

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。

详细介绍

1、我们的自变量是什么？因变量是什么？

解答：

自变量：两种不同的文字条件，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的墨色名称的时间。

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。

解答：

1) 适当假设集如下：

零假设 $H_0 : \mu_a = \mu_b$; (代表两种不同的文字条件的总体均值之间没有显著差异)

对立假设 $H_a : \mu_a \neq \mu_b$; (代表两种不同的文字条件的总体均值之间有显著差异)

其中： μ_a 表示在一致文字条件下的总体均值，此处的均值指总体的平均时间；

μ_b 表示在不一致文字条件下的总体均值，此处的均值指总体的平均时间；

取 α 水平为 0.05，执行双尾检测；

则若计算出的 t 在临界区，证明拒绝零假设；

若计算出 t 不在临界区，则接受零假设

2) 执行相依样本检测方式：配对 t 检测

使用此种检测方式的前提：两组样本中，每对数据的差值必须服从正态分布

使用 Python 计算出 $s = 7.5316978909875401$ ， $k = 0.023147952541481439$

由于 $sk > 0$ ，故每对数据的差值呈正偏态分布。

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。

解答：

描述性统计量

条件 a（在一致文字条件下）：

说出墨色名称的时间的均值为： $\mu_a = 14.05$;

值域为： $\text{range} = 13.70$;

条件 b（在不一致文字条件下）：

说出墨色名称的时间的均值为： $\mu_b = 22.01$;

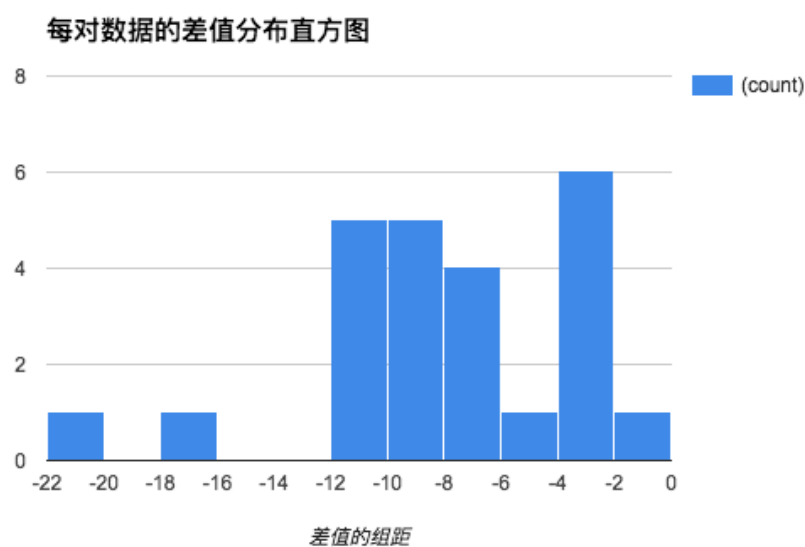
值域为： $\text{range} = 19.57$;

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。

解答：

1) 用直方图表示每对数据的差值如下：

（横轴表示差值的组距，纵轴表示相应的频数）



从图表中可以看出，差值始终为负数，即参与者在条件 a 下所用的时间始终小于条件 b，即 $a_i - b_i < 0$ ，故可以以此修改假设集如下：

修改适当假设集如下：

1) 零假设 $H_0 : \mu_a = \mu_b$; (代表两种不同的文字条件的样本均值之间没有显著差异)

2) 对立假设 $H_a : \mu_a \neq \mu_b$; (代表两种不同的文字条件的样本均值之间有显著差异)

