

实验12、非参数统计分析

年级：15级 专业：生信 学号：1513401013 姓名：郑磊

编号 一 二 三 四 总分 评阅人

得分

软硬件平台：

1. 硬件平台：（硬件配置）i5，2.9HZ处理器，16G内存，64位操作系统
2. 系统平台：（操作系统及其版本号）Windows10 企业版
3. 软件平台：（软件系统及其版本号，若是在线分析平台，还需要提供URL地址）R3.4.1 ， Rstudio

一、目的要求：

- 1、加深对秩和检验的理解；
- 2、熟悉并掌握秩和检验相关的R语言函数和脚本；
- 3、理解参数统计和非参数的异同之处。

二、实验内容：

1、Wilcoxon符号秩和检验

1.1、配对设计资料的符号秩和检验

1.1.1、数据描述：

某研究人员使用中药舒心散治疗21例冠心病患者，分别于治疗前和治疗后1个月检测优球蛋白（ELT）。

1.1.2、数据读取：

```
dir="D:/RFile/实验十二"
```

```
setwd(dir)
```

```
file="e12-data-1-1.txt"
```

```
data<-read.table(file,head=T,sep="\t")
```

```
data
```

```
x<-data[,2];y<-data[,3]
```

1.1.3、数据可视化观察：

```
png("e12_data-1-1_boxplot.png")
```

```
boxplot(data[,2:3])
```

```
dev.off()
```

1.1.4、正态性检验：

```
shapiro.test(x)
```

```
shapiro.test(y)
```

1.1.5、方差齐性检验：

```
data2<-data.frame(X<-c(data[,2],data[,3]),A<-factor(rep(1:2,c(21,21))))
```

#Bartlett检验 - 如果我们的数据服从正态分布，那么这种方法将是最为适用

的。对于正态分布的数据，这种检验极为灵敏；而当数据为非正态分布时，使用该方法则很容易导致假阳性误判。

```
bartlett.test(data[,2:3])
```

```
#或bartlett.test(X~A,data=data2)
```

#Levene检验 - 相较于Bartlett检验，这一方法更为稳健，这一方法被封装于car程序包中。

```
library(car)
```

```
leveneTest(X~A,data=data2)
```

#Fligner-Killeen检验 - 这是一个非参数的检验方法，完全不依赖于对分布的假

设

```
fligner.test(X~A,data=data2)
```

1.1.6、Wilcoxon配对符号秩和检验（双侧）：

#H0:Md=0, H1:Md \neq 0

```
wilcox.test(x, y, paired = TRUE, alternative = "two.sided")
```

1.1.7、配对t检验（等方差双侧检验）：

#H0: $\mu_1=\mu_2$, H1: $\mu_1\neq\mu_2$

```
t.test(x,y, paired = TRUE, var.equal=TRUE, alternative = "two.sided")
```

1.1.8、配对t'检验（异方差双侧检验-Welch t检验）：

#H0: $\mu_1=\mu_2$, H1: $\mu_1\neq\mu_2$

```
t.test(x,y, paired = TRUE, alternative = "two.sided")
```

1.1.9、综合上述统计计算结果，进行对比分析讨论。

2、Kruskal-Wallis H检验

2.1、完全随机设计多个独立样本的秩和检验【计量资料】

2.1.1、数据描述：

某研究组欲研究A、B两个菌种对小鼠巨噬细胞功能的激活作用，将57只小鼠随机分为三组，其中一组为生理盐水对照组，用常规巨噬细胞吞噬功能的检测方法，获得三组的吞噬指数。

2.1.2、数据读取：

```
dir="D:/RFile/实验十二"
```

```
setwd(dir)
```

```
file="e12-data-2-1.txt"
```

```
data<-read.table(file,head=T,sep="\t")
```

```
data
```

2.1.3、数据可视化观察：

```
png("e12_data-2-1_boxplot.png")
```

```
boxplot(data[,2:3])
```

```
dev.off()
```

2.1.4、正态性检验：

```
shapiro.test(data[,1])
```

```
shapiro.test(data[,2])
```

```
shapiro.test(data[,3])
```

