

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：两种不同的实验条件（即一致文字条件和不一致文字条件）

因变量：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？你想执行什么类型的统计测试？为你的选择提供正当理由。

以下为修改后内容：

该任务的适当假设集：

零假设： $\mu_1 \geq \mu_2$ 即 Congruent 下总体平均值大于等于 Incongruent 下总体平均值

备择假设： $\mu_1 < \mu_2$ 即 Congruent 下总体平均值小于 Incongruent 下总体平均值

其中： μ_1 表示的是 Congruent 下测试者所用时间平均值， μ_2 表示的是 Incongruent 下测试者所用时间平均值。

测试类型：相依样本 t 单侧检验，由于样本较小所以使用 t 检验。使用单侧检验的目的在于，我并不关系这个总体平均值是否相等，而是比较关系两个总平均值的大小关系。由于本测试是由同一个测试者对不同条件进行的测试，所以得到的两个样本是相依样本。

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。

| | 描述性 | Congruent | Incongruent |
|--------|-----------|-------------|---------------|
| 集中趋势测量 | 平均值 | 14.051125 | 22.01591667 |
| | 中位数 | 14.3565 | 21.0175 |
| | 最小值 | 8.63 | 15.687 |
| | 最大值 | 22.328 | 35.255 |
| | 极差 | 13.698 | 19.568 |
| | 相对极差 | 0.9748 | 0.888849 |
| | 观测值 | 24 | 24 |
| 变异性测量 | 样本标准差 | 3.559357958 | 4.797057122 |
| | 四分位差（IQR） | 3.925 | 5.153 |
| | 特异值 | 22.328 | 35.255,34.288 |

