

P1 检验心理学现象 Report

R语言分析

背景介绍

Stroop效应是心理学中的一个名词，**百度百科**中对其介绍如下：

斯特鲁普效应(Stroop effect)在心理学中指干扰因素对反应时间的影响。例如当测试者被要求说出某个颜色和其覆盖的文字字面意义不符的词语时，被测者往往会反应速度下降，出错率上升。

准确的表述是当人们对某一特定刺激作出反应时，由于某种因素的干扰，被刺激者难以集中精力对特定刺激作出反应的现象叫“斯特鲁普效应”，因为这个效应是斯特鲁普发现的，故以其名斯特鲁普效应。

本项目意图通过虚拟实验并对实验结果进行统计分析来验证Stroop效应是否真实存在，以及其影响程度。

实验设计

本次实验为虚拟实验，我们针对一组受试者展开实验。实验的**自变量**为对同一组受试者的不同色彩单词展示，**因变量**为受试者完成一组单词颜色识别所用的时间（秒）。自变量取值如下：

1. 向受试者展示的单词与其颜色一致；
2. 向受试者展示的单词与其颜色不一致；

实验过程如下：

- 向受试者展示第一组彩色单词，文字内容与其色彩一致；
- 受试者自行读出每个单词的颜色，优先保证正确率，读完后记录完成时间；
- 向受试者展示第二组彩色单词，文字内容与其色彩不一致；

- 受试者自行读出每个单词的颜色，优先保证正确率，读完后记录完成时间；
- 记录每一个受试者前后两组所用时间，**分别计入样本1 和 样本2**，然后进行统计分析。

可点此[Stroop效应实验](#)进行试验。实验中我们优先保证受试者对单词颜色识别的正确率，即记录受试者100%正确完成颜色识别所用的时间。如此可以避免某些受试者牺牲正确率以追求反应时间，造成实验数据的偏倚。

注：由于操作性的关系无法获取真实的实验数据，本次虚拟实验的数据采用Udacity提供的一个样本量为24的配对样本数据集。点此可见[数据集](#)。

初步分析

先读入数据并作出频率直方图，以对数据的整体分布有直观的了解：

```
1. pf <- read.csv('stroopdata.csv')
2. head(pf)
```

	Congruent <dbl>	Incongruent <dbl>
1	12.079	19.278
2	16.791	18.741
3	9.564	21.214
4	8.630	15.687
5	14.669	22.803
6	12.238	20.878

