

武汉科技大学

# 非参数统计实验报告

专业班级: \_\_\_\_\_

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。

## 一、 实验目的

- 1、熟悉 Spearman 秩相关检验、Kendall  $\tau$  相关检验的原理和检验步骤；
- 2、了解并掌握 Theil 回归的原理和方法；
- 3、熟练应用 Spearman 秩相关检验、Kendall  $\tau$  相关检验对相关问题进行检验；
- 4、能够应用 Theil 回归对相关问题进行回归拟合，并能够解释清楚拟合结果的意义；
- 5、能够结合具体实例编程进行 Spearman 秩相关检验、Kendall  $\tau$  相关检验及 Theil 回归拟合；
- 6、熟悉符号检验和 Wilcoxon 符号秩检验的原理、应用条件及检验步骤；
- 7、能够掌握符号检验和 Wilcoxon 符号秩检验的区别与联系；
- 8、熟练应用符号检验和 Wilcoxon 符号秩检验方法对相关问题进行检验；
- 9、使用 R 软件进行符号检验和 Wilcoxon 符号秩检验。

## 二、 检验原理和步骤：

### 1、Spearman 秩相关检验：

在给定一系列数对  $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$  后，为检验它们所代表的二元变量  $X$  和  $Y$  是否相关，我们引入 Spearman 秩相关检验。检验的问题为：

$$H_0: X \text{ 与 } Y \text{ 不相关} \Leftrightarrow H_1: X \text{ 与 } Y \text{ 相关 (或正相关或负相关)}.$$

设  $R_i$  是  $X_i$  在  $(X_1, X_2, \dots, X_n)$  中的秩， $S_i$  是  $Y_i$  在  $(Y_1, Y_2, \dots, Y_n)$  中的秩。然后对每对观测值的秩进行比较，记  $d_i^2 = (R_i - S_i)^2$ ，这可以看成是某种距离的度量。显然，如果这些  $d_i^2$  很大，说明两个变量可能是负相关，而如果它们很小

则可能是正相关。在这种考虑之下，和 Pearson 相关系数  $r$  类似，记  $\bar{R} = \frac{1}{n} \sum_{i=1}^n R_i$

及  $\bar{S} = \frac{1}{n} \sum_{i=1}^n S_i$ ，Spearman 检验统计量为

$$r_s = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2 \sum_{i=1}^n (S_i - \bar{S})^2}} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}.$$

和经典的相关系数一样， $r_s$  满足  $-1 \leq r_s \leq 1$ . 对于  $n \leq 100$ ,  $r_s$  在零假设下的分布有表可查.

在大样本情况下，可以采用正态近似进行检验：

$$Z = r_s \sqrt{n-1} \rightarrow N(0,1) \text{ 当 } n \rightarrow \infty.$$

注意：Spearman 秩相关关系度量的是变量之间的依存关系，而不是 Pearson 意义下的线性相关。

## 2、Kendall $\tau$ 相关检验

Kendall (1938) 提出一种类似于 Spearman 秩相关的检验方法，从两个变量  $(X, Y)$  是否协同来检验变量间的相关性.

若  $(X_j - X_i)(Y_j - Y_i) > 0, j \neq i$ ，则称数对  $(X_i, Y_i)$  和  $(X_j, Y_j)$  协同，

若  $(X_j - X_i)(Y_j - Y_i) < 0, j \neq i$ ，则称数对  $(X_i, Y_i)$  和  $(X_j, Y_j)$  不协同. 检验问题为：

$$H_0: X \text{ 与 } Y \text{ 不相关} \Leftrightarrow H_1: X \text{ 与 } Y \text{ 相关 (或正相关或负相关)}.$$

定义 Kendall  $\tau$  统计量为：

$$\tau_a = \frac{n_c - n_d}{n_c + n_d} = \frac{n_c - n_d}{C_n^2} = \frac{2K}{n(n-1)},$$

其中的  $n_c$  表示协同的数对的数目， $n_d$  表示不协同的数对的数目， $K = n_c - n_d$ .

显然， $-1 \leq \tau_a \leq 1$ .  $|\tau_a|$  较大时拒绝原假设. 具体检验时可查零分布表，大样

本时可用采用正态近似： $K \sqrt{\frac{18}{n(n-1)(2n+5)}} \rightarrow N(0,1) \text{ 当 } n \rightarrow \infty.$

## 3、Theil 回归

在给定一系列数据  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  时，如果认为它满足线性模型

$$y = \alpha + \beta x + \varepsilon,$$

则可用 Theil 方法拟合回归直线. 其思想类似于最小二乘法, 从残差  $e_i = y_i - (\alpha + \beta x_i)$  出发, 它要寻求斜率  $\beta$  使得所有观测对  $(x_j, y_j)$  与  $(x_i, y_i)$  拟合回归直线后的残差之差的正负符号的个数相等.

