

Project 1: 检验心理学现象 - Stroop 效应

Q1: 我们的自变量是什么？因变量是什么？

自变量是文字条件（一致文字，不一致文字）。

因变量是说出同等大小的列表中的墨色名称的时间。

Q2: 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）。

零假设：两种文字条件下，说出同等数量墨色名称所使用的时间无区别。

对立假设：两种文字条件下，说出同等数量墨色名称所使用的时间有较大区别。

μ_1 表示一致文字条件下的总体均值 μ_2 表示不一致文字条件下的总体均值

H_0 表示零假设 H_A 表示对立假设

$$H_0: \mu_1 - \mu_2 = 0$$

$$H_A: \mu_1 - \mu_2 \neq 0$$

我希望执行双尾检验 ($\alpha=0.05$)，因为我不确定两种不同条件下，哪一个使用的时间更少，或者更多，我并没有一个方向导向，因此我需要使用双尾检验。

在 z 检验和 t 检验中我会选择 t 检验，是因为 z 检验需要知道总体均值 μ ，总体标准偏差 σ ，但往往在统计数据中给出的都只是 n 个样本数据，例如本次 stroop 实验，这样 t 检验正好可以胜任，t 检验需要用到的都是样本数据，样本均值，样本标准偏差，样本标准误差。

本次 stroop 实验使用的是相依样本，因为两次不同条件下的数据来自同一群人。

综上所述，我们可以总结一下使用 t 检验是否满足了全部的前提：

(<http://www.csic.cornell.edu/Elrod/t-test/t-test-assumptions.html>)

- 1) 该实验的自变量确实是两种取值，即两种不同的文字条件，因此满足条件！
- 2) 因变量是时间数，属于连续变量（类似的还有尺寸，体积，高度），因此满足条件！
- 3) 因为本次实验是同一群人，属于相依样本，这群人中每个参与者之间都是互相独立的，因此满足条件！（btw：通过每个人得出的两个时间数并不独立，两个数据是依附在同一个人参与者身上的相依关系，这也是相依样本的体现）。
- 4) 通过 Q4 中绘制的直方图，可以清晰的看到两个不同文字条件下得出的时间数量都是正态分布的，因此满足条件！

1, 2, 3, 4 条件均满足！因此我们完全可以信赖本次 t 检验的结果报告！

Q3: 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。

经过对数据集的计算，得出如下描述性统计：

Congruent mean(一致性条件下的样本均值)： $\bar{X}_1 = 14.05$

Congruent median(一致性条件下的样本中位数)： $M_1 = 14.36$

Congruent sd(一致性条件下的样本标准偏差)： $S_1 = 3.56$

Incongruent mean(非一致性条件下的样本均值)： $\bar{X}_2 = 22.02$

Incongruent median (非一致性条件下的样本中位数)： $M_2 = 21.02$

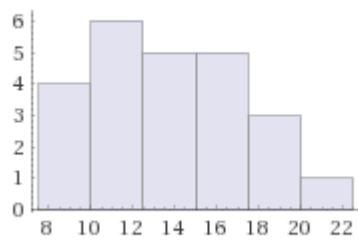
Incongruent sd(非一致性条件下的样本标准偏差)： $S_2 = 4.80$

附：使用 excel 计算：均值-》离均差-》平方偏差-》平方偏差之和-》方差-》标准偏差

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Congruent	Incongruent		8.63	15.687			Congruent						
2	12.079	19.278		8.987	17.394	14.051125	<=contruent 均值	3.889277016	12.66902907	7.496188	23.01175704			
3	16.791	18.741		9.401	17.425	14.3565	<=contruent 中位数	7.506915016	3.559357958	<=标准差	10.72508	4.797057122	<=标准差	
4	9.564	21.214		9.564	17.51			20.13429077			0.64307			
5	8.63	15.687		10.639	17.96	22.01591667	<=incongruent 均值	29.38859627			40.05519			
6	14.669	22.803		11.344	18.644	21.0175	<=incongruent 中位数	0.381769516			0.6195			
7	12.238	20.878		12.079	18.741			3.287422266			1.294854			
8	14.692	24.572		12.13	19.278			0.410720766			6.533562			
9	8.987	17.394		12.238	20.33			25.64536202			21.36211			
10	9.401	20.762		12.369	20.429			21.62366252			1.572307			
11	14.48	26.282		12.944	20.762			0.183933766			18.19947			
12	22.328	24.524		14.233	20.878			68.50665977			6.290482			
13	15.298	18.644		14.48	21.157			1.554697266			11.36982			
14	15.073	17.51		14.669	21.214			1.044228516			20.30329			
15	16.929	20.33		14.692	22.058			8.282164516			2.842315			
16	18.2	35.255		15.073	22.158			17.21316377			175.2733			
17	12.13	22.158		15.298	22.803			3.690721266			0.020188			
18	18.495	25.139		16.004	23.894			19.74802502			9.75365			
19	10.639	20.429		16.791	24.524			11.64259702			2.518305			
20	11.344	17.425		16.929	24.572			7.328525766			21.07652			
21	12.369	34.288		18.2	25.139			2.829544516			150.604			
22	12.944	23.894		18.495	26.282			1.225725766			3.527197			
23	14.233	17.96		19.71	34.288			0.033078516			16.45046			
24	19.71	22.058		22.328	35.255			32.02286627			0.001771			
25	16.004	21.157						3.813720766			0.737738			
26														
27														
28														
29														

