

第七次作业

1 第一题

表1是8名健康男子服用肠溶醋酸棉酚片前后的精液检查结果，服用时间为1~3个月，问服药后精液中的精子浓度有无下降？

- 1.请说明题中资料及设计类型，及所选取的方法
- 2.请按流程进行完整的假设检验，注意书写规范，如统计量斜体、下标等
- 3.可自由选择一种或结合使用多种软件，附上相关代码或必要步骤的截图
- 4.请勿仅粘贴软件结果图，需整理并报告结果，如秩次表、秩和、统计量、 P 值等，给出统计+专业结论
- 5.请装订好作业，写上姓名及学号

由题意，此为配对设计的资料比较，采用Wilcoxon符号秩和检验。

建立检验假设：

$$\begin{aligned} H_0 : \mu &= \mu_0, \text{即假设差值的总体中位数为等于} 0 \\ H_1 : \mu &\neq \mu_0, \text{即假设差值的总体中位数大于} 0 \\ \alpha &= 0.05 \end{aligned} \quad (1)$$

```
# Part1
data <- read.csv("Homework7/Data1.csv")
before <- data$before
after <- data$after

boxplot(before, after)
wilcox.test(before, after,
  paired = TRUE,
  alternative = "greater",
  exact = FALSE,
  correct = FALSE
)
```

由于R并未提供秩和计算功能，采用Python实现秩和计算

```
# -*- encoding: utf-8 -*-
#@Author : Arthals
#@File : Homework7.py
#@Time : 2022/11/14 16:29:24
#@Contact : zhuozhiyongde@126.com
#@Software: Visual Studio Code
```

```

def Part1():
    before = [6000, 22000, 5900, 4400, 6000, 6500, 26000, 5800]
    after = [660, 5600, 3700, 5000, 6300, 1200, 1800, 2200]
    diff = [after[i] - before[i] for i in range(len(before))]
    diff_index = [(index + 1, diff_value)
                  for index, diff_value in enumerate(diff)]
    diff_index = sorted(diff_index, key=lambda x: abs(x[1]))
    diff_index = [(index, diff_value) for index, diff_value in diff_index
                  if diff_value != 0]
    print(diff_index)

    # 给出秩和
    rank_dic = {}
    for index, diff_value in diff_index:
        if abs(diff_value) in rank_dic.keys():
            rank_dic[abs(diff_value)].append(diff_value / abs(diff_value))
        else:
            rank_dic[abs(diff_value)] = [diff_value / abs(diff_value)]

    rank_pos = 0
    rank_neg = 0
    temp_rank = 1
    for rank in rank_dic.keys():
        # print(f"{rank}:{temp_rank}~{temp_rank + len(rank_dic[rank]) - 1}")
        total_rank = sum(range(temp_rank, temp_rank + len(rank_dic[rank])))
        rank_pos += len([pos_num for pos_num in rank_dic[rank] if pos_num == 1])
        / len(rank_dic[rank]) * total_rank
        # print(f"正秩和:{rank_pos}")

        rank_neg += len([
            pos_num for pos_num in rank_dic[rank] if pos_num == -1
        ]) / len(rank_dic[rank]) * total_rank
        # print(f"负秩和:{rank_neg}")

        temp_rank += len(rank_dic[rank])
    print(f"rank_pos:{rank_pos}")
    print(f"rank_neg:{rank_neg}")

if __name__ == '__main__':
    Part1()

```

得到输出结果

$$\begin{aligned} rank_{neg} &: 33 \\ rank_{pos} &: 3 \\ p - value &= 0.01785 < \alpha = 0.05 \end{aligned} \quad (2)$$

故拒绝 H_0 ，接受 H_1 ，即认为差值的总体中位数不为0，也即认为服药后精液中精子浓度有所下降。

2 第二题

为研究某新药治疗贫血患者的疗效，将20名贫血患者随机分为两组，一组用新药治疗，一组用常规药治疗，测得血红蛋白增量如下表2，问两种药物疗效有无差别？

请用参数检验和非参数检验的方法分别对此题进行分析，比较两种方法的结果是否一致。若不同，请讨论原因。

2.1 参数检验

由题意，此为两组独立样本的比较，才用对立t检验，首先校验正态性和方差齐性。

2.1.1 正态性检验

设立检验假设：

$$\begin{aligned} H_0 &: \text{样本来自正态总体} \\ H_1 &: \text{样本来自非正态总体} \\ \alpha &= 0.05 \end{aligned} \quad (3)$$

```
# Part2
data <- read.csv("Homework7/Data2.csv")
new_drug <- data[which(data$drug == 1), "radio"]
regular_drug <- data[which(data$drug == 2), "radio"]
shapiro.test(data$radio)
```

得到输出结果 $W = 0.97186, p - value = 0.7937 > \alpha = 0.05$ ，从而不拒绝 H_0 ，即认为样本来自正态总体。

2.1.2 方差齐性检验

采用Levene检验

设立检验假设：

$$\begin{aligned} H_0 &: \sigma_1^2 = \sigma_2^2 \\ H_1 &: \sigma_1^2 \neq \sigma_2^2 \\ \alpha &= 0.1 \end{aligned} \quad (4)$$

```
var.test(data$radio ~ data$drug)
```

