

COMPARING THE MEANS

What will it depend on:

- Effect size (distribution-based)

e.g. r^2 , Cohen's $d = (\mu_1 - \mu_2)/\sigma$

- Sample size n

s.e.m. = σ / \sqrt{n}

Hence, we can have a fairly large effect size but no significant difference between the means or a very small effect size but a significant difference between the means

在 t 检验中，检查效应值（Effect Size）和标准误（Standard Error of the Mean, SEM）可以帮助你更全面地理解两组数据之间的差异。

效应值（Effect Size）

效应值是量化两组数据之间差异强度的数值。在 t 检验中，常见的效应值指标包括 Cohen's d （用于两组独立样本）或 Cohen's d 的变种（如用于配对样本或重复测量）。

对于两组独立样本，Cohen's d 的计算公式为： $d = \frac{\bar{X}_1 - \bar{X}_2}{s_p}$ ，其中 \bar{X}_1 和 \bar{X}_2 是两组的均值， s_p 是合并标准差（pooled standard deviation）。

对于配对样本或重复测量，你可以使用类似的公式，但需要将差值（每个配对的两个观测值之间的差异）作为分子，并将这些差值的标准差作为分母。

合并标准差（pooled standard deviation）或称为合并方差（pooled variance）在两组独立样本的 t 检验中用于估计总体的方差。当两个总体的方差被认为是相等（即方差齐性）时，可以使用合并标准差。

合并标准差的计算公式如下：

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

意义：

效应值提供了关于两组数据之间差异大小的直接信息。与 p 值不同，效应值不受样本大小的影响，因此可以在不同研究之间进行比较。

Cohen's d 的值越大，表示两组之间的差异越大。通常， d 值在 0.2 到 0.5 之间被认为是小的效应，0.5 到 0.8 是中等的效应，大于 0.8 是大的效应。

```
# install.packages("pwr")
library(pwr) # 真的好用！
model = df %>% cohens_d(value ~ treatment, paired = TRUE)
model$effsize
```

注意我记得强调过，说效应量大小的时候必须提到 p-value，这样两个结合起来说。

即 The effect size is ... with a p-value of ...

文心一言

效应量（effect size）是一个量化两组或多组之间差异大小的指标，它提供了关于实验结果大小和重要性的信息。效应量的大小可以帮助我们理解实验结果的实际意义，并判断该结果是否足够显著以支持研究假设。常见的效应量指标包括 Cohen's d （对于 t-test）、相关系数 r 、决定系数 R^2 （对于线性回归）等。

标准误 (Standard Error of the Mean, SEM)

标准误是样本均值与总体均值之间差异的标准差。它反映了样本均值作为总体均值估计值的精确度。

标准误的计算公式为: $(SEM = \frac{s}{\sqrt{n}})$, 其中 s 是样本标准差, n 是样本大小。

标准误越小, 表示样本均值越接近总体均值, 即样本均值的估计值越精确。

在 t 检验中, 标准误用于计算 t 值和置信区间。较大的标准误会导致 t 值较小, 从而降低检测到显著差异的可能性。

T-test 适合样本量大小 <30

但是不是说如果样本太大了就用 Wilcoxon test (also known as Wilcoxon rank sum test or Mann-Whitney U test), 两个检验在 sample size 方面的功效没有差异, 只是取决于数据是否正态分布。