其中：

μ_a 表示在一致文字条件下的样本均值，此处的均值指样本的平均时间；

μ_b 表示在不一致文字条件下的样本均值，此处的均值指样本的平均时间；

3) 取 α 水平为 0.05，执行单尾检测；

由于 $\alpha = 0.05$ 且 $DF = 23$ ，故

$$t_{critical} = -1.714$$

则 $t < t_{critical}$ ，则 t 在临界区，则证明拒绝零假设；

若 $t > t_{critical}$ ，则 t 不在临界区，则接受零假设

2) 通过 Python 绘制箱线图对比两组样本数据如下：

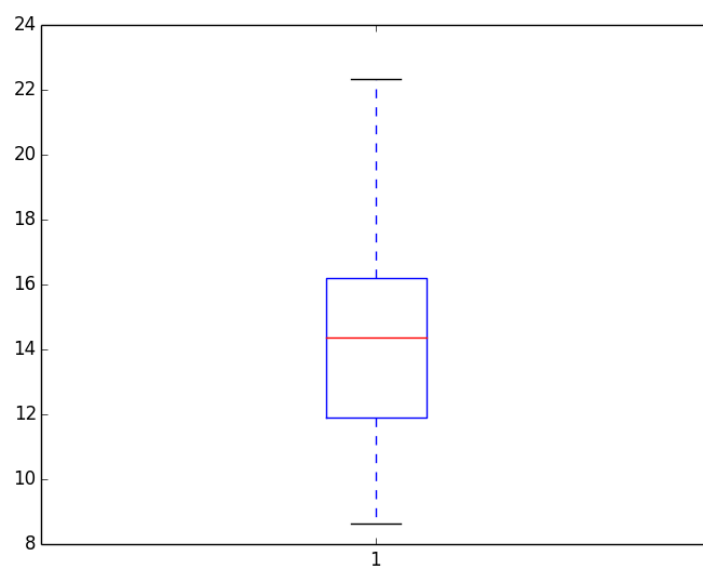


图 1 条件 a 样本箱线图

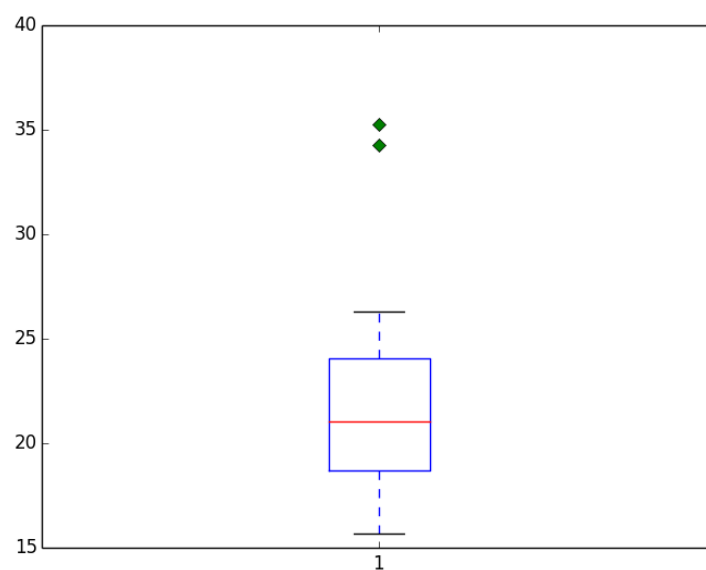
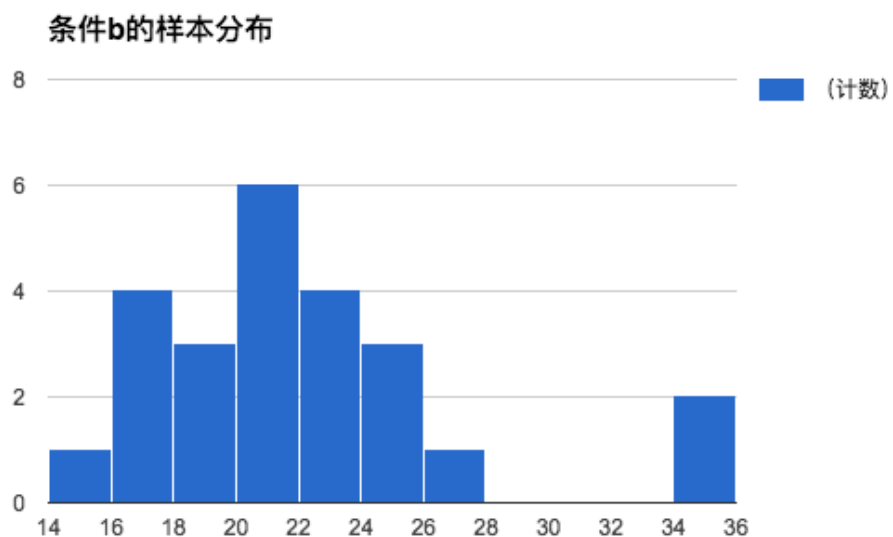
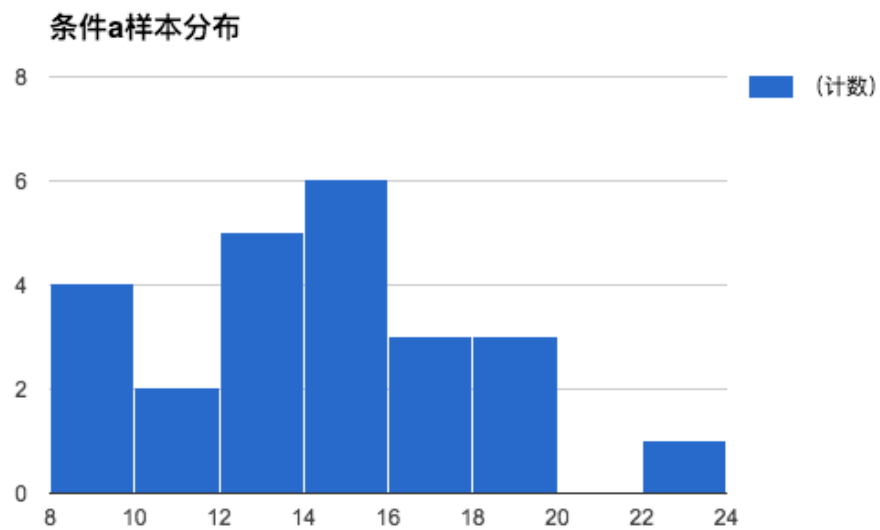