记  $b_{ij} = \frac{Y_j - Y_i}{X_j - X_i}$ ,  $1 \leq i < j \leq n$ , 则用  $b_{ij}$  中位数作为斜率  $\beta$  的估计, 从而有

$$\hat{\beta} = \text{median}\{b_{ij}, 1 \leq i < j \leq n\}, \quad \hat{\alpha} = \text{median}_{1 \leq i \leq n}\{Y_i - \hat{\beta} X_i\}.$$

#### 4、广义符号检验:

广义符号检验是非参数统计中最古老的检验法之一, 最早可追溯到 Arbuthnott 于 1701 年一项有关伦敦出生的男婴性别比是否超过 1/2 的研究. 它之所以被称为符号检验, 是因为它处理的数据只有两类观测值, 如果用符号 “+” 和 “-” 做区分, 符号检验就是通过符号 “+” 和 “-” 的个数来做统计推断, 故称为符号检验.

检验原理步骤

(1) 设置零假设和备选假设

$$H_0: Q_\pi = q_0 \leftrightarrow H_1: Q_\pi > q_0, H_1: Q_\pi < q_0 \text{ 或者 } H_1: Q_\pi \neq q_0.$$

(2) 构造检验统计量  $K(X_1, X_2, \dots, X_n)$

记样本中小于  $q_0$  的点数为  $S^-$ , 而大于  $q_0$  的点数为  $S^+$ , 则在零假设下,  $S^-$  应服从二项分布  $B(n, \pi)$ . 统计量的具体选择如下表:

对 $H_0: Q_\pi = q_0$ 的检验 下面变量 $K$ 的分布为 $B(n, \pi)$ , $\hat{Q}_\pi$ 为样本 $\pi$ 分位点		
备选假设	$p$ 值	使检验有意义的条件*
$H_1: Q_\pi > q_0$	$P_{H_0}(K \leq s^-)$	$\hat{Q}_\pi > q_0$
$H_1: Q_\pi < q_0$	$1 - P_{H_0}(K \leq s^- - 1)$	$\hat{Q}_\pi < q_0$
$H_1: Q_\pi \neq q_0$	$2 \min\{P_{H_0}(K \leq s^-), 1 - P_{H_0}(K \leq s^- - 1)\}$	

\*如果条件不满足, 不用计算也知道检验结果不会显著

前表的特例: 对  $H_0: M(= Q_{0.5}) = M_0$  的检验

变量  $K = \min(S^+, S^-)$  的分布为  $B(n, 0.5)$

备选假设	$p$ 值 (这里 $k = \min(s^+, s^-)$ )
$H_1: M > M_0$ 或 $H_1: M < M_0$	$P(K \leq k)$
$H_1: M \neq M_0$	$2P(K \leq k)$

(3) 数值实现  $k = K(X_1, X_2, \dots, X_n)$

(4) 计算零假设下, 随机变量  $T$  落入某区间的精确概率或近似概率.(小概率事件或  $p$  值)

(5) 根据  $p$  值下结论。

### 5、Wilcoxon 符号秩检验:

Wilcoxon 符号秩检验是对符号检验的一种改进, 符号检验只利用了样本差异方向上的信息, 并未考虑到差别的大小, Wilcoxon 符号秩检验弥补了符号检验的这一不足。其检验步骤如下:

(1) 设置零假设和备选假设

$$H_0: M=M_0 \leftrightarrow H_1: M \neq M_0,$$

$$H_0: M \leq M_0 \leftrightarrow H_1: M > M_0$$

$$\text{或者 } H_0: M \geq M_0 \leftrightarrow H_1: M < M_0.$$

(2) 对  $i=1, 2, \dots, n$ , 计算  $|X_i - M_0|$ , 它们代表这些样本点到  $M_0$  的距离.

(3) 找出  $|X_i - M_0|$  的秩, 如果有结时, 每个点取平均秩.

(4) 令  $W^+$  等于  $X_i - M_0 > 0$  对应的  $|X_i - M_0|$  的秩和, 而  $W^-$  等于  $X_i - M_0 < 0$  对应的  $|X_i - M_0|$  的秩和.

(5) 对双边检验  $H_0: M=M_0 \leftrightarrow H_1: M \neq M_0$ , 取  $W = \min(W^+, W^-)$ , 当  $W$  很小时拒绝零假设; 对单边检验  $H_0: M \leq M_0 \leftrightarrow H_1: M > M_0$ , 取  $W=W^-$ ; 对单边检验  $H_0: M \geq M_0 \leftrightarrow H_1: M < M_0$ , 取  $W=W^+$ .

