第5章 多总体的统计检验

CRD

CBD

BIB设计

- 比较数据，n=19类产品，用p=4种不同的广告方法进行一段时间后看销售是否受到促销方式不同的影响？
- 问题：四种方法是否不同？

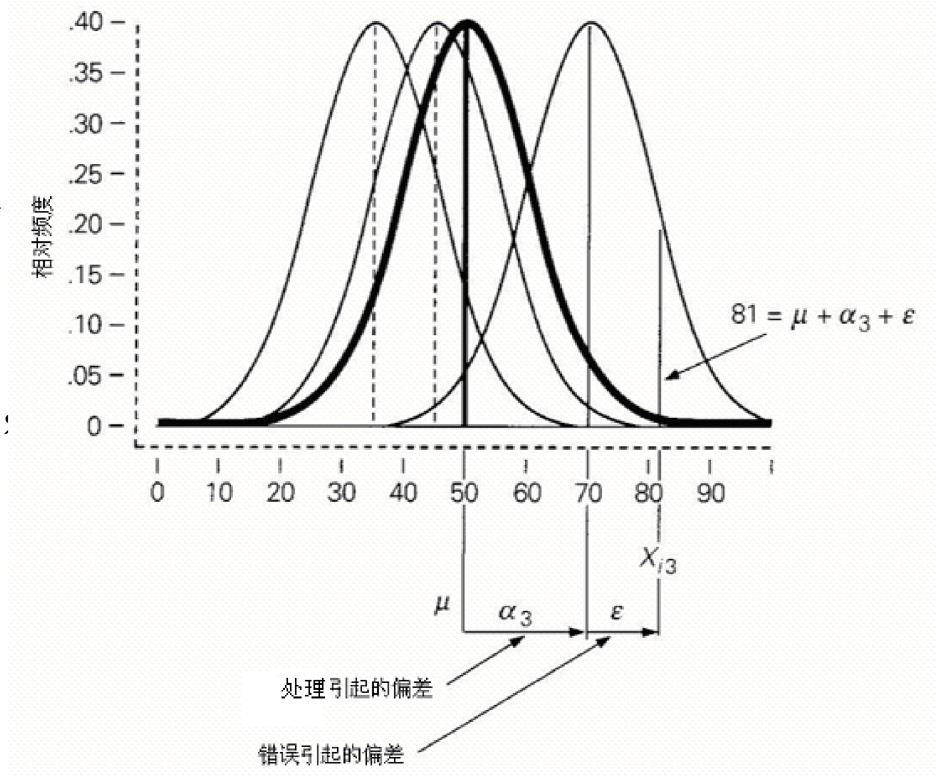
促销方法			
A	**B**	**C**	**D**
133.8	**151.2**	**193.4**	**225.8**
125.3	**149.0**	**185.3**	**224.6**
143.1	**162.7**	**182.8**	**220.4**
128.9	**143.8**	**188.5**	**212.3**
135.7	**153.5**	**198.6**	

方差分析

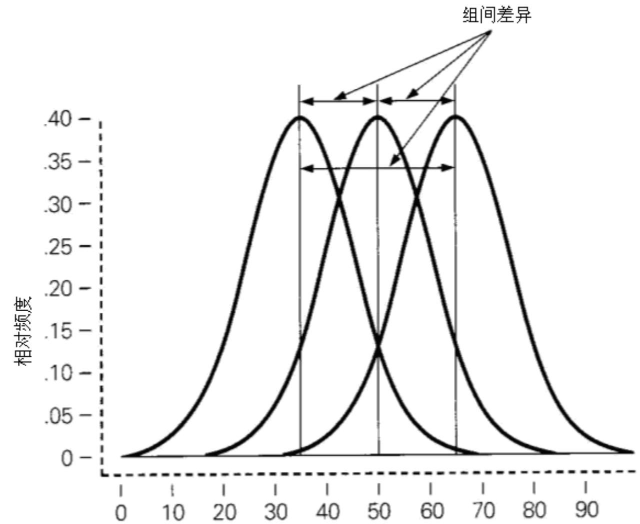
- 方差分析（analysis of variance，ANOVA）是分析定性自变量对因变量影响的一种方法。
- 自变量是定性变量，也称为因子。
- 分析结果是由一个方差分析表表示的。
- 原理为：因变量的值随着自变量的不同取值而变化。将这些变化按照自变量进行分解，使得每个自变量都包含一份贡献，不能分解的部分是随机误差的贡献。
- 将各自变量的贡献和随机误差的贡献进行比较（F检验），判断该自变量的不同水平是否对因变量的变化有显著贡献。输出就是F-值和检验的一些p-值。

几个术语

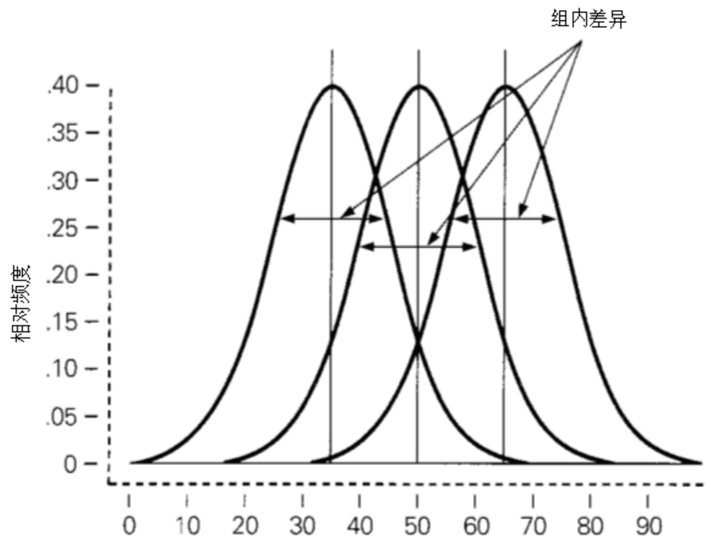
- **因子**　　　在分析中处于自变量的位置.
- **水平**　　　在一个自变量中的不同条件或数值.
- **总方差**　不考虑实验分组,所有数据的方差。



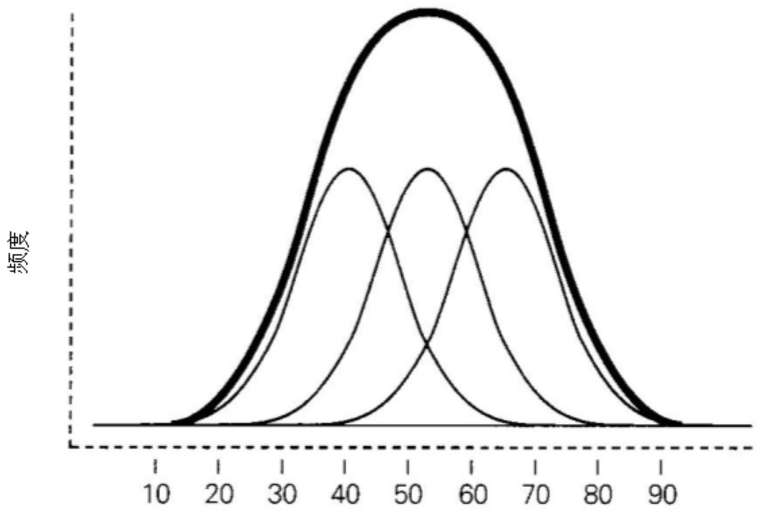
- **组间方差（Between-Groups Variance）**



- **组内方差（Within-Groups Variance）**



- 方差分析的作用是帮助我们计算方差是因误差产生的还是因处理产生的方差的比例



单因素方差分析的数据结构

水平 样本 序号	A_1	A_2	$\cdots$	A_k
1	Y_{11}	Y_{21}	$\cdots$	Y_{k1}
2	Y_{12}	Y_{22}	$\cdots$	Y_{k2}
3	Y_{13}	Y_{23}	$\cdots$	Y_{k3}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
n_i	Y_{1n_1}	Y_{2n_2}	$\cdots$	Y_{kn_k}
样本均值 $\overline{Y}_{i.}$	$\overline{Y}_{1.}$	$\overline{Y}_{2.}$	$\cdots$	$\overline{Y}_{k.}$

