

# 统计学：决策的科学项目说明

## 背景信息

在一个 Stroop（斯特鲁普）任务中，参与者得到了一系列文字，每个文字都用一种油墨颜色展示。参与者的任务是将文字的打印颜色大声说出来。这项任务有两个条件：一致文字条件，和不一致文字条件。在一致文字条件中，显示的文字是与它们的打印颜色匹配的颜色词，如“红色”、“蓝色”。在不一致文字条件中，显示的文字是与它们的打印颜色不匹配的颜色词，如“紫色”、“橙色”。在每个情况中，我们将计量说出同等大小的列表中的墨色名称的时间。每位参与者必须全部完成并记录每种条件下使用的时间。

## 调查问题

作为一般说明，请确保记录你在创建项目时使用或参考的任何资源。作为项目提交的一部分，你将需要报告信息来源。

1. 我们的自变量是什么？因变量是什么？

自变量是文字与颜色是否匹配

因变量是参与者说出列表墨色名称的反应时间

2. 此任务的适当假设集是什么？你需要以文字和数学符号方式对假设集中的零假设和对立假设加以说明，并对数学符号进行定义。你想执行什么类型的统计检验？为你的选择提供正当理由（比如，为何该实验满足你所选统计检验的前置条件）。

原假设用 $H_0$ 表示，即文字的颜色是否匹配对受试者的反应时间没有影响。

对立假设用 $H_A$ 表示，即文字和颜色匹配时，受试者的反应时间要小于不匹配时的反应时间。

用 $\mu_{co}$ 表示完成“一致”任务的总体平均时间，用 $\mu_{inc}$ 表示完成“不一致”任务的总体平均时间，公式表示如下：

$$H_0: \mu_{co} = \mu_{inc}$$

$$H_A: \mu_{co} < \mu_{inc}$$

z-test 适用于总体参数。在本试验中，我们无法获取总体参数，只有样本，并且需要通过比较样本之间的差别进而推断总体的情况，所以需要使用 t-test。

本实验所获取的样本，是同一组受试者分别进行两次测试所得到的，所以应当是相依样本。本试验所使用方法为相依样本中的 repeated measured design（重复衡量设计）。

应用 t-test 的前提条件：

- 1) 仅适用于相互独立的两个随机变量。
- 2) 因变量为连续型变量
- 3) 因变量的观测值之间相互独立。但对于相依样本来说，只需要每对数据的因变量差值之间相互独立即可。

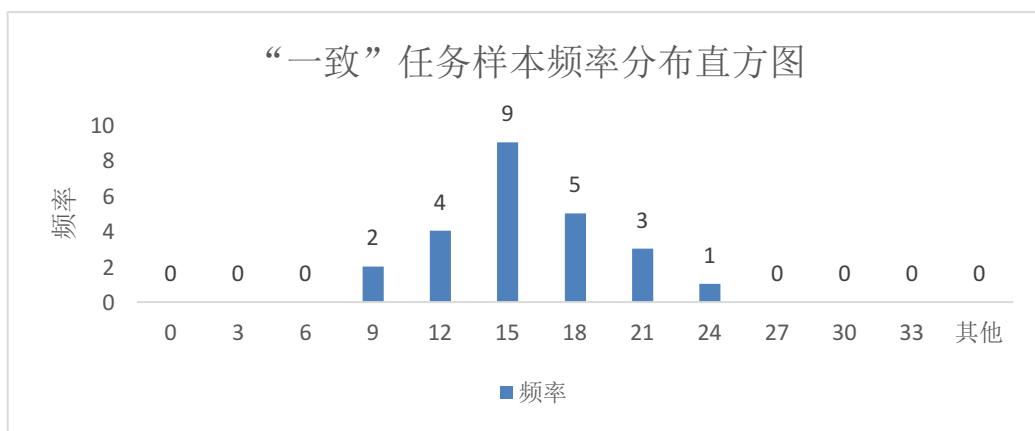
4) 满足方差齐性，因变量服从正态分布或着近似服从正态分布。  
因对立假设具有方向性，且 $\mu_{co} - \mu_{inc} < 0$ ，故采用负方向单尾 t 检验。

