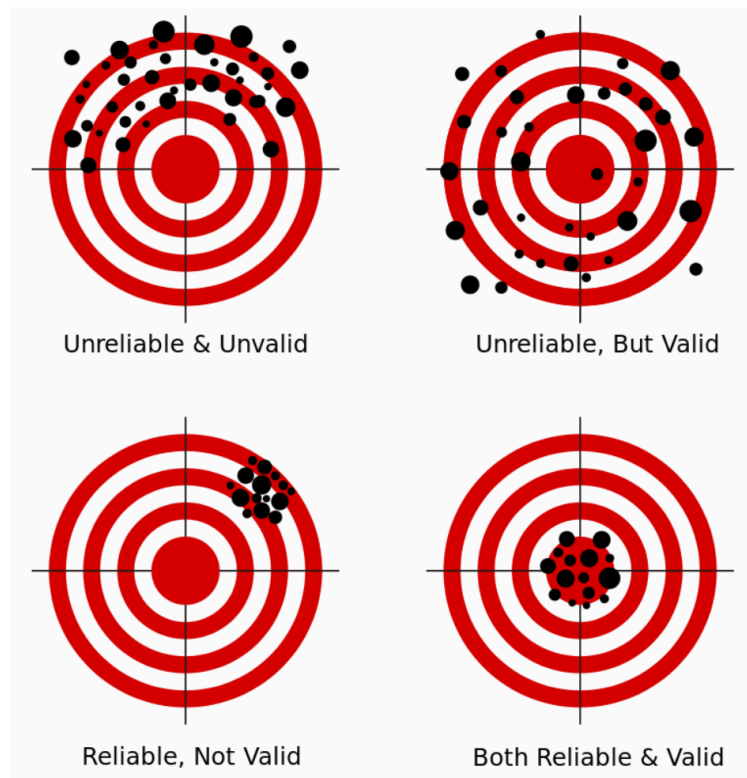


# 证据的强度和多重比较问题

## 精度与准确度 (Precision vs Accuracy)



## 样本量 (Sample Size)

- **定义:** 样本量是指从总体中随机抽取的样本数量。
- **重要性:** 样本量的大小直接影响统计推断的精度和准确性。样本量越大，估计的精度越高，统计信息的分布越可预测，统计检验的功效越高，置信区间的估计宽度越小。
- **影响因素:**
  - **效应大小 (Effect Size):** 效应大小是指总体中两个变量之间的差异程度。效应大小越大，所需的样本量越小。
  - **显著性水平 (Significance Level):** 显著性水平是指拒绝原假设的概率，通常设置为 0.05。显著性水平越低，所需的样本量越大。
  - **功效 (Power):** 功效是指在原假设为假的情况下，拒绝原假设的概率。功效越高，所需的样本量越大。
  - **总体方差 (Population Variance):** 总体方差越大，所需的样本量越大。

## 样本大小的影响 (The Effects of Sample Size)

由 **中心极限定理 (Central limitation theory)** 知，当样本数量足够大时，样本的分布会趋向于 **正态分布**，且概率密度在 **均值** 附近越来越密集，方差越来越小。由样本得出的估计与真实分布间差异越小，置信区间越小。

## 功效 (Power)

- 定义:** 功效是指在原假设为假的情况下，拒绝原假设的概率。
- 重要性:** 功效反映了统计检验的敏感性，即检测到真实效应的能力。功效越高，越能避免第二类错误 (假阴性)。
- 影响因素:**
  - 样本量 (Sample Size):** 样本量越大，功效越高。
  - 效应大小 (Effect Size):** 效应大小越大，功效越高。
  - 显著性水平 (Significance Level):** 显著性水平越低，功效越低。
  - 总体方差 (Population Variance):** 总体方差越大，功效越低。

### 总结:

样本量和功效是相互关联的。为了获得更高的功效，需要更大的样本量。在进行统计推断之前，需要根据研究目标和预期效应大小，进行样本量计算，以确保足够的样本量，并达到预期的功效水平。

### 举例:

假设我们要研究一种新药对治疗某种疾病的疗效。为了确定该药是否有效，我们需要进行一项临床试验。在设计试验时，我们需要考虑样本量和功效。

- 样本量:** 如果我们希望检测到较小的效应大小，或者希望更高的功效，就需要更大的样本量。
- 功效:** 如果我们希望在该药确实有效的情况下，能够检测到它的疗效，就需要更高的功效。

通过样本量计算，我们可以确定所需的样本量，以确保试验能够检测到预期的效应大小，并达到预期的功效水平。

## 多重比较

### 假阳性影响示例

在统计学中，多重比较 (Multiple Comparisons) 关注在进行多个假设检验时如何控制错误发现率 (如假阳性率)。在进行一次假设检验时，设定的显著性水平 (通常为  $\alpha = 0.05$ ) 表示在正确的情况下，错误拒绝原假设的概率是 5%。然而，当同时进行多次比较时，错误的概率会累积，导致整体的假阳性率增加。

## 示例解析

在您的例子中，我们有以下几个元素：

### 1. 比较的设置

有 **3个水平**：2个治疗组 (treatment)、1个对照组 (control)。在两个子组（成年人和儿童）中，总共可以进行 **最多6个比较**。

### 2. 结果测量

对于两个感兴趣的结果测量，总共可以进行 **最多12个比较**。

### 3. 独立检验

每个比较都是单独以  $\alpha = 0.05$  的水平进行检验。

## 假阳性概率

### 1. 单个测试的假阳性率

对于每个比较，假阳性（错误拒绝原假设）发生的概率是 5%。

### 2. 多重测试的累积效应

如果进行 12 次独立的比较，则计算至少一个比较出现假阳性的概率。计算至少一个假阳性的方法是计算所有比较都没有假阳性的概率，然后用 1 减去这个值。

设每个比较的假阳性率为 5%（即 0.05），那么没有假阳性的概率为：

$$P(\text{no false positives}) = (1 - 0.05)^{12} = 0.95^{12} \approx 0.54$$

因此，至少有一个假阳性的概率为：

$$P(\text{at least one false positive}) = 1 - P(\text{no false positives}) \approx 1 - 0.54 = 0.46$$

## 家族错误率 (Family-Wise Error Rate, FWER)

**家族错误率 (FWER)** 是进行多个比较时发生至少一个假阳性的概率。在您的示例中，FWER = 0.46 表示在进行 12 个比较时，有 46% 的概率至少会发现一个假阳性。这表明，随着比较数量的增加，管理假阳性风险变得愈加重要。

## 总结

多重比较是统计分析中一个复杂但重要的话题。由于在多次假设检验中可能累积的假阳性风险，研究者在设计实验或分析数据时，常常需要采取措施（如 Bonferroni 校正、Tukey 检验等）来控制 FWER，并确保结果的有效性和可靠性。