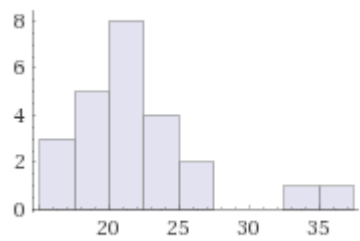
Q4: 提供显示样本数据分布的一个或两个可视化。用一两句话说明你从图中观察到的结果。

Histogram:



一致性条件下的直方图

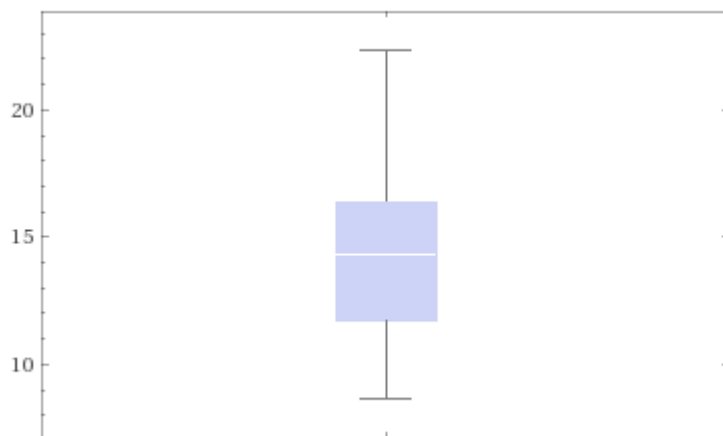
Histogram:



非一致性条件下的直方图

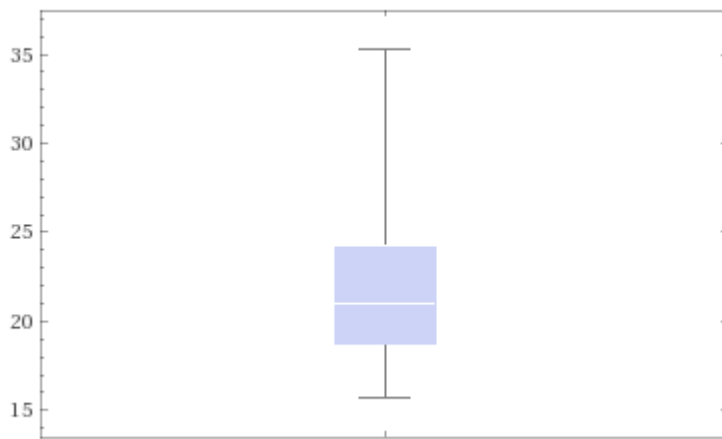
通过两种不同条件下的直方图，可以看到都呈现正态分布，并且是正偏斜分布，其中众数（mode）的差别较大，一致性的众数为 11，非一致性的众数为 21，初步的直观感受是在非一致性条件下，相依样本识别同样墨色数量的时间变长了

Box-and-whisker chart:



一致性条件下的箱线图

Box-and-whisker chart:



非一致性条件下的箱线图

通过箱线图可以看出两种条件下的分布都没有出现异常值。

Q5: 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？

统计报告如下：

1. 统计描述

两种条件的差异的样本均值： $\bar{X} = -7.96$

两种条件的差异的样本标准偏差： $S = 4.86$

两种条件的差异的样本标准误差： $SEM = 0.99$

2. 推论统计值

- 1) 相依样本的双尾检验, $\alpha=0.05$, t 临界值= -2.069 or +2.069
- 2) 该样本自由度 $df=23$
- 3) 计算得出 t 值=-8.02
- 4) 计算得出 p 值<0.0001
- 5) 95% CI (-10.02 , -5.91)
- 6) cohen $d = -1.64$
- 7) $r^2 = 0.7366$

APA Style :

$t(23)=-8.02, p<0.0001$, two-tailed

3. 推论结果

从 t 值可以看出已经越过-2.069 临界点 3 倍之多，以及从 p 值<0.0001 可以得

知此结果具有显著统计意义，因此可以拒绝零假设。

与我的预期一致。

95% CI 值表示相对于不一致文字条件，在一致文字条件下所使用的时间要少大约 10 到 5；

Cohen d 值表示差异的均值距离总体均值的距离大概是-1.64 个样本标准差，即效应量的一个度量指标（效应量度量类型分为差异度量和相关度量：cohen d 就是差异度量类型中的标准化差异，另外还有一个就是最简单的差异度量-均值差异）；

r^2 值表示对于 24 个人的样本来说有接近 74%的时间数值差异是由不一致文字条件引起的（效应量度量类型分为差异度量和相关性度量： r^2 就是相关度量类型中的一种，它表示某个变量的变化比例）；

Q6: 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！

Thinking...

附参考资料：

<http://www.csic.cornell.edu/Elrod/t-test/t-test-assumptions.html>

<http://discussions.youdaxue.com/t/topic/42566/4>