

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：自变量是一致或不一致的文字条件
因变量是参与者在每种条件下使用的时间

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：零假设：不同文字条件下，参与者使用时间没有差别。 $(H_0: \mu_1 = \mu_2)$
对立假设：不一致文字条件下，参与者使用时间更久。或不一致文字条件下的参与者使用时间与一致文字条件下参与者使用时间差值大于零。 $(H_A: \mu_1 < \mu_2)$

以上假设中，根据中心极限定理， μ_1 与 μ_2 分别等于文字条件一致和不一致情况下的总体平均时间。零假设文字条件一致的情况下，总体平均时间 μ_1 等于不一致的总体平均时间 μ_2 。而对立假设中，文字条件不一致所用总体平均时间 μ_2 比一致条件下总体时间 μ_1 更长。

此次调查为两种条件下同一被试者参与调查，相应数据将会有同一被试者在两种不同条件下的相依样本。因此应执行相依样本统计测试(Dependant t-test for paired-samples, within-subject design for two-conditions)。相依样本的标准偏差使用两个条件下的样本差值计算，计算时用到 Bessel's Correction。

之所以选择相依样本 t 检验是因为样本量小于 30，且总体标准差未知。假设前提是样本数据从总体中随机抽取，两个样本均经过相同的重复干预，其差值的均值在总体中成正态分布。

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。

参与者在文字条件一致和不一致情况下所用时间数据的中位数分别是 14.36 和 21.02，平均数分别为 14.05 和 22.02。两组数据差值 D 的中位数和均值分别为 7.67 和 7.96。以样本原数据无法测量众数。

文字条件一致，不一致和两者之差的全距分别为：13.70, 19.57, 19.97

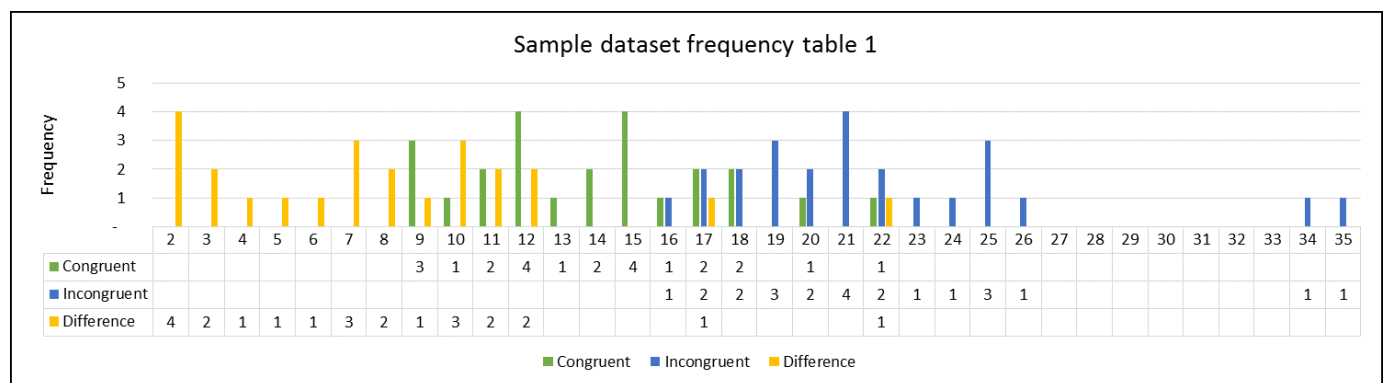
三组数据四分位间距分别为：4.67, 5.38, 7.24

三组数据标准差分别为：3.56, 4.80, 4.86

	Congruent	Incongruent	Difference
Median	14,36	21,02	7,67
Mean	14,05	22,02	7,96
Range	13,70	19,57	19,97
Q3 – Q1	4,67	5,38	7,24
Standard deviation	3,56	4,80	4,86

