第五章 非参数统计方法

- 第一节 非参数统计的一般问题
- 第二节 单样本非参数统计检验方法
- 第三节 两个相关样本的非参数统计方法
- 第四节 两个独立样本的非参数统计方法
- 第五节 多个相关样本的非参数检验方法*
- 第六节 多个独立样本的非参数检验方法*

学习要求

- 了解非参数统计的含义与内容
- 掌握单样本、双样本情形之下的非参数检验理论
- 了解多个样本情形之下的非参数统计检验思路
- 重点掌握每一种非参数统计检验方法对数据类型的要求及操作原理

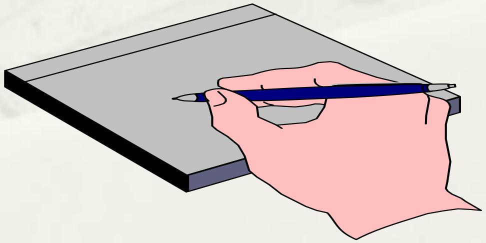
第一节　非参数统计的一般问题

- 在统计学中，如果总体的精确率分布形式已知，而只是其中的某些参数未知时，通常是从总体中随机取样本，根据样本信息对总体参数进行估计或假设检验，这就是一般所说的参数统计方法。
- 但在许多实际问题中，我们对总体分布的具体形式是未知或知之甚少的，只知道总体为连续分布还是离散分布，也不能对总体的分布形式作进一步的假定（如假定总体为近似正态分布等），这时要对总体的某些性质进行统计估计或假设检验，就要采用非参数统计方法。

非参数统计的历史

- **非参数统计的形成主要归功于20世纪40年代～50年代化学家F.Wilcoxon等人的工作。Wilcoxon于1945年提出两样本秩和检验，1947年Mann和Whitney二人将结果推广到两组样本量不等的一般情况；**
- **Pitman于1948年回答了非参数统计方法相对于参数方法来说的相对效率方面的问题；**

- 60年代中后期，Cox和Ferguson最早将非参数方法应用于生存分析。
- 70年代到80年代，非参数统计借助计算机技术和大量计算获得更稳健的估计和预测，以P.J.Huber以及 F.Hampel为代表的统计学家从计算技术的实现角度，为衡量估计量的稳定性提出了新准则。
- 90年代有关非参数统计的研究和应用主要集中在非参数回归和非参数密度估计领域，其中较有代表性的人物是Silverman和J. Fan。



非参数统计方法的特点

- 首先，在利用样本资料对总体进行估计或检验时，不必依赖于总体的分布形式。因此，也称之为“自由分布统计”。
- 其次，它与总体分布所具有的参数无关，所以通常不必对总体所特有的参数（如均值、标准差）进行估计或检验。
- 再次，它对变量的量化要求很低，不论是品质标志还是数量标志，均可以采用非参数统计方法进行估计或检验。

非参数统计的种类

样本 变量要求	单样本	双样本		多样本	
		相关样本	独立样本	相关样本	独立样本
定类变量	二项分布检验、卡方拟合优度检验、*K-S*拟合优度	麦克勒玛检验	费雪精确概率、检验、*K-S*检验	柯克伦*Q*检验	列联表检验
定序变量	游程检验	符号检验、威尔克逊*t*检验	中位数检验、*U*或*W*检验、*W-W*游程检验	费里德曼双向评秩方差分析、肯德尔协和系数检验	推广中位数检验、*K-W* 的H检验

* 3、非参数统计方法的主要优缺点
* 不受总体分布类性的限制，应用范围广泛，基本上，每一种参数统计方法都有相应非参数统计方法对应；对数据的要求不象参数统计方法那样严格。
* 其不足之处是对符合用参数统计的资料，用非参数统计方法时，犯第二类错误的概率β比参数统计方法要大，亦即(1－β)要小。若要使其相同，非参数的方法比参数的方法所需要得样本含量更多。故适合参数统计条件的资料，通常首选参数统计，若应用条件不能满足，才用非参数统计的方法。

第二节 单样本非参数统计检验方法

* χ^2 适应性检验
* 柯尔莫哥洛夫检验
* 单样本游程检验

一、χ^2适应性检验

* 分布在参数统计中可用于方差估计检验，但在非参数统计领域，它有更加广泛的应用。在单样本情况之下，它主要用于检验客观现象是否服从于某种理论分布（称为吻合性或拟合优度检验），或者检验某种理论分布是否正确（称一致性检验或同质性检验）。我们将两者合称为“适应性检验”。原假设及备择假设为：
* H_0:观察值的频数O_i与期望（理论）频数E_i相吻合
* H_i:观察值的频数O_i与期望（理论）频数E_i不相吻合

χ^2拟合优度检验原理以及计算

类别	1	2	....	K	总和
观测频数	O_1	O_2		O_K	n

假设检验问题：

$$H_0: F(X) = F_0(X) \leftrightarrow H_1: F(X) \neq F_0(X)$$

观测频数 O_i 和理论频数 E_i 的差别作为检验总体分布和理论分布是否一致的标准，定义**Pearson** χ^2 统计量：

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} = \sum \frac{O_i^2}{E_i} - n$$

