-1})$$

using the t-tables

If using R, the t.test command will automatically calculate the correct number of degrees of freedom using Welch's t-test.

Confidence Intervals.

C... confidence level

pt.estimate ± margin·of·error

t^* ·std error

$$\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

$$\boxed{\bar{x}_1-\bar{x}_2 \pm t^*\cdot\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$$

$$t^* = qt\left(\frac{1+C}{2},\ \boxed{df = \min(n_1,n_2)-1}\right)$$

Hypothesis Testing.

$H_0: \mu_1 = \mu_2$ (no effect) vs. $H_a: \begin{cases} \mu_1 < \mu_2 \\ \mu_1 \neq \mu_2 \\ \mu_1 > \mu_2 \end{cases}$

The test statistic, under the null, is

$$T = \frac{\bar{X}_1 - \bar{X}_2 - 0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t(df = \min(n_1, n_2) - 1)$$