

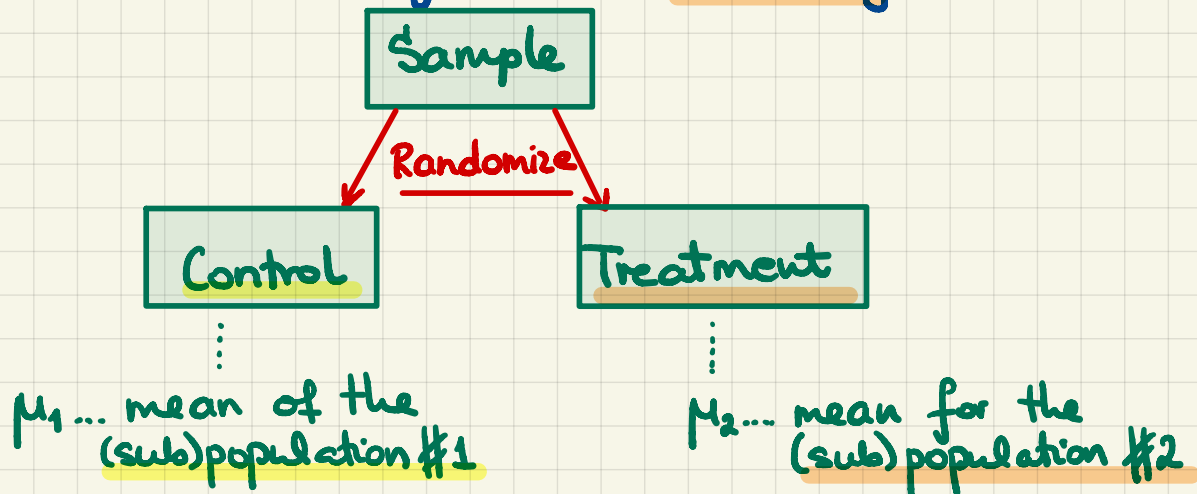
M358K : November 27th, 2023.

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"



Our focus:

$$\mu_1 - \mu_2$$

Goals:

- confidence intervals
- hypothesis testing

We should look at the statistic:

$$\bar{X}_1 - \bar{X}_2$$

Assumptions:

- both population dist's are normal or the samples are sufficiently large.
- the two samples are independent.

For $i = 1, 2$, we have

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

$\Rightarrow$

$$\bar{X}_1 - \bar{X}_2 \sim \text{Normal} \left(\text{mean} = \underline{\mu_1 - \mu_2}, \text{sd} = \underline{\quad ? \quad} \right)$$

$$\text{Var}[\bar{X}_1 - \bar{X}_2] = \text{Var}[\bar{X}_1] + \text{Var}[\bar{X}_2] = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

↑
independence

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

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

Caveat: We usually don't know σ_1 and σ_2 .

We use:

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

↑ using the t-tables

If using R, with entire data sets, the t-test command will automatically calculate the correct number of degrees of freedom using the Welch t-test.

Confidence Intervals.

C... confidence level

pt. estimate ± margin of error

↓

$\bar{x}_1 - \bar{x}_2$

±

$t^* \cdot \text{stderror}$

$t^* = qt((1+C)/2, df = \min(n_1, n_2) - 1)$

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

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

Hypothesis Testing.

$$H_0: \mu_1 = \mu_2 \text{ (no effect)}$$

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:

$$T = \frac{\bar{X}_1 - \bar{X}_2 - 0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \sim t(df = \min(n_1, n_2) - 1)$$