当 $\chi^2 < \chi^2_{\alpha,1-c}$，拒绝零假设。

* [例1] 某企业开发了一种新型的食品，初步设想出五种不同的包装方式（每种包装方式的含量相同），现欲了解消费者对这些不同包装方式的偏好是否有差异，经过市场实验，得到如表12-2所示的销售数据。

表2　各种包装方式的饮料销售量　　　　单位：瓶

包装方式	甲	乙	丙	丁	戊	合计
销售量	325	384	320	326	345	1700

- H_0：对不同包装方式的偏好无差异
- H_1：对不同包装方式的偏好有差异
- 在H_0成立之下,应有：
- $E_1=E_2=E_3=E_4=E_5=1700/5=340$
- 故统计量值为：

$$\chi^2=\sum_{i=1}^{5}\frac{\left(O_i-E_i\right)^2}{E_i}$$

$$=\frac{\left(325-340\right)^2}{340}+\frac{\left(384-340\right)^2}{340}+\frac{\left(320-340\right)^2}{340}+\frac{\left(326-340\right)^2}{340}+\frac{\left(345-340\right)^2}{340}$$

$$=8.1824$$

$$\chi^2=8.1824<\chi_\alpha^2(4)$$

故不拒绝H_0，即不能认为五种不同包装方式之间销售有显著差异。

二、Kolmogorov-Smirnov正态性检验

Kolmogorov-Smirnov正态性检验根据样本经验分布和理论分布的比较，检验样本是否来自于该理论分布（**R**语言ks.test {stats}）。假设检验问题：

$$H_0: \text{样本来自所给分布}$$

$$H_1: \text{样本不是来自该分布}$$

假设样本的经验分布函数为$F_n(x)$，定义

$$D = \max |F_n(x) - F(x)|$$

当 $D > D_\alpha$ 时，拒绝零假设。

* [例2] 某茶叶公司的产品灌装生产线在灌装过程中，会出现重量（份量）的偏差。根据质量要求，一定范围之内的误差是允许的。质量标准是：平均盒重（净）500g，允许极限误差（99.73%的可靠性）为12g。现随机抽取1000盒产品进行检验，结果重量资料如表12-3所示（已分组）。现欲想证明该灌装生产线所包装的产品重量是否服从于均值500g，方差为16g的正态分布。

表3 灌装产品重量的样本资料

按重量分组	盒数	累计盒数	累计频数	按正态分布计算Z值	理论累计频数	绝对差异
以下	1	1	0.001	-3.5	0.0002	0.0008
486-488	1	2	0.002	-3.0	0.0013	0.0007
488-490	4	6	0.006	-2.5	0.0062	0.0002
490-492	16	22	0.022	-2.0	0.0228	0.0008
492-494	47	69	0.069	-1.5	0.0668	0.0022
494-496	86	155	0.155	-1.0	0.1587	0.0037
496-498	137	292	0.292	-0.5	0.3085	0.0165
498-500	205	497	0.497	0.0	0.5000	0.0003
500-502	210	707	0.707	0.5	0.6915	0.0155
502-504	141	848	0.848	1.0	0.8413	0.0067
504-506	82	930	0.930	1.5	0.9332	0.0032
506-508	46	976	0.976	2.0	0.9772	0.0012
508-510	18	994	0.994	2.5	0.9938	0.0002
510-512	4	998	0.998	3.0	0.9987	0.0007
512-514	1	999	0.999	3.5	0.9998	0.0008
以上	1	1000	1.000	4.0	1.0000	0.0000
合计	1000					

- 此列原假设H_0为：产品包装净重服从均值为500g，标准差为4g的正态分布。有关中间过程列在表12-3中。
- 因本例理论分布的总体参数 μ 与 σ 均已知，故可计算出每一组上限为止的“理论频率”。
- D统计量值为：

$$D=\max\{|Sn(x)-Fn(x)|\}=0.0165$$

- 查D分布表。因本例n大大超过40，我们采用近似的公式计算临界值，即：

$$D_{0.05}(1000)=\frac{1.36}{\sqrt{1000}}=0.04301$$

- 由于$D=0.0165<D_{0.05}(1000)=0.04301$故不能拒绝$H_0$,即可认为该生产线产品的包装净重服从正态分布。

三、单样本随机游程检验

- 随机性是抽样调查方案设计中的一条重要原则。但在现实生活中，我们经常会遇到一些非随机的序列。游程检验（也称连贯检验）就是为了检验样本观察值出现次序的随机性而发展起来的一种非参数统计方法，有着十分广泛的应用。例如检验股票价格波动的随机性，检验样本的随机性，检验生产过程是否处于随机控制状态等等。
- 如果一个变量的取值只有两种情况（如记为M与F），即是非标志（若不是“是非标志”，我们可以将之转化成“是非标志”）。变量值按一定次序出现（即有顺序的），则就可能有如下形式的序列：

$$\underline{MMM}\quad \underline{FFF}\quad \underline{M}\quad \underline{FF}\quad \underline{MM}\quad \underline{FF}\quad \underline{M}\ \underline{FFF}\quad \underline{MMM}\quad \underline{FFFF}$$

- 所谓游程，就是由同类事物（符号，如M）连续构成的一个子序列，它的前面和后面有另外的事物（符号，如F），或前后根本没有别的事物。显然，上面列出的变量值序列就有十个游程。第一个游程是由3个M构成，第二个游程是由3个F构成，第三个游程则由一个M构成，第四个游程由两个F松成……
- 游程检验中最常用的方法是游程个数检验。其原假设及备择假设为：

  H0：现象（序列）是随机的

  H1：序列是非随机的

序列“++++————+++——++++++————” 包含6个游程, 其中三个是+‘s 另外三个是–’s.