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。

以下为第一次提交内容：

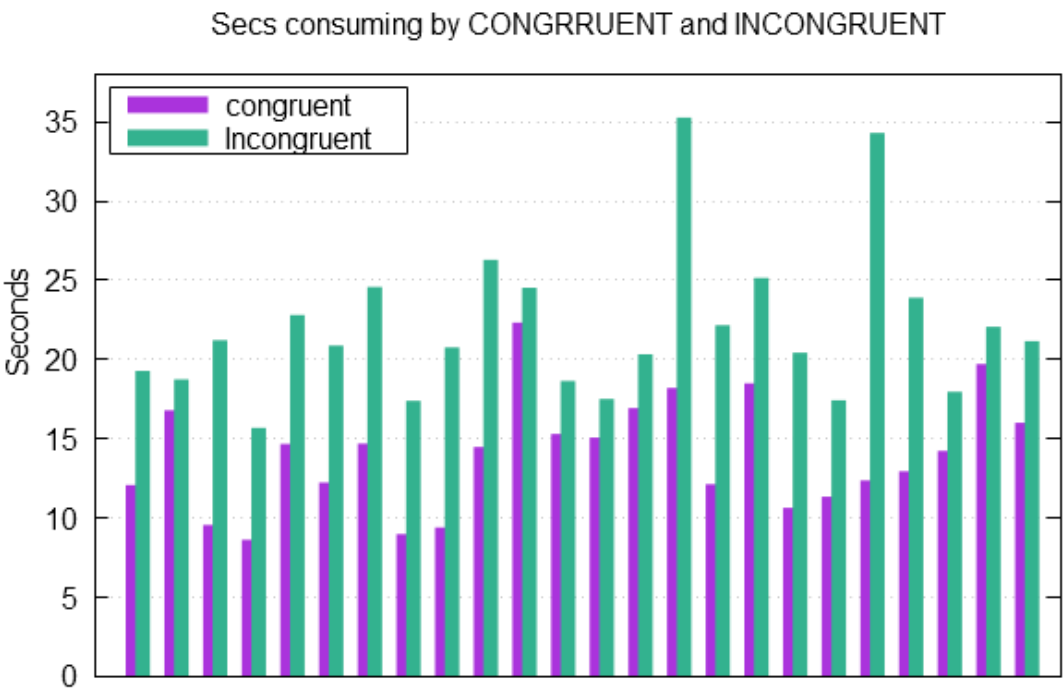


图 1 每个测试者在两种测试方案下的结果

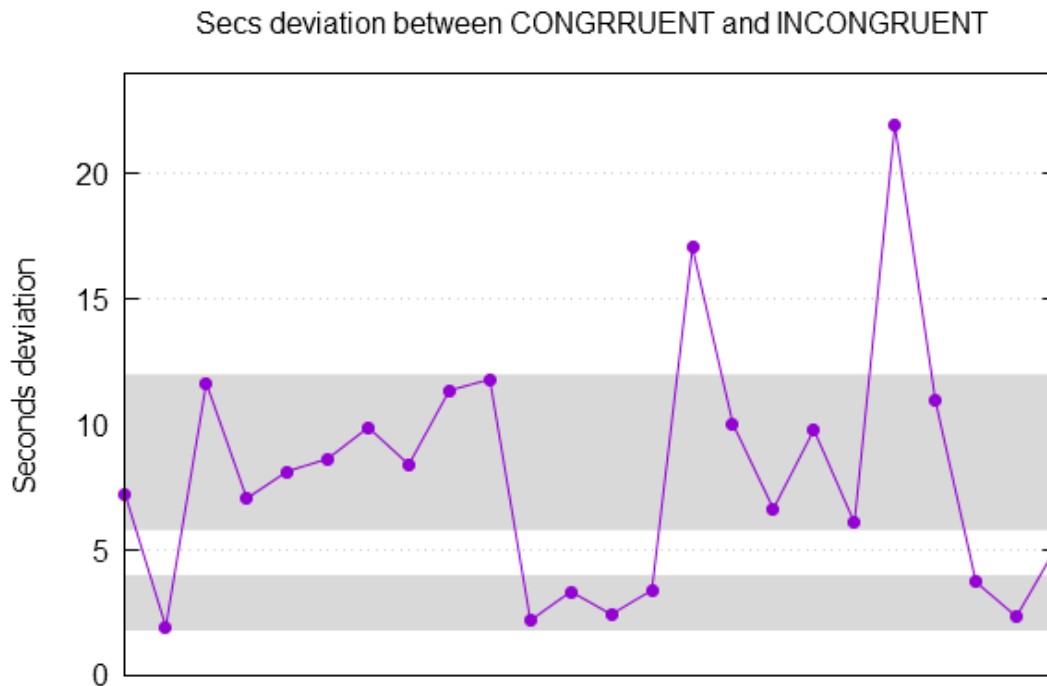
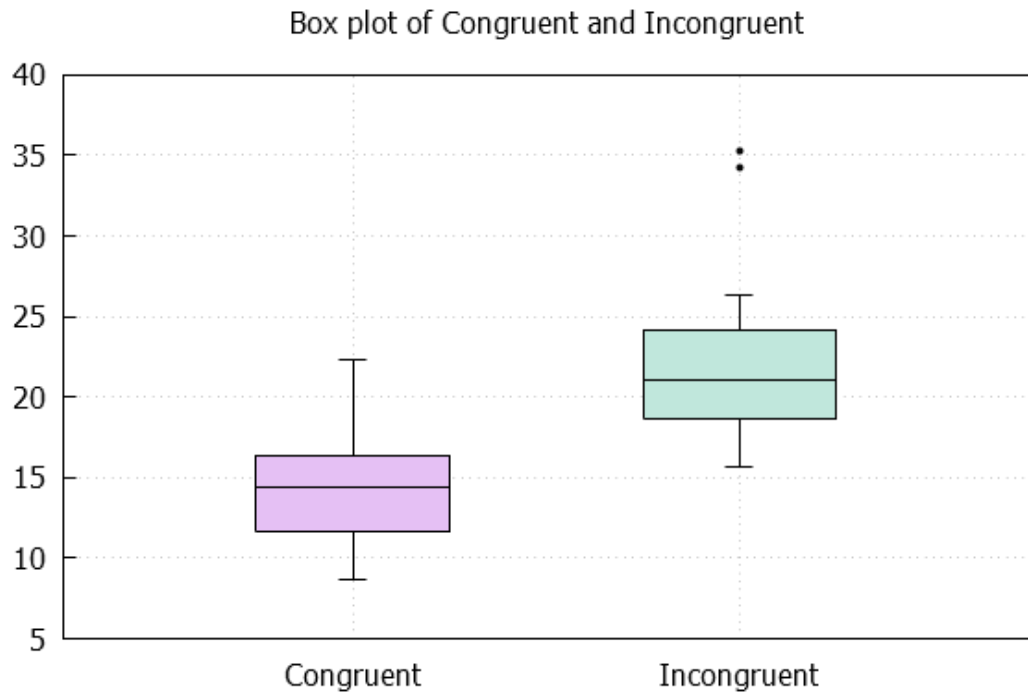


