南京郵電大學

实 验 报 告

（ 2022 / 2023 学年 第 一 学期）

课程名称 非参数统计

实验名称 非参数统计上机实验

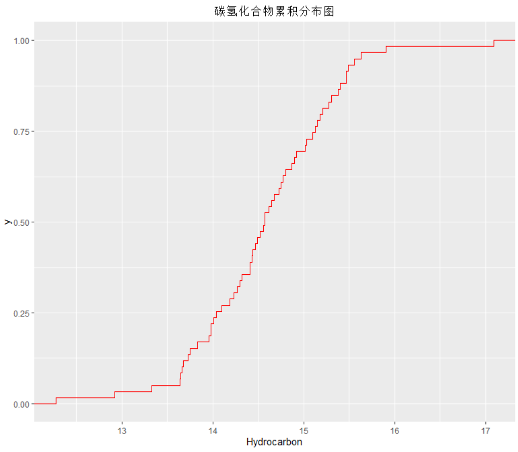
学生姓名 王畅 班级学号 B20070412

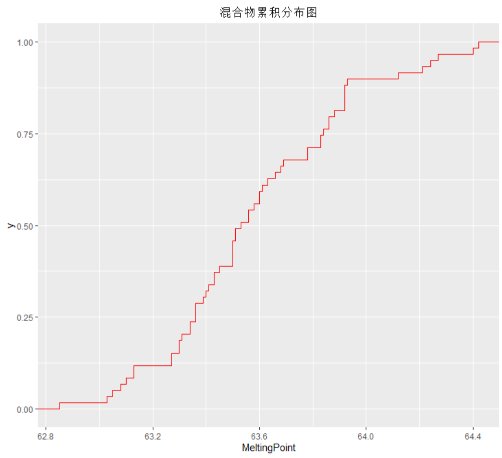
学院(系) 理学院 专 业 应用统计学

目录

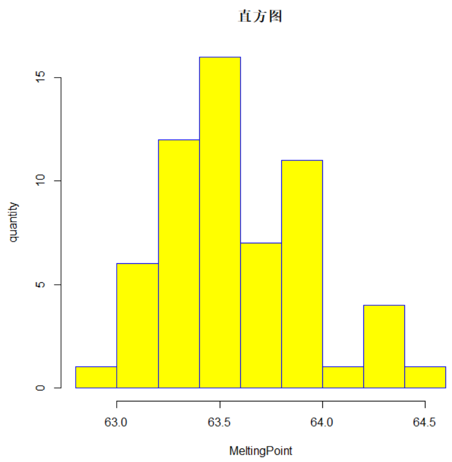
1.通过 R 语言绘制题目图片

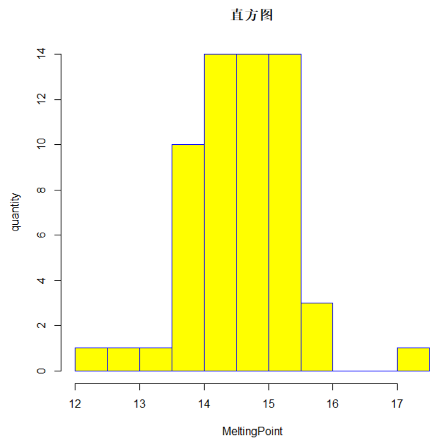
1) 应有累积分布图如下



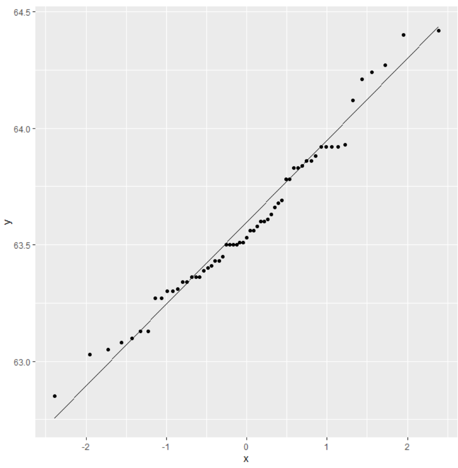


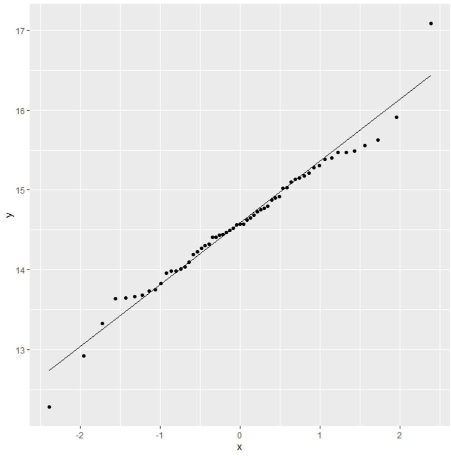
应有直方图如下





应有 Q-Q 图如下





2) 计算得到的分位数有：

碳氢化合物的分位数有：

```
+      quantile(rt$Hydrocarbon,probs = c(0.90,0.75,0.50,0.25,0.10))
+
+ )
    90%    75%    50%    25%    10%
15.470 15.115 14.570 14.070 13.676
```

混合物的分位数有：

```
+   quantile(rt$MeltingPoint,probs = c(0.90,0.75,0.50,0.25,0.10))
+
+ )
    90%    75%    50%    25%    10%
63.968 63.835 63.530 63.360 63.130
```

3) 通过 Q-Q 图判断，我认为原分布是 Gauss 分布

代码如下：

```
library(ggplot2)
library(dplyr)
setwd("C:\\Users\\耶梦加得\\Desktop\\R 语言实验\\非参数统计实验");
rt<-read.table("beenswax.txt",head=TRUE);
```

```
#x11(1)
ggplot(rt,aes(x = MeltingPoint))+
  stat_ecdf(color = "red")+
  labs(title = "混合物累积分布图")+
  theme(plot.title = element_text(hjust = 0.5))
#x11(2)
ggplot(rt,aes(x = Hydrocarbon))+
  stat_ecdf(color = "red")+
  labs(title = "碳氢化合物累积分布图")+
  theme(plot.title = element_text(hjust = 0.5))