如果+s 和 –s 是随机出现, 设序列个数为N 其中包含N+ 个+ 和 N– 个– (即 N = N+ + N–) ， 则游程个数渐进一个正态分布其均值和方差如下

$$\mu = \frac{2\,N_+\,N_-}{N} + 1,$$

$$\sigma^2 = \frac{2\,N_+\,N_-\,(2\,N_+\,N_- - N)}{N^2\,(N-1)} = \frac{(\mu-1)(\mu-2)}{N-1}.$$

* [例3] 在证券价格理论中，有一种叫“随机漫步”理论，认为股市价格变化是随机的。人们经常采用游程检验来验证这一理论。设某种股票在过去的38个交易日中价格变动情况如下（+表示价格上升，-表示价格下降）：

+++--+---++-+--++++---++++-++--++-+---

计算得 $n_1=20$, $n_2=18$, $R=18$。查游程总数临界值表，在0.05显著性水平下，$R_{1,\frac{a}{2}}=13$，$R_{2,\frac{a}{2}}=25$， 显然 $R_{1,\frac{a}{2}}<R<R_{2,\frac{a}{2}}$，即实际序列中游程个数“不多也不少”，故不能拒绝 H_0，即认为该股票价格变化是随机的。

第三节 两个相关样本的非参数统计方法

* 麦克勒玛检验
* 威尔柯克逊秩和检验
* 符号检验
* 威尔克逊配对符秩检验

一、麦克勒玛检验基本原理

- 麦克勒玛（*McNemar*）检验是适用于研究现象“前后”情况有无显著变化的一种非参数统计方法。
- 设*n*个样本单位在某一条件下（即变化前）的观察值为第一个样本（观察值为“是非标志”），在另一个条件下（即变化后）的观察值为第二个样本，则可以得到如表4所示的频数统计表。

表4　麦克勒玛检验频数表

		变化后	
变化前		0	1
	0	A	B
	1	C	D

- 这里，A是前后均为“非”的次数。D为前后均为“是”的次数，B是从“非”变为“是”的次数，C是从“是”变为“非”的次数。显然，前后情况有无变化，就是指C、B 两格子内次数的变动情况。麦克勒玛检验关心的也正是这一点，故统计假设为：

H_0：事件在两个方向上的变化可能性相同

H_1：事件在两个方向上的变化可能性不同

* [例4]　某高校欲研究某系学生专业态度的变化情况，以验证新生入学专业教育的效果。从整个专业的100名新生中随机抽取80名学生进行态度调查：在刚入校时，记载学生们对所学专业的态度（喜欢或不喜欢），经过一段时间的专业教育，在新生入学后第三个月对这80名学生的专业态度再次作访问调查，两次专业态度整理成下表5。

表5　大学生专业态度变化频数统计表

		入学三个月后的专业态度		合计
		不喜欢	喜欢	
入学初的专业态度	不喜欢	20（*A*）	40（*B*）	60
	喜欢	6（*C*）	14（*D*）	20
合计		26	54	80

* 计算卡方统计量值为：

$$\chi^2 = \frac{(|B-C|-1)^2}{B+C} = \frac{(|40-6|-1)^2}{40+6} = 23.674$$

* 在显著性水平 **0.05**时 。因为

$$\chi^2 = 23.674 > \chi^2_{0.05}(1) = 3.841$$

* 故拒绝H_0， 认为学生专业态度有明显变化（即更多的学生培养起了专业兴趣）。

二 威尔柯克逊秩和检验

我们处理两个正态总体均值检验时，可以用 t -检验 $(n_1 < 30, n_2 < 30 \cdots)$ 设从总体X中抽取了 n_1 个样本 $x_1, \cdots, x_{n_1}$ ，从总体Y中抽取了 n_2 个样本 $y_1, \cdots, y_{n_2}$ 。把 $x_1, \cdots, x_{n_1}$ 和 $y_1, \cdots, y_{n_2}$ 这 n个样本放在一起，按从小到大的次序排好。某样本在该合样本中的次序号称为该样本的秩。若这两个总体的均值 μ_1 和 μ_2 相等，则从这两个总体中取出的样本的秩(rank)是随机的。如果 $\mu_1 < \mu_2$ ，则从第二个总体Y中取出的样本在合样本中的秩会偏高。根据这一点，我们用总体中Y样本在合样本中的秩来检验两总体的均值是否相等，即 $\mu_1 = \mu_2$ 。这是一种常见的非参数统计方法，在SPSS, SAS等软件包中都有根据秩来作检验的方法。

1.小样本检验

如果从总体X 和Y 抽取的样本大小都不大（n<30)，且不假定正态性，此时t-检验已不能用。一种常用的非参数检验方法称为威尔柯克逊（$Wilcoxon$)秩和检验。直观上,若 $\mu_1 < \mu_2$，则Y样本倾向于取比样本X更大的值。即它们的秩倾向于取大值。反过来，若 $Y_1, \cdots, Y_{n_2}$ 的秩取大值者多，就反映 $\mu_1 < \mu_2$ 可能成立，为得到一个单值指标，我们用记 R_i 在合样本中Y_i 的秩。令 $R = \sum_{i=1}^{n} R_i$,即样本在合样本中的秩和。

