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"

Sample

Randomize

Control — Treatment

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

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

Our focus: $\boxed{\mu_1 - \mu_2}$

Goals:

- confidence intervals
- hypothesis testing

We should look at 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

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

$\Rightarrow$

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

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

independence

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

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

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

We use:

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

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 $\pm$ margin-of-error

$t^* \cdot$ stderror

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

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

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

Hypothesis Testing.

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

The Test Statistic, under the null, is:

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