

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？

自变量是：包括一致文字条件（颜色和文字一致）以及不一致文字条件（颜色和文字不一致）两种情况

因变量是：参与者说出同等大小列表中墨色名称的所需要的时间。

参考连接：<https://wenku.baidu.com/view/19b9aa3d5727a5e9856a61f2.html>

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）。

H0: $X_{c\text{spar}} = X_{i\text{spar}}$ (不一致组平均时间与一致组平均时间没有显著变化)

即 $X_{c\text{spar}} - X_{i\text{spar}} = 0$. 即两组测试差异的期望值 expected difference 为 0

Ha: $X_{c\text{spar}} \neq X_{i\text{spar}}$ (不一致组平均时间与一致组平均时间不相同)

注： $X_{c\text{spar}}$ is the mean value for Congruent group 一致组的样本均值

$X_{i\text{spar}}$ is the mean value for Incongruent group 不一致组的样本均值

观察差异值分布样本 T， 等于 $(X_c - X_i)$, 即每行参与者对应的两组测试需要时间的差值。

其中 X_c 为一致组对应所需时间， X_i 为不一致组对应所需时间

统计检验类型：显著性水平 $\alpha=0.05$ 配对 T 检验

采用 two-tail 的保守评估方式(不确定变化的趋势)。

受试者独立性：参与者是随机选取，互相之间保持独立性。但由于每一位测试者参与了两类测试，故两组测试结果存在相依性。

T 样本分布假设：此样本分布大致符合正态随机分布。

T 样本的分布可以估计总体的 variance. 方差，因为随着样本增加，总体样本的分布呈现正态分布。

总体方差 population variance 应近似等于该样本的方差

两组分组的关系属于 dependent test. 因为每人分别参加两组测试, 第一次测试的结果可能会对第二次测试产生影响。

T 检验而非 Z 检验, 因为我们不知道总体 population 的分布情况, 只能从有限的样本中去推论出总体 population 分布情况。

Comments by reviewer

Q2a: 请注意, 两个 μ 所代表的并不是样本均值, 而是总体均值。如果它代表样本均值, 那么我们根本无需假设, 直接根据表格提供的样本数据计算就可以得出结论。

疑问: 1) 两个分组测试的均值怎么是总体均值呢? 正是因为我们不知道总体的均值, 所以才采用样本的均值乃至求出 T statistics 来估计总体样本, 即两组测试差异性的假设是否存在, 来做判断。如果两组 μ 是总体均值, 恰恰就不需要假设, 正如你所说, 直接得出统计值即可。

2) 另外, 95%CI 的置信水平分布, 不就是利用样本来衡量总体(两组差异值的分布)的均值范围吗?(即两组均值有 6 到 10 秒的差异水平)

Q2a 从字面意义理解, “假设”针对什么? 自然是未知的事物。对于我们已知的事物, 没有假设的必要。如果你认为假设检验中的假设, 针对的是样本均值, 那么按照你的思路, 假设集是这样的:

原假设: 本次实验得出的两组样本均值相等, 即 $\bar{x}_i = \bar{x}_c$, 换句话说, 即 $22.02 = 14.05$;

备择假设: 本次实验得出的两组样本均值不相等, 即 $\bar{x}_i \neq \bar{x}_c$, 换句话说, 即 $22.02 \neq 14.05$ 。

这样的假设集是否有意义? 这就好比假设“ $1 + 1 = 2$ ”, 不具有任何现实意义。因此, 所有的假设检验, 针对的都是总体。正是因为我们不知道一致组和不一致组的总体均值, 才要去假设他们的关系, 即零假设 $\mu_i = \mu_c$, 备择假设 $\mu_i \neq \mu_c$ 。

解释: 感谢给出的建议, 确实自己也有表述不清楚的地方, 无论如何, 这样的讨论都是非常有意义的, 非常受益!

- 1) 假设确实都是针对总体参数, 如均值, 方差等。这个大原则是没有错误的。
- 2) 两个分组测试的样本均值以及根据相应差异时间的差异分布样本值, 确实首先是样本参数, 这个也是在我们做分析(即假设)之前的真实情况, 然后在我们做了假设的前提下, 这两组样本测试结果实际上转化了为总体测试结果, 但是必须是在我们 H_0 以及 H_a 的前提条件下。所以我的疑问和你的质疑并不矛盾, 我实际强调说明的是假设之前的情况, 而你需要做的是在做假设之后的准确描述。
- 3) 基于以上两点, 我们得到了假设前提下的差异分布的应该是总体分布对应的参数, 包括均值, 方差, 以及 T 统计值及其 P 值等等, 并且利用反正法, 通过经验建议 α 水平=0.05 验证 H_0 是否真正成立。最终得到关于该假设的总体分布评估的实际结论, 即 Reject H_0 , Accept H_a 。