2.1.5、方差齐性检验：

```
data2<-data.frame(X~c(data[,1],data[,2],data[,3]),A<-
```

```
factor(rep(1:3,c(24,24,24))))
```

```
bartlett.test(data)
```

```
library(car)
```

```
leveneTest(X~A,data=data2)
```

```
fligner.test(X~A,data=data2)
```

2.1.6、Kruskal-Wallis检验：

```
install.packages("agricolae")
```

```
library( agricolae)
```

```
#H0:M1=M2=M3, H1:三者不等
```

```
kruskal.test(X~A,data=data2)
```

2.1.7、单因素方差分析：

```
m<-aov(X~A,data=data2)
```

```
summary(m)
```

2.1.8、多重比较：

```
mm<-TukeyHSD(m)
```

```
mm
```

```
png("e12_data-2-1_TukeyHSD_plot.png")
```

```
plot(mm)
```

```
dev.off()
```

2.1.9、综合上述统计计算结果，进行对比分析讨论。

3、随机区组设计资料（多组）的秩和检验-Friedman检验

3.1、数据描述：

在某项实验中，9名受试对象对四种不同频率声音刺激的反应率（%）结果。

3.2、数据读取：

```
dir="D:/RFile/实验十二"
```

```
setwd(dir)
```

```
file="e12-data-3-1-win.txt"
```

```
data<-read.table(file,head=T,sep="\t")
```

```
data
```

3.3、数据可视化观察：

```
png("e12_data-3-1_boxplot.png")
```

```
boxplot(data[,2:5])
```

```
dev.off()
```

3.4、正态性检验：

```
apply(data[,2:5],2,shapiro.test)
```

3.5、方差齐性检验：

```
data2<-data.frame(X=-c(data[,2],data[,3],data[,4],data[,5]), A=-  
factor(rep(1:4,rep(9,4))))
```

```
bartlett.test(data[,2:5])
```

```
library(car)
```

```
leveneTest(X~A,data=data2)
```

```
fligner.test(X~A,data=data2)
```

3.6、Friedman检验：

```
dm<-as.matrix(data[,2:5])
```

```
dimnames(dm) <- list(1:9, c("A", "B", "C", "D"))
```

```
friedman.test(dm)
```

3.7、单因素方差分析：

```
m<-aov(X~A,data=data2)
```

```
summary(m)
```

3.8、多重比较：

```
mm<-TukeyHSD(m)
```

```
mm
```

```
png("e12_data-3-1_TukeyHSD_plot.png")
```

