

项目一 检验心理学现象

一. 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。

二. 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, 即文字和颜色是否一致

因变量: 测试者说出同等大小的列表中墨色名称的时间

(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)。

此任务其适当的假设集如下:

μ_{con} : 在一致文字条件下说出墨色名称时间的总体均值。

μ_{incon} : 在不一致文字条件下说出墨色名称时间的总体均值。

并做出如下的假设:

零假设 H_0 : 在一致文字条件和在不一致文字条件下, 说出墨色名称时间是一样的, 没有显著差异, 即 $\mu_{con}=\mu_{incon}$ 。

备择假设 H_1 : 在一致文字条件和在不一致文字条件下, 说出墨色名称时间是不同的, 即 $\mu_{con}\neq\mu_{incon}$ 。

根据提供的条件来说, 在此执行相依样本双尾 t 检验, 取显著性水平 $\alpha=0.05$, 其理由如下:

1. 样本是同样的人在不同条件下重复测试得来的。
2. 提供了两个不同测试条件下的样本数据, 样本含量很小, 而且总体标准差以及均值都是未知, 因此选择相依 t 检验。
3. 只是检验说出墨色名称时间是否一致, 并没有方向性, 因此采用双尾检验。
4. 满足使用配对样本 t 检验的前置条件:
 - (1) 这里参与试验的人都是随机选的, 所以得到的是随机样本。
 - (2) 两样本均数比较时, 样本数据可以估计总体的方差。
 - (3) 观察样本数据, 两样本呈现出正态分布的趋势。

(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。

集中趋势测量:

1. 均值:

一致文字条件: $\mu_{con}=14.051125$

不一致文字条件: $\mu_{incon}=22.015917$

2. 中位数:

一致文字条件: $median_{con}=14.3565$

不一致文字条件: $median_{incon}=21.0175$

3.众数:

一致文字条件: 所有数据都不同, 均匀分布

不一致文字条件: 所有数据都不同, 均匀分布

变异测量:

1.全距:

一致文字条件: $R_{con}=13.698$

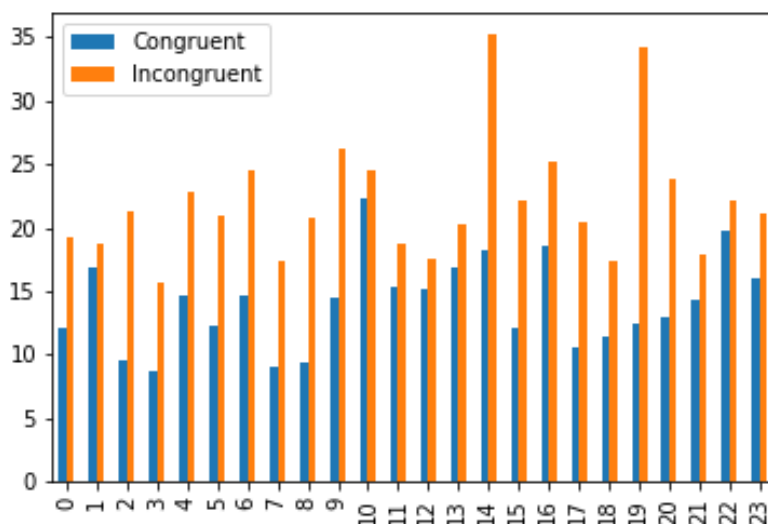
不一致文字条件: $R_{incon}=19.568$

2.样本标准偏差:

一致文字条件: $var_{con}=3.559358$

不一致文字条件: $var_{incon}=4.797057$

(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。



从图中可以看出不一致文字条件 Incongruent 下说出墨色名称时间均大于一致文字条件 Congruent 下所用的时间。

(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?

根据前面的分析, 在这里采用相依双尾 t 检验, 显著性水平 $\alpha=0.05$, 自由度 $v=23$, 则查表可得临界值 $t_c = 2.069$ 。

在这里设两个样本分别为 x_{1i} 和 x_{2i} , 则两个样本差值为:

$$d_i = x_{1i} - x_{2i}$$

相依样本差异均值为:

$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n}$$

经计算可得 $\bar{d} = 7.96479$, 具体计算程序见附件 Data.ipynb。

相依样本差值的标准偏差为:

$$s_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1}}$$

经计算可得 $s_d = 4.86483$ 。

t 统计量分布为：

$$t = \frac{\bar{d} - \mu_0}{s_d/\sqrt{n}} = \frac{\bar{d} - 0}{s_d/\sqrt{n}}$$

经计算可得 $t = 8.020706$ 。

由前面查表得到的临界值 $t_c = 2.069$ 可知， $t > t_c$ ，则有 $p < \alpha$ ，则拒绝零假设。

结论：

经过计算可知，在此拒绝零假设，即一致文字条件下与不一致文字条件下，说出墨色名称时间是不一样的，而且从数据上来看，不一致文字条件下说出墨色名称时间是普遍大于一致文字条件下。

三. 参考内容

1. 优达课程 Intro to Inferential Statistics

2. 假设检验

<https://baike.baidu.com/item/%E5%81%87%E8%AE%BE%E6%A3%80%E9%AA%8C/638320>

3. t 检验 <https://baike.baidu.com/item/t%E6%A3%80%E9%AA%8C>

4. 集中趋势与变异统计量 <https://wenku.baidu.com/view/5daa3bc9f705cc17552709b4.html>

5. 《深入浅出统计学》

四. 附件

数据处理的程序文件：Data.ipynb