统计&假设检验

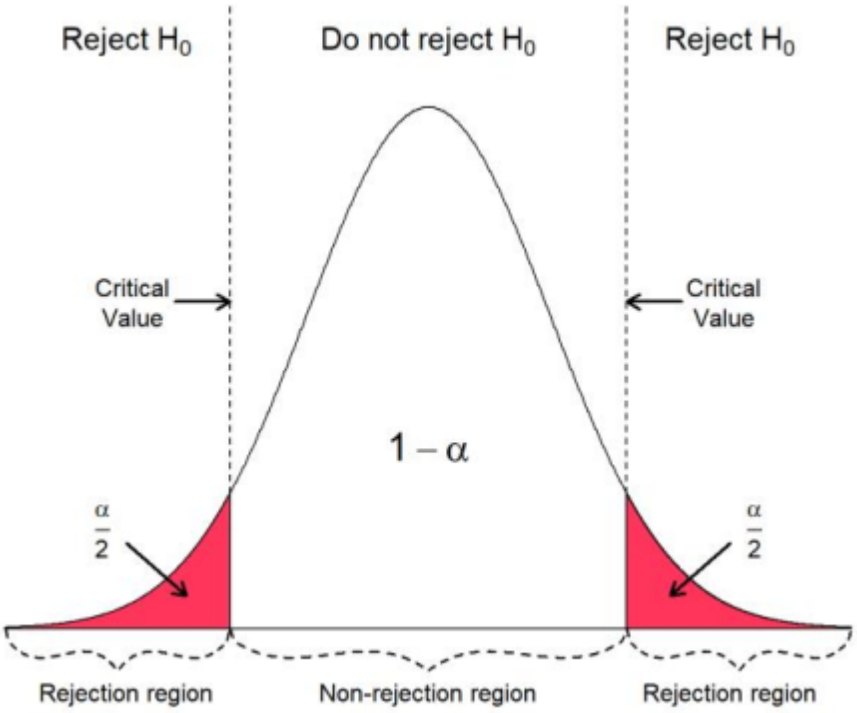
假设检验：

- **Null Hypothesis**H_0：实验无效的假设
- **Alternative Hypothesis**H_a：实验有效的假设
- **Statistical Test**：根据Sample量化与H_0, H_a距离来验证是否拒绝H_0
- **显著水平Significance Level**：有关证据强度的量化，记作α，代表允许拒绝H_0时出错的概率，通常$\alpha = 0.1, 0.5$
- **置信区间Confidence Level**：确定性的程度，代表我们做出决定的信心，记作C

$$\alpha = 1 - C$$

	Null Hypothesis is TRUE	Null Hypothesis is FALSE
Reject null hypothesis	Type I Error (False positive)	Correct Outcome! (True positive)
Fail to reject null hypothesis	Correct Outcome! (True negative)	Type II Error (False negative)

- **拒绝域Critical Region**：拒绝H_0的空间，也称作Rejection Region



- **拒绝域的设计：**根据问题设计two-tailed tests或one-tailed tests

p-value：H_0是正确的概率，p-value越小，代表更要拒绝H_0

样本的组数量：

- One-sample：仅有一组样本
- Two-sample：有两组样本，可以是unpair也可以是pair的。unpaired指两组样本不相关，paired指两组样本相关，比如同一人群，吃饭前和吃饭后的体重

计算p-value方法：

- 假设为数据是服从高斯分布的，考虑**均值**，则有p-value方法：
  - Z-Test：已知标准差，**需要样本数大于30**

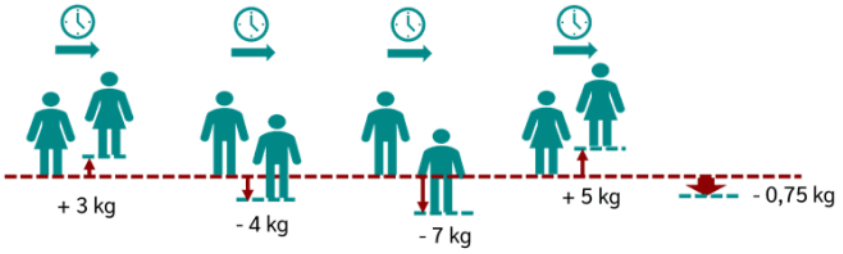
$$z = \frac{\mu_1 - \mu_0}{\sigma / \sqrt{n}}$$