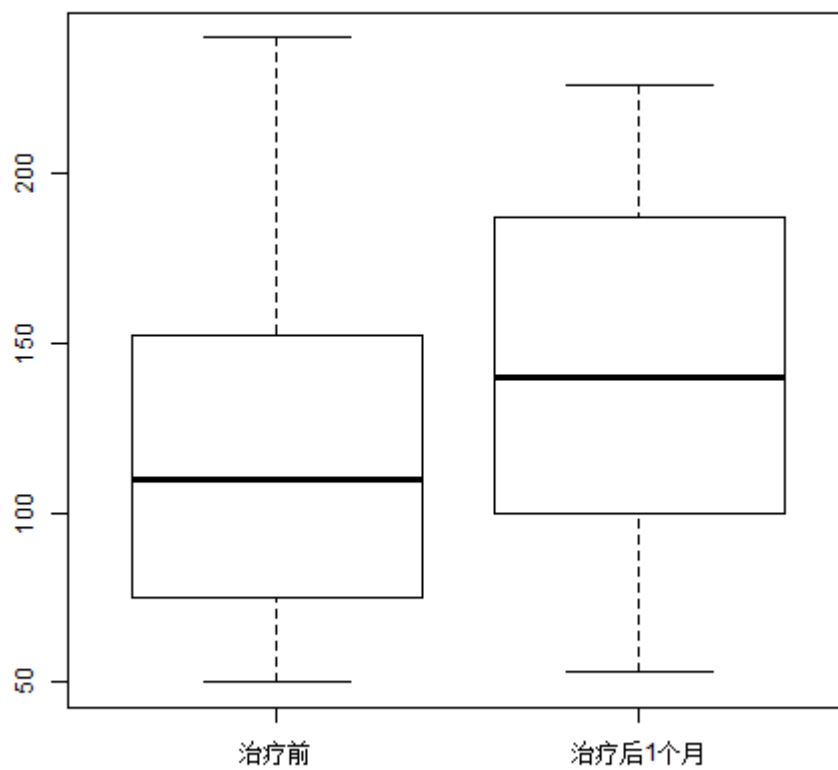
```
plot(mm)
```

```
dev.off()
```

3.9、综合上述统计计算结果，进行对比分析讨论。

三、实验结果

1.1.3



1.1.4

```
> shapiro.test(x)
```

Shapiro-Wilk normality test

data: x

W = 0.90004, p-value = 0.03505

```
> shapiro.test(y)
```

Shapiro-Wilk normality test

data: y

W = 0.94173, p-value = 0.2357

X即治疗前不是正太分布, y即治疗后是正太分布

1.1.5

```
> bartlett.test(data[, 2:3])
```

Bartlett test of homogeneity of variances

data: data[, 2:3]

Bartlett's K-squared = 0.0050007, df = 1, p-value = 0.9436

```
> leveneTest(X~A, data=data2)
```

Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
--	----	---------	--------

group	1	0.0024	0.9611
-------	---	--------	--------

	40		
--	----	--	--

```
> fligner.test(X~A, data=data2)
```

Fligner-Killeen test of homogeneity of variances

data: X by A

Fligner-Killeen:med chi-squared = 0.043842, df = 1, p-value = 0.8341

1.1.6

```
> wilcox.test(x, y, paired = TRUE, alternative = "two.sided")
```

Wilcoxon signed rank test with continuity correction

data: x and y

V = 57.5, p-value = 0.04525

alternative hypothesis: true location shift is not equal to 0

1.1.7

```
> t.test(x,y, paired = TRUE, var.equal=TRUE, alternative = "two.sided")
```

Paired t-test

```
data: x and y
t = -2.1572, df = 20, p-value = 0.04333
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -40.4632590 -0.6795982
sample estimates:
mean of the differences
      -20.57143
```

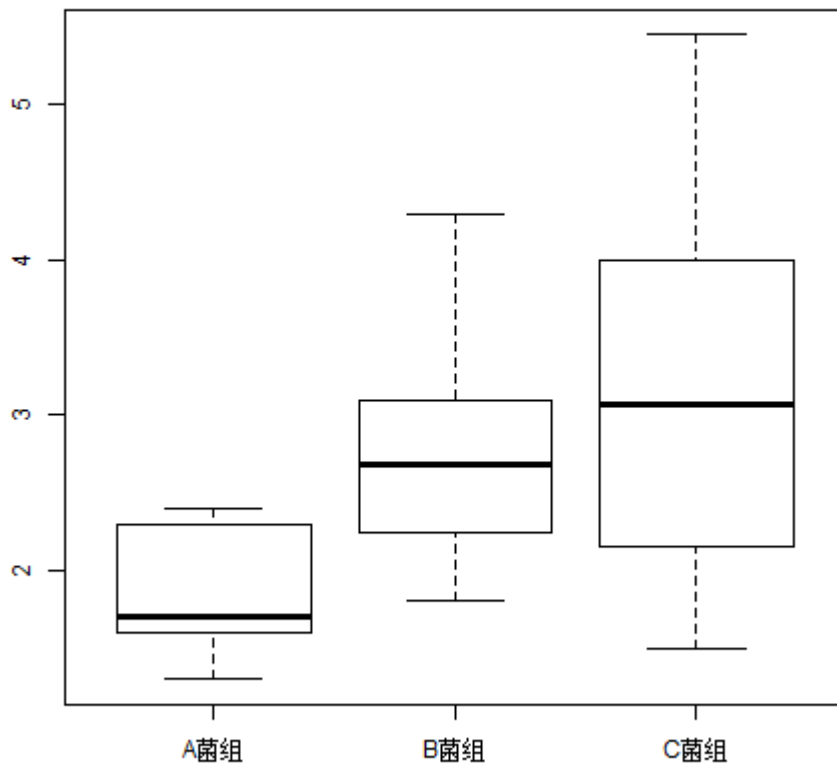
1.1.8

```
> t.test(x,y, paired = TRUE, alternative = "two.sided")
```

Paired t-test

```
data: x and y
t = -2.1572, df = 20, p-value = 0.04333
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -40.4632590 -0.6795982
sample estimates:
mean of the differences
      -20.57143
```

