

统计分析 - 斯特鲁普效应

一、调查背景

STROOP 效应是指字义对命名的干扰效应。一般认为，念字和命名是两个不同的认知过程。Stroop 于 1935 年做了一个实验，他使用的刺激字与书写它所用的颜色相矛盾，结果发现，说字的颜色时会受到字义的干扰。

一般认为，STROOP 效应是由于念字自动化造成的。人们对字加工快，而对颜色加工慢，因此，当要说颜色时，就会受到字义的干扰，而反过来，念字却不会受到字的颜色的干扰。

二、研究方法

斯特鲁普在相同的测试中使用不同的变化以创造不同的刺激：

1. 以相同于文字名字所代表的颜色显示的文字。
2. 以不同于文字名字所代表的颜色显示的文字。

这两种刺激中，第一种代表“一致的条件”（congruous condition），念出与颜色相同色标的名字。而相反的，在刺激 2 中，颜色的名字与所使用的颜色不一致（例如文字“红色”是以“绿色”油墨印出）

在每个情况中，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。

三、研究结果

变量

1. 自变量是文字含义与文字颜色是否一致
2. 因变量是受试者读出所有文字颜色所用的时间

假设

零假设是，无论文字的颜色和含义一致或是不一致时，对于受试者没有差别。

对立假设是，文字含义和颜色一致时，受试者的完成测试用的时间比不一致时更短

这里用 H_0 表示零假设，用 H_A 表示对立假设，用 μ_c 表示文字含义与颜色一致的总体的均值。用 μ_i 表示文字含义与颜色不一致的总体的均值。

$H_0: \mu_c = \mu_i$

$H_A: \mu_c < \mu_i$

统计测试类型

由研究方法所知，这是一组受试者参加两次测试所得到的两个样本，也就是所谓的相依样本。由于不知道总体参数 σ ，因此采用 t 检验。在相依样本 t 检验的测试类型中，有一种叫做重复测量设计，就是在两个处理中使用相同个体组成的同一个样本（一个被试样本，得到两个统计样本，这两个统计样本是相关样本）

这里，将采用负方向的单尾检验。

因为我们的对立假设是文字含义和颜色一致时，受试者的反应时间更短。所以检验必须具有方向性，不使用双尾检验，而使用单尾检验。从我们对立假设 $\mu_c < \mu_i$ 可以得知，检测的方向是负方向。

t-test 前提

两个样本的 t -test 前提是：

1. 数据是连续的，而非离散的
2. 数据遵循正态分布
3. 两个总体的方差相等(对于相依样本，自动认为其总体方差相等)
4. 两个样本是独立的，对于一个因变量的观察独立与对于其它因变量的观察。
5. 两个样本都是从他们对应总体中随机抽取的，总体中的每个个体都有同等的机会被选中。

样本数据

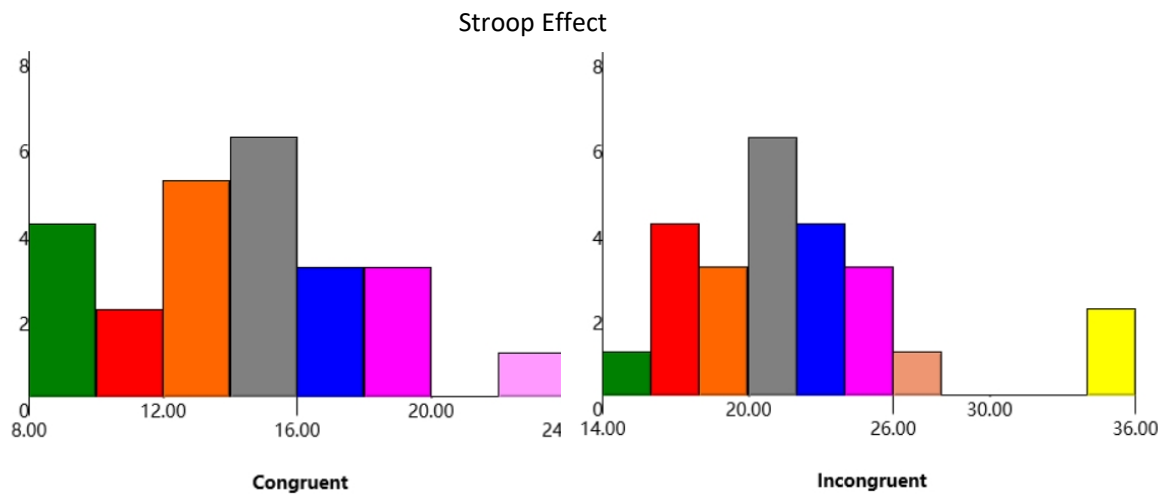
$n = 24$

Congruent	Incongruent	$X_c - X_i$
12.079	19.278	-7.199
16.791	18.741	-1.95
9.564	21.214	-11.65
8.63	15.687	-7.057
14.669	22.803	-8.134
12.238	20.878	-8.64
14.692	24.572	-9.88
8.987	17.394	-8.407

9.401	20.762	-11.361
14.48	26.282	-11.802
22.328	24.524	-2.196
15.298	18.644	-3.346
15.073	17.51	-2.437
16.929	20.33	-3.401
18.2	35.255	-17.055
12.13	22.158	-10.028
18.495	25.139	-6.644
10.639	20.429	-9.79
11.344	17.425	-6.081
12.369	34.288	-21.919
12.944	23.894	-10.95
14.233	17.96	-3.727
19.71	22.058	-2.348
16.004	21.157	-5.153

样本数据可视化

直方图



从直方图中可以观察到，文字含义和颜色一致时，受试者的用时集中在 12 到 20 之间
文字含义和颜色不一致时，受试者的用时集中在 16 到 26 之间

数据集的统计描述

均值

$$\bar{X}_c = 14.05$$

$$\bar{X}_I = 22.02$$

前后二者所用的时间均值的差别

$$\bar{X}_C - \bar{X}_I = -7.97$$

标准偏差

$$S_C = 3.56$$

$$S_I = 4.80$$

$$S_D = 4.86$$

标准误差

$$SEM = 0.99$$

t 统计量

$$t\text{-statistic} = -8.05$$

t 临界值

这里使用 α 级别为 0.05 的单尾检验，自由度是 23，t 临界值是 -1.714

置信区间

自由度是 23，95% 置信区间的 t 临界值是 2.069，误差界限是 $t\text{-critical} * SEM = 2.069 * 0.99 = 2.05$

$$\text{置信区间 CI: } \bar{x}_D \pm 2.05 = -7.97 \pm 2.05 = (-10.02, -5.92)$$

关于均值差异的置信区间: 95% CI = (-10.02, -5.92)

决策

$$t(23) = -8.05, P < .05, \text{one-tailed}$$

根据 t 统计量和 t 临界值，结果有统计上的显著性。

t 统计量落在临界区内， $P < .05$ ，拒绝零假设

实验证明，当文字颜色和文字含义一致时，受试者完成测试使用的时间更短

Cohen's d

$$d = -1.64$$

效应量 r^2

$$r^2 = 0.74$$

也就是说 74% 的差异，是由文字的颜色和含义一致与不一致所造成的。

四、参考文献

[斯特鲁普效应](#)

[When to use a t-score vs z-score](#)

[Assumptions for the t-test](#)