(6) 根据  $W$  的值查 Wilcoxon 符号秩检验分布表, 对  $n$  很大的时候, 可以采用正态近似。

## 案例一：（50 分）

（1、请详细写出您的检验步骤并附上程序源代码；

2、给出实验结果并对实验结果进行分析给出你的结论）

在一项身高与体重的关系研究中有 20 个人的身高（单位：厘米）和体重（单位：公斤）的数据为：

身高	136	160	135	131	156	181	145	190	174	162
体重	14	35	40	22	60	90	44	72	80	25
身高	172	177	180	166	175	184	186	170	155	161
体重	70	78	60	64	91	85	73	50	48	64

（1）Spearman 和 Kendall 检验统计量来检验身高和体重之间是否相关？是正相关还是负相关？请提出你的检验问题，并给出检验步骤过程和检验结果且分析结果。

（2）作出散点图和回归直线图（一般线性回归和 Theil 回归两种方法），写出两种方法下的回归直线，并进行检验且解释你的结果。

答：

根据题目作出假设： $H_0$ :身高和体重不相关 vs  $H_1$ :身高与体重呈正相关趋势。

由此进行检验：

#输入数据

```
x=c(136,160,135,131,156,181,145,190,174,162,172,177,180,166,175,184,186,170,155,161)
```

```
y=c(14,35,40,22,60,90,44,72,80,25,70,78,60,64,91,85,73,50,48,64)
```

#spearman 秩相关检验

```
cor.test(x,y,method="spearman",alt="greater");
```

#kendall tau a 检验

#在没有打结的情况下  $\tau_a = \tau_b$ ，因此本题检验也可以使用公式

```
cor.test(x,y,method="kendall",alt="greater");
```

#绘制散点图

```
plot(x,y,main="身高体重散点图",xlab="身高",ylab="体重");
```

#绘制一般线性回归直线图

```
lm_model<-lm(y~x);
```

# 绘制回归直线

```
abline(lm_model,col="blue");
```

# 获取回归方程系数

```
summary(lm_model); #查看详细回归结果
# 提取回归直线方程
jieju<-coef(lm_model)[1];
xielv<-coef(lm_model)[2];
cat("一般线性回归直线方程: y =",jieju,"+",xielv,"*x\n");
#theil 回归
library(mblm)
mblm_model=mblm(y ~ x,repeated=F);
abline(mblm_model,col="red");
summary(mblm_model); #查看详细回归结果
# 提取回归直线方程
jieju<-coef(mblm_model)[1];
xielv<-coef(mblm_model)[2];
cat("theil 回归直线方程: y =",jieju,"+",xielv,"*x\n");
```

检验结果:

```
> #spearman秩相关检验
> cor.test(x,y, meth="spearman", alt="greater");

Spearman's rank correlation rho

data:  x and y
S = 255.19, p-value = 8.148e-06
alternative hypothesis: true rho is greater than 0
sample estimates:
      rho
0.8081266

> #kendall tau a 检验
> #在没有打结的情况下tau a=tau b, 因此本题检验也可以使用公式
> cor.test(x,y, meth="kendall", alt="greater");

Kendall's rank correlation tau

data:  x and y
z = 3.7025, p-value = 0.0001067
alternative hypothesis: true tau is greater than 0
sample estimates:
      tau
0.603183
```



```
> summary(lm_model); #查看详细回归结果

Call:
lm(formula = y ~ x)

Residuals:
    Min       1Q   Median       3Q      Max
-30.308 -12.974   4.337  10.056  22.033

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -114.9041    30.0797  -3.820   0.00125 **
x              1.0507     0.1816   5.787 1.75e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.87 on 18 degrees of freedom
Multiple R-squared:  0.6504,    Adjusted R-squared:  0.631
F-statistic: 33.49 on 1 and 18 DF,  p-value: 1.747e-05

> cat("一般线性回归直线方程: y =",jieju,"+",xielv,"*x\n");
一般线性回归直线方程: y = -114.9041 + 1.050692 *x

> summary(mblm_model); #查看详细回归结果

Call:
mblm(formula = y ~ x, repeated = F)

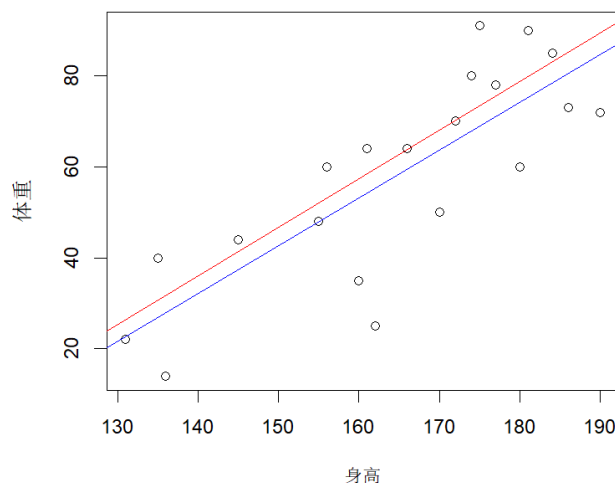
Residuals:
    Min       1Q   Median       3Q      Max
-34.522 -17.512   0.000   5.882  17.590

Coefficients:
            Estimate      MAD V value Pr(>|V|)
(Intercept) -113.5405    12.5875      0 1.91e-06 ***
x              1.0683     0.9908  14064 4.10e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.58 on 18 degrees of freedom

> cat("theil回归直线方程: y =",jieju,"+",xielv,"*x\n");
theil回归直线方程: y = -113.5405 + 1.068287 *x
```