2.1.3



2.1.4

```
> shapiro.test(data[,1])
```

Shapiro-Wilk normality test

```
data: data[, 1]
```

```
W = 0.8653, p-value = 0.01852
```

```
> shapiro.test(data[,2])
```

Shapiro-Wilk normality test

```
data: data[, 2]
```

```
W = 0.89125, p-value = 0.0141
```

```
> shapiro.test(data[,3])
```

Shapiro-Wilk normality test

```
data: data[, 3]
```

```
W = 0.90834, p-value = 0.1094
```

A, B 不满足正太分布, C 组是正太分布

2.1.5

```
> bartlett.test(data)
```

Bartlett test of homogeneity of variances

```
data: data
```

```
Bartlett's K-squared = 21.385, df = 2, p-value = 2.271e-05
```

```
> library(car)
```

```
> leveneTest(X~A, data=data2)
```

Levene's Test for Homogeneity of Variance (center = median)

	Df	F value	Pr(>F)
group	2	19.31	4.716e-07 ***
	54		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> fligner.test(X~A, data=data2)
```

Fligner-Killeen test of homogeneity of variances

```
data: X by A
```

```
Fligner-Killeen:med chi-squared = 26.674, df = 2, p-value = 1.613e-06
```

2.1.6

```
> kruskal.test(X~A, data=data2)
```

Kruskal-Wallis rank sum test

```
data: X by A
```

```
Kruskal-Wallis chi-squared = 20.265, df = 2, p-value = 3.977e-05
```

2.1.7

```
> m<-aov(X~A, data=data2)
```

```
> summary(m)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	2	15.68	7.838	12.2	4.25e-05 ***
Residuals	54	34.69	0.642		

