

- 检测心理学现象
  - 样本表格
  - 回答标准
- 问题回答
  - 问题 1：确认试验中的变量
    - 自变量X
    - 因变量Y
  - 问题 2a：建立假设
    - 零假设
    - 对立假设
  - 问题 2b：建立统计检验
    - 计算公式
    - t检验临界区
    - 拒绝零假设
  - 问题 3：报告描述性统计分析
    - 相依样本a和b差值中心测量-平均值和中位数
    - 相依样本a和b差值变异测量-四分位距法判别异常值
  - 问题 4：绘制数据图
    - 绘制方法: 使用python的可视化库Matplotlib
    - 直方图:
    - 箱线图:
  - 问题 5：执行统计检验并解读结果
    - 计算
    - 校验
    - 结论

# 检测心理学现象

## 样本表格

Congruent（样本a）	Incongruent（样本b）	样本差值 $\mu_{ab}$
12.079	19.278	7.199
16.791	18.741	1.95
9.564	21.214	11.65
8.63	15.687	7.057
14.669	22.803	8.134

12.238	20.878	8.64
14.692	24.572	9.88
8.987	17.394	8.407
9.401	20.762	11.361
14.48	26.282	11.802
22.328	24.524	2.196
15.298	18.644	3.346
15.073	17.51	2.437
16.929	20.33	3.401
18.2	35.255	17.055
12.13	22.158	10.028
18.495	25.139	6.644
10.639	20.429	9.79
11.344	17.425	6.081
12.369	34.288	21.919
12.944	23.894	10.95
14.233	17.96	3.727
19.71	22.058	2.348
16.004	21.157	5.153

回答标准

标准	符合要求
问题 1：确认试验中的变量	Q1：问题回复正确确认了试验中的自变量和因变量。
问题 2a：建立假设	Q2a：零假设和对立假设均以文字和数学方式进行了明确说明。数学陈述中的符号也进行了定义。
问题 2b：建立统计检验	Q2b：通过统计检验来辩证提出的假设，并针对统计检验的假设前提进行说明
问题 3：报告描述性统计分析	Q3：为数据集组计算了描述性统计分析，包括至少一项中心性测量和一项可变性测量。

问题 4：绘制数据图	Q4：创建了展示数据的一项或两项可视化，包括用注释说明图中可观察到的信息。
问题 5：执行统计检验并解读结果	Q5：正确执行并报告了一项统计检验，包括检验统计量、P 值和检验结果。检验结果针对执行的试验任务进行了解释。

## 问题回答

### 问题 1：确认试验中的变量

自变量X

文字的颜色和文字的名字是否一致

因变量Y

识别文字所需要的反应时间

### 问题 2a：建立假设

零假设

$$H_0: \mu_a = \mu_b (\alpha < 0.05)$$

假设测试人员在文字名字和颜色一致和文字名字和颜色不一致的情况下，反应时间没有显著变化

对立假设

$$H_a: \mu_a \neq \mu_b$$

假设测试人员在文字名字和颜色一致和文字名字和颜色不一致的情况下，反应时间会明显加大

### 问题 2b：建立统计检验

计算公式

样本大小n:  $n = 24$

样本a平均值:  $\mu_a = \frac{\sum_{i=1}^n x_a^i}{n}$

样本b平均值:  $\mu_b = \frac{\sum_{i=1}^n x_b^i}{n}$

相依样本a和b平均值差值: $\mu_d = \mu_b - \mu_a$

相依样本a和b差值的标准误差:  $S = \sqrt{\frac{\sum_{i=1}^n (x_b^i - x_a^i)^2}{n - 1}}$

t检验的统计量:  $t = \frac{\mu_d}{\frac{S}{\sqrt{n}}}$

t检验临界区

样本自由度根据公式 $n_f = n - 1$ 计算得到:  $n_f = 23$

从样本数据我们是想了解实不一致的情况下比一致时间要长的多，那么我们使用双尾检测

如果满足 $\alpha < 0.05$ ,那么查询t-table表后得到t的临界区范围  $\pm 2.074$

拒绝零假设

t统计量	拒绝零假设	结论
$t < -1.714$	是	不一致的情况下反应时间显著降低
$t > 1.714$	是	不一致的情况下反应时间显著加大
$-1.714 \leq t \leq 1.714$	否	不一致的情况下反应时间不明显

问题 3：报告描述性统计分析

相依样本a和b差值中心测量-平均值和中位数

平均值计算公式:  $\bar{\mu}_{ab} = \frac{\sum_{i=1}^n (x_b^i - x_a^i)}{n - 1}$

平均值:  $\bar{\mu}_{ab} = 7.96$

按照升序排列 $\mu_{ab}$ :

序号	样本差值 $\mu_{ab}$
1	1.95
2	2.196

3	2.348
4	2.437
5	3.346
6	3.401
7	3.727
8	5.153
9	6.081
10	6.644
11	7.057
12	7.199
13	8.134
14	8.407
15	8.64
16	9.79
17	9.88
18	10.028
19	10.95
20	11.361
21	11.65
22	11.802
23	17.055
24	21.919

