

统计学：决策的科学项目说明

说明：[点此查看此文档的英文版本](#)。

背景信息

在一个 Stroop（斯特鲁普）[1][2]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？

答：

自变量：文字与它们的打印颜色匹配或不匹配。

因变量：被试说出同等大小的列表中的墨色名称的时间。

2. 此任务的适当假设集是什么？你想执行什么类型的统计测试？为你的选择提供正当理由。

答：

- 适当假设集：

- 第一组试验：文字与它们的打印颜色匹配，样本的总体时间均值为 μ_1 。

- 第二组试验：文字与它们的打印颜色不匹配，样本的总体时间均值为 μ_2 。

- 零假设： $\mu_1 = \mu_2$ ，文字与打印颜色不匹配不会导致识别墨色名称时间变长。

- 对立假设： $\mu_1 < \mu_2$ ，文字与打印颜色不匹配会导致识别墨色名称时间变长。

- 统计测试类型：

- 相依样本 t 检验 —— 预期检验和后期检验(Pretest Posttest)

- 选择理由：

- 本题目中，样本由随机抽样方式取得，试验样本少于 30 个，且总体的标准差未知。这里针对同一受试者进行两次测试，分别在两种不同的条件下先后进行测试，可以看做对同一受试者样本实施了处理措施进行先后两次试验，属于受试者内设计，且样本中每组配对的差值相互独立，这里使用相依样本 t 检验，检验在文字与它们的打印颜色变成不匹配后，受试者的平均所用时间减少量是否显著到足以断定是因为文字与打印颜色不匹配导致的。

现在轮到你自行尝试 Stroop 任务了。前往[此链接](#)，其中包含一个基于 Java 的小程序，专门用于执行 Stroop 任务。记录你收到的任务时间（你无需将时间提交到网站）。现在[下载此数据集](#)，其中包含一些任务参与者的结果。数据集的每行包含一名参与者的表现，第一个数字代表他们的一致任务结果，第二个数字代表不一致任务结果。