线性模型：

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i$$

假设：

$$Y_{i1}, Y_{i2}, \cdots, Y_{in_i} \sim N(\mu_i, \sigma^2), i = 1, \ldots, k,$$

检验：

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$$

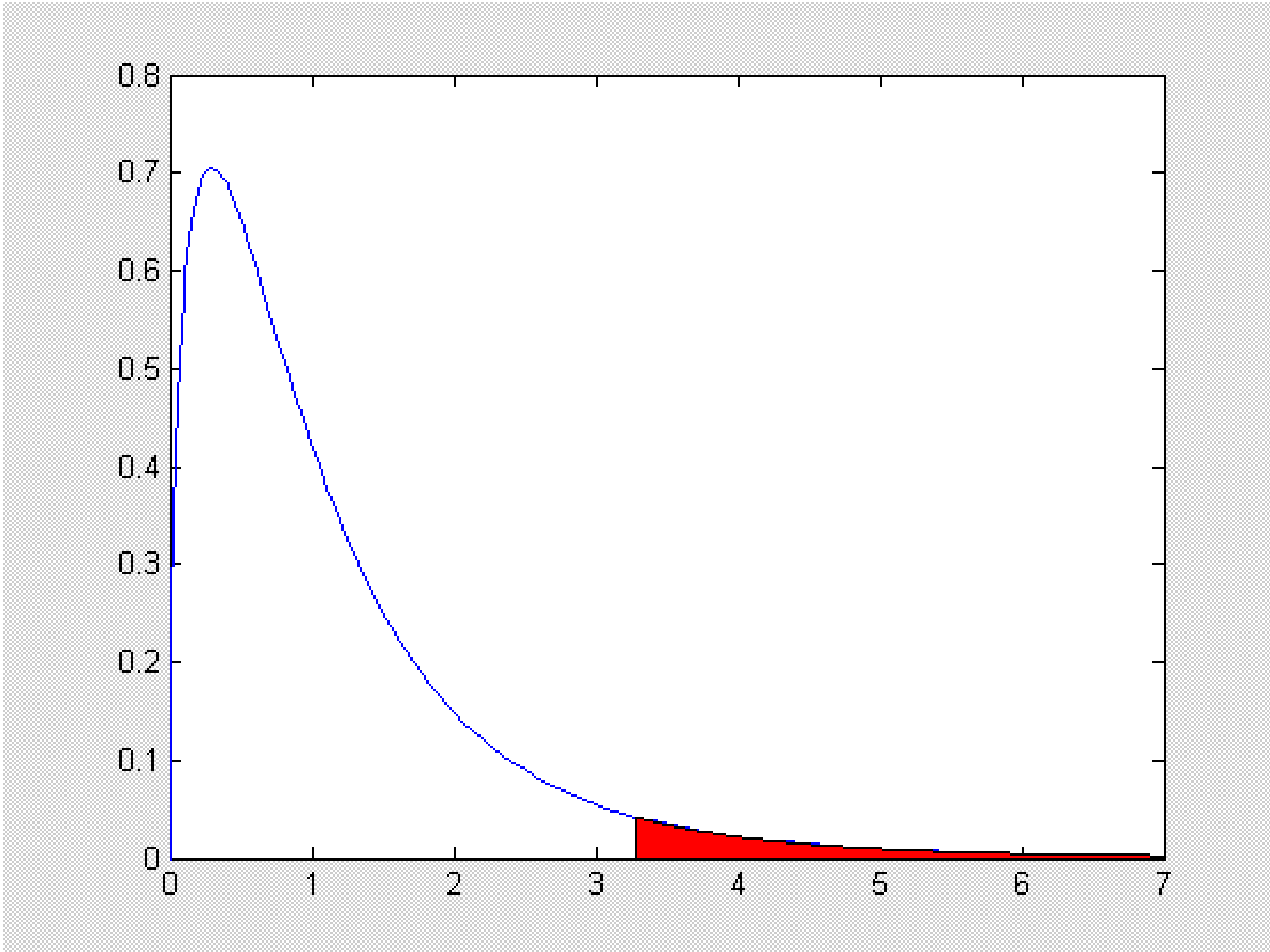
公式：总平方和=组间平方和+组内平方和

$$SST = SSB + SSE = \sum_{i=1}^{k} n_i (\bar{Y}_{i.} - \bar{Y}_{..})^2 + \sum_{i=1}^{k}\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2$$

其中，*SST* 有自由度 *n-1*，*SSB*有自由度 *k-1*，*SSE* 有自由度 *n-k*，在正态分布的假设下，如果各组增重均值相等(零假设)，则

$$F = \frac{MSB}{MSE} = \frac{SSB/(k-1)}{SSE/(n-k)}$$

有自由度为 *k-1* 和*n-k* 的*F* 分布。


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
7

方差分析表：

(比较一元总体的) ANOVA

	Sum of Squares(平方和)	Df 自由度	Mean Square(均方)	F	Sig.
Between Groups(处理)	SSB	k-1	MSB=SSB/(k-1)	F= MSB/MSE	$p = P(F(k-1, n-k) > F)$
Within Groups (误差)	SSE	n-k	MSE=SSE/(n-k)		
Total(总和)	SST	n-1			

这里n 为观测值数目，k 为水平数，F_α是自由度为$k-1$和$n-k$的F 分布的上α分位数

$$p = P(F(k-1, n-k) > F)$$

这个参数方法依赖着方差相等的正态总体假设，如果正态假设不成立，需要使用非参数的 Kruskal－Wallis 检验方法。

将 k 组样本合在一起，计算每一个数据在合样本中的秩。记 x_{ij} 在合样本 $\{x_{i1}, x_{i2}, \cdots, x_{in_i}, i=1,2,\cdots,k\}$ 中的秩为 R_{ij}，$R_{ij}=1,2,\cdots,N$。**Kruskal－Wallis** 检验的基本思想是用 x_{ij} 的秩 R_{ij} 代替 x_{ij}，然后用 ANOVA 方法作统计分析。

Kruskal-Wallis单因素方差分析

完全随机设计数据形态

序号 \ 样本 \ 水平	A_1	A_2	$\cdots$	A_k
1	Y_{11}	Y_{21}	$\cdots$	Y_{k1}
2	Y_{12}	Y_{22}	$\cdots$	Y_{k2}
3	Y_{13}	Y_{23}	$\cdots$	Y_{k3}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
n_i	Y_{1n_1}	Y_{2n_2}	$\cdots$	Y_{kn_k}
样本均值 $\overline{Y}_{i.}$	$\overline{Y}_{1.}$	$\overline{Y}_{2.}$	$\cdots$	$\overline{Y}_{k.}$

Kruskal-Wallis单因素方差分析

基本原理：类似处理两个样本相关性位置检验的W-M-W方法类似，将多个样本混合起来求秩，如果遇到打结的情况，采用平均秩，然后再按样本组求秩和。

完全随机设计数据的秩

序号 \ 样本 \ 水平	A_1	A_2	$\cdots$	A_k
1	R_{11}	R_{21}	$\cdots$	R_{k1}
2	R_{12}	R_{22}	$\cdots$	R_{k2}
3	R_{13}	R_{23}	$\cdots$	R_{k3}
$\vdots$	$\vdots$	$\vdots$		$\vdots$
n_i	R_{1n_1}	R_{2n_2}	$\cdots$	R_{kn_k}
样本均值 $\bar{R}_{i.}$	$\bar{R}_{1.}$	$\bar{R}_{2.}$	$\cdots$	$\bar{R}_{k.}$

检验方法

$R_{i+}=\sum_{j=1}^{n_i}R_{ij}$ ， $\bar{R}_i=\sum_{j=1}^{n_i}R_{ij}\Big/n_i$ 是第 i 个总体的样本的秩的平均，

$\bar{R}=\sum_{i=1}^{k}\sum_{j=1}^{n_i}R_{ij}\Big/N=\left(1+2+\cdots+N\right)\Big/N=\left(N+1\right)\Big/2$,

$$SST=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(R_{ij}-\bar{R}\right)^2=\sum_{i=1}^{k}\sum_{j=1}^{n_i}R_{ij}^2-N\bar{R}^2$$

$$=1^2+2^2+\cdots+N^2-N\left((N+1)/2\right)^2$$

$$=N(N+1)(2N+1)/6-N(N+1)^2/4$$

$=N\left(N^2-1\right)\Big/12$ 为总平方和

$SSB=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(\bar{R}_i-\bar{R}\right)^2=\sum_{i=1}^{k}n_i\left(\bar{R}_i-\bar{R}\right)^2=\sum_{i=1}^{k}n_i\left(\bar{R}_i-(N+1)/2\right)^2$ 为组间平方和， $SSW=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(R_{ij}-\bar{R}_i\right)^2$ 为组内平方和。

若没有结，

$$SST=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(R_{ij}-\bar{R}\right)^2=N\left(N^2-1\right)\Big/12$$

$$df(SST)=N-1, MST=SST/df(SST)=N\left(N+1\right)\Big/12$$

定义Kruskal-Wallis的H统计量：

$$H=SSB/MST=\frac{12}{N(N+1)}SSB=\frac{12}{N(N+1)}\sum_{i=1}^{k}n_i(\bar{R}_i-\frac{N+1}{2})^2$$

$$df(SSB)=k-1, E(SSB\mid H_0)=(k-1)N\left(N+1\right)\Big/12$$

$$df(H)=df(SSB), E(H\mid H_0)=(k-1)$$

在零假设情况下，H近似服从$\chi^2(k-1)$, 当$H>\chi^2_\alpha(k-1)$的时候拒绝零假设。

下面计算在原假设 H_0 为真，即 k 个总体 $X_1, X_2, \cdots, X_k$ 为同一连续型分布时，Kruskal－Wallis 检验统计量 H 的期望。

$E(R_{i+}) = n_i(N+1)/2$， $D(R_{i+}) = n_i(N-n_i)(N+1)/12$，（Wilcoxon 秩和统计量 $E(W_Y) = m(N+1)/2, D(W_Y) = mn(N+1)/12$）

$E(\bar{R}_i) = (N+1)/2$， $D(\bar{R}_i) = (N-n_i)(N+1)/(12n_i)$。

$$E\left(SSB\right) = \sum_{i=1}^{k} n_i E\left(\bar{R}_i - (N+1)/2\right)^2$$

$$= \sum_{i=1}^{k} n_i \frac{(N-n_i)(N+1)}{12n_i} = \frac{N+1}{12}\sum_{i=1}^{k}(N-n_i) = \frac{N(N+1)(k-1)}{12}$$

$$E\left(H\right) = \frac{12}{N(N+1)}E\left(SSB\right) = \frac{12}{N(N+1)}\frac{N(N+1)(k-1)}{12} = k-1$$