考虑检验问题：$H_0: \mu_1 = \mu_2 \leftrightarrow H_1: \mu_1 > \mu_2$

由上面的分析知可以用如下的检验：选取适当的常数 C_1，当 $R < C_1$ 时拒绝 H_0

对另一类单侧检验问题：$H_0: \mu_1 = \mu_2 \leftrightarrow H_1: \mu_1 < \mu_2$

选取适当的常数 C_2，当 $R > C_2$ 时拒绝 H_0。上述的 C_1 和 C_2 可以在附表"*Wilcoxon*"秩和检验上 α 分位点"表中根据显著性水平 α 和样本大小 (n_1, n_2) 查出，

其中 α 为显著性水平。

对双侧检验：$H_0: \mu_1 = \mu_2 \leftrightarrow H_1: \mu_1 \neq \mu_2$

我们有如下的检验规则：选取适当的常数 C_1 和 C_2，当 $C_1 < R < C_2$ 时接收 H_0。常数 C_1 和 C_2 可在 *Wilcoxon* 表中根据显著性水平 $\alpha/2$ 和样本大小 (n_1, n_2) 查出。

注意：由于是考虑 μ_1 是否等于 μ_2，故总体 X 和总体 Y 地位是等价的。

我们用一个实例来说明这种方法。

例5.观察两个连锁店周转金是否有区别（$\alpha = 0.05$）数据如下：

分店1	235	255	355	195	244	240	236	259	260
分店2	240	198	220	215	245				

解.这是一个双侧检验：

$$\mathrm{H}_0 : \mu_1 = \mu_2 \quad \leftrightarrow \quad \mathrm{H}_1 : \mu_1 \neq \mu_2$$

先把所有的数据按从小到大的次序排序（数据相同时把所处秩平均），见下表。

秩	值	来自第 i 个样本
1	195	1
2	198	2
3	215	2
4	220	2
5	235	1
6	236	1
7．5	240	1
7．5	240	2
9	244	1
10	245	2
11	255	1
12	259	1
13	260	1
14	355	1

第二个样本的秩和为 $R=2+3+4+7.5+10=26.5$ 。查 $Wilcoxon$ 秩和检验临界值表，$\alpha/2=0.025$，

$(9,5)$ 对应的一对数为 $(22,53)$，

由 $26.5 \in (22,53)$，接受 H_0，即两个分店周转金无差异。

若把提法改为分店1的周转金是否比分店2的周转金多，则检验为单侧的：

$$H_0: \mu_1=\mu_2 \leftrightarrow H_1: \mu_1>\mu_2$$

查 $Wilcoxon$ 秩和临界值表 $\alpha=0.05$，$(9,5)$ 对应的一对数为 $(24, 51)$，因为 $26.5>24$ 接受 H_0，即分店1的周转金不比分店2的周转金多。

例6．抽查两个单位职工工资情况：

单位1（X）：1960，2240，1710，2410，

1620，1930

单位2（Y）：2110，2430，2070，2710，

2500，2840，2880

问单位1的职工工资是否比单位2的职工工资低？

解.这是一个单侧检验问题：

用 $Wilcoxon$ 秩和统计量检验之。先计算单位2的职工工资在合样本中的秩（见下表）：

$$H_0:\ \mu_1=\mu_2 \leftrightarrow \mathrm{H}_1:\ \mu_1<\mu_2$$

秩	值	来自第i个样本
1	1620	1
2	1710	1
3	1930	1
4	1960	1
5	2070	2
6	2110	2
7	2240	1
8	2410	1
9	2430	2
10	2500	2
11	2710	2
12	2840	2
13	2880	2

第二个样本的秩和为 $R=5+6+9+10+11+12+13=66$

查 Wilcoxon 秩和检验临界值表 $(\alpha = 0.05)$ (6,7)得一对数 (36,62) ，因为 66>62 ，拒绝 H_0 ，即单位1的职工工资比单位2的职工工资低。

如果把问题提为单位1的职工工资是否与单位2的职工工资相同？则是一个双侧检验问题： $H_0: \mu_1 = \mu_2 \leftrightarrow H_1: \mu_1 \neq \mu_2$

查Wilcoxon 秩和检验临界值表 $\alpha=0.025, n_1=6, n_2=7$ 得(34-64)。因为66>64 ，拒绝 H_0 ，即单位1的职工工资与单位2的职工工资不相同。

2. 大样本检验

当 n_1,n_2 较大（ $n_1>30,\ n_2>30$ ）时，已经证明统计量

$$\left(R-\frac{n_2(n+1)}{2}\right)\bigg/\sqrt{\frac{(n+1)n_1n_2}{12}}$$

近似服从标准正态分布 $N(0,1)$，其中 $n=n_1+n_2$ 。利用该统计量我们可以作大样本 *Wilcoxon* 秩和检验。对双侧检验 $\mathrm{H}_0:\mu_1=\mu_2\leftrightarrow\mathrm{H}_1:\mu_1\neq\mu_2$

我们有如下的检验规则：

$$\left|\left(R-\frac{n_2(n+1)}{2}\right)\bigg/\sqrt{\frac{(n+1)n_1n_2}{12}}\right|>u_{\alpha/2}$$