6 rows

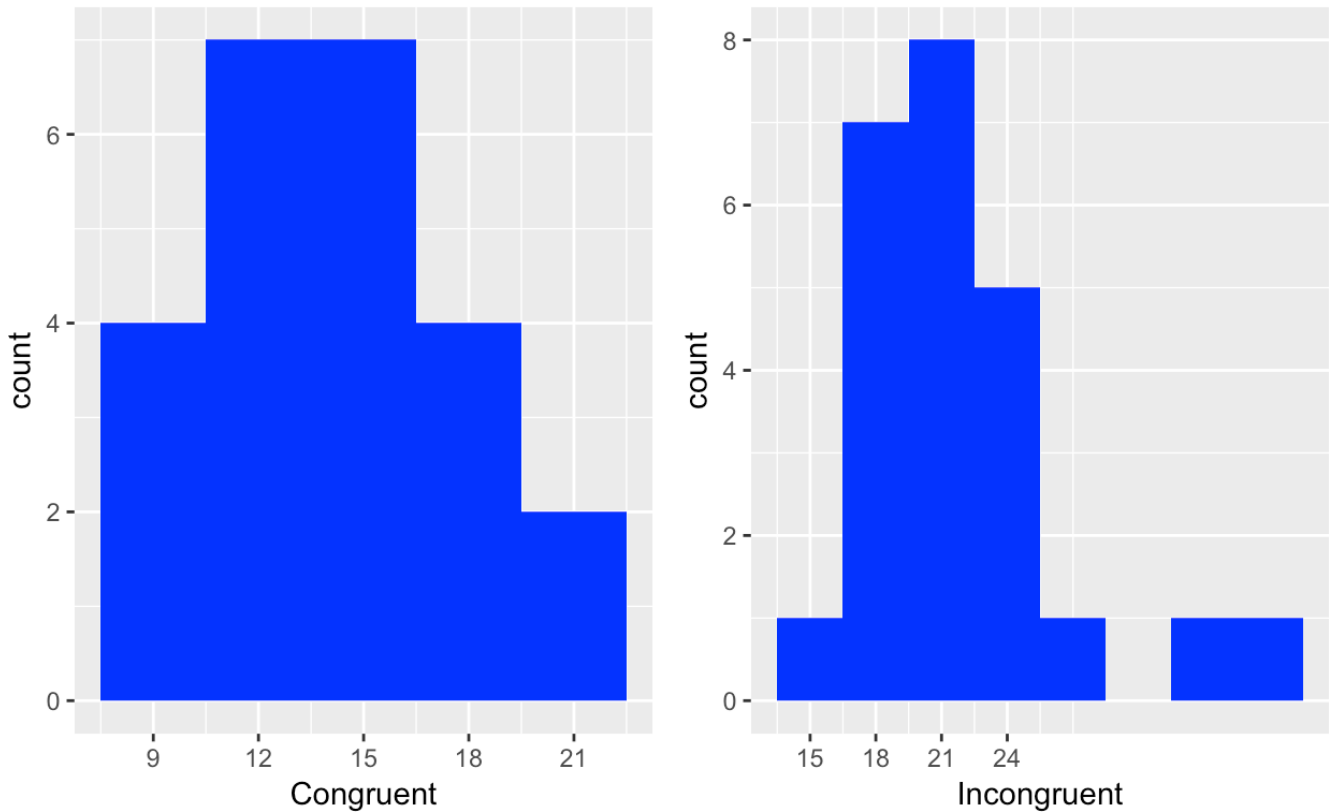
可以看出，数据集共有两个变量，Congruent和Incongruent。分别记录了受干扰和不受干扰的两组受试者完成测试所用的时间，单位是秒。

```
1. library(ggplot2)
2. library(dplyr)
3. library(gridExtra)
4.
5. p1 <- ggplot(data = pf, aes(x = Congruent)) +
6.   geom_histogram(binwidth = 3, fill = 'blue') +
7.   scale_x_continuous(breaks = seq(9, 24, 3))
```

```

8.
9.   p2 <- ggplot(data = pf, aes(x = Incongruent)) +
10.     geom_histogram(binwidth = 3, fill = 'blue') +
11.     scale_x_continuous(breaks = seq(9,24,3))
12.
13.   grid.arrange(p1,p2,ncol = 2)