身高体重散点图



根据结果可得：

(1) spearman 检验结果  $p$  值=8.148e-6 远小于显著性水平  $\alpha=0.05$ ,  $r_s=0.8081266$  接近 1, 由此可以拒绝原假设, 认为身高与体重呈正相关。Kendall tau a 检验结果  $p$  值=0.0001067 同样远小于显著性水平  $\alpha=0.05$ ,  $\tau_a=0.603183$  接近 1, 由此可以拒绝原假设, 认为身高与体重呈正相关。

(2) 一般线性回归是一种基于最小二乘法的回归分析方法, 对异常值非常敏感, 一个或少数几个异常值可能显著影响回归线的位置和斜率, 从而影响模型的预测准确性; theil 回归是一种基于中位数的回归分析方法, 对异常值具有较高的抵抗力。由于它是基于中位数的估计方法, 因此即使存在异常值, 也不会对回归线的估计产生显著影响。根据数据作图如上, 同时计算可得一般线性回归直线方程:  $y=-114.9041+1.050692*x$ , theil 回归直线方程:  $y=-113.5405+1.068287*x$ 。

### 案例二：（50 分）

(1、请详细写出您的检验步骤并附上程序源代码；

2、给出实验结果并对实验结果进行分析给出你的结论)

超市经理想了解每位顾客在该超市购买的商品平均件数是否为 10 件, 随机观察 12 位顾客, 得到如下数据:

顾客	1	2	3	4	5	6	7	8	9	10	11	12
件数	21	10	3	4	2	17	16	25	48	9	30	8

(1) 采用符号检验进行决策。

(2) 采用 Wilcoxon 符号秩检验进行决策, 比较它和符号检验的结果。

(3) 请分析上面两个检验, 哪个结果更可靠? 为什么?

答: 根据题目进行假设:  $H_0: \mu=10$  vs  $H_1: \mu<10$

由此进行检验:

#输入数据

data=c(21,10,3,4,2,17,16,25,48,9,30,8)

#广义符号检验

```
sign.text=function(x,p,q0){
```

```
  s1=sum(x<q0);s2=sum(x>q0);n=s1+s2
```

```
  p1=pbinom(s1,n,p);p2=1-pbinom(s1-1,n,p)
```

```
  if(p1>=p2)m1="one tail test:H1:Q>q0"
```

```
  else m1="one tail test:H1:Q<q0"
```

```
  p.value=min(p1,p2);m2="two tails test";p.value2=2*p.value
```

```
  if(q0==median(x)){p.value=0.5;p.value2=1}
```

```
list(Sign.test1=m1,p.values.of.one.tail.test=p.value,
      p.value.of.two.tail.test=p.value2))
sign.text(data,0.5,10);
#wilcoxon 符号秩检验
wilcox.test(data-10,alt="greater")
```

检验结果:

```
> sign.text(data,0.5,10);
$Sign.test1
[1] "one tail test:H1:Q<q0"

$P.values.of.one.tail.test
[1] 0.5

$P.value.of.two.tail.test
[1] 1

> wilcox.test(data-10,alt="greater")

      Wilcoxon signed rank test with continuity correction

data:  data - 10
V = 47, p-value = 0.1148
alternative hypothesis: true location is greater than 0
```

根据检验

(1) 广义符号检验可以得到  $p$  值=0.5 远大于显著性水平  $\alpha=0.05$ ，由此选择接受原假设。

(2) wilcoxon 检验结果  $p$  值=0.1148 远大于显著性水平  $\alpha=0.05$ ，由此选择接受原假设。两个假设结果相同。

(3) 可靠性分析:

符号检验主要关注样本数据的符号（即数据是正还是负），并不对数据的分布或形状做出具体假定，它常用来检验两平均值的一致性，且对样本是否来自正态总体没有严格规定。

Wilcoxon 符号秩检验除了关注样本数据的符号外，还考虑了数据的大小（即差的绝对值），它要求样本数据来自连续对称的总体分布，在这个假定下，总体中位数等于均值，该检验适用于比较两个相关样本的中位数是否存在差异。

根据两种检验的性质，Wilcoxon 符号秩检验要求数据呈对称分布，而本题数据并不对称，由此可得，本题结果符号检验更为可靠。