#直方图
#x11()
with(rt,hist(MeltingPoint, main="直方图",xlab = "MeltingPoint", ylab="quantity",col = "yellow",border = "blue"));
#x11()
with(rt,hist(Hydrocarbon, main="直方图",xlab = "MeltingPoint", ylab="quantity",col = "yellow",border = "blue"));
#Q-Q 图
#()
tibble(y = rt$MeltingPoint)%>%
  ggplot(aes(sample = y)) +
  geom_qq() + geom_qq_line()
#x11()
tibble(y = rt$Hydrocarbon)%>%
  ggplot(aes(sample = y)) +
  geom_qq() + geom_qq_line()
#计算分位数
with(rt,
  quantile(rt$MeltingPoint,probs = c(0.90,0.75,0.50,0.25,0.10))

)
with(rt,
     quantile(rt$Hydrocarbon,probs = c(0.90,0.75,0.50,0.25,0.10))

)


```

2.判断城市死亡率是否有逐年增加的趋势

本题通过 Cox-Staut 方法检查数据是否有趋势

```
#2.3 题，通过 Cox-staut 法检验数据的趋势性质
chs<-c(17.3,17.9,18.4,18.1,18.3,19.6,18.6,19.2,17.7,20.0,19.0,18.8,19.3,20.2,19.9);
lchs<-length(chs);
#已知共有 15 个数据，则应该在数组中增补第 16 个数据于第 7 个数据之后
#增补的数据与第七个数据相同
chs1<-c(17.3,17.9,18.4,18.1,18.3,19.6,18.6,19.2,19.2,17.7,20.0,19.0,18.8,19.3,20.2,19.9);
lchs1<-length(chs1);
sf=0;sz=0;
for (i in 1:7) {
  panduan<-chs1[i]-chs1[i+9];
  #求取 S-的值
  if(panduan<0){
    sf=sf+1;
  };
  #求取 S+的值
  if(panduan>0){
    sz=sz+1;
  };

}
K=min(sf,sz);
#假设 H0：该城市死亡率有逐年上升的趋势，则有备择假设为 H1：其没有上升的趋势
#此时有 P 值为
p=1/2^15;
#P 值远小于显著性质数 0.05
#认为这个时候接受原假设，既这 15 年死亡率有上升的趋势
```