得到输出结果 $Pr(> F) = 0.9893 > \alpha = 0.1$ ，从而不拒绝 H_0 ，即认为方差齐。

2.1.3 校正t检验

由于方差齐，我们采用t检验来完成参数检验。

设立检验假设：

$$\begin{aligned} H_0 : \mu_1 &= \mu_2, \text{即认为两种药物疗效没有差别} \\ H_1 : \mu_1 &\neq \mu_2, \text{即认为两种药物疗效有差别} \end{aligned} \quad (5)$$

```
t.test(new_drug, regular_drug, var.equal = TRUE)
```

得到输出结果，95%CI : 0.2155422 ~ 8.7644578, $p - value = 0.04055 < \alpha = 0.05$ ，从而拒绝 H_0 ，接受 H_1 ，即认为两种药物疗效有差别。

2.2 非参数检验

```
wilcox.test(new_drug, regular_drug,  
            paired = FALSE,  
            exact = FALSE  
)
```

由于R并未提供秩和计算功能，采用Python实现秩和计算

```
def Part2():  
    new_drug = [34.5, 33.0, 32.5, 30.5, 29.5, 25.5, 25.0, 24.4, 23.6, 21.4]  
    regular_drug = [30.0, 28.5, 28.0, 26.0, 25.0, 21.0, 20.5, 19.9, 19.0, 17.1]  
    drug_total = []  
    drug_total.extend([(1, new_drug_value) for new_drug_value in new_drug])  
    drug_total.extend([(2, regular_drug_value)  
                      for regular_drug_value in regular_drug])  
  
    drug_total = sorted(drug_total, key=lambda x: x[1])  
    print(drug_total)  
    print(len(drug_total))  
    rank_total = []  
    pass_num = 0  
    for index, (group, value) in enumerate(drug_total):  
        if (pass_num):  
            pass_num -= 1  
            continue  
        # print(f"start:{index}")  
        rank = index + 1  
        same = 0  
        for i in range(index, len(drug_total)):
```

```

        if drug_total[i][1] == value:
            same += 1
        else:
            break
    # print(f"{value}:{rank}~{rank + same - 1},same:{same}")
    for i in range(index, index + same):
        rank_total.append(
            (drug_total[i][0],
             sum([j for j in range(rank, rank + same)]) / same))
    pass_num = same - 1
print(rank_total)

# 计算秩和
rank_group = [sum([rank for group, rank in rank_total if group == i]) for i
in range(1, 3)]
print(len(new_drug), len(regular_drug))
print(rank_group)

if __name__ == '__main__':
    Part2()

```

得到输出结果

```

Wilcoxon rank sum test with continuity correction

data:  new_drug and regular_drug
W = 75.5, p-value = 0.05869
alternative hypothesis: true location shift is not equal to 0

```

```
[130.5, 79.5]
```

也即

$$W = 79.5, p - value = 0.05869 > \alpha = 0.05 \quad (6)$$

从而不拒绝 H_0 ，即认为两种药物疗效没有差别。

3 第三题

某研究者欲比较A、B两种菌对小鼠巨噬细胞吞噬功能的激活作用，将59只小鼠随机分成三组，其中一组为生理盐水对照组，用常规巨噬细胞功能的监测方法，获得三组吞噬率(%)结果见表3，试判断三组吞噬率有无差异？

由题意，此为多个独立样本比较，故采用多个独立样本的Kruskal-Wallis H检验。

建立检验假设：

$$\begin{aligned}H_0 &: 3组吞噬率的总体分布相同 \\ H_1 &: 3组吞噬率的总体分布不全相同 \\ \alpha &= 0.05\end{aligned}\quad (7)$$

```
# Part3
data <- read.csv("Homework7/Data3.csv")
g_a <- data$A
g_b <- data$B
g_c <- data$C

group <- factor(
  c(
    rep("1", length(g_a)),
    rep("2", length(g_b)),
    rep("3", length(g_c))
  )
)
pha_rate <- c(g_a, g_b, g_c)
kruskal.test(pha_rate ~ group)
```