，拒绝 H_0

例7. 调查某公司产品在两个不同国家的认可程度,被调查人员对该产品打分结果如下:

国家 A 21 34 56 45 58 80 32 46 50 21 11 18 38 52 47 19
60 57 72 82 29 25 89 46 39 29 67 75 31 48 45

国家 B 68 77 51 51 64 43 41 20 44 57 60 82 86 92 54 33 18
39 52 66 78 77 54 63 48 40 29 56 45 21 50 48 20

取 $\alpha = 0.05$ ，问该公司产品在两个国家认可程度有无差别。

解. 设 $H_0: \mu_A = \mu_B \leftrightarrow H_1: \mu_A \neq \mu_B$ 为检验此假设，把合样本按低分到高分排序得

国家A的秩和=909, $n_1 = 31$ ，

国家B的秩和 =1171 $n_2 = 33$

均值= $\dfrac{(n+1)n_2}{2} = \dfrac{65 \times 33}{2} = 1072.5$

$$R - \frac{n_2(n+1)}{2} = 1171 - 1072.5 = 98.5$$

,

$$u = \frac{R - \dfrac{(n+1)n_2}{2}}{\sqrt{\dfrac{(n+1)n_1 n_2}{12}}} = \frac{98.5}{\sqrt{\dfrac{65 \times 31 \times 33}{12}}} = 1.3232 < 1.96$$

接受 H_0 ，即在水平 $\alpha = 0.05$ 下两国家对此产品认同程度无差别。

三符号检验

我们再来讨论成对数据均值的比较问题。那儿假定同一对样本数据之差服从正态分布。如果这不能假定，或同一对样本之差不能得到一个明确的数值，只知道一个比另一个好或无法区别。如何考虑? 用例子来说明。

有两种品牌啤酒A和B，为比较它们优劣，随机抽取了n个啤酒爱好者，每个爱好者品尝啤酒A和B各一份，请打分评判。这儿每一位爱好者对两种品牌的评判构成一个对子，因此是一个典型的成对比较模型。

如果第 i 位爱好者给啤酒品牌A和B的打分为 x_i 和 y_i，则可以对差 $y_i - x_i$ 用成对比较的方法来检验（大样本或假定 $y_i - x_i$ 服从正态分布）。另一方面这个人品尝后不一定能正确打分, 只能有如下感觉：甲比乙好，乙比甲好或差不多。令S_i 表示第 i 个人对两种啤酒甲和乙的评价，其中 S_i 是一个符号：

$$S_i = \begin{cases} + & A\text{比}B\text{好} \\ - & B\text{比}A\text{好} \\ 0 & \text{差不多} \end{cases} \qquad i = 1, \cdots, n$$

即这些爱好者的品尝结果化为n个符号 $S_1, S_2, \cdots, S_n$

如果这两种品牌的啤酒质量无区别，则可设原假设：

H_0：啤酒 A 和 B 无优劣之分。

我们用（ $S_1,\cdots,S_n$ ）来检验 H_0 是否成立。称为**符号检验**。具体做法为：令 m 是（ $S_1,\cdots,S_n$ ）中不为0的个数，S_m^+ 是($S_1,\cdots,S_n$)中“+”号个数。当 H_0 成立时，即啤酒 A 和 B 实际上无优劣之分，表态的每人以相同的可能性取“+”或“-”，即取“+”和“-”的概率都为 $1/2$，因此 $S_m^+ \sim B(m,0.5)$

在原假设 H_0 成立时，S_m^+ 应接近于 $m/2$ ，故可选取适当的常数C，

当 $\left|S_m^+ - m/2\right| \le C$ 接受 H_0

当 $\left|S_m^+ - m/2\right| > C$ 拒绝 H_0

为定常数C，对给定的检验水平α，需要求x使$P(|S_m^+ - m/2| \geq x) \leq C$。由于二项分布只能取有限个非负整数，因此不能针对给定的检验水平α定出临界值。故更恰当方法是计算p值：设$X \sim B(m, 0.5)$ p值定义为：

$$p = P(X \leq x') + P(X \geq m - x')$$

其中$x' = \min(S_m^+, m - S_m^+)$。（可从二项分布表中对$p=0.5$查出对应的$p$值）。对给定的$\alpha$，当$p < \alpha$时，拒绝$H_0$。

符号检验方法的优点一是不用任何分布的假定，二是数据量化精度要求不是很高。

例8．某厂日夜班产量的比较。数据如下：

观察日期	日班产量	夜班产量	符号
1	84	78	+
2	85	82	+
3	69	74	-
4	75	68	+
5	87	79	+
6	73	84	-
7	92	90	+
8	70	59	+
9	74	71	+
10	79	85	-
11	70	66	+
12	65	69	-
13	79	83	-
14	89	89	0
15	80	75	+

观察日，夜班生产效率有无差异 $(\alpha = 0.05)$

解：这儿 $m=14$，$S_m^+ = 9$，$x' = \min(9, 14-9) = 5$.查累积二项分布表 $(n=14, p=0.5)$，得

$p = 0.2120 + (1-0.7880) = 0.424 > 0.05$ 接受 H_0

这个值离0差得远，说明上述结果不足以支持日夜班效率有差别。如果我们用成对比较方法，则得 $\bar{d} = 1.929$，$S_d = 6.17$，$\sqrt{n}\bar{d}/S_d = 1.2109 < t_{0.025}(14)$

