Recap: One-sample vs. Two-sample t-test

Dataset – research question – design

One-sample t-test / Paired t-test - one-sample t-test on difference (Captopril)

- Is there an effect of administering captopril on the blood pressure of patients with hypertension?
- 15 persons at random
- blood pressure before and after treatment on the same patient (pretest/posttest, paired measurements)

$\Rightarrow$ difference $Y = \text{BP}_{\text{after}} - \text{BP}_{\text{before}}$.

Two-sample t-test (Armpit microbiome)

- Is there an effect of microbiome transplantation on the relative abundance of *Staphylococcus*?
- 20 persons at random
- divided over 2 **independent** groups (10 placebo, 10 transplantation).

Translating the research question into an effect size in the population

$$\delta = \mu - \mu_0$$

- difference between the population mean μ and a pre-specified value μ_0
- for captopril / paired data
  - μ: average difference after - before in the population
  - $\mu_0 = 0$

$$\delta = \mu_1 - \mu_2$$

- difference between group means

Point estimator

$$\hat{\delta} = \hat{\mu} - \mu_0, \qquad \hat{\mu} = \bar{Y}$$

$$\hat{\delta} = \hat{\mu}_1 - \hat{\mu}_2 = \bar{Y}_1 - \bar{Y}_2$$

Model / distribution

$$Y \sim N(\mu, \sigma^2) \implies \bar{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$Y_{ij} \sim N(\mu_j, \sigma^2) \implies \bar{Y}_j \sim N\left(\mu_j, \frac{\sigma^2}{n_j}\right)$$

$$\implies \bar{Y}_1 - \bar{Y}_2 \sim N\Big(\mu_1 - \mu_2,\ \sigma^2\left(\tfrac{1}{n_1} + \tfrac{1}{n_2}\right)\Big)$$

Assumptions

- Representative sample: $E[\bar{Y}] = \mu$
- Normality
- Independence (between subjects)

- Representative sample: $E[\bar{Y}_j] = \mu_j$
- Normality
- Independence **and** equal variances between the 2 groups ($\sigma_1^2 = \sigma_2^2 = \sigma^2$)

Variance/SE estimator

$$\hat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}$$

$$SE_{\bar{Y}} = \frac{S}{\sqrt{n}}$$

$$S_p^2 = \frac{\sum_{j=1}^{2}\sum_{i=1}^{n_j}(Y_{ij} - \bar{Y}_j)^2}{n_1 + n_2 - 2}$$

$$SE_{\bar{Y}_1 - \bar{Y}_2} = S_p\sqrt{\tfrac{1}{n_1} + \tfrac{1}{n_2}}$$

Pivotal statistic

$$T = \frac{\bar{Y} - \mu}{SE_{\bar{Y}}} \sim t_{n-1}$$

$$T = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{SE_{\bar{Y}_1 - \bar{Y}_2}} \sim t_{n_1+n_2-2}$$

Confidence interval

$$\bar{Y} \pm t_{n-1,\alpha/2}\, SE_{\bar{Y}}$$

$$(\bar{Y}_1 - \bar{Y}_2) \pm t_{n_1+n_2-2,\alpha/2}\, SE_{\bar{Y}_1 - \bar{Y}_2}$$

Hypothesis test

$$H_0 : \mu = \mu_0 \quad \text{vs} \quad H_1 : \mu \neq \mu_0$$

(or one-sided: $\mu < \mu_0$ or $\mu > \mu_0$)

Test statistic under H_0:

$$T = \frac{\bar{Y} - \mu_0}{SE_{\bar{Y}}} \sim t_{n-1}$$

for captopril $\mu_0 = 0$

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs} \quad H_1 : \mu_1 - \mu_2 \neq 0$$

(or one-sided: $\mu_1 - \mu_2 < 0$ or $\mu_1 - \mu_2 > 0$)

Test statistic under H_0:

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{SE_{\bar{Y}_1 - \bar{Y}_2}} \sim t_{n_1+n_2-2}$$

- Both statistics are a **signal-to-noise ratio**: $T = \hat{\delta}/SE_{\hat{\delta}}$, with $\hat{\delta} = \bar{Y} - \mu_0$ (one-sample) and $\hat{\delta} = \bar{Y}_1 - \bar{Y}_2$ (two-sample).
- The p-value is the probability to obtain, under the null hypothesis (H_0), in a random sample with the same design, a test statistic that is as extreme or more extreme than the one observed in the experiment (more extreme $\rightarrow$ in the direction of H_1).

R function

`t.test(x, mu = mu0)`
or paired: `t.test(x1, x2, paired = TRUE)`

`t.test(y ~ group, var.equal = TRUE)`

When the assumptions are not met

- No normality, large n: central limit theorem
- No normality, small n: permutation test (permute the sign of $Y - \mu_0$, 2^n possibilities)
- Unequal variances: Welch two-sample t-test
- No normality, large n in both groups: central limit theorem
- No normality, small n: nonparametric test (Wilcoxon-Mann-Whitney, Chapter 9)

Design aspects: what does the power depend on?

$$T = \frac{\bar{Y} - \mu_0}{S/\sqrt{n}}$$

- S: not under our control, depends on the variability in the population
- n: most important design parameter
- significance level (usually 5%)
- true effect size $\mu - \mu_0$: not under our control, so we have to assess it upfront to determine the minimal n needed to pick up the minimal effect size δ that is scientifically relevant.

$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- S_p: not under our control, depends on the variability in the population
- n_1, n_2: most important design parameters; for a fixed total $n_1 + n_2$ the SE is minimal when $n_1 = n_2$
- significance level (usually 5%)
- true effect size $\mu_1 - \mu_2$: not under our control, so we have to assess it upfront to determine the minimal n needed to pick up the minimal effect size δ that is scientifically relevant.