因为R没有提供秩和计算功能，采用Python实现秩和计算：

```
def Part3():
    g_a = [
        45, 56, 57, 57, 60.3, 63, 64, 64, 64, 66, 66, 66, 66, 67, 70, 70, 70,
        71, 74, 74, 76, 73, 93, 95
    ]
    g_b = [
        51, 51, 54, 54, 59, 61, 61, 61, 62, 68, 68, 70, 70, 71, 70, 87, 89, 91,
        93
    ]
    g_c = [46, 31, 56, 48, 43, 24, 18, 36, 44, 36, 36, 24, 18, 36, 44, 36]
    g_total = []
    g_total.extend([(1, g_a_value) for g_a_value in g_a])
    g_total.extend([(2, g_b_value) for g_b_value in g_b])
```

```

g_total.extend([(3, g_c_value) for g_c_value in g_c])
g_total = sorted(g_total, key=lambda x: x[1])
print(g_total)
rank_total = []
pass_num = 0
for index, (group, value) in enumerate(g_total):
    if (pass_num):
        pass_num -= 1
        continue
    rank = index + 1
    same = 0
    for i in range(index, len(g_total)):
        if g_total[i][1] == value:
            same += 1
        else:
            break
    for i in range(index, index + same):
        rank_total.append(
            (g_total[i][0],
             sum([j for j in range(rank, rank + same)]) / same))
    pass_num = same - 1
g_rank = []
for i in range(1, 4):
    g_rank.append(sum([rank for group, rank in rank_total if group == i]))
print(f"group_rank:{g_rank}")
h_statistic = 12 / (len(g_total) * (len(g_total) + 1)) * sum([
    g_rank[i]**2 / len([j for j in g_total if j[0] == i + 1])
    for i in range(3)
]) - 3 * (len(g_total) + 1)
print(f"h_statistic:{h_statistic}")

if __name__ == '__main__':
    Part3()

```

得到结果:

Kruskal-Wallis rank sum test

data: pha_rate by group

Kruskal-Wallis chi-squared = 33.128, df = 2, p-value = 6.403e-08

```
group_rank:[928.0, 698.5, 143.5]
h_statistic:33.046603664882554
```

也即：

$$\begin{aligned}rank_A &= 928 \\rank_B &= 698.5 \\rank_C &= 143.5 \\H &= 33.05 \\p - value &= 6.403e - 08\end{aligned}\tag{8}$$

由于 $p - value = 6.403e - 08 < \alpha = 0.05$ ，所以拒绝 H_0 ，接受 H_1 ，也即认为吞噬率的差异有统计学意义，3组吞噬率的总体分布不完全相同。

4 第四题

按照性别相同、体重相近原则将蟾蜍配成10个区组，每个区组包括4只蟾蜍，随机将其分配到4个处理组中：对照组和实验1~3组，分别在蟾蜍上颌黏膜滴加0.5ml不同溶液并保持30min。4种溶液分别为PBS、PNS、PNS脂质体和脂质体。观察记录离体上颌黏膜纤毛运动持续的时间（min），结果见下表4。试问，4种溶液对纤毛运动时长的影响有无不同？

由题意，此为随机化区组设计资料，故采用Friedman检验

建立检验假设：

$$\begin{aligned}H_0 &: \text{不同溶液纤毛运动时间的总体分布位置相同} \\H_1 &: \text{不同溶液纤毛运动时间的总体分布位置不全相同} \\ \alpha &= 0.05\end{aligned}\tag{9}$$

```
# Part4
data <- read.csv("Homework7/Data4.csv")
score <- data$score
group <- data$group
block <- data$block

friedman.test(score ~ group | block)
```

因为R没有提供秩和计算功能，采用Python实现秩和计算：

```
def Part4():
    import numpy as np
    data = np.array([[630, 487, 720, 619], [621, 387, 601, 567],
                     [546, 316, 539, 531], [498, 257, 264, 367],
                     [523, 286, 310, 432], [531, 367, 431, 422],
```



```

        [520, 345, 492, 489], [532, 324, 335, 316],
        [623, 321, 620, 611], [664, 432, 656, 597]])
group, solution = data.shape
for g in range(group):
    if len(set(data[g])) != len(data[g]):
        print(f"第{g+1}组数据有重复")

# 校验没有重复后直接按大小编秩
data_change = []
for g in range(group):
    inside = [(index, value) for index, value in enumerate(data[g])]
    print(f"第{g+1}组数据:{inside}")
    inside = sorted(inside, key=lambda x: x[1])
    inside = [(index, rank + 1)
               for rank, (index, value) in enumerate(inside)]
    print(f"第{g+1}组数据:{inside}")
    data_change.append(inside)

print(data)
# 计算秩和
g_rank = []
for s in range(solution):
    s_rank = 0
    for g in range(group):
        s_rank += sum(
            [data_change[g][i][1] for i in range(solution) if data_change[g]
             [i][0] == s])
    g_rank.append(s_rank)

    print(f"Rank{s+1}: {s_rank}")
m_stastic = sum([(g_rank[i] - sum(g_rank) / solution)**2
                  for i in range(solution)])
print(f"m_stastic:{m_stastic}")

if __name__ == '__main__':
    Part4()

```

得到输出结果

```
Friedman rank sum test
```

```
data: score and group and block
```

```
Friedman chi-squared = 14.544, df = 3, p-value = 0.002251
```

```
Rank1: 39
```

```
Rank2: 11
```

```
Rank3: 29
```

```
Rank4: 21
```

```
m_stastic:424.0
```

也即：

$$rank_{PBS} = 39$$

$$rank_{PNS} = 11$$

$$rank_{PNS_{\text{脂质体}}} = 29$$

$$rank_{\text{脂质体}} = 21$$

$$M = 424$$

$$p - value = 0.002251$$

(10)

由于 $p - value = 0.002251 < \alpha = 0.05$ ，所以拒绝 H_0 ，接受 H_1 ，也即认为不同溶液纤毛运动时间的总体分布位置不全相同。