2.1.8

```
> mm<-TukeyHSD(m)
```

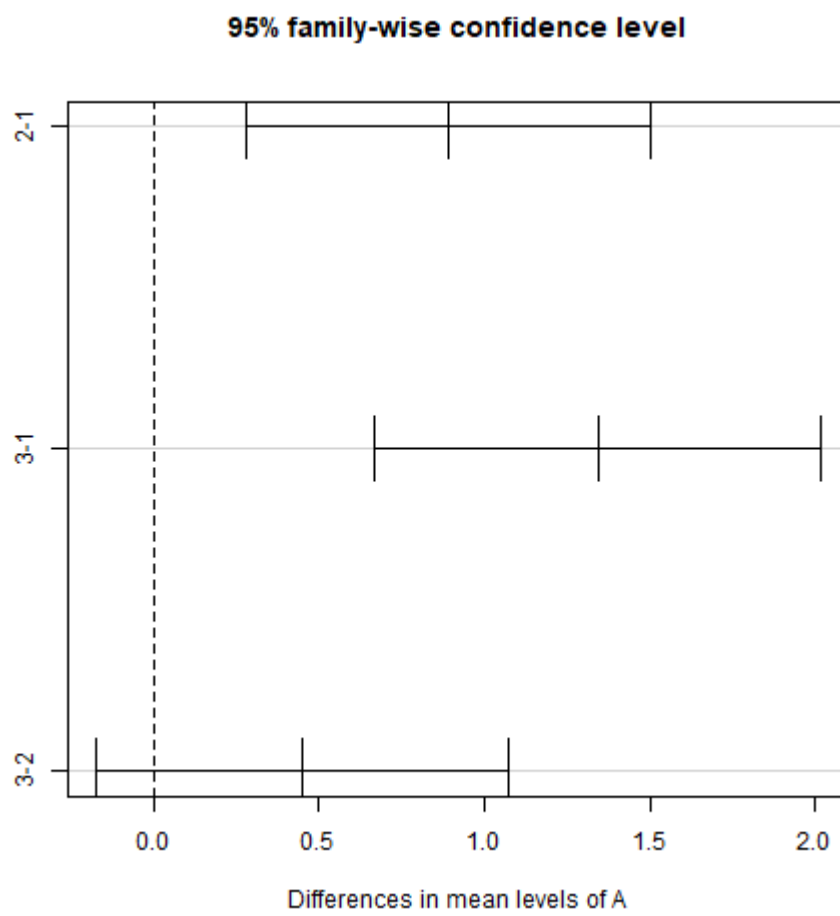
```
> mm
```

Tukey multiple comparisons of means
95% family-wise confidence level

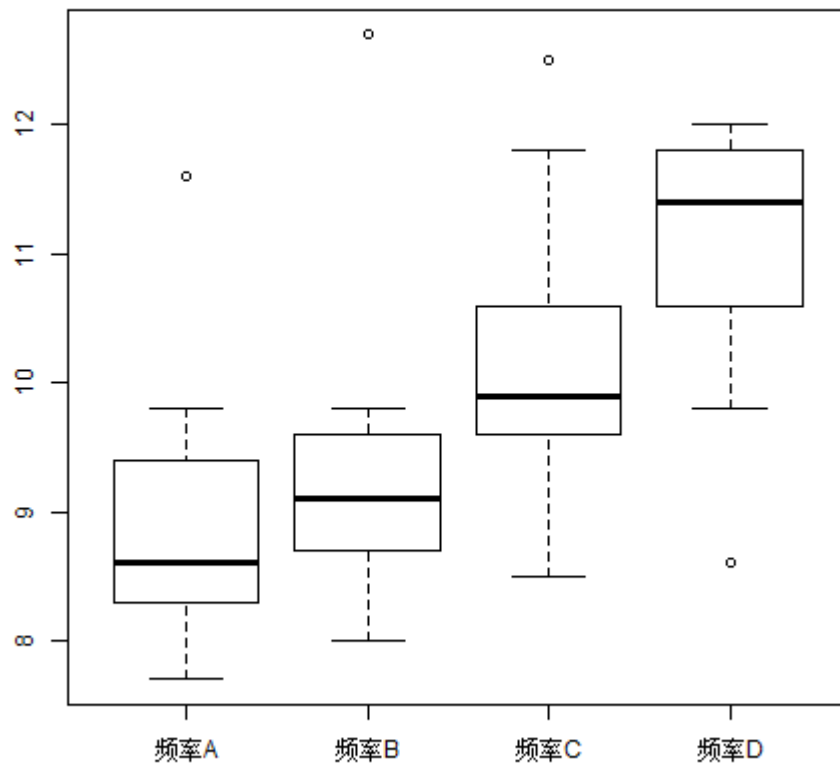
Fit: aov(formula = X ~ A, data = data2)

\$A

	diff	lwr	upr	p adj
2-1	0.8930392	0.2806852	1.505393	0.0025517
3-1	1.3428309	0.6699887	2.015673	0.0000369
3-2	0.4497917	-0.1736633	1.073247	0.2003202



3.3



3.4

```
> apply(data[, 2:5], 2, shapiro.test)
```

\$频率 A

Shapiro-Wilk normality test

data: newX[, i]

W = 0.88197, p-value = 0.1647

\$频率 B

Shapiro-Wilk normality test

data: newX[, i]

W = 0.80909, p-value = 0.02592

\$频率 C

Shapiro-Wilk normality test

data: newX[, i]

W = 0.94302, p-value = 0.614

\$频率 D

Shapiro-Wilk normality test

data: newX[, i]

W = 0.86434, p-value = 0.1069

除了频率 B 以外，其余均符合正态分布

3.5

> bartlett.test(data[, 2:5])

Bartlett test of homogeneity of variances

data: data[, 2:5]

Bartlett's K-squared = 0.30677, df = 3, p-value = 0.9587

> library(car)

> leveneTest(X~A, data=data2)

Levene's Test for Homogeneity of Variance (center = median)

Df F value Pr(>F)

group 3 0.0443 0.9874

32

> fligner.test(X~A, data=data2)

Fligner-Killeen test of homogeneity of variances

data: X by A

Fligner-Killeen:med chi-squared = 0.14651, df = 3, p-value = 0.9857

3.6

```
> friedman.test(dm)
```

Friedman rank sum test

data: dm

Friedman chi-squared = 17.225, df = 3, p-value = 0.0006354

3.7

```
> m<-aov(X~A, data=data2)
```

```
> summary(m)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	3	21.44	7.145	4.422	0.0104 *
Residuals	32	51.71	1.616		

3.8