图 2 每个测试者两种情况下的所用时间差

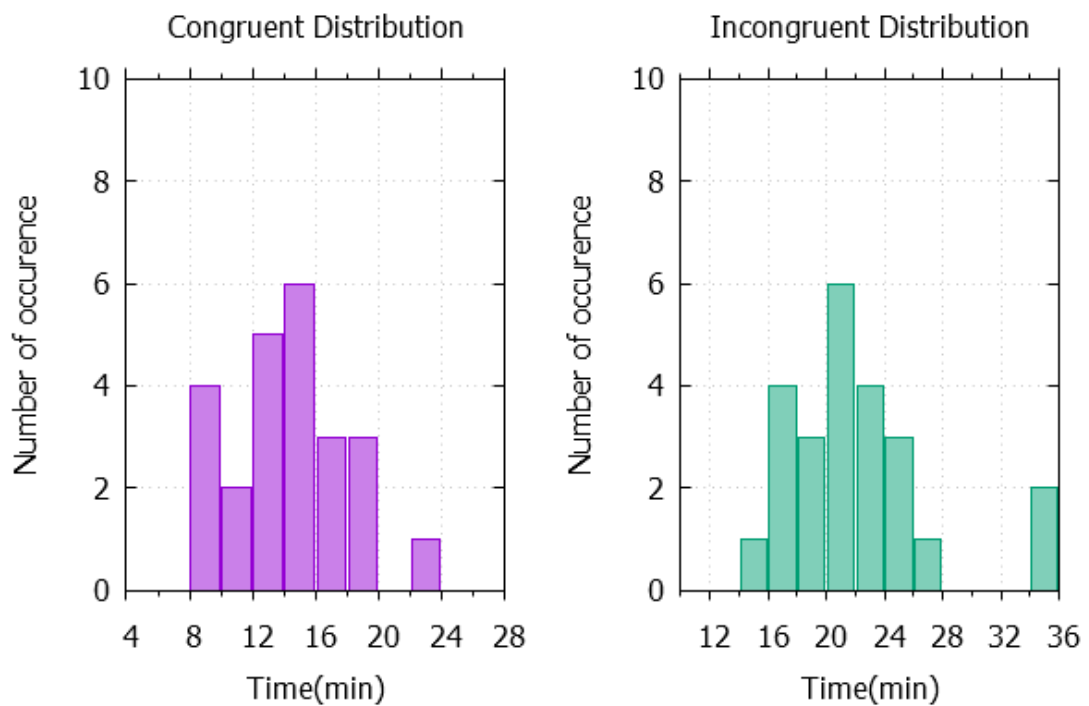
从图看出，对于所有测试者来说，在 incongruent 下的反应时间都要比 congruent 下的反应时间长，当然这点还是比较好理解的。但是我们想知道到底长多少呢，于是乎有了图 2，图 2 将每个测试者 incongruent 和 congruent 下的所用时间的差绘制在一条曲线上。从图 2 可以看出，不同测试者两种不同情况下所用测试时间差的变化很大，基本上趋于两个区间，一个是在 (5.8, 12) 一个是 (1.8, 4)，这两个区间一共落入了 21 个点，但是这两个区间本来就挺大，这说明不同的测试者对于这两种不同情况的反应时间变化区别较大，这一点可以从另外两个差值很大处得到验证（这两个异常值也直接导致了 Incongruent 的标准差要大于 congruent 的标准差）。个人觉得要想弄清楚测试时间差的趋势范围可能还需要更多的样本数据才行。

以下为修改后内容：

按照老师的修改意见，以上两副都是讲原始数据绘制在上面，这在数据量很大的时候是不能完成的（很同意），因此按照老师的建议绘制了以下的图形。



箱线图的确很有意思，从这两个图中可以清楚看到中位数、最大最小值以及异常值的分布情况，并且可以进行对比，而且在尽管在数据量很大的时候也能够方便进行分析。从图中可以看出 **Incongruent** 的数值明显高于 **Congruent**，而且在 **Incongruent** 下存在两个异常值。**Congruent** 中位数位于上半部分，说明其为负偏态分布，而 **Incongruent** 中位数位于下半部分，说明这种分布为正偏态分布。



这是对两种不同条件测试结果的频数统计，从图中可以看出，**congruent** 下的众数范围在 14-16 之间，由之前计算可知其平均值为 14.05，中位数为 14.35，其之间的大小关系，符合负偏态分布的特征，而 **Incongruentde** 众数在 20-22 之间，由之前计算可

知平均值为 22, 中位数为 21, 其大小关系符合正偏态分布, 和上一副图的结论一致。但是个人觉得从频数分布图上并不能很清楚地看出其为哪种偏态分布形式。

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?