通过验证数据，认为这 15 年内该城市死亡率有上升趋势

3.检验篮球赛三分球得分次数是否有显著性差异

1）通过符号检验检验差异

```
number of successes = 5, number of trials = 10, p-value = 0.623
alternative hypothesis: true probability of success is less than 0.5
95 percent confidence interval:
 0.0000000 0.7775589
95 percent confidence interval:
 0.0000000 0.7775589
```

2）通过配对 Wilcoxon 符号秩检验和 T 检验得到的结果如下

Wilcoxon 符号秩检验：

```
data:  data1$联赛 1
V = 24.5, p-value = 0.3798
alternative hypothesis: true location is less than 95
data:  data1$联赛 2
V = 29, p-value = 0.5609
alternative hypothesis: true location is less than 57
```

T 检验结果：

```
data:  data1$联赛 1 and data1$联赛 2
t = 2.2377, df = 11.997, p-value = 0.04499
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  0.8863231 66.5136769
sample estimates:
mean of x mean of y
     92.2      58.5
```

3）综上所述，这些数据中 Wilcoxon 检验更好，能够更加直观的反映数据与中位数之间的检验差异，而符号秩检验过分的关注中位数与数据的偏差值，不能良好反映数据的情况

总代码如下：

```
联赛 1<-c(91,46,108,99,110,105,191,57,34,81);
联赛 2<-c(81,51,63,51,46,45,66,64,90,28);
data1<-data.frame(联赛 1,联赛 2);
zhw1<-median(data1$联赛 1);
zhw2<-median(data1$联赛 2);
#对两组数据做符号检验 binom.test
binom.test(sum(data1$联赛 1>zhw1),length(data1$联赛 1),al="l");
binom.test(sum(data1$联赛 2>zhw2),length(data1$联赛 2),al="l");
#对两组数据作 Wilcoxon 符号秩检验
#对于已经计算出其中位数，直接检验其是否存在显著性差异
wilcox.test(data1$联赛 1,mu=zhw1,alternative = "less",exact = FALSE,correct = FALSE,conf.int = TRUE)
wilcox.test(data1$联赛 2,mu=zhw2,alternative = "less",exact = FALSE,correct = FALSE,conf.int = TRUE)
#用 T 检验检验样本数据
t.test(data1$联赛 1,data1$联赛 2)

#综上所述，这些数据中 Wilcox 检验更好，能够更加直观的反映数据与中位数之间的检验差异
```

4.判断信号是否为随机干扰

本实验通过游程检验判断信号是否为随机干扰

代码如下：

```
library(lawstat)
```

```
xinhao<-c(0,1,0,1,1,1,0,0,1,1,0,0,0,0,1,1,1,1,1,1,1,1,1,0,1,0,0,1,1,1,0,1,0,1,0,1,0,0,0,0,0,0,0,0,1,0,1,1,0,0,1,1,1,0,1,0,1,0,0,0,1,0,0,1,0,1,0,1,0,0,0,0,0,0,0,0);
xinhao1=xinhao-0.5
x11()
plot(1:76,xinhao1)
x11()
#H0,假设原信号是随机干扰
runs.test(xinhao1,plot.it=T)
#p-value < 2.2e-16,p 值不到0.05，可以拒绝是随机的假设。
```