例 5.1: 为研究 4 种不同的药物对儿童咳嗽的治疗效果，将 25 个体质相似的病人随机分为 4 组，各组人数分别为 8 人， 4 人， 7 人和 6 人，各自采用 A, B, C, D 四种药进行治疗，假定其他条件均保持相同， 5 天后测量每个病人每天的咳嗽次数如下表：（单位：次数），试比较这 4 种药物的治疗效果是否相同？

表 5.3. 4 种药物治疗效果比较表

	A	秩	B	秩	C	秩	D	秩
重复	80	1	133	3	156	4	194	7
	203	8	180	6	295	15	214	9
	236	10	100	2	320	16	272	12
	252	11	160	5	448	21	330	17
	284	14			465	23	386	19
	368	18			481	25	475	24
	457	22			279	13		
	393	20						
处理内秩和 $R_{,j}$		104		16		117		88
处理内平均秩 $\bar{R}_{,j}$		13		4		16.7		14.7

统计分析：由 (5.2) 式

$$\begin{aligned} H &= \frac{12}{25(25+1)}\left[\frac{104^2}{8}+\frac{16^2}{4}+\frac{117^2}{7}+\frac{88^2}{6}\right]-3(25+1) \\ &= 8.072088 \end{aligned}$$

结论： $H = 8.072088 > \chi^2_{0.05,3} = 7.814728$ ，故接受 H_1 ，显示四种药物疗效不等。

有相等的观察值时 Kruskal－Wallis 检验统计量的修正

有结时，R_{i+} 的期望和方差分别为

$$E(R_{i+})=n_i(N+1)/2$$

$$D(R_{i+})=n_i(N-n_i)(N+1)/12-n_i(N-n_i)\sum_{t=1}^{g}(\tau_t^3-\tau_t)\Big/\big(12N(N-1)\big)$$

$\bar{R}_i=R_{i+}/n_i$ 的期望和方差分别为

$$E(\bar{R}_i)=(N+1)/2$$

$$D(\bar{R}_i)=(N-n_i)(N+1)/(12n_i)-(N-n_i)\sum_{t=1}^{g}(\tau_t^3-\tau_t)\Big/\big(12n_iN(N-1)\big)$$

Kruskal－Wallis 检验统计量 H 的期望为

$$E\left(H\right)=\frac{12}{N(N+1)}\sum_{i=1}^{k}n_iE\left(\bar{R}_i-(N+1)/2\right)^2$$

$$=(k-1)\left(1-\frac{\sum_{t=1}^{g}(\tau_t^3-\tau_t)}{N^3-N}\right)$$

秩取平均后，若将 H 修正为 H_a，其中

$$H_a=\frac{H}{1-\frac{\sum_{t=1}^{g}(\tau_t^3-\tau_t)}{N^3-N}}$$

则 H_a 的期望仍等于 $k-1$。

有结秩取平均后，经过修正的 Kruskal－Wallis 检验统计量 H_a 也渐近服从

$\chi^2(k-1)$ 分布。

趋势的秩检验方法

单调上升趋势性检验问题的原假设和备择假设分别为

$$H_0: \theta_1 = \theta_2 = \cdots = \theta_k\text{，}\quad H_1: \theta_1 \le \theta_2 \le \cdots \le \theta_k\text{，且}\theta_1 < \theta_k$$

有单调上升趋势时的“理想”的情况为

$$x_{11} = x_{12} = \cdots = x_{1n_1} < x_{21} = x_{22} = \cdots = x_{2n_2} < \cdots < x_{k1} = x_{k2} = \cdots = x_{kn_k}$$

同 Kruskal－Wallis 检验，记 x_{ij} 在合样本 $\left\{x_{i1}, x_{i2}, \cdots, x_{in_i},\ i = 1, 2, \cdots, k\right\}$ 中的秩为 R_{ij}， $R_{ij} = 1, 2, \cdots, N$ 。有单调上升趋势时的“理想”的情况下，

x_{ij} 的秩 $R_{ij} = r_i$，其中

$$r_i = \sum_{t=1}^{i-1} n_t + (n_i + 1)/2\text{，}\quad i = 1, 2, \cdots, k\text{，}\quad j = 1, 2, \cdots, n_i$$

计算下列成对数据的相关系数。

$$\begin{pmatrix} R_{11} \\ r_1 \end{pmatrix}, \cdots, \begin{pmatrix} R_{1n_1} \\ r_1 \end{pmatrix}, \cdots, \begin{pmatrix} R_{k1} \\ r_k \end{pmatrix}, \cdots, \begin{pmatrix} R_{kn_k} \\ r_k \end{pmatrix}$$

$$r = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(R_{ij} - \bar{R}\right)\left(r_i - \bar{r}\right)}{\sqrt{\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(R_{ij} - \bar{R}\right)^2} \sqrt{\sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(r_i - \bar{r}\right)^2}}$$

$$= \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} R_{ij} r_i - N(N+1)^2 / 4}{\sqrt{N(N^2-1)/12} \sqrt{N(N^2-1)/12 - \sum_{i=1}^{k} (n_i^3 - n_i)/12}}$$

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} R_{ij} r_i = \frac{1}{2}\sum_{i=1}^{k}\left(2\sum_{t=1}^{i} n_t - n_i + 1\right) R_{i+}$$

$$= \frac{1}{2}\sum_{i=1}^{k}\left(2\sum_{t=1}^{i} n_t - n_i\right) R_{i+} + \frac{N(N+1)}{4}$$

取 $T = \sum_{i=1}^{k} w_i R_{i+}$， $w_i = 2\sum_{t=1}^{i} n_t - n_i$， $i = 1, 2, \cdots, k$

在原假设为真，即 k 个总体 $X_1, X_2, \cdots, X_k$ 为同一连续型分布时，T 的期望和方差分别为

$$E(T) = N^2(N+1)\big/2$$

$$D(T) = N(N+1)\left(\sum_{i=1}^{k} n_i w_i^2 - N^3\right)\Big/12$$

并且 T 服从对称分布，其对称中心为 $N^2(N+1)\big/2$。

这时 T 有渐近正态性，简记为

$$T \dot{\sim} N\left(N^2(N+1)/2,\ N(N+1)\left(\sum_{i=1}^{k} n_i w_i^2 - N^3\right)\Big/12\right)$$

有结，即有重复观察值秩取平均时，T 的期望保持不变：$E(T)=N^2(N+1)/2$，而方差需修正为

$$D(T)=\left(N(N^2-1)-\sum_{t=1}^{g}(\tau_t^3-\tau_t)\right)\left(\sum_{i=1}^{k} n_i w_i^2 - N^3\right)\Big/\left(12(N-1)\right)$$

例 一般来说，年龄越大的人的β脂蛋白的含量越大。现观察三组人，他们都是男性。第 1 组人的年龄在 20 到 30 岁之间，第 2 组人的年龄在 30 到 40 岁之间，第 3 组人的年龄在 40 到 50 岁之间。他们的β脂蛋白的测量值见下表。问这三组人的测量值是否符合人们的经验：年龄越大的人的β脂蛋白的含量越大？

表 不同年龄组的男性的β脂蛋白的含量

第 1 组	第 2 组	第 3 组	第 1 组	第 2 组	第 3 组
260	310	320	205	210	380
200	310	260	190	280	240
240	190	360	200	210	295
170	225	310	250	280	260
270	170	270	200	240	250

首先计算每一个男性β脂蛋白含量的秩。

$n_1 = n_2 = n_3 = 10$,，　$w_1 = 10$，$w_2 = 30$，$w_3 = 50$，检验统计量 $T = 16160$。

由于 $N = n_1 + n_2 + n_3 = 30$，$E\left(T \mid H_0\right) = 13950$。观察值中有 $g = 11$ 个结，其中长度为 2 的结有 7 个，长度为 3 的结有 4 个，$D\left(T\right) = 616827.586$。由渐近正态性知，检验的 p 值为

$$p = P\left(N(13950, 616827.586) \geq 16160\right) = 0.002447$$

p 值很小，所以我们认为年龄越大的人的β脂蛋白的含量越大。

Jonckheere-Terpstra检验

检验原理以及方法

假设k个独立的样本：$X_{11},\cdots,X_{1n_1};\cdots;X_{k1},\cdots,X_{kn_k}$ 分别来自于k个形状相同的分布：$F(x-\theta_1),\cdots,F(x-\theta_k)$.