以下为第一次提交内容:

置信水平为 95%, 关键统计量为 $Y-X$ (其中 Y 表示的是 incongruent 下的样本结果, X 表示的是 congruent 下的样本结果)。

非参数检验:

零假设: 认为每个人在两种不同条件下的测试时间差服从正态分布

为了验证其是否为正态分布, 本题令 $Z=Y-X$, 那么该问题即变化成了判断 Z 是否服从正态分布。下表数据源

| Congruent | Incongruent | Incon-Con | Z_i | F_{Z_i} | Z_{Fi} |
|-----------|-------------|-----------|--------|-----------|----------|
| 12.079 | 19.278 | 7.199 | 1.95 | 0.04 | -1.75 |
| 16.791 | 18.741 | 1.95 | 2.196 | 0.08 | -1.4 |
| 9.564 | 21.214 | 11.65 | 2.348 | 0.13 | -1.12 |
| 8.63 | 15.687 | 7.057 | 2.437 | 0.17 | -0.95 |
| 14.669 | 22.803 | 8.134 | 3.346 | 0.21 | -0.8 |
| 12.238 | 20.878 | 8.64 | 3.401 | 0.25 | -0.67 |
| 14.692 | 24.572 | 9.88 | 3.727 | 0.29 | -0.55 |
| 8.987 | 17.394 | 8.407 | 5.153 | 0.33 | -0.43 |
| 9.401 | 20.762 | 11.361 | 6.081 | 0.38 | -0.3 |
| 14.48 | 26.282 | 11.802 | 6.644 | 0.42 | -0.2 |
| 22.328 | 24.524 | 2.196 | 7.057 | 0.46 | -0.1 |
| 15.298 | 18.644 | 3.346 | 7.199 | 0.50 | 0 |
| 15.073 | 17.51 | 2.437 | 8.134 | 0.54 | 0.1 |
| 16.929 | 20.33 | 3.401 | 8.407 | 0.58 | 0.2 |
| 18.2 | 35.255 | 17.055 | 8.64 | 0.63 | 0.33 |
| 12.13 | 22.158 | 10.028 | 9.79 | 0.67 | 0.43 |
| 18.495 | 25.139 | 6.644 | 9.88 | 0.71 | 0.55 |
| 10.639 | 20.429 | 9.79 | 10.028 | 0.75 | 0.67 |
| 11.344 | 17.425 | 6.081 | 10.95 | 0.79 | 0.8 |
| 12.369 | 34.288 | 21.919 | 11.361 | 0.83 | 0.95 |
| 12.944 | 23.894 | 10.95 | 11.65 | 0.88 | 1.17 |
| 14.233 | 17.96 | 3.727 | 11.802 | 0.92 | 1.4 |
| 19.71 | 22.058 | 2.348 | 17.055 | 0.96 | 1.75 |
| 16.004 | 21.157 | 5.153 | 21.919 | 0.96 | 1.75 |

说明: Z_i 列是对 Incon-Con 列排序后的结果, F 是密度函数, 也就是左边对应的 Z_i 列元素在其左边 (包括) 的元素个数占样本总数的比例, 所以是依次递增的, Z_{Fi} 对应的是 F_i 的概率值下的 Z 分位数, 由查阅标准正态函数分布表得到。