$= 2.1448$

接受 H_0，即结论和符号检验方法的结论相同。

在符号检验中，由于 $S_m^+ \sim B(m, 0.5)$，故当 $m>30$ 时，可用正态逼近定理，即统计量 $(S_m^+ - m/2)/\sqrt{m/4}$ 近似服从正态分布 $N(0, 1)$，我们可用此近似分布来检验原假设。

例9. 观察甲乙饮料受欢迎程度有无差异，调查总人数 1000 人，其中认为甲饮料优于乙饮料者 510 人，差不多的有 80 人。由调查你可得出什么结论？$(\alpha = 0.01)$

解： 本题中，$m = 920$，$S_m^+ = 510$，H_0：甲乙两种饮料无差异，

$$\left(S_m^+ - m/2\right)\Big/\sqrt{m/4} = 2(510-460)/\sqrt{920} = 3.2969$$

$$u_{\alpha/2} = u_{0.005} = 2.575$$

$3.2969 > u_{\alpha/2} = 2.575$，拒绝 H_0，即甲乙两种饮料欢迎程度有差异。如果检验甲饮料是否优于乙饮料，用单侧检验：H_0：$\mu_1 = \mu_2 \leftrightarrow H_1 : \mu_1 > \mu_2$

$u_\alpha = u_{0.01} = 2.327$，$\left(S_m^+ - m/2\right)\Big/\sqrt{m/4} = 3.2969 > 2.327$

拒绝 H_0，即甲饮料欢迎程度优于乙饮料。

四威尔柯克逊符号秩和检验

例10. 为评判甲乙两种油漆牌号的优劣，征求了 13 个用户的意见，评分结果及符号如下：

用户	1	2	3	4	5	6	7	8	9	10	11	12	13
甲	55	32	41	50.5	60	48	39	45	48	46	52.2	45	44
乙	35	37	43.1	55	34	50.3	43	46.1	51	47.3	55	46.5	44
符号	+	-	-	-	+	-	-	-	-	-	-	-	0

由上表看出，一共有12个非0符号，有2个正号，显示多数用户倾向于乙牌号优。在符号检验中，我们就只能根据“+”“—”号的数目去做结论。

但细看一下结果，我们发现“乙比甲优”的10户中，乙的分比甲高得不很多，而在认为“甲比乙优”用户中，甲得分远远超过乙，这一事实给 $2:10$ 这个表面结果打了一份折扣。这启发我们除符号外，还应当把它们得分之差也考虑进去。把上表扩充，把甲与乙差的绝对值考虑进去，再根据甲与乙差的绝对值从小到大排序：

|甲 – 乙|	20	5	2.1	4.5	26	2.3	4	1.1	3	1.3	2.8	1.5	0
秩	11	10	4	9	12	5	8	1	7	2	6	3	不定秩
符号	+	-	-	-	+	-	-	-	-	-	-	-	

记 R^+ = 符号为“+”的样本的秩和，称为符号秩和。本题中符号秩和 $R^+ = 11 + 12 = 23$

若甲优于乙，则不仅“+‘号多，且”+“号观察值的秩也偏大，故总效果是 R^+ 应偏大。反之，乙优于甲，则 R 将偏小，故当原假设

H_0 ：{甲，乙无优劣之分} 成立时，R^+ 应当不大不小，故可选择适当的常数 C_1, C_2 ，当

$$C_1 \le R^+ \le C_2$$

时，接受 H_0 。这儿 C_1, C_2 取决于非零符号个数及 显著性水平，这种检验称为**威尔柯克逊符号秩和检验**。查威尔柯克逊符号秩和检验临界值表(见附表15)，这儿 $\alpha = 0.05$， $m = 12$， $R^+ = 23$

取 $\alpha/2=0.025$，$m=12$,查表得一对数(13,65) ,由于 $13 < R^+ < 65$ ，接受 H_0 ，即所得结果尚不构成甲，乙两产品有优劣之分的充分证据。

对本例，如果用威尔柯克逊符号检验，

有，$m=12$ ，$S_m^+=2$ ，对 $m=12$ ，$p=0.5$，查累积二项分布表，取 $x'=\min(2,10)=2$ ，得 p 值

$$p=\mathrm{P}(X\le 2)+\mathrm{P}(X\ge 10)=0.0193+(1-0.9807)$$

$$=0.0386<0.05$$

因此结论为拒绝 H_0，即甲，乙两种产品有优劣之分。

两种不同的检验方法，有时会得到两种不同结果，何者为好？不能一概而论说这种方法比另一种方法好。一般而言，用哪种检验要看数据来源。在符号检验中我们只看符号；在威尔柯克逊符号秩和检验中，我们看符号加秩(不忽视数值，但仅利用它决定秩)，而 t-检验用到正态性及其数值。

例11. 评价推销产品中“试用”的效果。对试用效果用10分制打分。调查8人，试用前后打分为：

被调查者	前	后	符号	秩
1	8	9	-1	2
2	3	4	-1	2
3	6	4	+2	4
4	5	6	-1	2
5	7	7	0	不排
6	4	1	+3	5．5
7	6	9	-3	5．5
8	7	2	+5	7