中位数计算公式:  $M = \frac{\mu_{ab}^{12} + \mu_{ab}^{13}}{2}$

中位数:  $M = 7.67$

相依样本**a**和**b**差值变异测量-四分位距法判别异常值

序号	样本差值 $\mu_{ab}$	$Q_n$
1	1.95	

2	2.196	
3	2.348	
4	2.437	
5	3.346	
6	3.401	$Q_1$
7	3.727	
8	5.153	
9	6.081	
10	6.644	
11	7.057	
12	7.199	$Q_2$
13	8.134	
14	8.407	
15	8.64	
16	9.79	
17	9.88	
18	10.028	$Q_3$
19	10.95	
20	11.361	
21	11.65	
22	11.802	
23	17.055	
24	21.919	

异常值计算判断:  $exp < Q_1 - 1.5 * (Q_3 - Q_1) || exp > Q_3 + 1.5 * (Q_3 - Q_1)$

根据表格我们可以得到:

$$IQR = Q_3 - Q_1 = 9.94$$

$$exp < -7.99 || exp > 19.968$$

因此第24条数据，我们认为21.919是异常的数据

## 问题 4：绘制数据图

---

绘制方法: 使用python的可视化库Matplotlib

绘制代码:

```
# -*- coding:UTF-8 -*-

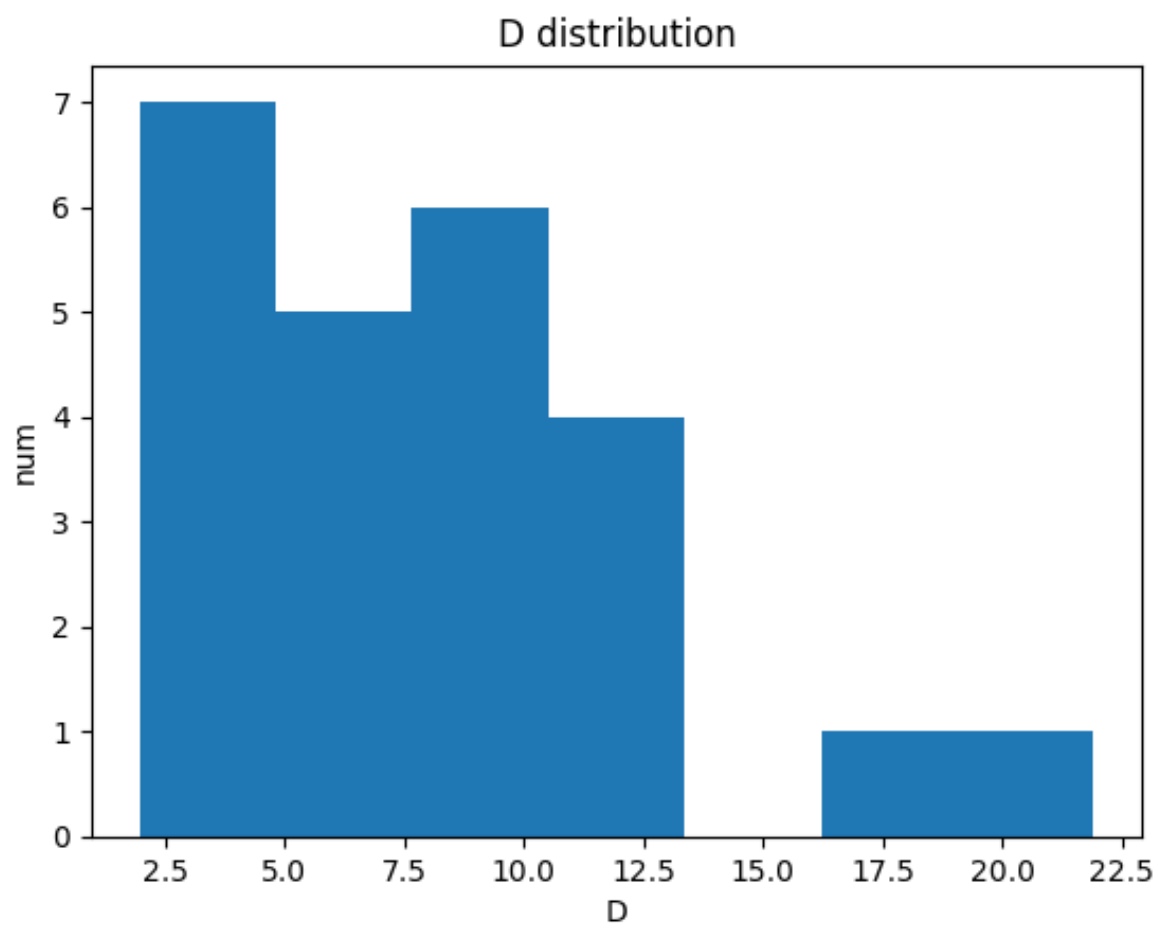
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
import numpy as np

# 0、导入数据集
df = pd.read_excel('stroopdata.xlsx', 'stroopdata')

# 1、直方图
fig = plt.figure()
ax = fig.add_subplot(111)
ax.hist(df['D'], bins=7)
plt.title('D distribution')
plt.xlabel('D')
plt.ylabel('num')
plt.show()

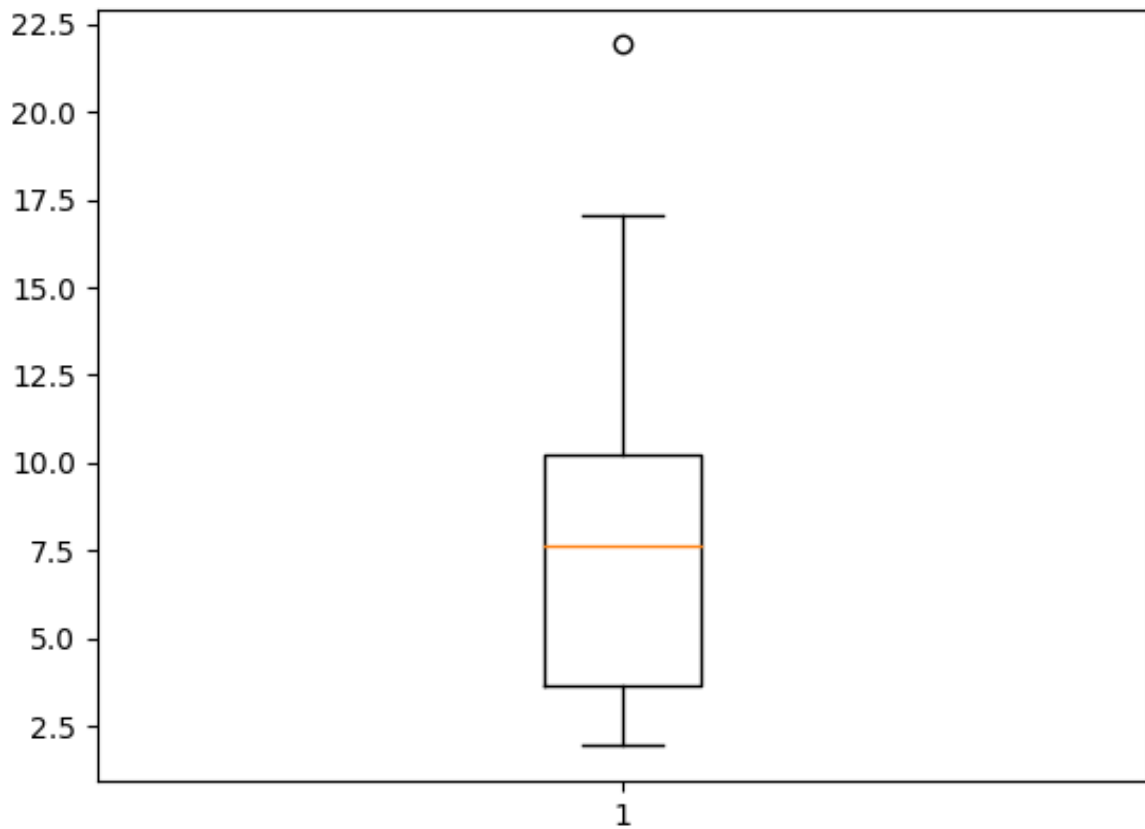
# 2、箱线图
fig = plt.figure()
ax = fig.add_subplot(111)
ax.boxplot(df['D'])
plt.show()
```