假设检验问题：

$$H_0:\theta_1=\cdots=\theta_k, \qquad H_1:\theta_1\le\cdots\le\theta_k \quad \theta_1<\theta_k$$

计算步骤

1. 计算 $W_{ij}=\sum_{r=1}^{n_i}\sum_{s=1}^{n_j} I(X_{ir}<X_{js})$

2. 计算Jonckheere-Terpstra统计量： $J=\sum_{i<j} W_{ij}$

3. 当J取大值的时候，考虑拒绝零假设，J精确分布可以查零分布表，对于大样本，可以考虑正态近似。

打结的情况时，采用变形的公式：

$$W_{ij}^{*}=\#(X_{iu}<X_{jv},\quad u=1,\cdots,n_i;v=1,\cdots,n_j)+\frac{1}{2}\#(X_{iu}=X_{jv},\quad u=1,\cdots,n_i;v=1,\cdots,n_j)$$

$$J^{*}=\sum_{i<j} W_{ij}^{*}$$

在原假设成立，即所有的$\theta_1, \theta_2, \cdots, \theta_k$ 全都相等时，这k 个连续型随机变量总体$X_1, X_2, \cdots, X_k$ 同分布。与 Mann－WhitneyU 统计量服从对称分布(见性质 5.5)相类似，在原假设成立时 Jonckheere－Terpstra 检验统计量 J 也服从对称分布，对称中心为

$$\sum_{1\le i<j\le k}\left(n_i n_j/2\right)=\frac{1}{4}\left[\left(\sum_{t=1}^{k} n_t\right)^2-\sum_{t=1}^{k} n_t^2\right]=\frac{1}{4}\left[N^2-\sum_{t=1}^{k} n_t^2\right]$$

在原假设成立时 J 有渐近正态性：若$\min\left\{n_1, \cdots, n_k\right\}\to\infty$，且对所有的$i=1, \cdots, k$ 都有$n_i/N\to\lambda_i\in(0,1)$，则

$$\frac{J-E(J)}{\sqrt{D(J)}}\xrightarrow{L}N(0,\, 1)$$

原假设成立，即所有的$\theta_1, \theta_2, \cdots, \theta_k$全都相等时，

$$E(W_{ij}) = n_i n_j \big/ 2$$

所以J的期望为

$$E(J) = \sum_{1 \le i < j \le k} \left(n_i n_j \big/ 2 \right) = \frac{1}{4}\left[N^2 - \sum_{t=1}^{k} n_t^2 \right]$$

$$D(W_{ij}) = n_i n_j (n_i + n_j + 1) \big/ 12,$$

$$D(W_{ij} + W_{tj}) = (n_i + n_t) n_j (n_i + n_t + n_j + 1) \big/ 12,$$

$$D(W_{ij} + W_{tj}) = D(W_{ij}) + D(W_{tj}) + 2Cov(W_{ij}, W_{tj}),$$

$$Cov(W_{ij}, W_{tj}) = \frac{1}{2}\Big(D(W_{ij} + W_{tj}) - D(W_{ij}) - D(W_{tj}) \Big) = n_i n_j n_t / 12$$

在$i \neq t$，$i \neq s$，$j \neq t$和$j \neq s$时，

$Cov(W_{ij},\ W_{ts}) = 0$，　　在$i \neq t$，$i \neq s$，$j \neq t$和$j \neq s$时，

$$Cov(W_{ij},\ W_{it}) = Cov(W_{ij},\ W_{tj}) = n_i n_j n_t \big/ 12$$

$$Cov(W_{ij},\ W_{jt}) = Cov(W_{ij}, W_{ti}) = -n_i n_j n_t \big/ 12,$$

$$D(J) = \sum_{i<j} D\left(W_{ij}\right) + 2\sum_{i<j<s}\left(\operatorname{cov}\left(W_{ij}, W_{is}\right) + \operatorname{cov}\left(W_{ij}, W_{js}\right) + \operatorname{cov}\left(W_{is}, W_{js}\right)\right)$$

$$= \sum_{i<j} n_i n_j (n_i + n_j + 1) \big/ 12 + \sum_{i<j<s} n_i n_j n_s / 6$$

$$= \frac{1}{12}\left(\sum_{i \neq j} n_i^2 n_j + \sum_{i<j} n_i n_j\right) + \frac{1}{36}\left(N^3 - \sum_i n_i^3 - 3\sum_{i \neq j} n_i^2 n_j\right)$$

$$= \frac{1}{36}\left(N^3 - \sum_i n_i^3 + 3\left(N^2 - \sum_i n_i^2\right) / 2\right)$$

$$= \frac{1}{72}\left[N^2\left(2N + 3\right) - \sum_{t=1}^{k} n_t^2\left(2n_t + 3\right)\right]$$

J 的渐近正态性可简记为

$$J \dot{\sim} N\left(\left(N^2-\sum_{t=1}^{k} n_t^2\right)\Big/4,\ \left(N^2(2N+3)-\sum_{t=1}^{k} n_t^2(2n_t+3)\right)\Big/72\right)$$

有结时 J 的方差应修正为

$$D(J)=\frac{1}{72}\left[N^2\left(2N+3\right)-\sum_{t=1}^{k} n_t^2\left(2n_t+3\right)-\sum_{s=1}^{g}\tau_s(\tau_s-1)(2\tau_s+5)\right]$$

$$+\frac{1}{36N(N-1)(N-2)}\left[\sum_{t=1}^{k} n_i(n_i-1)(n_i-2)\right]\left[\sum_{s=1}^{g}\tau_s(\tau_s-1)(\tau_s-2)\right]$$

$$+\frac{1}{8N(N-1)}\left[\sum_{t=1}^{k} n_i(n_i-1)\right]\left[\sum_{s=1}^{g}\tau_s(\tau_s-1)\right]$$

例5.3

例 5.3: 为测试不同的医务防护服的功能，让三组体质相似的受试者分别着不同的防护服装，记录受试者每分钟心脏跳动的次数，每人试验 5 次，得到 5 次平均数列在下表，医学理论判断：这三组受试的心跳次数可能存在如下关系：第一组 $\leq$ 第二组 $\leq$ 第三组，下面用这些数据验证这一论断是否可靠？

表 5.6. 三组受试心跳次数测试数据

第一组	125	136	116	101	105	109		
第二组	122	114	131	120	119	127		
第三组	128	142	128	134	135	131	140	129

例5.3解

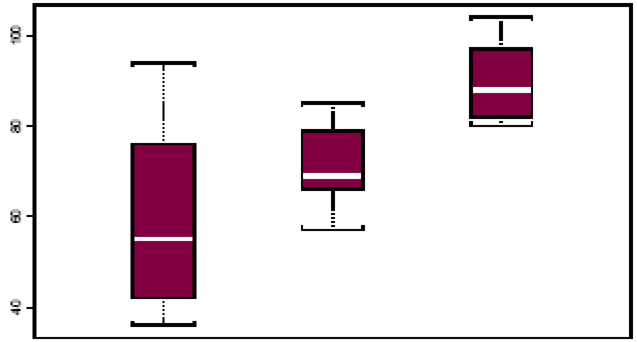


图 5.1: 医学防护服的效果比较箱线图

解: 设 $\theta_i, i=1,2,3$ 表示第 i 组的位置参数，则假设检验问题是:

$$H_0:\ \theta_1=\theta_2=\theta_3 \leftrightarrow H_1:\ \theta_1 \le \theta_2 \le \theta_3,$$

因此采用 Jonckheere-Terpstra 检验，计算 W_{ij} 如下:

$$W_{12}=25,\quad W_{13}=42\quad W_{23}=45$$

因此， $J=W_{12}+W_{13}+W_{23}=112$ ，由附表 7 查 $J \ge 90=0.051$ ，因此有理由拒绝零假设 H_0, 认为医学临床经验是可靠的。

课堂练习

1. 证明当$k=2$时，K-W 检验与 W-M-W 检验等价。

2. 证明当$k=2$时，J-T 检验与 W-M-W 检验等价。

3. $Y_{i1},Y_{i2},\cdots,Y_{in_i} \sim N(\mu_i,\sigma^2), i=1,\ldots,k,$用似然比检验法求假设问题$H_0:\mu_1=\mu_2=\cdots=\mu_k$的否定域。

4. (1) 试证明趋势性秩检验统计量T的期望$E(T)=N^2(N+1)/2$和方差$D(T)=N(N+1)\left(\sum_{i=1}^{k} n_i w_i^2 - N^3\right)/12$。

   (2) 试证明趋势性秩检验统计量T服从对称中心为$N^2(N+1)/2$的对称分布。

5. 设有 k 个连续型随机变量总体：$X_1, X_2, \cdots, X_k$。设第 i 个总体 X_i 的分布函数为 $F\left(x-\theta_i\right)$，$i=1, 2, \cdots, k$。$x_{i1}, x_{i2}, \cdots, x_{in_i}$ 是来自第 i 个总体 X_i 的样本，其容量为 n_i，$i=1, 2, \cdots, k$。总的样本容量为 $N=\sum_{i=1}^{k} n_i$。所有 N 个样本单元都是互相独立的。对有方向检验问题为

$$H_0: \theta_1=\theta_2=\cdots=\theta_k,$$

$$H_1: \theta_1 \leq \cdots \leq \theta_{t-1} \leq \theta_t \geq \theta_{t+1} \geq \cdots \geq \theta_k，\text{且} \theta_1<\theta_t \text{或} \theta_k<\theta_t \text{成立},$$

取 Jonckheere–Terpstra 型检验统计量：

$$J=\sum_{1 \leq i<j \leq t} W_{ij}+\sum_{t \leq i<j \leq k} W_{ji}。$$

试求在原假设成立，即所有的 $\theta_1, \theta_2, \cdots, \theta_k$ 全都相等时，计算 J 的期望和方差。

区组设计数据回顾

- 在有区组的数据中，总的变化可以分解到以下几个方面:
  - 处理造成的不同
  - 区组内的变化
  - 区组之间的变化
- 当有区组存在时, 代表处理的样本的独立性就不存在了.