- 报告关于此数据集的一些描述性统计。包含至少一个集中趋势测量和至少一个变异测量。

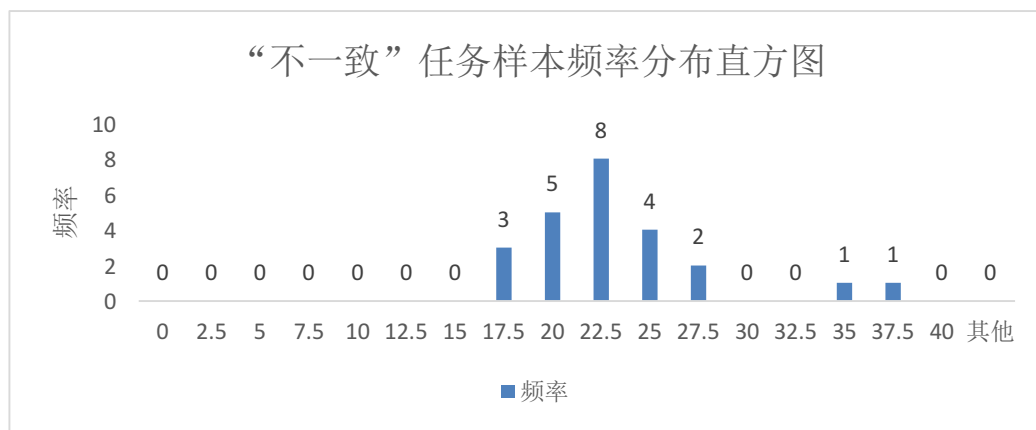
“一致”任务样本量 $n_{co}$ 为 24，样本均值 $\overline{X_{co}} = 14.05$ ，样本标准差 $\overline{S_{co}} = 3.56$ 。

“不一致”任务样本量 $n_{inc}$ 为 24，样本均值 $\overline{X_{inc}} = 22.02$ ，样本标准差 $\overline{S_{inc}} = 4.80$ 。

- 提供显示样本数据分布的一个或两个可视化。用一两句话说明你从图中观察到的结果。



从“一致”任务样本的频率分布直方图可以看出，该样本的众数位于[12,15)区间内，与样本均值位于同一区间。



从“不一致”任务样本频率分布直方图可以看出，该样本的众数位于[20,22.5]区间内，与样本均值位于同一区间。

从两个样本的频率分布直方图来看，“不一致”任务样本众数所在区间较“一致”任务样本众数所在区间出现明显右移。

- 现在，执行统计测试并报告你的结果。你的置信水平和关键统计值是多少？你是否成功拒绝零假设？对试验任务得出一个结论。结果是否与你的期望一致？

- 1) 计算两个样本均值的差异:

$$\overline{X_{co}} - \overline{X_{inc}} = 14.05 - 22.02 = -7.97$$

- 2) 计算两个样本差异的标准差:

$$S = \sqrt{\frac{\sum_{i=1}^n [(X_{coi} - X_{inci}) - (\overline{X_{co}} - \overline{X_{inc}})]^2}{n - 1}} \approx 4.86$$

- 3) 计算 SEM

$$SEM = \frac{S}{\sqrt{n}} = 0.99$$

- 4) 计算 t-statistic

$$t - statistic = \frac{\overline{X_{co}} - \overline{X_{inc}}}{S/\sqrt{n}} \approx -8.03$$

- 5) 计算 t-critical, 因采用单尾检测, 为降低犯 Type I error 的概率, 故设置  $\alpha = 0.01$ , 自由度为 23, 则:

$$t - critical = -2.5$$

- 6) 效应量:

$$r^2 = \frac{t^2}{t^2 + df} = .73$$

表明 73% 的差异是由于文字与颜色是否匹配所造成的。

- 7) 置信区间 (CI)

99% 置信区间的 t-critical value 为 2.807, 则 margin of error 是:

$$t - critical * SEM = 2.807 * 0.99 = 2.78$$

令  $\overline{X_D} = \overline{X_{co}} - \overline{X_{inc}}$ , 则:

$$CI: (\overline{X_D} + SEM, \overline{X_D} - SEM) = (-10.75, -5.19)$$

即在 99% 的置信水平下, 样本均值差异  $\overline{X_D}$  会落在  $(-10.75, -5.19)$  区间内。

- 8) 决策

经计算, t-statistic 对应的 p 值远小于 0.0001, 即  $p < \alpha$ , 拒绝原假设  $H_0$ , 说明在置信度为  $\alpha = 0.01$  时, 受试者完成“一致”任务的平均时间小于“不一致”任务的平均时间具有统计显著性。

从而说明, 斯特鲁普效应中, 若新的刺激特征与原先的刺激不相同, 则会干扰人们的认知, 使人们所需的反应时间变长。(来源于 wikipedia 对斯特鲁普效应的描述)