武漢科技大学

非参数统计实验报告

专业班级：统计

学　　号：

姓　　名：

成　　绩：

武汉科技大学理学院

数学与统计系

填 写 说 明

1. 排版要求：正文小四号字体，中文使用宋体，西文使用 Time New Roman 字体，行距使用 1.25 倍行距。

2. 内容要求：按题目要求填写内容。运行结果截图大小适当。

3. 双面打印。

实验四　Cox-Stuart 趋势检验和随机性游程检验

实验名称	Cox-Stuart 趋势检验和随机性游程检验		
指导教师	宋硕	实验日期	2024.10.27

一、　实验目的

1、熟悉 Cox-Stuart 趋势检验和随机性游程检验的应用条件及检验步骤；

2、熟练应用 Cox-Stuart 趋势检验和随机性游程检验方法对相关问题进行检验；

3、使用 R 软件进行 Cox-Stuart 趋势检验和随机性游程检验.

二、　检验原理和步骤：

1、Cox-Stuart 趋势检验：

设数据序列 X_1，X_2,L X_n 独立，

1.1 双边假设检验问题: H_0: 数据序列无趋势 $\leftrightarrow H_1$: 数据序列有增长或减少趋势，

令
$$c=\begin{cases}\dfrac{n}{2}, & n\text{为偶数}\\ \dfrac{n+1}{2}, & n\text{为基数}\end{cases}$$

取 x_i 和 x_{i+c} 组成数对 $\left(x_i，x_{i+c}\right)$.当 n 为偶数时，共有 $n'=c$ 对,当 n 为奇数时，共有 $n'=c-1$ 对.计算每一数对前后两值之差：$D_i=x_i-x_{i+c}$.用 D_i 的符号度量增减.令 S^+ 为正 D_i 的数目，令 S^- 为正负 D_i 的数目，$S^+ + S^- = n'$.令 $K=\min\left(S^+,S^-\right)$，显然当正号或负号太多时，即 K 过小的时候，认为数据存在趋势.

在零假设下，$K=\min\left(S^+,S^-\right)$: $B\left(n',0.5\right)$，从而转化为符号检验问题.

1.2 对于单边检验问题：H_0: 数据序列有下降趋势 $\leftrightarrow H_1$: 数据序列有上升趋势，

H_0: 数据序列有上升趋势 $\leftrightarrow H_1$: 数据序列有下降趋势，

结果类似，S^+ 很大（或 S^- 很小时），有下降趋势；反之，有上升趋势.

和符号检验几乎类似，Cox-Stuart 趋势检验过程总结于下表：

零假设：H_0	备择假设：H_1	检验统计量(K)	p 值
H_0:无上升趋势	H_1:有上升趋势	$S^+ = \sum \text{sign}(D_i)$	$P(S^+ \leqslant k)$
H_0:无下降趋势	H_1:有下降趋势	$S^- = \sum \text{sign}(-D_i)$	$P(S^- \leqslant k)$
H_0:无趋势	H_1:有上升或下降趋势	$\min(S^-, S^+)$	$2P(K \leqslant k)$

小样本时，用近似正态统计量 $Z = (K \pm 0.5 - n'/2) / \sqrt{n'/4}$

（在 ± 处，$K < n'/2$ 时取加号，$K > n'/2$ 时取减号）

大样本时，用近似正态统计量 $Z = (K - n'/2) / \sqrt{n'/4}$

对水平 α，如果 p 值 < α，拒绝 H_0；否则不能拒绝

2、随机性游程检验：

设是由 0 或 1 组成的序列 $X_1, X_2, \text{L}\ X_N$，假设检验问题：

$$H_0: \text{数据出现顺序随机} \leftrightarrow H_1: \text{数据出现不随机},$$

R 为游程个数，假设有 m 个 0，n 个 1，$m+n=N$. Mood(1940)证明了如果随机性的原假设成立，则 R 在 m 和 n 给定下的条件分布为：

$$P(R=2k)=\frac{2\binom{m-1}{k-1}\binom{n-1}{k-1}}{\binom{N}{n}}, P(R=2k+1)=\frac{\binom{m-1}{k-1}\binom{n-1}{k}+\binom{m-1}{k}\binom{n-1}{k-1}}{\binom{N}{n}}$$

建立了抽样分布后，在零假设成立时，可以计算 $P(R \le c_1)$ 或 $P(R \ge c_2)$ 的值进行检验. 通常的表是给出水平 α=0.025,0.05 及 m、n 时临界值满足 $P(R \le c_1) \le \alpha$ 或 $P(R \ge c_2) \le \alpha$.