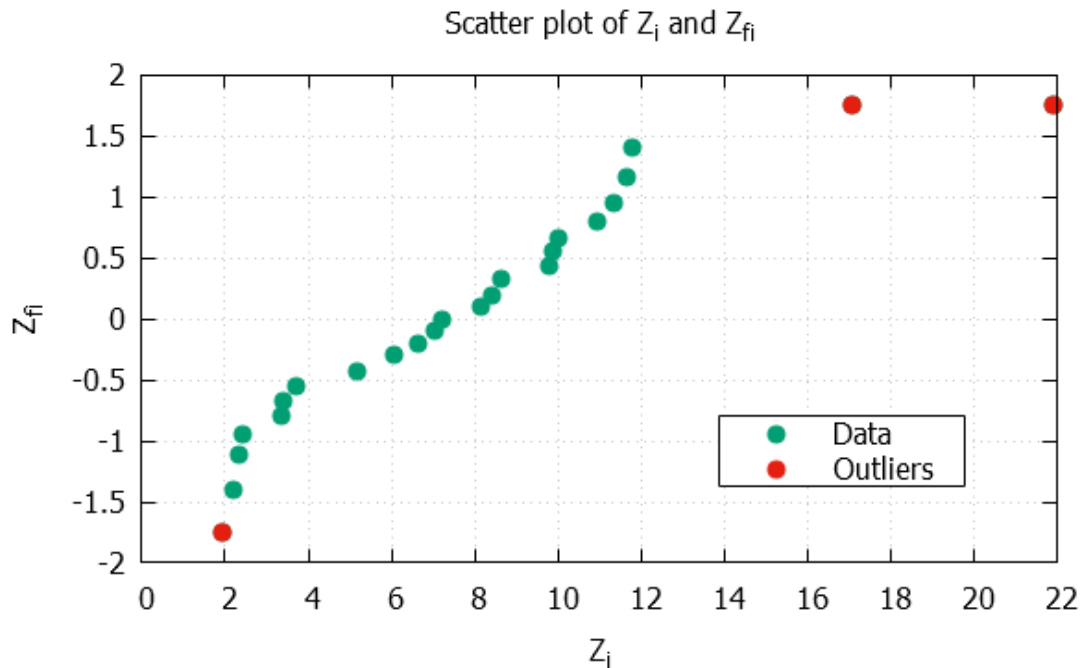
一. 散点图直观法检验 Z 值分布情况

在此先用散点图法对数据进行直观的检验。如果 Z_i 服从正态分布,那么 $y_i = \frac{z_i - \mu}{\sigma}$ 可

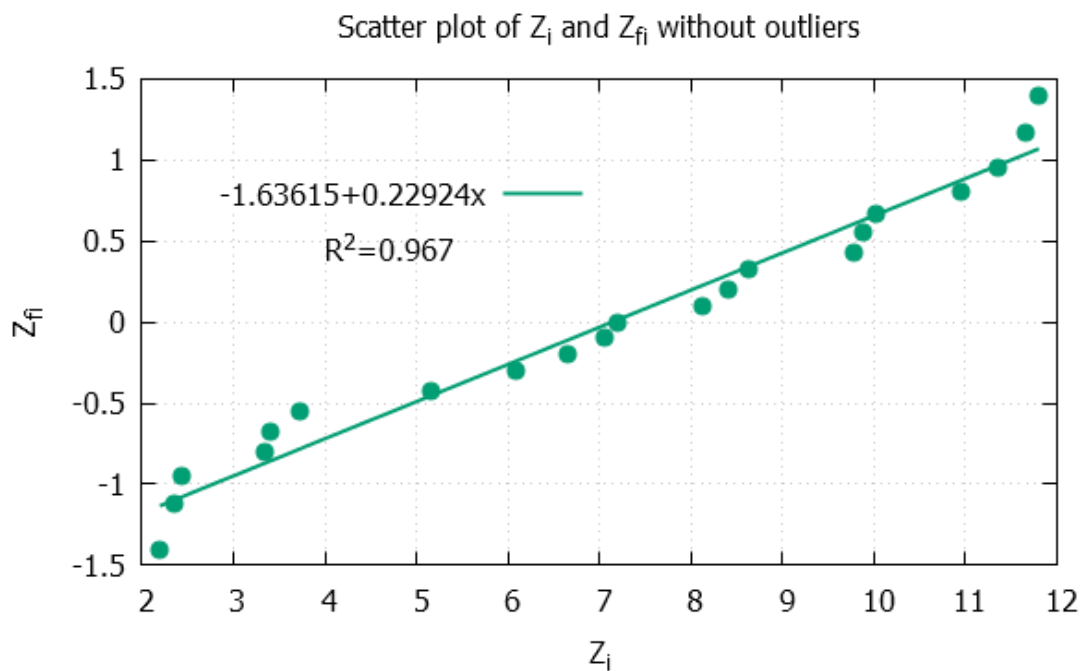
以将每个变量标准化,而 $F_i = \phi\left(\frac{z_i - \mu}{\sigma}\right) = \phi(z_{F_i})$ 说明 $y_i = z_{F_i}$, 所以 y_i 和 z_{F_i} 之间是