检验试用产品对推销产品有无效果。($\alpha = 0.05$)

解.我们用威尔柯克逊符合秩和检验，由上表首先计算符号值。这儿有三个1，分别占有1，2和3这三个秩，取秩平均值 $\frac{1+2+3}{3}=2$ 分别作为这三个1 的秩。同样 -3 和3相应的符号秩为 $(5+6)/2=5.5$，因此威尔柯克逊符合秩和

$$R^{+}=4+5.5+7=16.5$$

去掉第5个被调查者（因为他的试用效果为零），有效人数 $m=7$，由题义原假设为“试用”对推销产品无效。这是一个双侧检验问题:

$$H_0: \mu_1=\mu_2 \leftrightarrow \mathrm{H}_1: \mu_1\neq\mu_2 \quad \text{（双侧）}$$

对 $\alpha/2=0.025$ 查秩和临界值表得 $(2,26)$ ，因为 $2<16.5<26$ ，接受 H_0 ，即“试用”对推销产品无效。

如果提法为检验“试用”对推销产品是否有效，则检验为单侧的：

H_0：$\mu_1=\mu_2 \leftrightarrow H_1$：$\mu_1<\mu_2$ （试用后感觉好）

取 $\alpha=0.05$ ，查秩和临界值表得一对数 $(3,25)$
由于 $16.5<25$ ，接受 H_0 。
如果 m 很大，则可以用大样本检验统计量

$$W=\left(R_m^+-\frac{m(m+1)}{4}\right)/\sqrt{\frac{1}{24}m(m+1)(2m+1)}$$

可以证明当 n 很大时 W 近似服从正态分布 $N(0,1)$

例12. 一组工人把货物装箱，考察他们的装箱能力。工作一段时间后对他们进行培训，一个月后再观察装箱情况，结果如下：

编号	1	2	3	4	5	6	7	8	9	10	11	12	13
前	10	20	30	25	27	19	8	17	14	18	21	23	32
后	21	19	30	26	21	22	20	16	25	16	24	24	31

编号	14	15	16	17	18	19	20	21	22	23	24	25	26
前	40	21	11	19	27	32	41	33	18	25	24	16	25
后	41	25	25	16	25	33	40	39	22	24	30	12	24

在水平 $\alpha = 0.05$ 下，检验培训是否提高工人的装箱水平。

解:这可以用 *Wilcoxon* 符号秩和检验方法来检验:

H_0: 培训不起作用 $\leftrightarrow$ H_1: 培训起作用。

计算得 $R^+ = 218$ $\quad \dfrac{m(m+1)}{4} = \dfrac{25(25+1)}{4} = 162.5$

$$\sqrt{\frac{m(m+1)(2m+1)}{24}} = \sqrt{\frac{25(25+1)(50+1)}{24}} = 37.165$$

$$W = (218 - 162.5)/37.165 = 1.493 < 1.645 = u_{0.05}$$

接受 H_0，即一个月的培训对提高工人装箱能力没有效果。

编号	前	后	差	秩
1	10	21	+11	23.5
2	20	19	-1	5.5
3	30	30	0	不计
4	25	26	+1	5.5
5	27	21	-6	21
6	19	22	+3	14.5
7	8	20	+12	25
8	17	16	-1	5.5
9	14	25	+11	23.5
10	18	16	-2	12
11	21	24	+3	14.5
12	23	24	+1	5.5
13	32	31	-1	5.5

编号	前	后	差	秩
14	40	41	+1	5.5
15	21	25	+4	17
16	11	16	+5	19
17	19	17	-2	12
18	27	25	-2	12
19	32	33	+1	5.5
20	41	40	-1	5.5
21	33	39	+6	21
22	18	22	+4	17
23	25	24	-1	5.5
24	24	30	+6	21
25	16	12	-4	17
26	25	24	-1	5.5

第四节 两个独立样本的非参数统计方法

* 曼—惠特尼*U*检验
* 中位数检验
* 斯米尔诺夫检验
* 双样本游程检验
* 独立双样本卡方检验

一、曼—惠特尼U 检验(Mann-Whitney U test)

- 这是检验两个独立样本是否来自具有相同均值的总体的非参数检验方法，又称秩和检验法。它与配对*Wilcoxon*检验相类似，要考虑到每一个样本中各观察值所处的次序（秩），故为一种功效较强的检验方法。
- 设第一样本n_1个观察值为x_i，$i=1,2,\ldots,n_2$；第二样本n_2个观察值为x_j，$j=1,2,\ldots,n_2$。则其基本步骤为

（1）将两个样本合并成一个样本再评秩。可以按升序评秩，也可按降序评秩。若多个观察点数值相同，则取其平均秩次。

（2）计算每个样本观察点所得的秩和。记为TR_1与TR_2。

（3）计算*U*统计量。如果两个样本的确抽自同一个总体（H_0），则可以设想样本1所得到的平均秩次与样本2所得到的平均秩次大致相同。故定义统计量*U*为：

$$U=\min\{U_1,U_2\}$$

其中 $U_1=n_1n_2+\dfrac{n_1(n_1+1)}{2}-TR_1$　$U_2=n_1n_2+\dfrac{n_2(n_2-1)}{2}-TR_2$