直方图:



箱线图:





## 问题 5：执行统计检验并解读结果

计算

因为存在一个异常值，所以删除异常值的一行表格如下：

Congruent（样本a）	Incongruent（样本b）	样本差值 $\mu_{ab}$
12.079	19.278	7.199
16.791	18.741	1.95
9.564	21.214	11.65
8.63	15.687	7.057
14.669	22.803	8.134
12.238	20.878	8.64
14.692	24.572	9.88
8.987	17.394	8.407
9.401	20.762	11.361

14.48	26.282	11.802
22.328	24.524	2.196
15.298	18.644	3.346
15.073	17.51	2.437
16.929	20.33	3.401
18.2	35.255	17.055
12.13	22.158	10.028
18.495	25.139	6.644
10.639	20.429	9.79
11.344	17.425	6.081
12.944	23.894	10.95
14.233	17.96	3.727
19.71	22.058	2.348
16.004	21.157	5.153

根据问题2a和2b中的计算公式，我们可以很快得出结论：

总体个数:  $n = 23$

自由度:  $n_f = 22$

相依样本a和b差值的标准误差:  $S = 8.49$

相依样本a和b平均值差值:  $\mu_d = 7.36$

t检验的统计量:  $t = 8.76$

t检验在  $\alpha < 0.05$ , 查询t-table表后得到t的临界区范围  $\pm 2.074$

### 校验

那么根据表

t统计量	拒绝零假设	结论
$t < -2.074$	是	不一致的情况下反应时间显著降低
$t > 2.074$	是	不一致的情况下反应时间显著加大
$-2.074 < t < 2.074$	否	不一致的情况下反应时间不明显

实验一：文字识别	实验二：文字识别	实验三：文字识别
----------	----------	----------

## 结论

我们得到结果是拒绝零假设，当文字名字和文字颜色不一致时，时间明显加大