Q2b: 因为我们不知道总体 population 的分布情况”说法并不准确。虽然我们确实不知道总体的分布, 但由于 t 检验的条件限制和日常经验, 我们一般会认为总体符合正态分布。不能使用 z 检验的根本原因, 在于我们不知道总体的某些除了分布以外的其他信息, 请指出这些信息具体是什么。

解释：确实需要具体一点，虽然总体分布理论是正态分布，但我们不知道其均值和方差，所以需要 CI 95% 置信水平，大致确定其总体差异分布的均值水平范围。另外，通过计算差异样本得到方差或者标准差，也可以近似去估计其总体测试差异的标准差或方差。这样一个总体的正态分布就基本了解了。（其他的相关描述值，比如包括 R^2 , Cohn'd, low limit, up limit, confidence level 等等，都可以得到了）（checked by reviewer）

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。（四舍五入，保留小数点后两位）

均值 $X_{cspar} = 14.05$ ，标准差 $S_c = 3.56$, $n_1 = 24$

均值 $X_{ispar} = 22.02$ ，标准差 $S_i = 4.80$, $n_2 = 24$

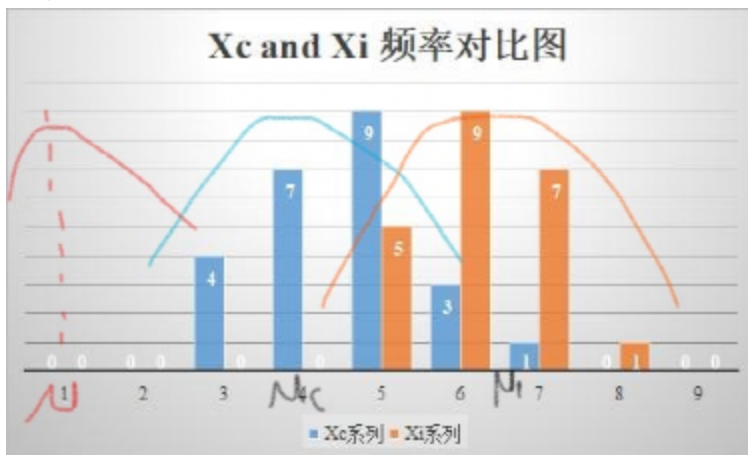
两组对应样本数据均值差异 $\text{average of } (X_{cspar} - X_{ispar}) = -7.96$

Standard Deviation of the $(X_c - X_i) = 4.86$

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。

$T \text{ statistics} = (X_{cspart} - X_{ispart}) / (SD / \sqrt{n}) = -7.96 / (4.86 / \sqrt{24}) = -8.02$

观察一致组的测试平均比不一致组的时间平均要少用 6 到 10 秒。（根据 95%CI）



说明：Xc 系列为一致组测试用时频率结果，样本均值为 14.05

Xi 系列为不一致组测试用时频率结果，样本均值为 22.02

注：X 轴为单位 4 长度,左侧红色分布图为大概相对位置。

可以观察到两组为近似正态分布，左侧红色的分布是 $X_c - X_i$ 的分布，理论上应该也符合正态分布，均值为 -7.96。

另外，三组分布的标准差比较接近都在 4 左右，其中 $X_c - X_i$ 分布的标准差最大，为 4.86，形状会稍微宽一些。另外两个为 3.56 和 4.80。

所以随着样本数的增加，三个分布都会呈现正态分布。

另外，也可以归一化表示，即差值分布使用公式 $(X - X_{spar}) / SD$ ，曲线会更加明显一些。

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？

$T_{critical} (\alpha=0.05, DF=23, two-tailed) = + or - 2.069$

双尾假设对应的两个置信水平分布如下：

置信水平负方向分布 95% CI = $(-7.96 - 2.069 * 4.86 / \sqrt{24}), -7.96 + 2.069 * 4.86 / \sqrt{24}) = (-10.01, -5.91)$

置信水平正方向分布 95% CI = $(7.96 - 2.069 * 4.86 / \sqrt{24}), 7.96 + 2.069 * 4.86 / \sqrt{24}) = (5.91, 10.01)$

计算 P 值，根据 T table，我们观察到当我们得到的 T statistics 值为 -8.02，其位置远大于 T 表中左侧的数值，可以看出，此值对应的概率值 P 几乎为 0。

结论：reject H0. 因为 T 统计值为 $-8.02 < -2.069$ (左侧 2.5% 范围之内)，即落入总体 5% 显著变化的区间中。所以结果与期望是一致的，即 Xc 一组的测试引入了显著的变化。根据 P 值接近为 0 的结论，观察我们的 H0 假设，即测试前后的用时完全没有变化的假设的可能性 P 接近为 0。

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！

观察者在读取的文字和图像根据常识进行某种关联，如果符合常识，则分别很快结束。如果不能，则要有一个反复或者重新确认的时间。经典自动化理论解释是：字词的语义属于自动加工，故速度较快，而颜色的命名属于控制加工，故速度较慢。

常见的类似任务比如根据提示，比如右手，则需要抬起左手等等！

优达学城

2016 年 9 月