存在线性关系的,实际上也就 z_i 和 z_{F_i} 之间存在线性关系,于是乎做了上面这个表,

现在对 z_i 和 Z_{fi} 进行作图,结果如下。



从图中可以看出,如果将 3 个异常值(outliers)去掉,剩下的点会很好地分布在一条直线上,值得一提的是这三个异常值中的两个就是第 3 题表中所提到的异常值,因此也进一步证实了这 2 个点的异常性,所以之后的分析将删除这 3 个点,源数据变成 **21 组**。按照同样的方法,再进行一次散点图分析,并且进行直线拟合其图形如下。



由此可以看出，去掉异常值之后的线性拟合很强，因此接受 H_0 ，认为其为正态分布，但是这种方法是比较粗糙简单的分析方法，下面使用 χ^2 检验法分析：

二、 χ^2 检验法分析法检验 Z 值分布情况

设 $Z \sim N(\mu, \sigma^2)$ ，由于 2 个参数均未知，故使用极大似然估计估计这两个参数。易知 μ 的极大似然估计就是 Z 的平均值即为 7.154， σ^2 的极大似然估计是总体方差

$$\sigma^2 \approx S_n^2 = \frac{1}{n} \sum_{i=1}^n (Z_i - \bar{Z})^2 = 10.333$$

$$\mu \approx \bar{Z} = 7.154$$

| | zi (sorted) | Standard | 概率 | 理论频数 | 实际频数 |
|----|-------------|--------------|--------|---------|------|
| 1 | 2.196 | -1.542386828 | 0.0618 | 1.2978 | 1 |
| 2 | 2.348 | -1.495101068 | 0.0668 | 1.4028 | 2 |
| 3 | 2.437 | -1.467414011 | 0.0708 | 1.4868 | 3 |
| 4 | 3.346 | -1.184632723 | 0.119 | 2.499 | 4 |
| 5 | 3.401 | -1.167522744 | 0.113 | 2.373 | 5 |
| 6 | 3.727 | -1.066107233 | 0.157 | 3.297 | 6 |
| 7 | 5.153 | -0.622492143 | 0.2676 | 5.6196 | 7 |
| 8 | 6.081 | -0.333800134 | 0.3707 | 7.7847 | 8 |
| 9 | 6.644 | -0.158656168 | 0.4384 | 9.2064 | 9 |
| 10 | 7.057 | -0.030175781 | 0.488 | 10.248 | 10 |
| 11 | 7.199 | 0.013999074 | 0.504 | 10.584 | 11 |
| 12 | 8.134 | 0.304868716 | 0.617 | 12.957 | 12 |
| 13 | 8.407 | 0.389796429 | 0.6517 | 13.6857 | 13 |
| 14 | 8.64 | 0.462280522 | 0.6772 | 14.2212 | 14 |
| 15 | 9.79 | 0.820034627 | 0.7939 | 16.6719 | 15 |

| | | | | | |
|----|--------|-------------|--------|---------|----|
| 16 | 9.88 | 0.848032774 | 0.8023 | 16.8483 | 16 |
| 17 | 10.028 | 0.894074172 | 0.8133 | 17.0793 | 17 |
| 18 | 10.95 | 1.180899637 | 0.881 | 18.501 | 18 |
| 19 | 11.361 | 1.308757843 | 0.9049 | 19.0029 | 19 |
| 20 | 11.65 | 1.398663005 | 0.9192 | 19.3032 | 20 |
| 21 | 11.802 | 1.445948765 | 0.9251 | 19.4271 | 21 |

说明：zi 列来自于上个表中，只是删除了 3 个异常点，将其按照正态分布的方法进行标准化之后得到 standard 列，然后从标准正态分布表中读出对应标准值的左侧概率值得到‘概率’列,按照总组数和概率的乘积得到理论频数列，而由于 zi 是以及排好顺序的，所以实际频数就是依次递增的。

按照含有未知参数 χ^2 检验法分析法，有以下结论：

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - \hat{n} p_i)^2}{\hat{n} p_i} \sim \chi^2(k-l-1), \text{ 在本例中, } k \text{ 为样本类别为 } 21, l \text{ 表示估计的样本参}$$

数为 2。 n_i 表示的是实际的频数， $\hat{n} p_i$ 表示的是按照正态分布下估计出来的频数。

因此本例计算结果为 $\chi^2 = 8.74$ ，而 $\chi_{0.05}^2(18) = 28.869$ ，由于 $8.74 < 28.869$ ，故认为其结果落在了置信度为 95% 的区间里。接受原假设，认为 Z 值服从正态分布。且利用极大似然估计结果可知其服从 $N(7.154, 10.333)$ 的正态分布。