- Friedman检验也称Friedman χ^2检验，是1937年Friedman提出的检验方法。它是检验K个总体的分布中心是否有差异。Friedman提出的检验方法是独立地在每一个区组内各自对数据进行排秩。

- 例如美国通用、福特与克莱斯勒汽车公司5种不同车型的某年产品油耗情况如表所列，数据分析关心的问题之一是三个公司汽车耗油有无差异，

- 3个汽车公司5种不同车型某年产品油耗情况

3个汽车公司5种不同车型某年产品油耗情况

公司	超小型	小型	中型	大型	运动型
通用	20.3	21.2	18.2	18.6	18.5
福特	18.6	18.5	25.6	24.7	19.3
克莱斯勒	19.3	20.7	24.0	19.8	21.4

$k=3,b=5$

Friedman秩方差分析

假设检验问题：

$$H_0: \theta_1 = \cdots = \theta_k : H_1: \exists i, j \in 1, \cdots, k, \theta_i \neq \theta_j$$

完全随机区组设计表

	样本1	样本2	…	样本k
区组1	x_{11}	x_{12}	…	x_{1k}
区组2	x_{21}	x_{22}	…	x_{2k}
…	…	…	…	…
区组b	x_{b1}	x_{b2}	…	x_{bk}

在同一区组内，计算样本的秩, 并求出第j种处理的秩和与平均秩：

$$R_{+j} = \sum_{i=1}^{b} R_{ij}, \ j = 1, \ldots, k \qquad \bar{R}_{.j} = \frac{R_{+j}}{b}$$

	样本1	样本2	…	样本k
区组1	R_{11}	R_{12}	…	R_{1k}
区组2	R_{21}	R_{22}	…	R_{2k}
…	…	…	…	…
区组b	R_{b1}	R_{b2}	…	R_{bk}
秩和	R_{+1}	R_{+2}	…	R_{+k}

倘若 k 种条件（方法）不存在差异，那么无论从哪一个区组去观察，每一种处理所得到的数据在该区组内可能地排秩为 1 至 k 中的任何一个数。因此，假如原假设为真的话，对每一处理 j，$\bar{R}_{.j}$ 应与 $\bar{R}=\dfrac{(k+1)}{2}$ 相距不远。仿照方差分析，由处理产生的“秩变异平方和”

$$SSB=\sum_{j=1}^{k}\sum_{i=1}^{b}(\bar{R}_{.j}-\frac{k+1}{2})^2=\sum_{j=1}^{k}b(\bar{R}_{.j}-\frac{k+1}{2})^2$$

当原假设为真，SSB 应该比较小。反之，若该平方和较大的话，则拒绝原假设。这个平方和究竟怎样算大怎样又算小，统计学的常规处理手法之一还是将它与另外的平方和或平均平方和来比较，Friedman 检验统计量就是将这个平方和除以秩的整体平均平方和。

$$SSB=\sum_j\sum_i\left(\overline{R}_{.j}-\overline{R}\right)^2=\sum_j b\left(\overline{R}_{.j}-\overline{R}\right)^2=\sum_j R_{+j}{}^2/b-kb\overline{R}^2,$$

$$\overline{R}=\left(k+1\right)/2$$

若原假设成立, $P(R_{ij}=r)=1/k, r=1,\cdots,k$

$$E(R_{ij})=\left(k+1\right)/2, D(R_{ij})=(k^2-1)/12$$

R_{ij}与$R_{i'j}$独立，故$E(\overline{R}_{.j})=\left(k+1\right)/2,\quad D(\overline{R}_{.j})=(k^2-1)/12b$

$$E(SSB)=k(k^2-1)/12$$

$$Q=SSB/(k(k+1)/12), E(Q)=k-1, Q\to\chi^2(k-1)$$

检验统计量

利用普通类似方差分析构造统计量：

$$Q = SSB/(k(k+1)/12) = \frac{12}{bk(k+1)}\sum R_{+j}^2 - 3b(k+1)$$

在零假设成立下$Q \sim \chi^2(k-1)$，如果 Q 偏大，那么就考虑拒绝原假设。如果存在打结的情况，则可采用修正公式计算。

$$Q_a = \frac{Q}{1 - \frac{\sum_{i=1}^{b}\sum_{t=1}^{g_i}(\tau_{i,t}^3 - \tau_{i,t})}{bk(k^2-1)}}$$

$$E(\overline{R}_{\cdot j})=E\left(\sum_{i=1}^{b}a(R_{ij})\Big/b\right)=\sum_{i=1}^{b}E\left(a(R_{ij})\right)\Big/b=\left(k+1\right)\Big/2$$

$$D(\overline{R}_{\cdot j})=D\left(\sum_{i=1}^{b}a(R_{ij})\Big/b\right)$$

$$=\left(k^{2}-1\right)\Big/\left(12b\right)-\sum_{i=1}^{b}\sum_{t=1}^{g_i}(\tau_{i,t}^{3}-\tau_{i,t})\Big/(12b^{2}k)$$

从而，秩取平均后，Friedman 检验统计量 Q 的期望为

$$E\left(Q\right)=\frac{12b}{k(k+1)}\sum_{j=1}^{k}E\left(\overline{R}_{\cdot j}-(k+1)/2\right)^{2}$$

$$=\frac{12b}{k(k+1)}\sum_{j=1}^{k}D\left(\overline{R}_{\cdot j}\right)=\frac{12b}{k(k+1)}\left(k\frac{k^{2}-1}{12b}-k\frac{\sum_{i=1}^{b}\sum_{t=1}^{g_i}(\tau_{i,t}^{3}-\tau_{i,t})}{12b^{2}k}\right)$$

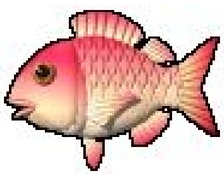
$$=(k-1)\left(1-\frac{\sum_{i=1}^{b}\sum_{t=1}^{g_i}(\tau_{i,t}^{3}-\tau_{i,t})}{bk(k^{2}-1)}\right)$$

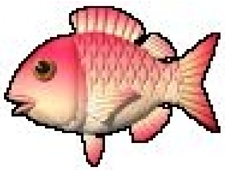
例5.5

例 5.5: 设有来自 A,B,C,D 四个地区的四名厨师制作名菜京城水煮鱼，想比较它们的品质是否相同，经四位美食评委评分结果如下表，试测验四个地区制作的水煮鱼这道菜品质有无区别。

表 5.10. 评委对四名厨师的评分数据表

美食	地	区			
评委	A	B	C	D	
1	85(4)	82(2.5)	82(2.5)	79(1)	
2	87(4)	75(1)	86(3)	82(2)	
3	90(4)	81(3)	80(2)	76(1)	
4	80(3)	75(1.5)	81(4)	75(1.5)	
秩和 $R_{.j}$	15	8	11.5	5.5	$R_{..} = 40$





上表中括号内数据为每位评委品尝四种菜后所给评分的秩。

统计分析： $b=4$（区组数）， $k=4$（处理数）

表 5.11. 结点校正计算表

相同的秩	1.5	2.5	
τ_i	2	2	
$\tau_i^3-\tau_i$	6	6	$\sum(\tau_i^3-\tau_i)=12$

$$
\begin{aligned}
Q &= \frac{12}{4\times 4\times(4+1)}[15^2+8^2+11.5^2+5,5^2]-3\times 4(4+1)\\
&= 7.7250\\
Q_c &= 8.1316
\end{aligned}
$$

结论：实际测量 $Q_c=8.1316>\chi^2_{0.05,3}=7.82$ ，接受 H_1 ，认为四个地区的菜品质上存在显著差异。

Hollander-Wolfe两处理比较检验

当用Friedman秩方差分析，检验出认为处理之间表现出差异的时候，那么可以进一步研究处理两两之间是否存在差异。Hollander-Wolfe检验公式：

$$D_{jj'} = | R_{+j} - R_{+j'} | / SE$$

其中 $SE = \sqrt{bk(k+1)/6}$，在打结的情况下可使用修正的公式。当 $|D_{jj'}| \geq Z_{1-\alpha^*}$ 时认为两个处理之间存在差异，其中 $\alpha^* = \alpha / k(k-1)$，$\alpha$ 是显著性水平。

例 5.6: 由例 5.5 知，四个地区所制的水煮鱼品质上有显著差异，成对样本比较有 $k(k-1)/2=4(4-1)/2=6$ 种，四种水煮鱼的秩和分别为

$$R_{.1}=15,\quad R_{.2}=8,\quad R_{.3}=11.5,\quad R_{.4}=5.5$$

设 $\alpha \ = \ 0.10,\ \ \alpha^*=0.10/4(4-1)=0.0167$

$$Z_{1-0.0167} \ = \ Z_{0.9833}=2.13$$

$$SE=\sqrt{\frac{4\times 4(4+1)}{6}-\frac{4\times 12}{6(4-1)}}=3.266$$

表 5.12. 两两处理的 Hollander-Wolfe 计算表

比较式	$\lvert R_{.i}-R_{.j}\rvert$	SE	D_{ij}	$Z_{0.9833}$
$A\ \ VS\ \ B$	15-8=7	3.266	2.1433*	2.13
$A\ \ VS\ \ C$	15-11.5=3.5	3.266	1.0716	2.13
$A\ \ VS\ \ D$	15-5.5=9.5	3.266	2.9088*	2.13
$B\ \ VS\ \ C$	8-11.5=3.5	3.266	1.0716	2.13
$B\ \ VS\ \ D$	8-5.5=2.5	3.266	0.7655	2.13
$C\ \ VS\ \ D$	11.5-5.5=6	3.266	1.8371	2.13