得到的结论如下：

```
> runs.test(xinhao1,plot.it=T)

 Runs Test - Two sided

data:  xinhao1
Standardized Runs Statistic = -Inf, p-value < 2.2e-16
```

p-value < 2.2e-16,p 值不到 0.05，可以拒绝是随机的假设。

既认为原信号不是随机的

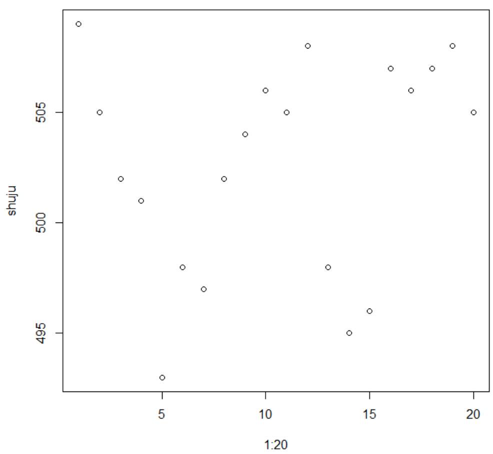
5.检查机器装多装少是否是随机的

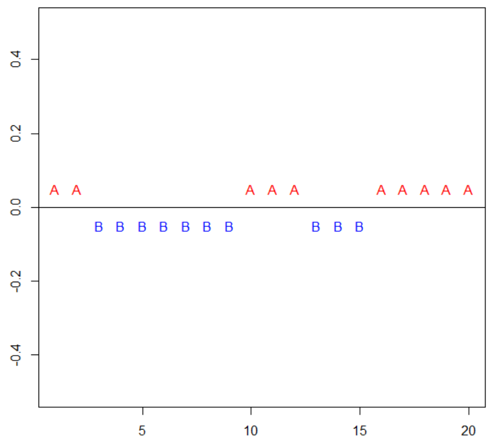
本实验通过游程检验判断装多少的数据是否为随机的

代码如下：

```
shuju<-c(509,505,502,501,493,498,497,502,504,506,505,508,498,495,496,507,506,507,508,505);
x11()
plot(1:20,shuju)
#H0,假设装多装少是随机的
runs.test(shuju,plot.it=T)
# p-value = 0.005837,p 值不到0.05，可以拒绝是随机的假设。
```

应有数据分布图如下：





得到的结论如下：

```
 Runs Test - Two sided

data:  shuju
Standardized Runs Statistic = -2.7568, p-value = 0.005837
```

既认为机器出现了问题，装多少并不是随机的

6.检验捕获的鱼的长度

由题意，应该有：

```
changdu<-c(64,65,65,66,67,rep(68,4),rep(69,3),rep(70,4),rep
(71,5),72,72,72,73,73,73,75,rep(77,6),78,83);
print(changdu);
#求样本中位数的置信区间
#求鱼长度的中位数
fish.median<-median(changdu);
splus<-sum(changdu>fish.median);
sminus<-sum(changdu<fish.median);
k=min(splus,sminus);
n=splus+sminus;
binom.test(k,n,0.5);
wilcox.test(changdu-fish.median);
plot(density(changdu));
ks.test(changdu,pnorm,maen(changdu))
```

```
#则有中位数的置信区间有：
```