结论：认为同一个人在 congruent 和 incongruent 下的测试时间差服从 $N(7.154, 10.333)$ 正态分布，置信水平为 95%，其结果和我的预期结果基本一致。这里的均值为 7.154，接近于 congruent 的均值 14.05 的一半，因此个人觉得可以说统计结果表明对于 Incongruent 下所用时间是 Congruent 下时间的 1.5 倍。

以下为修改后内容：

按照老师的修改意见：应该使用相依样本的 t 单尾检验，我很同意老师的建议，不过我想强调的是我没有采用 t 检验是因为我的检验是时间差的分布类型，是对分布类型进行的假设检验，可能是由于我的关注点不太正确吧，下面是按照老师建议采用 t 单尾检验对两个总体的均值进行检验：

零假设： $\mu_1 \geq \mu_2$ 即 Congruent 下总体平均值大于等于 Incongruent 下总体平均值

备择假设： $\mu_1 < \mu_2$ 即 Congruent 下总体平均值小于 Incongruent 下总体平均值

其中： μ_1 表示的是 Congruent 下总体平均值， μ_2 表示的是 Incongruent 下总体平均值。之所

以这样设置，是因为我希望证明 $\mu_1 < \mu_2$ ，故将其对立面 $\mu_1 \geq \mu_2$ 放在了零假设的位子，这样我可以有理由去拒绝 H_0 ，接受 H_1 ，从而达到想要验证的目的。

令 $Z = \text{Incongruent} - \text{Congruent}$ ，则假设变成

$$H_0: \mu_3 \leq 0$$

$$H_1: \mu_3 > 0$$

其中 μ_3 表示的是 Z 的总体均值。

由于总体方差未知，故只能采用 t 检验，统计量为 $T = \frac{\bar{z} - 0}{S} \sqrt{n}$ （其中， \bar{z} 表示 Z 的样本平均值，S 表示样本标准差，n 为样本数量），计算可得：

$$\bar{Z} = 7.965, S = 4.865, T = \frac{7.965 - 0}{4.865} \sqrt{24} = 8.02$$

显著性水平为 0.05，则拒绝域为 $T > t_{\alpha}(n-1)$ ，读表可得 $t_{0.05}(24-1) = 1.7139$ ，由于 8.02

要远大于 1.7139，所以认为其落入了拒绝域。即拒绝原假设，认为 Incongruent 下的时间不小于等于 Congruent 下的时间，即 Incongruent 下的所用时间大于 Congruent 下的时间。这个结论和我预测的结果一致的。

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！

原因：

这个现象有点意思。按照维基百科中的解释^[1]，通过这个实验可以得出 3 个结论：

- (1) 语义干扰：即完成不一致任务所花时间要长于一致的，说明字体的含义对测试者的识别过程产生了干扰，这可能是由于在测试者接受了不一致的信息时大脑需要更多的时间反应和处理，从而导致时间延长。这一点挺容易理解的。
- (2) 语义加强：即完成一致任务的时间要少于 neutral 组，（neutral 组的字显示颜色或者是字体），说明当颜色和字体含义一致时，人们的反应时间要小于那些只有颜色或者只有字体的测试方法，从而说明了语义对此过程有加强的作用。应该是在大脑识别过程中，不同的大脑区域负责不同的功能，当字体一致时，大脑中的读字和识别颜色功能同时启动，且反应结果从语言上看是一致的（例如，‘红色’）这两种功能有可能就有协同作用了吧，但是在 neutral 组，读字和识别颜色并不一致（‘**udacity**’），但是字体本身含义和颜色没多大关系，因此可能就没有了协同作用。
- (3) 当要求测试者识别字体时候，这种干扰效应就会消失，这估计是由于在大脑的不同区域去识别颜色和读出字体，而这两个区域对这种干扰的敏感度不同，有的区域对这种干扰很敏感，而有的则不。

类似效应的替代有：

- 将这个测试中的颜色换成形状，也就是说可以画一个方形上面写上“圆形”字体来干扰读者，测试者需要说出形状名称或者是说出上面的名字

优达学城

2016 年 9 月

参考来源：

1. https://en.wikipedia.org/wiki/Stroop_effect