```
> mm<-TukeyHSD(m)
```

```
> mm
```

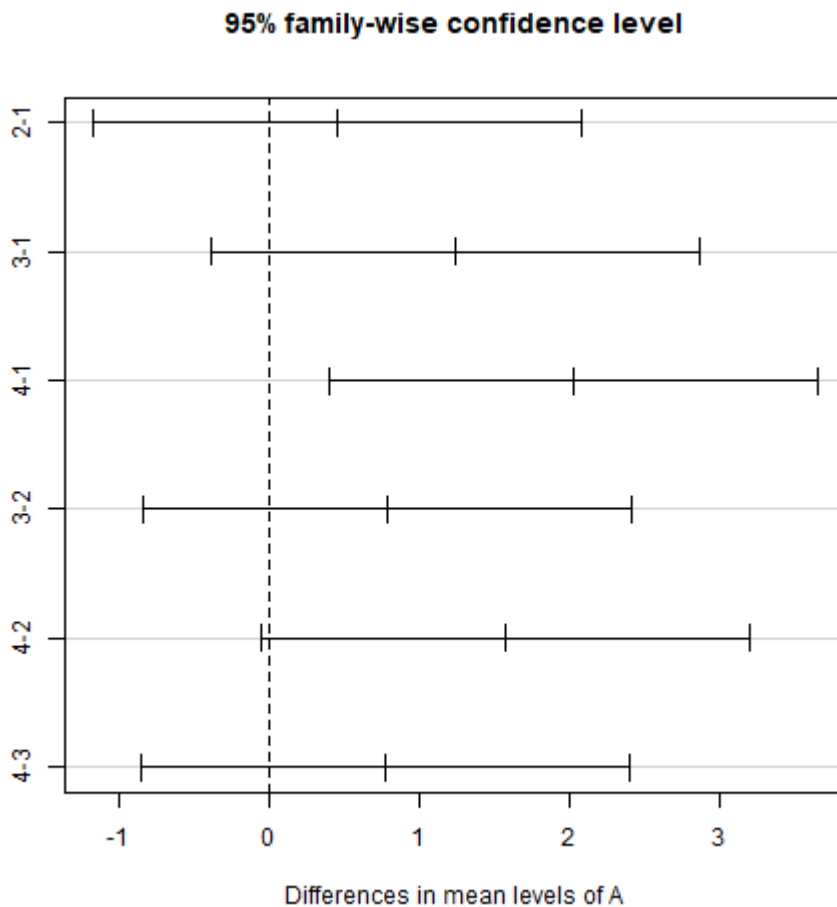
Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = X ~ A, data = data2)

\$A

	diff	lwr	upr	p adj
2-1	0.4555556	-1.1680393	2.079150	0.8716145
3-1	1.2444444	-0.3791504	2.868039	0.1824383
4-1	2.0222222	0.3986274	3.645817	0.0100093
3-2	0.7888889	-0.8347060	2.412484	0.5595152
4-2	1.5666667	-0.0569282	3.190262	0.0617587
4-3	0.7777778	-0.8458171	2.401373	0.5709292



四、讨论：

当样本满足正态分布时，用参数检验 barlett 更为合适，而当数据不满足分布或方差不齐时，参数检验可能会产生错误，而非参数检验则较为准确。

1.1.9

由于样本数据一个符合正太分布，一个不符合。但并不那么极端，所以两种检验结果有一定的不同，但结论一致。经过 wilcox 检验可知拒绝原假设 $M_d=0$ ，所以两组数据有显著性差异。和下面的参数检验 t 检验所得结论一致。

2.1.9

Kruskal-Wallis H检验是类似单因素方差分析的非参数方法可以认为是独立样本Wilcoxon秩和检验在数据组超过两组情况下的扩展。

本例中应拒绝原假设，三组数据具有显著性差异。使用参数检验得到的结论是一样的。由多重比较图可以看出，1和3,2和3有明显差异，而2和3有较大差异。

3.9

Friedman检验是类似于重复测量资料ANOVA检验的非参数检验，可以认为是配对样本Wilcoxon秩和检验在数据组超过两组情况下的扩展。本例中拒绝原假设说明四组数据具有显著性差异。后面的参数检验avo结论与之一致。有多重比较图可以看出，4-1和4-2差距较大其余的差异一般。