得到其 95%的置信区间应该为（65，75）

7.判断两个讲师的课时量是否相同

通过 Brown-Mood 方法和 Wilcoxon 方法检验其差别，应有代码如下：

```
BMq.test=function(x,y,q,alt)
{
  xy=c(x,y)
  quantile.xy=quantile(xy,q)
  t=sum(xy>quantile.xy)
  lx=length(x[x!=quantile.xy])
  ly=length(y[y!=quantile.xy])
  lxy=lx+ly
  A=sum(x>quantile.xy)
  z=(A-lx*t/(lx+ly))/(lx*ly*t*(lx+ly-t)/(lx+ly)^3)^0.5
  if(A>(min(lx,t)/2)){
    z1=(A+0.5-lx*t/(lx+ly))/(lx*ly*t*(lx+ly-t)/(lx+ly)^3)^0.5
  }
  else{
    z1=(A-0.5-lx*t/(lx+ly))/(lx*ly*t*(lx+ly-t)/(lx+ly)^3)^0.5
  }
  if(alt=="greater"){
    pv1=1-phyper(A,lx,ly,t)
    pv2=1-pnorm(z)
    pv3=1-pnorm(z1)
  }
  if(alt=="less"){
    pv1=phyper(A,lx,ly,t)
    pv2=pnorm(z)
    pv3=pnorm(z1)
  }
  if(alt=="two.side"){
    pv1=2*min(1-phyper(A,lx,ly,t),phyper(A,lx,ly,t))
    pv2=2*min(1-pnorm(z),pnorm(z))
    pv3=2*min(1-pnorm(z1),pnorm(z1))
  }
```

```
  conting.table<-matrix(c(A,lx-A,lx,t-A,ly-(t-A),ly,t,lxy-t,lxy),3,3)
  col.name<-c("X","Y","X+Y");
  row.name<-c(">MQXY","<MQXY","TOTAL")
  dimnames(conting.table)=list(row.name,col.name)
  list(contingency.table=conting.table,p.value=pv1,pvnorm=pv2,pvnr=pv3)
}
#假设两组数据的中位数的相同的，既有H0：讲课课时没有差异，从中位数判断应该相等
a<-c(321,266,256,386,330,329,303,334,299,221,365,250,258,342,243,298,238,317);
b<-c(488,593,507,428,807,342,512,350,672,589,665,549,451,492,514,391,366,469);
BMq.test(a,b,0.5,"two.side")
#$p.value= 5.230556e-06，其P 值过低，既置信区间很小，拒绝原假设
#用Wilcox 检验两组数据
#假设两组数据没有中位数的差距
wilcox.test(a,b)
#p-value = 7.977e-07，P 值过小，认为置信区间过小，拒绝原假设
```

假设两组数据的中位数的相同的，既有 H0：讲课课时没有差异，从中位数判断应该相等

但是有结果：

```
$p.value
[1] 5.230556e-06
```

p.value= 5.230556e-06，其 P 值过低，既置信区间很小，拒绝原假设

用 Wilcox 检验两组数据：

```
Wilcoxon rank sum test with continuity correction

data:  a and b
W = 5.5, p-value = 7.977e-07
```