随机区组调整秩和检验Hodges-Lehmnan检验

假设检验问题:

$$H_0: \theta_1 = \cdots = \theta_k \leftrightarrow H_1: \exists i, j \in 1, \cdots, k \quad \theta_i \neq \theta_j$$

完全随机区组设计表

	样本1	样本2	…	样本k
区组1	x_{11}	x_{12}	…	x_{1k}
区组2	x_{21}	x_{22}	…	x_{2k}
…	…	…	…	…
区组b	x_{b1}	x_{b2}	…	x_{bk}

计算步骤

1. 计算每一区组的位置估计，中位数或平均值等，如:

$$\bar{X}_{i.} = \frac{1}{k}\sum_{j=1}^{k} X_{ij}$$

2. 计算$AX_{ij} = X_{ij} - \bar{X}_{i.}$，被称为调整观察值。

3. 将全部调整观测值混合求秩，设AX_{ij}对应的混合秩为 R_{ij} ，者称为调整秩。

$$\bar{R}_{.j} = \frac{1}{b}\sum_{i=1}^{b} R_{ij}$$

$$SSB = \sum_{j}\sum_{i}\left(\bar{R}_{.j} - \bar{R}\right)^2 = \sum_{j} b\left(\bar{R}_{.j} - \frac{kb+1}{2}\right)^2$$

若原假设 H_0 成立，即同一个区组内的 k 个观察值独立同分布时，SSB 在区组 i 里 k 个秩的值分别为 $R_{i1}, R_{i2}, \cdots, R_{ik}$， $i=1, \cdots, b$ 都给定的条件下的条件期望为 $E(SSB)=\sum_{i=1}^{b}\sum_{j=1}^{k}\left(R_{ij}-R_{i+}/k\right)^2\Big/b$

$(k-1)\big/E\left(SSB\right)=(k-1)b\Big/\sum_{i=1}^{b}\sum_{j=1}^{k}\left(R_{ij}-R_{i+}/k\right)^2$，令

$$L=\frac{(k-1)b}{\sum_{i=1}^{b}\sum_{j=1}^{k}\left(R_{ij}-R_{i+}/k\right)^2}SSB$$

$$E(L)=\frac{(k-1)b}{\sum_{i=1}^{b}\sum_{j=1}^{k}\left(R_{ij}-R_{i+}/k\right)^2}E(SSB)=k-1$$

L 渐近服从 $\chi^2(k-1)$ 分布，从而算得 Hodges－Lehmann 检验的 p 值为 $P\left(\chi^2(k-1)\geq L\right)$。

检验

出现打结时，需要用修正的公式。

有结时，L 需要修正为 L_a，

$$L_a=\frac{(k-1)\left(\sum_{j=1}^{k}R_{+j}^2-kb^2(bk+1)^2\big/4\right)}{bk(bk+1)(2bk+1)/6-\sum_{t=1}^{g}(\tau_t^3-\tau_t)\big/12-\sum_{i=1}^{b}R_{i+}^2\big/k}$$

其中 g 为 AX_{ij} ($i=1,2,\cdots,b$；$j=1,2,\cdots,k$)中结的个数，τ_t 是第 t 个结的长度，$t=1,2,\cdots,g$。L_a 仍渐近服从 $\chi^2(k-1)$ 分布。

例5.7

研究某项催眠技术与病人所处的环境有没有关系，挑选 8 个病人做试验。共有四种环境：害怕、愉快、抑郁和平静。让每一位病人在这四种环境下都用这项催眠技术催眠。催眠一段时间后，测量试验者的皮肤电压值。测量数据见下表：

		病人							
		1	2	3	4	5	6	7	8
环境	害怕	23.1	57.6	10.5	23.6	11.9	54.6	21.0	20.3
	愉快	22.7	53.2	9.7	19.6	13.8	47.1	13.6	23.6
	抑郁	22.5	53.7	10.8	21.1	13.7	39.2	13.7	16.3
	平静	22.6	53.1	8.3	21.6	13.3	37.0	14.8	14.8

解答

表 5.14. Friedman 检验区组内秩表

处理	秩								$R_{.j}$
A	4	4	3	4	1	4	4	3	27
B	3	2	2	1	4	3	1	4	20
C	1	3	4	2	3	2	2	2	19
D	2	1	1	3	2	1	3	1	14

由此可计算得 Friedman 检验统计量 $Q = 6.45$，查表知，此时的 p 值为 0.091，如果取 $\alpha = 0.05$，则不能拒绝原假设。
于是使用用用 HL 检验如下：

解答（续）

$X_{ij}-X_{\cdot j}$ 的秩：Hodges-Lehmmann 秩数据表

处理	平均秩								秩和
A	21	29	24	27	10	32	31	26	200
B	18	11	16.5	7	23	28	5	30	138.5
C	15	13	25	14	22	2	6	4	101
D	16.5	9	8	19.5	19.5	1	12	3	88.5

计算 HL 检验统计量的值为 8.53 。由 χ^2 近似知，其检验的 p 值为 0.036，对于 $\alpha=0.05$, 因而拒绝零假设，即认为病人的不同情绪状态会影响催眠效果。Friedman 检验与 HL 检验是有不同的结论。

有重复观察值的Hodges-Lehmnan检验

下面我们以2个处理的有重复观察值的区组设计的一个实际问题为例引入Hodges–Lehmann检验。

下表是三位秘书分别用电子打字机和计算机的打字速度(每分钟的单词数)。这批数据能否说明用计算机的打字速度快?

表 1　有重复观察值的 2 个处理的区组设计

	秘书		
	1	2	3
电子打字机	72，58	59，52	52，51，59
计算机	76，66	57，60	65，57，64
均值	68	57	58

本例有两个处理和三个区组。

用电子打字机打字为处理 1，用计算机打字为处理 2。这三位秘书分别表示区组 1、2 和 3。原假设和备择假设分别为

H_0：用电子打字机和用计算机打字一样快；

H_1：用电子打字机打字比用计算机打字慢。

不同的秘书的打字速度可能有所不同，有的甚至有很大的差别。例如，由表可以看到，第 1 位秘书不论用电子打字机，还是用计算机，打字的速度都比第 2、3 位秘书快。区组之间有差异，所以直接比较不同区组的观察值是没有意义的。

Hodges－Lehmann 检验方法首先设法消除观察值的区组效应，它取区组的均值作为区组效应的估计值(各个区组的均值见表的最后一行)。然后对观察值进行调整，将每一个观察值都减去该观察值所在区组的均值，所得到的差称为残差。

表 2 表 1 中消除观察值的区组效应后的残差

	区组		
	1	2	3
处理 1	4，－10	2，－5	－6，－7，1
处理 2	8，-2	0，3	7，－1，6

将残差放在一起并由小到大排列，求出每一个残差的秩。残差的秩列于表 3。

表 3 残差的秩

	区组			秩和
	1	2	3	
处理 1	11，1	9，4	3，2，8	$R_{1+}=38$
处理 2	14，5	7，10	13，6，12	$R_{2+}=67$

计算第 i 个处理里残差的秩和 R_{i+}， $i=1, 2$ (见表 3 的最后一列)。

取 R_{1+} 为检验统计量，在它比较小的时候拒绝原假设，认为处理 1(即用电子打字机)打字慢，处理 2(即用计算机)打字快。本例中虽然残差消除了区组的效应，但同一个处理里来自不同区组的残差只能说它们相互独立，而不能说它们同分布。所以我们不能用两样本的 Wilcoxon 秩和检验方法来解决这个问题。

Hodges－Lehmann 检验可用来解决这类问题。这个检验方法的基本思想就是在区组 1 里 4 个秩的值分别为 11、1、14 和 5，区组 2 里 4 个秩的值分别为 9、4、7 和 10，以及区组 3 里 6 个秩的值分别为 3、2、8、13、6 和 12 等给定的条件下计算 $R_{1+}=38$ 的 p 值。所以严格地说，Hodges－Lehmann 检验是条件检验，它的 p 值是条件 p 值。

在原假设为真，即用电子打字机和用计算机打字一样快时，同一个区组里的残差的秩可以认为服从均匀分布。如果我们将区组 1 里 4 个残差的秩分别记为 $R_{11,1}$，$R_{11,2}$，$R_{21,1}$ 和 $R_{21,2}$，其中 $R_{ij,k}$ 表示第 i 个处理第 j 个区组的第 k 个秩，则在原假设为真，即用电子打字机和用计算机打字一样快时，$\left(R_{11,1}, R_{11,2}, R_{21,1}, R_{21,2}\right)$ 服从均匀分布，即

$$P\left(R_{11,1}=a_1, R_{11,2}=a_2, R_{21,1}=a_3, R_{21,2}=a_4\right)=1/4!=1/24$$

区组 1 的 4 个秩中有 2 个在处理 1。这 2 个秩共有 $\mathrm{C}_4^2=6$ 个不同的可能，每一可能出现的概率都等于 $1/6$。

类似地，区组 2 的 4 个秩(9、4、7 和 10)中有 2 个在处理 1，这 2 个秩也有 $C_4^2 = 6$ 种不同的可能，每一个可能出现的概率也都等于 $1/6$；区组 3 的 6 个秩(3、2、8、13、6 和 12)中有 3 个在处理 1，这 3 个秩共有 $C_6^3 = 20$ 种不同的可能，每一个可能出现的概率都等于 $1/20$。由此可见，处理 1 的 7 个秩共有 $6 \cdot 6 \cdot 20 = 720$ 种不同的可能，每一个可能出现的概率都等于 $1/720$。