```



从图中可以看出：

- 样本1的数据分布近似与正态分布；
- 样本2数据如果刨除最右端的极值，也近似于正态分布；
- 样本2的数据普遍大于样本1；

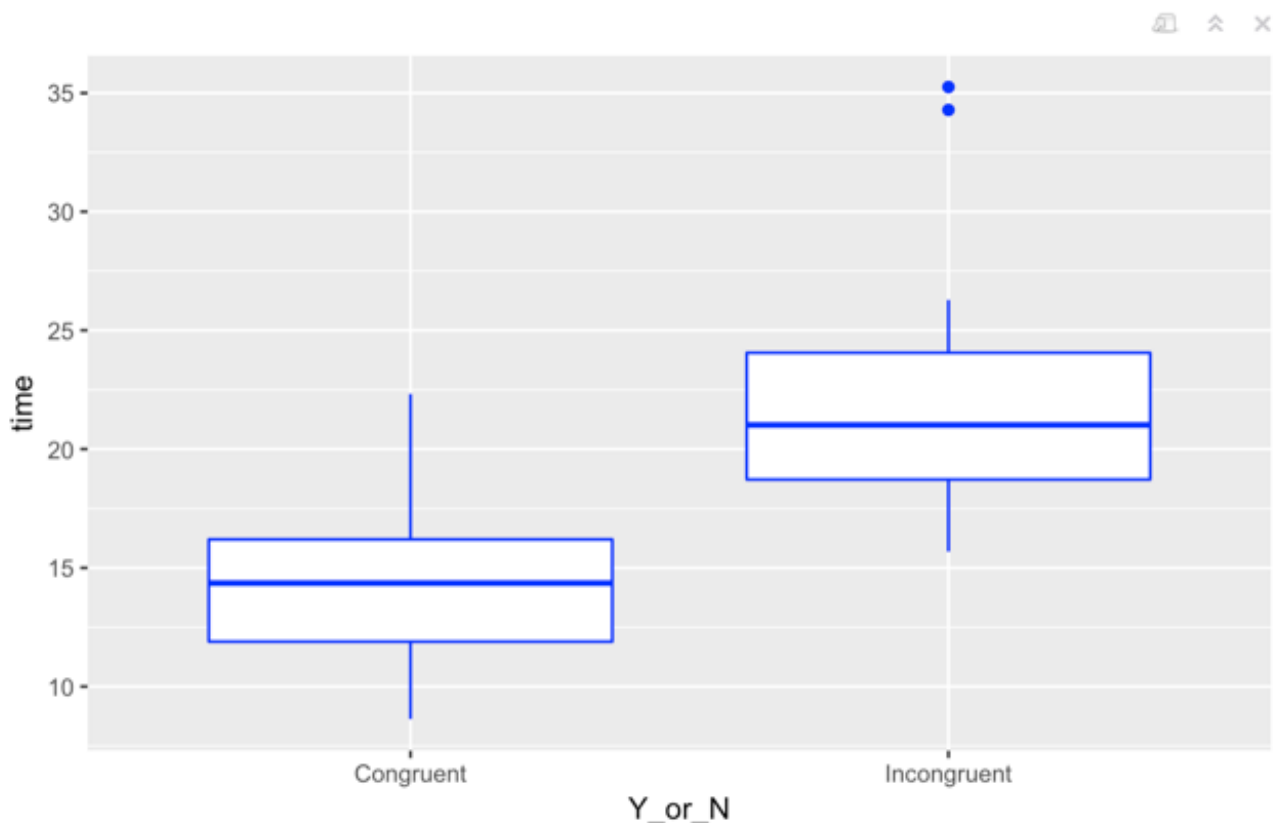
`summary(pf)`

做简单的统计和观察，可以发现样本2的中位数和均值均高于样本1:这意味着**我们有理由怀疑当存在干扰因素时，人们的反应时间确实有所延长**

Congruent	Incongruent
Min. : 8.63	Min. :15.69
1st Qu.:11.90	1st Qu.:18.72
Median :14.36	Median :21.02
Mean :14.05	Mean :22.02
3rd Qu.:16.20	3rd Qu.:24.05
Max. :22.33	Max. :35.26

此外，还可以绘制出两个样本数据的箱线图

```
1. library(tidyr)
2. pf_long <- gather(pf, key = 'Y_or_N', 'time', 1:2)
3. ggplot(data = pf_long, aes(x = Y_or_N, y = time)) +
4.   geom_boxplot(color = 'blue') +
5.   scale_y_continuous(breaks = seq(5,35,5))
```



从图中可以看出样本1的数据集中在15s附近，而样本2的数据集中在20s附近，也就是样本2的数据要普遍大于样本1，这与我们在直方图中观察到的现象一致。另外我们还可以发现，样本2的极大值距离中位数较远，可能为异常值。

对样本2的数据进行计算得出其 $IQR = 5.33$ ，而样本2的极大值 $X_{max} = 35.255$ ，上四分位

数 $Q_3 = 24.0515$, 由于 $X_{max} > Q_3 + IQR * 1.5$, 我们可以判定样本2的极大值为异常值。

统计检验

对于本次实验，我们具体的实验目的是检验干扰因素的存在是否确实会延长人们的反应时间（*Stroop*效应），并观察*Stroop*的影响程度。由此我们应当对样本的均值进行假设检验。根据我们在频率直方图中的观察，两个样本的分布都非常接近正态分布的形态，我们可以大胆的猜测样本所属总体服从正态分布。考虑到我们无法获得所有人反应时间的总体数据，只能通过实验的记录样本数据对总体进行估计，且样本量比较小（ <30 ），因此首先考虑对样本均值进行 **t检验**。另外，我们是针对同一组受试者进行两组实验，因此两次实验的结果属于相关样本，所以我们对本次实验采用相关样本t检验。

我们可以做如下假设：

- **零假设 H_0** : 单词与其颜色不一致对反应时间没有显著影响，即 $\mu_1 - \mu_2 = 0$;
- **备择假设 H_a** ：单词与其颜色不一致会导致反应时间显著加长，即 $\mu_1 - \mu_2 < 0$; (其中 μ 表示样本所在总体的均值，这同时意味着我们要进行单尾检验，检验方向为左尾)