而当样本很大时，在零假设下，

$$Z=\frac{R-\mu_R}{\sigma_R}=\frac{R-\left(\frac{2mn}{m+n}+1\right)}{\sqrt{\frac{2mn(2mn-m-n)}{(m+n)^2(m+n-1)}}} \longrightarrow N(0,1).$$

于是可以借助于正态分布表来得到 p 值和检验结果. 这时，在给定水平 α 后，可以用此近似分布来得到临界值 c_1 和 c_2.

关于随机性的游程检验的过程总结于下表之中:

零假设 H_0	备选假设 H_1	检验统计量(K)	p值
H_0 有随机性	H_1 无随机性(有聚类倾向)	游程R	$P(\|K\| \le k)$
m 和n 较大时, 用近似正态统计量$Z = (R - \frac{2mn}{m+n} - 1)/\sqrt{\frac{2mn(2mn-m-n)}{(m+n)^2(m+n-1)}}$			
对水平α, 如果p值$< \alpha$, 拒绝H_0, 否则不能拒绝			

三、 案例分析（请详细写出您的检验步骤并附上程序源代码）

案例一

（数据 2.6.6.txt）下面是某村 1975-2004 年，每年收入 5000 元以上的户数：

33 32 46 36 40 40 40 36 41 39 43 35 45 39 42

43 47 51 45 45 46 59 47 51 55 42 51 49 69 57

请用 Cox-Stuart 检验来看该村的高于 5000 元的人群是否有增长趋势.

案例二

（数据 2.6.7.txt）一个监听装置收到如下信号：

0 1 0 1 1 1 0 0 1 1 0 0 0 0 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

能否说该信号是纯粹随机干扰？.

四、 实验结果

（1）

代码：

```
data=c(33,32,46,36,40,40,40,36,41,39,43,35,45,39,42,43,47,51,45,45,46,59,47,51,
       55,42,51,49,69,57)
D=data[1:15]-data[16:30]
sum(sign(D)==1)
sum(sign(D)==-1)
pbinom(1,15,0.5)
```

```
> sum(sign(D)==1)
[1] 1
> sum(sign(D)==-1)
[1] 14
> pbinom(1,15,0.5)
[1] 0.0004882812
```

由 p 值远小于 0.05 可得，拒绝原假设：无增长趋势，不拒绝备择假设：有增长趋势。

（2）

代码：

```
run.test=function(y,cut=0){if(cut!=0)x=(y>cut)*1 else x=y
N=length(x);k=1;for(i in 1:(N-1))if (x[i]!=x[i+1])k=k+1;r=k;
m=sum(1-x);n=N-m;
p1=function(m,n,k){
   2*choose(m-1,k-1)/choose(m+n,n)*choose(n-1,k-1)}
p2=function(m,n,k){choose(m-1,k-1)*choose(n-1,k)/choose(m+n,n)
   +choose(m-1,k)*choose(n-1,k-1)/choose(m+n,n)}
r2=floor(r/2);if(r2==r/2){pv=0;for(i in 1:r2)pv=pv+p1(m,n,i);
for(i in 1:(r2-1))pv=pv+p2(m,n,i)} else{pv=0
for(i in 1:r2)pv=pv+p1(m,n,i)
for(i in 1:r2)pv=pv+p2(m,n,i)};if(r2==r/2)
   pv1=1-pv+p1(m,n,r2) else pv1=1-pv+p2(m,n,r2);
z=(r-2*m*n/N-1)/sqrt(2*m*n*(2*m*n-m-n)/(m+n)^2/(m+n-1));
ap1=pnorm(z);ap2=1-ap1;tpv=min(pv,pv1)*2;
list(m=m,n=n,N=N,R=r,Exact.pvalue1=pv,
        Exact.pvalue2=pv1,Aprox.pvalue1=ap1,Aprox.pvalue2=ap2,
        Exact.2sided.pvalue=tpv,Approx.2sided.pvalue=min(ap1,ap2)*2)}
data=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)
run.test(data)
```

```
> run.test(data)
$m
[1] 42

$n
[1] 34

$N
[1] 76

$R
[1] 37

$Exact.pvalue1
[1] 0.3103491

$Exact.pvalue2
[1] 0.7415319

$Aprox.pvalue1
[1] 0.3561322

$Aprox.pvalue2
[1] 0.6438678

$Exact.2sided.pvalue
[1] 0.6206983

$Approx.2sided.pvalue
[1] 0.7122645
```

由检验结果可以知道，p 值都大于 0.05，因此不拒绝原假设，认为监听设置有随机性。

五、　结果分析

（1）由 p 值远小于 0.05 可得，拒绝原假设：收入高于 5000 的人群无增长趋势，不拒绝备择假设：有增长趋势。

（2）由检验结果可以知道，p 值都大于 0.05，因此不拒绝原假设，认为监听设置有随机性。