- 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。

答：

第一组试验：条件为词汇与颜色匹配(Congruent)，样本均值 $\mu_1 = 14.05$ ，样本的标准偏差 $S_1 = 3.56$ 。

第二组试验：条件为词汇与颜色不匹配(Incongruent)，样本均值 $\mu_2 = 22.02$ ，样本的标准偏差 $S_2 = 4.80$ 。

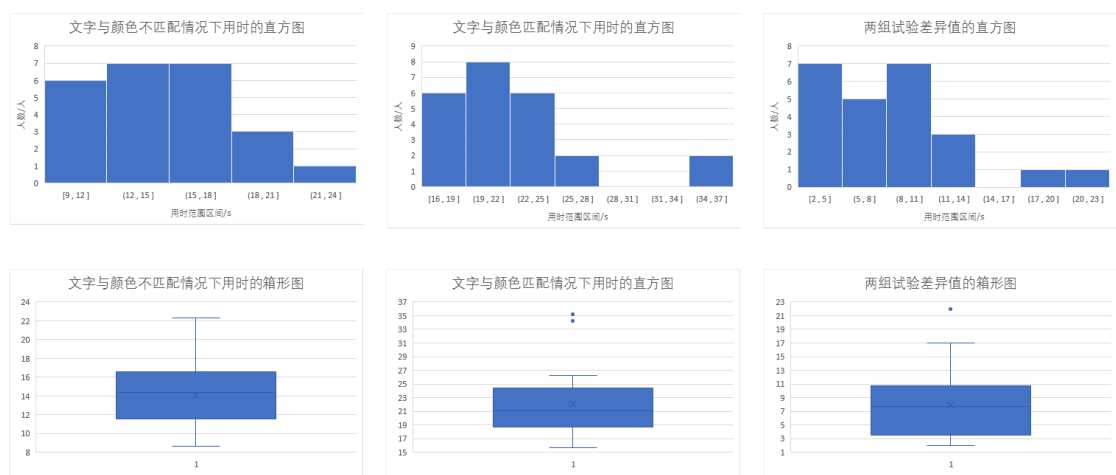
差别均值： $\mu_D = \mu_2 - \mu_1 = 7.96$ 。

差别值集合的样本估算标准差： $S_{DS} = 4.86$ 。

差别值集合的样本标准误差： $S_{DP} = 4.76$

- 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。

答：



在词汇与颜色匹配的条件下，样本的用时呈现左偏态分布，大多样本所用时间集中在 11 s ~ 17 s 之间，无异常值；在词汇与颜色不匹配的条件下，样本的用时呈现右偏态分布，大多样本用时集中在 18 s ~ 25 s 之间，存在 2 个异常值；样本的差别值呈现左偏态分布，大多数样本的差别值集中在 3 s ~ 11 s 之间，存在 1 个异常值。

- 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？

答：

设置 α Level: $\alpha = 0.05$, one-tailed。

样本大小: $n = 24$; 自由度: $DF = n - 1 = 23$ 。

t 临界值: $t\text{-critical} = 1.714$ 。

误差范围: $\text{margin of error} = t\text{-critical} * S_{DP} = 8.16$ 。

置信区间: $(\mu_D - \text{margin of error}, \mu_D + \text{margin of error}) = (-0.20, 16.12)$ 。

$$t\text{-statistic} = \frac{\mu_D}{S_{DS}/\sqrt{n}} = 8.02。$$

P value[3]:

P Value Results

t=8.02 DF=23

The two-tailed P value is less than 0.0001

By conventional criteria, this difference is considered to be extremely statistically significant.

因为 $t\text{-statistic} > t\text{-critical}$, $P < \alpha$, 所以我们拒绝零假设, 该差异具有统计显著性, $\mu_1 < \mu_2$ 成立。

效应量: Cohen's d: $d = \frac{\mu_D}{S_{DS}} = 1.64$, 这表明两组试验的均值相差 1.64 个标准偏差, 大于一个标准偏差, 两组试验的差异程度较大。

相关度量: $r^2 = \frac{t\text{-statistic}^2}{t\text{-statistic}^2 + DF} = 0.7366 = 73.66\%$, 这表明在这个样本中, 有 73.66% 的被试在第二组试验中花费更多时间是因为词汇与颜色不匹配导致的。

综上所述, 相比词汇与颜色匹配的情况下, 在词汇与颜色不匹配时, 读出墨色名称所需的时间更长。所以, 词汇的意义对人的颜色识别具有较大的干扰性。

这个结果与我的期望一致。

6. 可选: 你觉得导致所观察到的效应的原因是什么? 你是否能想到会取得类似效应的替代或类似任务? 进行一些调查研究将有助于你思考这两个问题!

答:

我觉得导致所观察到的效应的原因是, 相对于颜色, 人对文字更加敏感。人的大脑在处理信息时, 首先会对文字进行处理, 再处理其它特征(如颜色、形状等)。

取得类似效应的替代或类似任务: 让参与者先获得一系列图像, 由众多不同颜色的特殊字符组成(如: 某些象形文字、火星文、甲骨文等, 保证被试者不认识这些文字), 参与者需要逐个大声念出这些特殊字符的颜色, 记录参与者所用的时间。接着, 再让参与者获得一系列文字词义为颜色但是和油墨颜色不匹配的文字, 参与者逐个大声念出这些文字的油墨颜色, 记录参与者所用的时间。

同样使用相依样本 t 检验:

零假设: $\mu_1 = \mu_2$ 。

对立假设: $\mu_1 < \mu_2$ 。

通过该试验可以检验是否是因为人对文字更加敏感导致了该效应。

参考或使用的网站:

[1] 斯特鲁普效应的维基百科: https://en.wikipedia.org/wiki/Stroop_effect

[2] 斯特鲁普效应的百度百科:

http://baike.baidu.com/link?url=uCBYCBCFGSpY8B_PfbAZWjCVj1Cn8l32bibbV6Kd8l8CN8E-MNclb5jh47RdPeiarb5lYvkY87xqO7sA8QLzrsenQvnWap-f9UKjdJARlTKaoZvCu58SEeDRuDMYAbM6Cw2zjJlpSuv_5qgFUUg9a

[3] P value 的计算使用的网站: <http://www.graphpad.com/quickcalcs/>

优达学城

2016 年 9 月