利用统计函数，将两个样本的相关统计参数计算出来，如下：

```
1. s <- pf$Congruent - pf$Incongruent
2.
3. mean_s <- mean(s)
4. std_s <- sd(s)
5.
6. mean_s
7. std_s
```

```
[1] -7.964792
[1] 4.864827
```

样本	μ (均值)	S_d (标准差)	df (自由度)
样本1 - 样本2	-7.96	4.86	23

设定置信水平 $1 - \alpha = 0.95$ （即显著性水平 $\alpha = 0.05$ ），并且我们所做的t检验为单尾检验，代入自由度后查表得t界值： $t_{critical} = -1.714$

根据相关样本的t检验公式：

$$t_{statistic} = \frac{\bar{X} - \mu_0}{SE} ; SE = \frac{S_d}{\sqrt{n}}$$

得到样本统计量t值 $t_{statistic} = -8.02$

因为 $t_{statistic} < t_{critical}$ ，所以 $p < \alpha$ （如果通过在线工具进行计算，可以得知p值小于0.0001），所以拒绝 H_0 假设，接受 H_a 假设。

即，在置信水平 $1 - \alpha = 0.95$ 上，备择假设 H_a 是统计显著的。由此可以认为，Stroop效应确实存在。这恰好符合了实验设计时我们对实验结果的预期。

之后我们通过计算决定系数 R^2 和均值差异 Cohen's d 来判断Stroop效应的影响程度：

根据公式：

$$\text{Cohen's } d = \frac{\bar{X}_1 - \bar{X}_2}{S_p \text{ (合并标准差)}} \quad \left(\text{此处可用两样本差值的标准差 } S_d \text{ 来代替 } S_p \right)$$

得到Cohen's d = -1.64，即样本1和样本2均值的差异为1.64个标准差。

根据公式：

$$R^2 = \frac{t^2}{t^2 + df} ; df = n - 1$$

得到决定系数 $R^2 = 0.74$

这说明样本2和样本1之间的统计性差异中有74%的部分是由于干扰因素引起，即样本的统计性差异中约有74%是Stroop效应引起。

检验结论

对于样本集 ($\mu = -7.96$, $S_d = 4.86$), 我们有以下结论：

- 统计推论： $t(23) = -8.02, p < .05$, 单尾检验
 - 效应量测量： $d = -1.64, R^2 = 0.74$
-

总结与思考

根据以上分析与检验，我们可以得出以下结论：

1. 在显著性水平 $\alpha = 0.05$ 上，Stroop效应是确实存在的；
2. 对于已知的两个样本，它们统计性差异为1.64个标准差；
3. 对于样本的统计性差异而言，有约74%的部分是由Stroop效应引起的；

对于以上结论，我认为其内在原因是这样的：

由于思维习惯的作用，当人在看到单词等文字内容时，会自发的将更多的注意力集中在文字内容上，而不是颜色、形状、大小等其他因素。这一过程是自发的、潜意识的。但当我们需要受试者将注意力集中在文字颜色上时，文字内容本身就成了干扰因素。在进行实验的时候，受试者为了准确说出单词的颜色，就必须排除掉这个干扰因素，也就是要有意识的控制自己的思维，将注意力集中在单词的颜色上而非其内容上。这将会耗费更多精力，也必然需要更多反应时间。

虽然以上实验的样本数据从统计上说明了Stroop效应的存在，但是由于实验并未对受试者的年龄、性别、情绪、测试时间等其他潜在因素进行控制，因此并不能排除其他因素对Stroop效应的影响。

为了更好的验证Stroop效应和其影响，我们可以通过一些手段来排除干扰因素对实验的影响，比如：

- 可以针对不同性别设计对照实验，以排除性别差异；
- 可以针对不同年龄段设计对照实验，以排除年龄因素的影响；

- 比较麻烦的是情绪因素带来的干扰，这个比较难以排除。但是考虑到情绪通常是主观的、个人化的因素，一般很少出现多个受试者在同时段情绪不稳定。因此我们可以增加样本容量（例如达到100）以降低个别受试者的情绪波动对实验数据的影响。