p-value = 7.977e-07，P 值过小，认为置信区间过小，拒绝原假设

8.对土壤数据进行方差检验

通过 Mood 检验法和 Moses 检验法对数据的方差进行检验

代码如下：

```
a<-c(8.8,8.2,5.4,4.9,8.9,4.2,5.6,7.1,5.5,8.6,6.3,3.9);
b<-c(13.0,14.5,16.5,22.8,20.7,19.6,18.4,21.3,24.2,19.6,11.7
,18.9,14.6,19.8,14.5);
#用 Mood 方差检验法
#不妨假设 H0: 两种样品的方差相同
length.a<-length(a);
length.b<-length(b);
m=n=(12+15+1)/2;
c<-c(a,b);
#用平均秩法求秩
d<-rank(c,ties.method = "average")
data1<-data.frame(cbind(c,d))
M<-0;
for (i in 1:12){
  br<-(data1[i,2]-m)^2;
  M<-M+br;
}
E<-m*(m+n+1)*(m+n-1)/12;
varM<-m*n*(m+n+1)*(m+n+2)*(m+n-2)/180;
Z<-(M-E+0.5)/sqrt(varM)
print(Z)
# -0.6053207<Z_{0.05/2}=1.96  得到结论是：不能拒绝原假设 既两种
样品方差相同

#用 Moses 中位数检验法
#将数据 a 分成 4 组，将数据 b 分成 5 组
#求各小组的离差平方和应有
amean<-c(mean(a[1:3]),mean(a[4:6]),mean(a[7:9]),mean(a[10:1
2]));
SSA1=SSA2=SSA3=SSA4=SSB1=SSB2=SSB3=SSB4=SSB5=0
for (i in 1:3){
  br<-(a[i]-amean[1])^2
  SSA1=SSA1+br;
}
for (i in 4:6){
  br<-(a[i]-amean[2])^2
  SSA2=SSA2+br;
```

```
}
for (i in 7:9){
  br<-(a[i]-amean[3])^2
  SSA3=SSA3+br;
}
for (i in 10:12){
  br<-(a[i]-amean[4])^2
  SSA4=SSA4+br;
}
bmean<-c(mean(b[1:3]),mean(b[4:6]),mean(b[7:9]),mean(b[10:12]),mean(b[13:15]));
for (i in 1:3){
  br<-(b[i]-bmean[1])^2
  SSB1=SSB1+br;
}
for (i in 4:6){
  br<-(b[i]-bmean[1])^2
  SSB2=SSB2+br;
}
for (i in 7:9){
  br<-(b[i]-bmean[1])^2
  SSB3=SSB3+br;
}
for (i in 10:12){
  br<-(b[i]-bmean[1])^2
  SSB4=SSB4+br;
}
for (i in 13:15){
  br<-(b[i]-bmean[1])^2
  SSB5=SSB5+br;
}
shujuA<-c(SSA1,SSA2,SSA3,SSA4);
shujuB<-c(SSB1,SSB2,SSB3,SSB4,SSB5);
zongshuju<-c(shujuA,shujuB);
rankshuju<-rank(zongshuju,ties.method = "average")
#则有第一组平方和的秩和 S 为
S<-0;
for(i in 1:4){
  S=S+rankshuju[i]
}
TM<-S-4*(4+1)/2
print(TM)
#TM=3 查询Moses 的附表应有: W_{0.05，4，5}=3 拒绝H0
#既认为两组数据的方差不相等 ， 存在差异
```