处理 1 的秩和 $R_{1+} = R_{11,1} + R_{11,2} + R_{12,1} + R_{12,2} + R_{13,1} + R_{13,2} + R_{13,3}$。$R_{1+}$ 的观察值为

$$R_{1+} = 11 + 1 + 9 + 4 + 3 + 2 + 8 = 38$$

由于我们是在 R_{1+} 比较小的时候拒绝原假设，所以检验的 p 值是 R_{1+} 小于等于 38 的概率。

R_{1+} 小于等于 38 的概率为 $P\left(R_{1+} \le 38\right) = 33/720 = 0.0458$。由于 p 值较小，所以我们认为用电子打字机打字慢，而用计算机打字快。

若备择假设为双边假设，则 $p = 2P\left(R_{1+} \le 38\right) = 2 \times 33/720 = 0.0916$。

Durbin不完全区组分析

原理：

可能存在处理非常多，但是每个区组中允许的样本量有限的时候，每一个区组中不可能包含所有的处理，比如重要的均衡不完全区组BIB设计。Durbin(1951)检验便是针对这种问题。

BIB设计

以上介绍的完全随机区组设计要求每一个处理都出现在每一个区组中，但在实际问题中，不一定能够保证每一个区组都能有对应的样本出现。此时

1.每个处理在同一区组中最多出现一次；

2.每个区组中出现的处理数t小于处理个数k；

3.每个处理出现在b个区组中相同多的r个中；

4.每两种方法在同一个区组中都出现则相遇，任何两个处理相遇的次数 λ 一样

1. $kr = bt$;

2. $\lambda(k-1) = r(t-1)$; $bC_t^2 = \lambda C_k^2$ 。

3. $b \geq k$ 或 $r \geq t$. 如果 $t = k, r = b$, 则为完全区组设计.

表 2.8. 不同城市保险公司的绩效的 BIB 设计

	城市			(区组)
保险公司 (处理)	I	II	III	IV
A	34	28		59
B		30	36	45
C	36	44	48	
D	40		54	60

很容易看出 BIB 设计的均衡性质. 这里 $(k, b, r, t, \lambda) = (4, 4, 3, 3, 2)$.

X_{ij}表示第j个处理第i个区组中的观测值， R_{ij}为在第i个区组中第j个处理的秩，计算：

$$R_{ij} = 1, \cdots, t$$

$$R_{+j} = \sum_{i \in B_j} R_{ij}, \ j = 1, \ldots, k$$

$$\bar{R}_{.j} = R_{+j} / r$$

$$\bar{R} = \sum_{j} \bar{R}_{.j} = R_{+j} / r = \frac{t+1}{2}$$

当H_0成立时,

$$E(R_{+j}) = \sum_i E(R_{ij}) = \frac{r(t+1)}{2}, j = 1, \ldots, k$$

构造统计量:

$$SSB = \sum_j \sum_{i \in B_j} \left(\bar{R}_{.j} - \bar{R}\right)^2 = \sum_j r\left(\bar{R}_{.j} - \frac{t+1}{2}\right)^2 = \frac{1}{r}\sum_j [R_{+j} - r(t+1)/2]^2 = \sum_j R_{+j}{}^2 / r - rk(t+1)^2 / 4$$

$$D = \frac{12(k-1)}{k(t^2-1)} SSB = \frac{12(k-1)}{rk(t^2-1)}\left(\sum_j R_{+j}{}^2 - r^2 k(t+1)^2 / 4\right) = \frac{12(k-1)}{rk(t^2-1)}\sum_j R_{+j}{}^2 - \frac{3r(k-1)(t+1)}{t-1}$$

当D值较大的时候，可以考虑拒绝零假设，认为处理之间存在差异。在零假设成立时，大样本情况下，D近似服从分布$\chi^2(k-1)$,打结的时候，只要长度不大，对结果影响不太大。

例5.9

例 5.9: 设需要对四种饲料（处理）的养猪效果进行试验，用以比较饲料的质量。选 4 胎母猪所生的小猪进行试验，每头所生的小猪体重相当者 3 头。3 个月后测量所有小猪增加的体重（磅）如下，试比较四种饲料品质有无差别。

表 5.18. 四种饲料的养猪效果数据表

		区组（胎别）				
		I	II	III	IV	和 $n_{.j}$
	A	73(1)	74(1)	-	71(1)	3
饲	B	-	75(2.5)	67(1)	72(2)	5.5
料	C	74(2)	75(2.5)	68(2)	-	6.5
	D	75(3)	-	72(3)	75(3)	9

括号内的数为各区组内按 4 处理观测值大小分配的秩。

解答

解：假设检验问题：

$$\begin{aligned} H_0 &: \text{四种饲料质量相同} \\ H_1 &: \text{四种饲料质量不同} \end{aligned}$$

统计分析：由 (5.4) 式， $t = 4$, $k = 3$, $r = 3$, $v = 4 - 1 = 3$

$$\begin{aligned} Q &= \frac{12(4-1)}{3 \times 4(3+1)(3-1)}(3^2 + 5.5^2 + 6.5^2 + 9^2) \\ &\quad - \frac{3 \times 3(4-1)(3+1)}{3-1} \\ &= 60.9375 - 54 \\ &= 6.9375 \end{aligned}$$

结论：实测 $Q = 6.9375 < \chi^2_{0.05,3} = 7.82$，不拒绝 H_0，没有明显迹象表明四种饲料质量之间存在差异。

Page检验

设有一个k个处理b个区组的区组设计，其第i个处理第j个区组的观察值x_{ij}的分布函数为$F_j\left(x-\theta_i\right)$， $i=1, 2, \cdots, k$， $j=1, 2, \cdots, b$。Friedman 检验和 Hodges－Lehmann 检验讨论区组设计的无方向检验问题。Page 检验可用于区组设计的趋势性检验问题。单调上升趋势性检验问题的原假设和备择假设分别为

$$H_0: \theta_1=\theta_2=\cdots=\theta_k\text{，}\quad H_1: \theta_1 \leq \theta_2 \leq \cdots \leq \theta_k\text{，}\quad \text{且}\,\theta_1<\theta_k$$

而单调下降趋势性检验问题的原假设和备择假设分别为

$$H_0: \theta_1=\theta_2=\cdots=\theta_k\text{，}\quad H_1: \theta_1 \geq \theta_2 \geq \cdots \geq \theta_k\text{，}\quad \text{且}\,\theta_1>\theta_k$$

有单调上升趋势时的“理想”的情况为

$$x_{11} < x_{21} < \cdots < x_{k1}\text{，}\ x_{12} < x_{22} < \cdots < x_{k2}\text{，}\ \cdots\text{，}\ x_{1b} < x_{2b} < \cdots < x_{kb}$$

为此我们计算下列成对数据的相关系数：

$$\begin{pmatrix} 1 \\ R_{11} \end{pmatrix}, \cdots, \begin{pmatrix} 1 \\ R_{1b} \end{pmatrix}, \cdots, \begin{pmatrix} k \\ R_{k1} \end{pmatrix}, \cdots, \begin{pmatrix} k \\ R_{kb} \end{pmatrix}$$

R_{ij} 为在第j个区组中第i个处理的秩，计算：

$$r = \frac{\sum_{i=1}^{k}\sum_{j=1}^{b}\left(i-(k+1)/2\right)\left(R_{ij}-(k+1)/2\right)}{\sqrt{\sum_{i=1}^{k}\sum_{j=1}^{b}\left(i-(k+1)/2\right)^2}\sqrt{\sum_{i=1}^{k}\sum_{j=1}^{b}\left(R_{ij}-(k+1)/2\right)^2}}$$

$$= \frac{\sum_{i=1}^{k}\sum_{j=1}^{b} iR_{ij} - bk(k+1)^2/4}{bk(k^2-1)/12} = \frac{\sum_{i=1}^{k} iR_{i+} - bk(k+1)^2/4}{bk(k^2-1)/12},$$

E.B.Page 于 1963 年提出用 $\sum_{i=1}^{k} iR_{i+}$ 作为检验统计量。为此人们把这个检验方法称为 Page 检验，并把这个检验统计量记为 $P=\sum_{i=1}^{k} iR_{i+}$

在原假设 H_0 为真，$\theta_1=\theta_2=\cdots=\theta_k$ ，同一个区组内的 k 个观察值独立同分布时，Page 检验统计量 P 服从对称中心为 $bk(k+1)^2/4$ 的对称分布。检验统计量 P 的期望和方差分别为

$$E(P)=b\sum_{i=1}^{k} i\,E(R_{i1})=b\frac{k+1}{2}\sum_{i=1}^{k} i=\frac{bk(k+1)^2}{4}$$

$$D(P)=b\left(\frac{k^2-1}{12}\frac{k(k+1)(2k+1)}{6}-\frac{k+1}{12}\frac{k(k^2-1)(3k+2)}{12}\right)$$

$$=\frac{bk^2(k+1)^2(k-1)}{144}$$

在原假设H_0为真，即同一个区组内的k个观察值独立同分布时，若k固定，$b \to \infty$，则 Page 检验统计量 P 有渐近正态性：

$$\frac{P-E(P)}{\sqrt{D(P)}}=\frac{P-bk(k+1)^2/4}{k(k+1)\sqrt{b(k-1)}/12}\xrightarrow{L}N(0,1)$$

