

机器学习统计基础习题P1——测试心理学现象

前提条件：

自变量：文字语义与其打印颜色的匹配程度

因变量：说出所有文字打印颜色的时间

检测类型：t单尾检验（只有单份样本）、相依样本（一个对象进行不同条件下的两种测试）

可视化



结论：一致文字条件下大部分测试用时小于20；不一致文字条件下大部分测试用时大于20

数学假设：

μ_1 : 一致文字条件

μ_2 : 不一致文字条件

\bar{x}_1 : 一致文字抽样均值

\bar{x}_2 : 不一致文字抽样均值

σ_{x_1} : 一致文字标准误差

σ_{x_2} : 不一致文字标准误差

检验过程：

检测条件

$H_0: \mu_1 = \mu_2$ (一致文字和不一致文字的检测时间相同)

$H_1: \mu_1 < \mu_2$ (一致文字比不一致文字的检测时间少)

在 $\alpha = 0.05$ 水平下进行统计检验

自由度 $df = 24 - 1 = 23$

$$\bar{X}_1 = 14.05 \quad \sigma_{x_1} = 3.56$$

$$\bar{X}_2 = 22.02 \quad \sigma_{x_2} = 4.8$$

计算得到，差异的样本标准误差 $S = 4.86$

$$\text{Cohen's } d = \frac{M_D}{S_D} = \frac{14.05 - 22.02}{4.86} = -1.64$$

$$t_{\text{临界值}} = \frac{\bar{X}_1 - \bar{X}_2}{S / \sqrt{df}} = \frac{14.05 - 22.02}{4.86 / \sqrt{23}} = -7.85$$

检测结论

$\alpha = 0.05$ ，单尾 $t_{\text{statistic}} = -1.714$

$$\because t_{\text{statistic}} > t_{\text{临界值}}$$

\therefore 在 $\alpha = 0.05$ 水平下，拒绝 H_0 ，一致文字比不一致文字的检测时间少

$\alpha = 0.0005$ ，单尾 $t_{\text{statistic}} = -3.768$

$$\because t_{\text{statistic}} > t_{\text{临界值}}$$

\therefore 在 $\alpha = 0.0005$ 水平下，拒绝 H_0 ，一致文字比不一致文字的检测时间少

总结：

导致的可能原因是：文字语义与文字颜色不匹配的情况下，测试者需要使用思考的时间增加；增加的时间是清除惯性思维，重新识别颜色的时间；

类似任务：石头剪刀布游戏，改成一定要输；