图 2 条件 b 样本箱线图

从图中可以看出，条件 a 数据正常，条件 b 存在 2 个异常值。

3) 描述条件 a 和条件 b 下的时间分布直方图如下：



从图中可以看出，条件 a 和条件 b 的分布都比较集中。

条件 a 大部分都集中在 8 到 20 之间，均值 14.06 位于 14~16 之间

条件 b 大部分都集中在 14 到 28 之间，均值 22.01 位于 22~24 之间

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？

解答：

推理统计量如下

1) 执行了相依样本检测；

2) 按照 APA 格式

$$t(23) = -8.02, P = .0001, \text{ 单尾检测}$$

3) α 水平为 0.05

4) 两个相依样本差异的置信区间按照 APA 格式如下:

$$95\%CI = (-6.26, \infty)$$

5) Cohen's $d = -1.63$

6) $r^2 = 0.74$

由于 $P = .0001 < 0.05$, 故成功拒绝零假设。

结论: 从相依样本的置信区间可以看出, 两种不同的条件有显著差异。

结果与期望一致。

计算过程如下:

1)

μ_a (样本 a 的均值: 样本 a 为在一致文字条件下的样本时间)	14.05
μ_b (样本 b 的均值: 样本 b 为在一致文字条件下的样本时间)	22.01
$\mu_a - \mu_b$ (样本每对差值的数据集的均值)	-7.96
$S = \sqrt{\frac{\sum x_i^2}{n-1}}$, $x_i = a_i - b_i$ (样本差值的标准差)	4.86
$SEM = \frac{S}{\sqrt{n}}$ (样本差值的标准误差)	0.99
n (样本量)	24
df = n - 1 (自由度)	23
$t = \frac{\mu_a - \mu_b}{SEM}$	-8.02
p (t = -8.02, DF = 23 时)	< 0.0001
$t_{critical}$ ($\alpha = 0.05$, 单尾检测的临界值)	-1.714
95%CI: 95%的置信区间	(-6.26, ∞)
$d = \frac{\mu_a - \mu_b}{S}$ (差异性度量)	-1.63
r^2 (相关性度量)	0.74

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！

解答：

1) 原因是使用不同的文字条件对说出同等大小的列表中的墨色名称影响很大，占 74% 的影响因素；

2) 类似任务有：

自变量：人在上午 9 点，同一个人在下午 3 点

因变量：测量人的认知分数和短期记忆分数（BBC 人脸测试）

探索点：人在上午和下午的记忆力是否有显著差异

H_0 ：没有显著差异

H_a ：有显著差异