

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）任务中，参与者得到了一列文字，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：文字条件

因变量：参与者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。

（1）零假设：在 Stroop 任务中，不同文字条件对参与者说出同等大小的列表中的墨色名称的时间没有影响

$H_0: U_A - U_B = 0$ ；（ H_0 代表零假设， U_A 代表一致文字条件总体平均值， U_B 代表不一致文字条件的总体平均值）

（2）对立假设：在 Stroop 任务中，不同文字条件对参与者说出同等大小的列表中的墨色名称的时间有影响

$H_A: U_A \neq U_B$ ；（ H_A 代表对立假设， U_A 代表一致文字条件的总体平均值， U_B 代表不一致文字条件的总体平均值）

1. 在本次 Stroop 任务中，我们需要统计参与者分别在两种不同条件下说出墨色名称时间的均值是否有差异，我们得到参与者在两种条件下的少于 30 的样本，未知总体的标准差，所以我选择 dependent-sample-t-test 进行统计测试，基于对立假设，不一致文字条件下说出同等大小的列表中的墨色名称的时间有可能比一致文字条件下的时间减少或者增加，所以选择了 two-tailed test。

2. Two-tailed dependent t-test 的假设前提--①参与者说出墨色名称的时间服从高斯分布；②同一参与者一致文字条件下测得的时间与不一致文字条件下测得的时间相对应，所以样本中参与者在一致文字条件下测得的时间与不一致文字条件下测得的时间的差值是配对且相互独立的。

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。

3. 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。

Mean (Congruent) =14.051; Mean (Incongruent) =22.016; $M_D = -7.965$

Median (Congruent) =14.357; Median (Incongruent) =21.018

Standard deviation (Congruent) =3.559; Standard deviation (Incongruent) =4.797;

$S_D = 4.865$

4. 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。

Congruent 样本频率表

Congruent

最大值 22.328 最小值 8.63 差 13.698
组距 2

分组	频数	频率
8-12	6	0.25
12-16	11	0.46
16-20	6	0.25
20-24	1	0.04

表 1-1



在 stroop 任务测试中样本，一致文字条件下，参与者说出同等大小的列表中的墨色名称的时间大部分在 12~16 之间，占比 0.46；时间在 20~24 之间的占比最少，为 0.04。从 Congruent 直方图可以看出时间在 12~16 之间人数比较多，平均值、中位数落在 12~16 之间

Incongruent 样本频率表

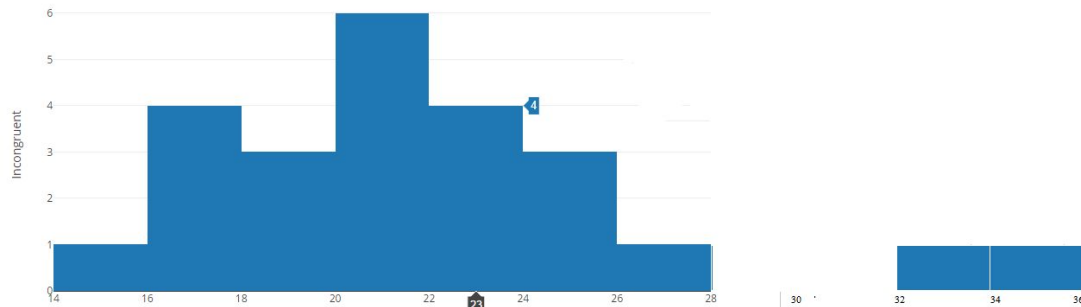
Incongruent

最大值 35.255 最小值 15.687 差 19.568
组距 2

分组	频数	频率
15-19	7	0.29
19-23	10	0.42

23-27	5	0.21
27-31	0	0.00
31-35	1	0.04
35-39	1	0.04

表 1-2



在 stroop 任务测试中样本，不一致文字条件下，参与者说出同等大小的列表中的墨色名称的时间大部分在 19~23 之间，占比 0.42；极少数参与者时间超过 30，仅为 0.08。从 Incongruent 直方图可以看出，参与者说出同等大小的列表中的墨色名称的时间处于 20~22 的最多，中位数在 20~22 之间。

5. 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？

$\alpha = 0.05$; $M_D = -7.965$ $S_D = 4.865$; $t_{critical} = \pm 2.069$; $r^2 = 0.737$; Cohen's $d = -1.637$;
 $t(23) = -8.021$, $p < .025$, two-tailed test

Confidence interval on the mean difference: 95% CI = (-5.910, -10.019) (所有结果保留三位小数)

$t(23) = -8.021 > t_{critical} = -2.069$, fall in the critical region, $p < .025$, 具有统计显著性，成功拒绝零假设。一致文字条件下，参与者说出同等大小的列表中的墨色名称的时间比不一致文字条件下少花时间，不一致文字条件影响了参与者测试时间，与预期结果一致。

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！

1. 导致观察到的效应可能是因为一致文字条件下，大脑把其中一个条件作为默认值，做测试的时候只需要快速识别另一个条件变化的内容就可以完成了（如果突然改变条件，大脑反应不快的话，马上就会出错）；相反，在不一致文字条件下，大脑很小心地识别处理两个条件后再作出判断，所以大多数情况下花的时间相对会比较长。

2. 类似任务 -- 用‘圆形’，‘正方形’，‘三角形’，‘椭圆’几个基本的图案作为测试集。

一致条件--图形内嵌一个相同的图案，如，圆形内嵌一个圆形；

不一致条件--图形内嵌一个不一样的图案，如，圆形内嵌一个正方形；
 记录每个参与者不同条件下的测试时间。

(本次项目参考资料全部来自于 Udacity 机器学习入门视频内容和本人知识积累) -- 二次修改答案参考了审阅老师的意见及其提供的参考资料。-- 三次修改参考了审阅老师的意见及其提供的参考资料 (t-test 前提, t-test Assumptions) 和百度百科

优达学城

2016 年 9 月