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。

将所用数据化成整数，整理成频率直方图（图 1 和图 2）：



图一展示了文字条件一致及不一致情况下参与者所用时间和两者之差的频率图。

横轴代表时间从 2 秒到 35 秒不等，纵轴表示该时间在几个不同测试者身上发生。

由于样本量有限，此图时间划分过细，不能很好代入分布类型，故改变组距，由原来每秒一组划分成每五秒一组，产生图 2：

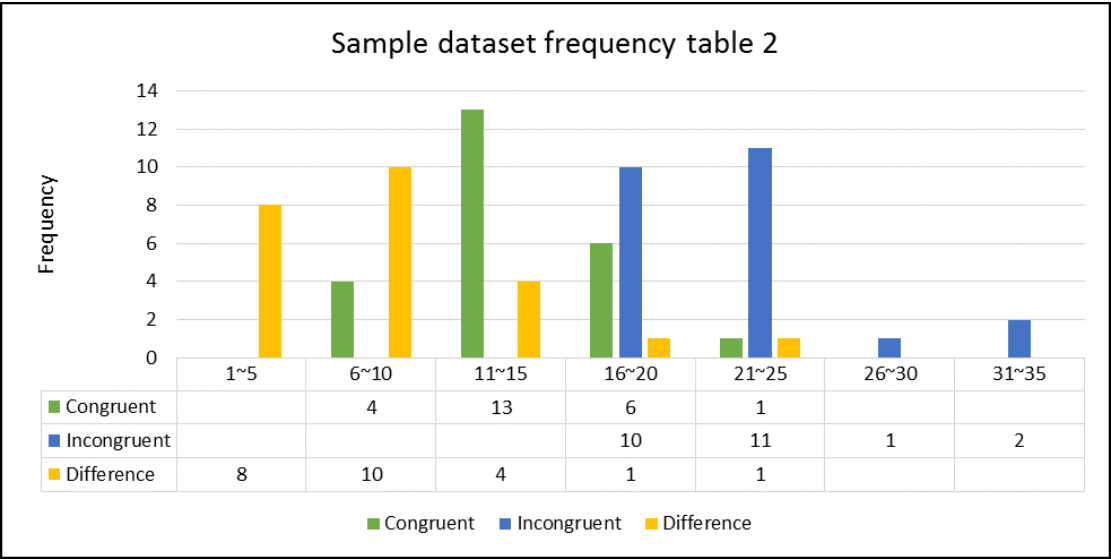


图 2 中三组数据均以正偏斜 (positively skewed) 形态分布：

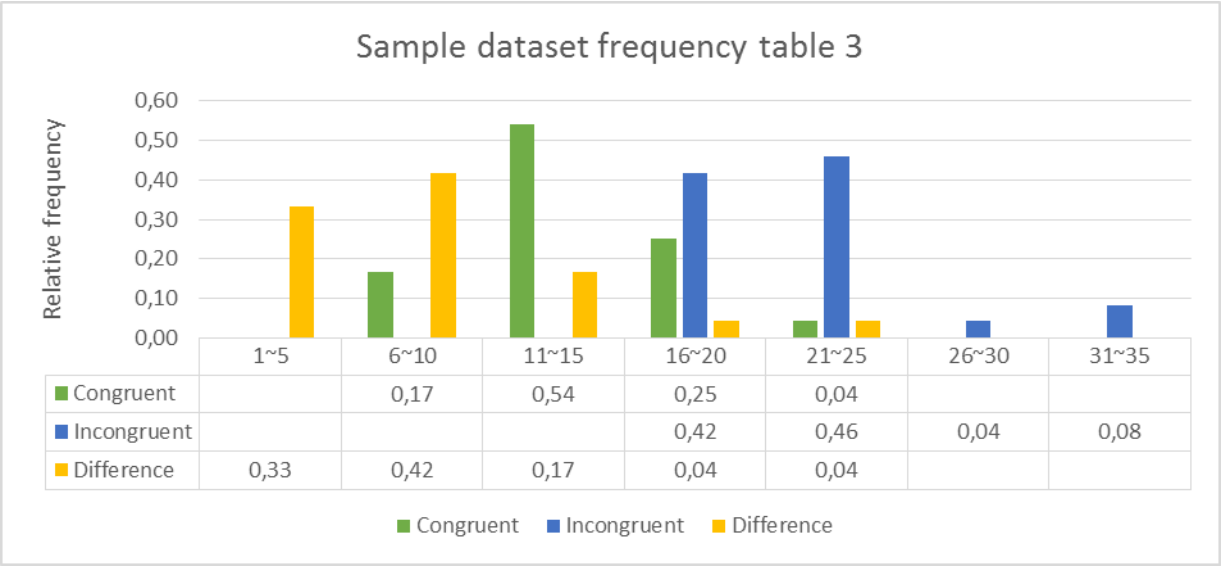
文字条件一致时，时间分布集中于 11~15 秒，包含其中位数 14.36 和平均值 14.05；