  - T-Test：当不知道总分布的标准差，且样本数量少的时候（$n < 30$）

$$\text{One Sample: } t = \frac{\bar{x} - \mu}{S / \sqrt{n}}$$

$$\text{Two Unpaired Sample : } t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

$$\text{Two Paired Samples: } t = \frac{\bar{x}_{diff} - 0}{s_{diff}/\sqrt{n}}$$



Number of cases

$n = 4$

Degrees of freedom

$df = n - 1 = 3$

Mean

$$\bar{x} = \frac{+3 - 4 - 7 + 5}{4} = -0.75$$

Standard deviation

$$s = \sqrt{\frac{(3 + 0.75)^2 + (-4 + 0.75)^2 + (-7 + 0.75)^2 + (5 + 0.75)^2}{4 - 1}} = 5.68$$

Standard error of the mean

$$s_e = \frac{s}{\sqrt{n}} = \frac{5.68}{2} = 2.84$$

t-Value

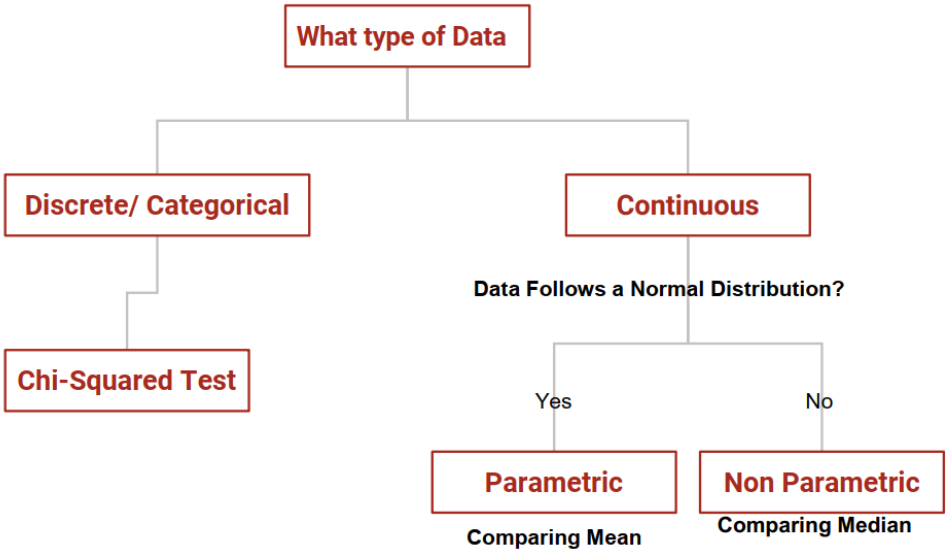
$$t = \frac{\bar{x} - 0}{s_e} = \frac{\bar{x}}{s_e} = \frac{-0.75}{2.84} = -0.26$$

- Chi-Square Test：检查两个种类的集合是否来自同一分布

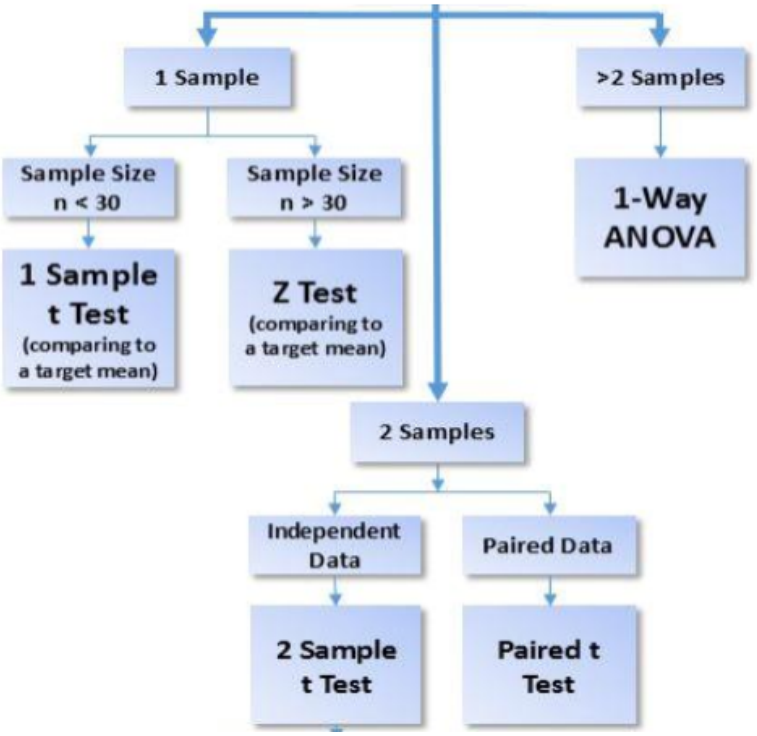
$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

- Anova Test：有多个样本组。零假设定义为在任何组中没有区别，替代假设定义存在至少一组由区别

- 若数据不服从高斯分布，则使用nonparametric test，可以是median



Parametric Tests (Comparing Means)



Nonparametric Tests (Comparing Medians)