不妨假设 H0：两种样品的方差相同

在 Mood 检验法中，有：

```
print(Z)
[1] -0.6053207
```

-0.6053207<Z_{0.05/2}=1.96 得到结论是：不能拒绝原假设 既两种样品方差相同

在 Moses 检验法中，有：

```
print(TM)
[1] 3
```

TM=3 查询 Moses 的附表应有：W_{0.05，4，5}=3 拒绝 H0

既认为两组数据的方差不相等 ，存在差异

9.检验三种汽车公司油耗是否存在差异

本题通过 Frieldan 检验和 Hodge-lehmman 检验方法进行检验，应有代码如下：

```
youhao<-c(20.3,21.2,18.2,18.6,18.5,25.6,24.7,29.3,19.3,20.7,24.0,23.1,20.6,19.8,21.4);
treat.YH<-c(1,2,3,4,5,1,2,3,4,5,1,2,3,4,5);
block.YH<-c(1,1,1,1,1,2,2,2,2,2,3,3,3,3,3);
#通过 Fieldman 检验应有：H0：其品质不存在差异
friedman.test(youhao,treat.YH,block.YH)
#五类数据，自由度应该为 4
#此时有：  chi-squared = 6.1333<X^2_{0.05,4}=9.488
#认为可以接受 H0，既油耗没有差异
meanyh<-rep(0,3)
meanyh[1]<-mean(youhao[1:5]);meanyh[2]<-mean(youhao[6:10]);meanyh[3]<-mean(youhao[10:15]);
tzyh<-rep(0,15)
for(i in 1:5){
  tzyh[i]<-youhao[i]-meanyh[1]
}
```

```
for(i in 6:10){
  tzyh[i]<-youhao[i]-meanyh[2]
}
for(i in 11:15){
  tzyh[i]<-youhao[i]-meanyh[3]
}
rank.tzyh<-rank(tzyh,ties.method = "average")
data1<-data.frame(tzyh,rank.tzyh)
meanRj=rep(0,5)
meanRj[1]<-(1/3)*(rank.tzyh[1]+rank.tzyh[6]+rank.tzyh[11])
meanRj[2]<-(1/3)*(rank.tzyh[2]+rank.tzyh[7]+rank.tzyh[12])
meanRj[3]<-(1/3)*(rank.tzyh[3]+rank.tzyh[8]+rank.tzyh[13])
meanRj[4]<-(1/3)*(rank.tzyh[4]+rank.tzyh[9]+rank.tzyh[14])
meanRj[5]<-(1/3)*(rank.tzyh[5]+rank.tzyh[10]+rank.tzyh[15])
meanRi=rep(0,3)
meanRi[1]<-mean(rank.tzyh[1:5])
meanRi[2]<-mean(rank.tzyh[6:10])
meanRi[3]<-mean(rank.tzyh[11:15])
R.j<-rep(0,3);
for(i in 1:5){
  br<-rank.tzyh[i]
  R.j[1]=R.j[1]+br;
}
for(i in 6:10){
  br<-rank.tzyh[i]
  R.j[2]=R.j[2]+br;
}
for(i in 11:15){
  br<-rank.tzyh[i]
  R.j[3]=R.j[3]+br;
}
totalR.j<-R.j[1]^2+R.j[2]^2+R.j[3]^2;
R.i<-rep(0,5)
R.i[1]<-rank.tzyh[1]+rank.tzyh[6]+rank.tzyh[11]
R.i[2]<-rank.tzyh[2]+rank.tzyh[7]+rank.tzyh[12]
R.i[3]<-rank.tzyh[3]+rank.tzyh[8]+rank.tzyh[13]
R.i[4]<-rank.tzyh[4]+rank.tzyh[9]+rank.tzyh[14]
R.i[5]<-rank.tzyh[5]+rank.tzyh[10]+rank.tzyh[15]
totalR.i<-R.i[1]^2+R.i[2]^2+R.i[3]^2+R.i[4]^2+R.i[5]^2;
#计算检验统计量Q
Q=((5-1)*(totalR.j-5*3^2*((5*3+1)^2)/4))/((1/6)*3*5*(5*3+1)
*(2*5*3+1)-(1/5)*totalR.i)
print(Q)
#得到统计量为Q= 13.43117>X^2_{0.05,4}=9.488
```

```
#既通过HL 检验认为油耗存在差异
```

通过 Fieldman 检验应有假设：H0：其品质不存在差异

通过 Fieldman 检验有结果：

```
 Friedman rank sum test

data:  youhao, treat.YH and block.YH
Friedman chi-squared = 6.1333, df = 4, p-value = 0.1894
```

此时有： chi-squared = 6.1333<X^2_{0.05,4}=9.488

认为可以接受 H0，既油耗没有差异

通过 HL 检验应有结果：

```
> print(Q)
[1] 13.43117
```

得到统计量为 Q= 13.43117>X^2_{0.05,4}=9.488

既通过 HL 检验认为油耗存在差异

10. 分析推销员推销结果

本实验通过 Cochran 检验法进行检验，有代码如下：

```
#Cochran 检验法
#不妨假设原假设 H0：三种推销效果相同
candid1<-c(rep(1,6),0,0,1,1,1,0);
candid2<-c(0,1,0,1,0,0,0,1,0,0,0,0)
candid3<-c(1,0,1,0,0,1,0,1,0,0,0,1)
candid<-matrix(c(candid1,candid2,candid3),nrow = 12,ncol =
3);
nidot.candid=apply(candid, 1, sum)
ndotj.candid=apply(candid, 2, sum)
k=ncol(candid)
Q=(k-1)*((k*sum(ndotj.candid^2)-(sum(ndotj.candid))^2))/+(k
*sum(nidot.candid)-sum(nidot.candid^2))
pvalue.candid=pchisq(Q,k-1,lower.tail = F)
pvalue.candid
# pvalue=0.07843739>0.05，接受原假设 H0，既效果相同
```

不妨假设原假设 H0：三种推销效果相同

得到的结果有：

```
1.  pvalue.candid
2. [1] 0.07843739
```

pvalue=0.07843739>0.05，接受原假设 H0，既效果相同