文字条件不一致时，时间分布集中于 21~25 秒，包含其中位数 21.02 和平均值 22.02；

二者差值时间分布集中于 6~10 秒，包含其中位数 7.67 和平均值 7.96。

从以上观察结果可知，当文字条件一致时，被试者反应速度比不一致时要快，所用时间更短。文字条件一致时平均比不一致时所用时间短 7.96 秒

同理可得该样本的相对频率图，如下图 3:



相对频率图中三组数据仍以正偏斜形态分布。

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？

本题总体数据未知，执行统计测试用 t 检测。

假设中不一致所用时间长于一致所用时间，即 $\mu_2 - \mu_1 > 0$ ，以下测试使用正单尾 t 检测。

差异点估计 point estimate = $\bar{D} = \mu_2 - \mu_1 = 7.96$

差异的标准偏差 $s = \sqrt{\sum(Di - \bar{D})^2 / (n - 1)} = 4.86$

自由度 $df = n - 1 = 23$

$t \text{ statistic} = \bar{D} / (s / \sqrt{n}) = 8.02$

对于 $\alpha = 0.05$, $t \text{ critical} = 1.714$,

$p < 0.0005 < \alpha$

p 值远小于 α , 文字条件一致与不一致使被试者测试得出显著不同的结果, 因此拒绝零假设。文字条件不一致时被试者所用时间更长。

置信区间 $\bar{D} \pm t \text{ critical} * s_D / \sqrt{n}$ 即 (6.26, 9.67) 95% 的被试者(总体)在文字条件不一致的情况下用时长于一致用时 6 秒到 10 秒。

Cohen's $d = \bar{D} / s_D = 1.64$ 文字条件一致与不一致的均值相距 1.64 个差异标准差

$r^2 = t \text{ statistic}^2 / (t \text{ statistic}^2 + df) = .74$ 因为 $0 < r^2 < 1$, r^2 更接近 1, 因此样本中被试者受自变量的影响很大。

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!

我认为导致文字条件不一致情况下被试者使用较长时间判断文字颜色的主要原因是被试者受文字意义的干扰而需要更长时间判断选项。根据 Neuroscience for kids¹ 网站, 有两个理论可以解释实验现象:

反应速度理论: 阅读速度要快于判断颜色的速度

选择性注意力理论: 判断颜色比阅读要投入更多精力。

该网站还给出了许多类似任务。包括旋转文字角度, 使用非颜色词汇, 使用表示情绪的词汇, 只给单词第一个字母或最后一个字母涂色等等。

我想到的类似任务有广告包装常用伎俩: 在垃圾食品包装上多采用给人 « 健康 » 印象的词汇, 以美化垃圾食品形象。实例有某零食包装上的 « 无反式脂肪 » 标志, 等等。

优达学城

2016 年 9 月

1. <https://faculty.washington.edu/chudler/words.html#seffect>