

P1: 检验心理学现象——项目报告

学员: Jerry Shi (shichengjie1996@gmail.com)

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?

答:

自变量: 两个条件, 即“一致文字条件”和“不一致文字条件”。

因变量: 完成任务所需要的时间。

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。

答:

适当的假设集:

零假设: “一致文字条件”下参与者完成任务时间的均值多于“不一致文字条件”下参与者完成任务时间, 即 $H_0: \mu_c \geq \mu_i$;

对立假设: “一致文字条件”下参与者完成任务时间的均值少于“不一致文字条件”下参与者完成任务时间, 即 $H_a: \mu_c < \mu_i$;

(其中 μ_c 为“一致文字条件”下参与者完成任务时间的总体均值;
 μ_i 为“不一致文字条件”下参与者完成任务时间的总体均值。)

执行的统计测试: 对参与者完成任务时间的均值进行相依样本单尾 t 检验。

理由:

- 1、选取 t 检验而不是 z 检验的原因: 在数据集中总体参数是未知的; 两组样本量较小 (小于 30)。因此适用 t 检验。
- 2、样本为相依样本的理由: 参与实验的是同一批人, 改变的是实验条件, 前后两次的实验结果是相互影响而非独立的, 因此两个样本是相依样本。
- 3、选择单尾检验的理由: 通过自己的实践和经验判断, 可以认为参与者在“不一致文字条件”下, 在说出颜色时会受到文字显示内容的干扰, 从而延长参与者完成任务的时间, 因此用单尾检验是较为可取的。

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。

答：

描述性统计：

集中趋势：选择测量中位数和均值。

变异性：选择测量四分位差和标准偏差。

统计结果如下表所示（单位：秒，保留小数点后两位）：

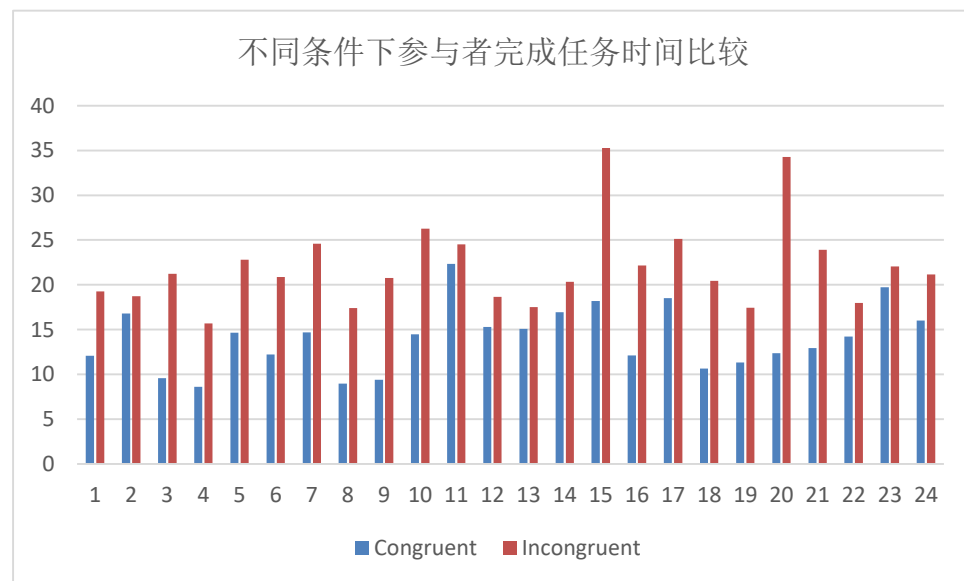
	一致	不一致
均值	14.05	22.02
中位数	14.36	21.01
四分位差	4.31	5.33
标准差	3.56	4.80

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

答:

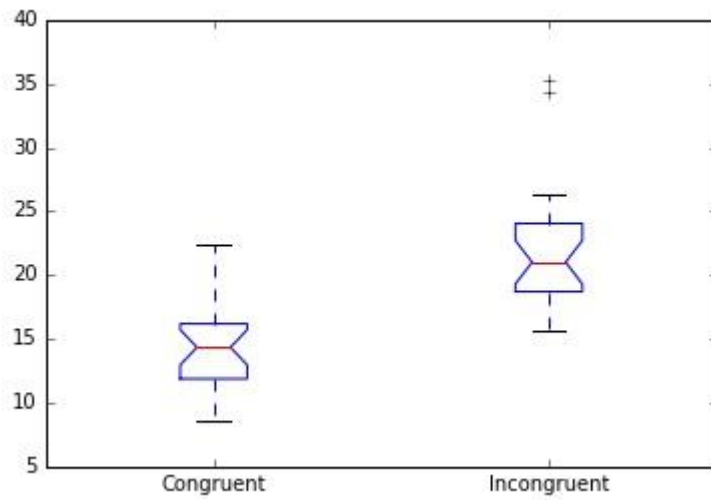
可视化内容选择:

1、两组样本数据分布的柱形图，图表如下:



结果: 在“一致文字条件”下, 参与者完成任务的时间总体上少于“不一致文字条件”下参与者完成任务的时间。

2、两组样本数据分布的箱线图，图表如下：



结果：在“一致文字条件”下，参与者完成任务的时间总体较少，但变异性较大；在“不一致文字条件”下，参与者完成任务的时间总体较多，变异性较小，但有两个异常值。

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？

答：

统计测试：相依样本的均值 t 检验（单尾）

零假设 H_0 : $\mu_c = \mu_i$

对立假设 H_a : $\mu_c < \mu_i$

置信水平： $\alpha = 0.95$

t 值： -8.02

p 值： 2.05×10^{-8}

结论：

通过查表，在样本自由度 df 为 24、单尾置信水平为 0.95 的情况下， t 的负临界值为 -1.711，大于计算出的 t 值 -8.02。

因此可以拒绝零假设，接受对立假设。结果表明在“一致文字条件”下，参与者完成任务的时间在统计上显著少于“不一致文字条件”下参与者完成任务的时间。

使用的资源:

1. Excel 表格 (直方图、描述性统计、t 检验函数);
2. python matplotlib 绘图函数库和 pandas 数据分析库 (绘制箱线图)。
2. stackoverflow 技术问答网站。
3. 优达学城课程视频、讲义。