例5.8

某药物有镇痛效果。一般来说，镇痛药物的剂量越大，镇痛时间就越长。为检验该药物有没有这样的效果，取它的 6 个剂量：1mg、2mg、3mg、4mg、5mg 和 6mg 做试验。有 9 个病人参加试验。对每一个病人先后分别按随机排列的次序用该药物的这 6 个剂量镇痛，镇痛时间(单位：分)见下表。

某药物的剂量和镇痛时间

		病人								
		1	2	3	4	5	6	7	8	9
剂量	1mg	36	62	53	105	36	118	42	51	114
	2mg	51	91	81	63	46	65	108	63	51
	3mg	71	40	67	49	62	126	123	55	30
	4mg	63	51	75	65	63	96	32	86	109
	5mg	82	33	116	107	42	122	69	41	97
	6mg	128	81	38	33	104	112	102	121	86

这是一个 $k=6$ 个处理， $b=9$ 个区组的区组设计。计算每一个数据在其所在区组的 6 个观察值中的秩，所得到的秩如表。

某药物镇痛时间的秩

		病人									
		1	2	3	4	5	6	7	8	9	
剂量	1mg	1	4	2	5	1	4	2	2	6	$R_{1+}=27$
	2mg	2	6	5	3	2	1	5	4	2	$R_{2+}=30$
	3mg	4	2	3	2	4	6	6	3	1	$R_{3+}=31$
	4mg	3	3	4	4	5	2	1	5	5	$R_{4+}=32$
	5mg	5	1	6	6	3	5	3	1	4	$R_{5+}=34$
	6mg	6	5	1	1	6	3	4	6	3	$R_{6+}=35$

在每一个区组内都没有相等的观察值。

经计算，Page 检验统计量 $P=688$ 。查 Page 检验临界值表（见本书附表 8），得到 $k=6$ ，$b=9$ 时水平 $\alpha=0.05$ 的临界值为 701。这说明若取水平 $\alpha=0.05$ 则根据已有数据我们不能认为该镇痛药物的剂量越大，镇痛时间就越长。我们也可以使用 Page 检验统计量 P 的渐近正态性计算 p 值。经计算，P 的期望和方差分别为 $E(P)=661.5$ 和 $D(P)=551.25$ 。由此算得检验的 p 值为 $P\left(N(661.5,551.25)\ge 688\right)=0.13$ 。

看上表的最后一列，我们发现 $R_{1+}<R_{2+}<\cdots<R_{6+}$ 。这说明随着剂量的增加，病人的镇痛时间的秩和在增加。这又似乎告诉我们该镇痛药物的剂量越大，镇痛时间就越长。这和 Page 检验的结论相矛盾。这个矛盾实际上说明了基于平均数的相关系数有缺陷，它往往夸大了相关程度。有关这个问题的详细说明见王静龙《非参数统计分析》附录 15。

Cochran检验

有些数据经常以序数面目出现，尤其是政治方面的民意调查或者市场调查中顾客的信息反馈，需要被调查者在某个问题中圈定等级，回答“是”或“否”。本节只考虑完全区组设计的一个极重要的特殊情况——观察值仅取两个值之一。例如，“是”与“否”，“+”与“—”，“成功”与“失败”等等。通常以 1 和 0 表示，于是每一个区组由 k 个 0 或 1 构成。若想检验各种处理的反应是否有差异，用类似于 Friedman 检验这样的方法将这些 0、1 数据转换为秩统计量的话，相当于几乎每个区组内排秩时都存在着“结”。

Cochran检验

检验原理以及计算：

当完全区组设计，并且观测只是二元定性数据时，Cochran Q检验方法进行处理。数据形式见下表。其中

$O_{ij} \in \{0,1\}, O_{ij} \sim b(1, p_{ij})$

		处理				
		1	2	$\cdots$	k	和
区	1	O_{11}	O_{12}	$\cdots$	O_{1k}	$n_{1.}$
	2	O_{21}	O_{22}	$\cdots$	O_{2k}	$n_{2.}$
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
组	b	O_{b1}	O_{b2}	$\cdots$	O_{bk}	$n_{b.}$
	和	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.k}$	N

检验

假设检验问题：

$$H_0: k\text{中方法相同} \leftrightarrow H_1: k\text{种方法有差异}$$

记 $n_{i.} = \sum_{j=1}^{k} O_{ij}, n_{.j} = \sum_{i=1}^{b} O_{ij}, n = \sum_{j=1}^{k}\sum_{i=1}^{b} O_{ij}, \bar{n} = \frac{1}{k}\sum_{j=1}^{k} n_{.j} = n/k$,

$O_{ij} \sim b(1, p_{ij}), i = 1, \cdots, b, j = 1, \cdots, k$。

若原假设 H_0 为真，处理之间无差异，即 $p_{i1} = p_{i2} = \cdots = p_{ik} = p_{i.}$，则同一个区

组内的 k 个观察值独立同分布，$O_{ij} \sim b(1, p_{i.}), j = 1, \cdots, b$， $\hat{p}_{i.} = n_{i.}/k$，

$$D(O_{ij}) = p_{ij}(1 - p_{ij}) \approx \hat{p}_{i.}(1 - \hat{p}_{i.}) = n_{i.}(k - n_{i.})/k^2$$

$$D(n_{.j}) = \sum_{i=1}^{b} D(O_{ij}) = \sum_{i=1}^{b} n_{i.}(k - n_{i.})/k^2$$

$$SSB=\sum_{j=1}^{k}(n_{.j}-\overline{n})^2, Q=(k-1)SSB/E(SSB),$$

$$E(SSB\,|\,n_{1.},\cdots,n_{b.})=E(\sum_{j=1}^{k}(n_{.j}-\overline{n})^2\,|\,n_{1.},\cdots,n_{b.})$$

$$=\sum_{j=1}^{k}E((n_{.j}-\overline{n})^2\,|\,n_{1.},\cdots,n_{b.})=\sum_{j=1}^{k}D(n_{.j}\,|\,n_{1.},\cdots,n_{b.})$$

$$=k\sum_{i=1}^{b}n_{i.}(k-n_{i.})/k^2=\sum_{i=1}^{b}n_{i.}(k-n_{i.})/k$$

Cochran Q 检验统计量；

$$Q=(k-1)SSB/E(SSB)=k(k-1)\sum_{j=1}^{k}(n_{.j}-\overline{n})^2/\sum_{i=1}^{b}n_{i.}(k-n_{i.})$$

$$=k(k-1)\sum_{j=1}^{k}(n_{.j}-n/k)^2/\left(kn-\sum_{i=1}^{b}n_{i.}{}^2\right)$$

$E(Q\,|\,n_{1.},\cdots,n_{b.})=k-1,$ 若原假设 H_0 成立， $Q\sim\chi^2(k-1)$ 。

当 Q 值偏大时，考虑拒绝零假设。

例：设有 A、B、C 三种榨汁机分给 10 个家庭使用，用以比较三种榨汁机的受喜爱程度是否相同。对于喜欢的品牌给 1 分，否则给 0 分，调查结果如下表：

家庭对榨汁机的喜爱程度统计表

		家庭										
		1	2	3	4	5	6	7	8	9	10	$n_{.j}$
榨汁机	A	0	0	0	1	0	0	0	0	0	1	2
	B	1	1	0	1	0	1	0	0	1	1	6
	C	1	1	1	1	1	1	1	1	1	0	9
	$n_{i.}$	2	2	1	3	1	2	1	1	2	2	17

假设检验：

$$H_0 \ : \ \text{三种榨汁机受喜爱程度相同}$$

$$H_1 \ : \ \text{三种榨汁机受喜爱程度不同}$$

统计分析：由于各主妇每人饮食和做家务的习惯不同，对各榨汁机的功能使用情况也有差异，故应以主妇为区组，

$$\begin{aligned}
\sum n_{.j} &= 17, \qquad\qquad k = 3 \\
\sum n_{i.}^2 &= 2^2 + 2^2 + \cdots + 2^2 = 33 \\
\sum n_{.j}^2 &= 2^2 + 6^2 + 9^2 = 121 \\
Q &= \frac{(3-1)(121 - 17^2/3)}{17 - 33/3} = \frac{49.3333}{6} \\
&= 8.2222
\end{aligned}$$

结论：现在实际测得 $Q = 8.2222 > \chi^2_{0.05,2} = 5.991$ ，接受 H_1 ，表示三种榨汁机受喜爱程度不同，以 C 榨汁机较受欢迎。实际上，从三种榨汁机受喜爱的概率点估计 $(\hat{p}_{.,1} = 0.12, \hat{p}_{.,2} = 0.35, \hat{p}_{.,3} = 0.53)$ 也支持了这一论断。

课堂练习

1. 证明当$k=2$时，Page 检验与配对试验中符号检验等价。
2. 在原假设H_0为真，即$\theta_1=\theta_2=\cdots=\theta_k$，同一个区组内的$k$个观察值独立同分布时，试证明 Page 检验统计量 P 服从对称中心为$bk(k+1)^2/4$ 的对称分布。
3. 王静龙《非参数统计分析》习题六之 1、 3、 6、 7、 8、10，习题七之 4、 5、 7、 9