（4）查*U*统计量分布表。若 $U\le U_{临}(\alpha,n_1,n_2)$ ，则拒绝H_0，认为两个样本的均值有显著差异，即抽自不同的总体。

[例13] 对两所大学入学新生的智能进行测验，结果如表8所示。现要检验这两所大学新生的智能水平是否有显著差异。

表8　两所大学新生智能抽样测验分数

甲大学学生编号	智能分数	乙大学学生编号	智能分数
1	75	1	90
2	81	2	97
3	87	3	82
4	95	4	79
5	90	5	86
6	86	6	87
7	65	7	94
8	78	8	91
9	92	9	80
10	97	10	84
11	86	11	88
12	82		

- 取显著性水平0.05。则统计假设为：
- H_0：两校新生智能水平无显著差异
- H_1：两校新生智能水平有显著差异
- $n_1=12$，$n_2=11$。将这两个样本混合之后评秩，结果如表9所示。

表9 两所大学新生智能分数抽样评秩

秩次	分数	学校	秩次	分数	学校
1	65	甲	12.5	87	乙
2	75	甲	14.5	88	甲
3	78	甲	14.5	88	乙
4	79	乙	16.5	90	甲
5	80	乙	16.5	90	乙
6	81	甲	18	91	乙
7.5	82	甲	19	92	甲
7.5	82	乙	20	94	乙
9	84	乙	21	95	甲
10.5	86	甲	22.5	97	甲
10.5	86	乙	22.5	97	乙
12.5	87	甲			

可计算知$TR_1 = 136$，$TR_2 = 140$。U统计量为：

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - TR_1 = 210 - 136 = 74$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2-1)}{2} - TR_2 = 187 - 140 = 47$$

$$U = \min\{U_1, U_2\} = 47$$

查曼—惠特尼U检验表，临界值

$$U_{临}(0.05/2, 12, 11) = 30$$

$$U = 47 > 30$$

故不拒绝H_0，即没有充分理由认为两所大学新生智能分数有显著差异

二、中位数检验

（一）基本原理

* 这是检验两个彼此独立样本是否来自有相同中位数的总体。由于在社会经济统计中，我们遇到的变量可能是“定序变量”，若检验两个样本在该变量值上的“一般水平”（统计平均数）是否相同，采用参数统计中“两个均值差异性”的 检验可能行不通，这时可采用中位数检验法，因为中位数也是一种平均数。

中位数检验的原假设及备择假设为：

H_0：两个独立样本来自有相同中位数的总体

H_1：两个独立样本来自的有不同中位数的总体

* [例14] 设有两批不同厂家的灯泡，经质量检验，它们的寿命如下（小时）：

  甲厂家：1208，1406，1250，1622，1326，1414，1500，1480，1251，1262，1365，1462，1518，1610，1285，1382

  乙厂家：1428，1579，1325，1328，1685，1476，1490，1588，1442，1578，1369，1479，1465，1672，1587，1592，1581

要求检验两厂该灯泡寿命的中位数是否相同。

* 此例若假定该灯泡的寿命服从正态分布，则就可用参数统计中的t检验法进行检验。我们现在采用中位数法进行检验。
* 由所给资料可计算知，n_1=16，n_2=17,混合中位数的中数为M_e=1465小时。则x（甲厂灯泡寿命超过混合中位数的个数）为5，y（乙厂灯泡寿命超过混合中位数的个数）为11。于是可计算出累积的一伴随概率为：

$$P=\frac{C_{16}^{5}C_{17}^{11}+C_{16}^{4}C_{17}^{12}+C_{16}^{3}C_{17}^{13}+C_{16}^{2}C_{17}^{14}+C_{16}^{1}C_{17}^{15}+C_{16}^{0}C_{17}^{16}}{C_{37}^{16}}$$

$$=0.51832\%$$

* 显然，$P<\alpha=0.01$ 。故我们认为两个厂的灯泡寿命中位数显著不同。

三、斯米尔诺夫检验

（一）基本原理

- 这是在柯尔莫洛夫检验（单样本，见第二节）的基础之上推广到两个独立样本之间的比较，判断两个总体分布是否相等的方法。有时也称 *K-S* 双样本检验。第一个样本有 n_1 个观察值，随机抽自某一分布函数为 $F(x)$(但未知具体形式)的总体，第二个样本有 n_2 个观察值，随机抽自另一分布函数为 $G(y)$(也未知其具体形式)的总体。现要通过两个样本的比较，对以下假设进行检验

$H_0: F(x) = G(y)$，即两总体分布相同 （$-\infty < x, y < \infty$）

$H_1: F(x) \neq G(y)$，即两总体分布不同 （$-\infty < x, y < \infty$）

* [例16] 设男、女两类消费者对某餐厅风味的评分（10分制）资料如下表11所示。现欲知两类消费者的评分分布是否相同。

表11　男、女两类消费者的评分

男消费者评分		女消费者评分	
序号	评分	序号	评分
1	8．0	1	7．5
2	10．0	2	6．5
3	9．0	3	6．0
4	9．0	4	6．0
5	8．5	5	7．5
6	9．5	6	8．0
7	7．5	7	8．5
8	7．0	8	9．0
9	8．5	9	8．5
10	6．5	10	6．5
11	6．0	11	7．0
12	9．5	12	9．0
13	6．0	13	9．5

- 先将上述样本资料混合编制单项式分布数列，如表12所示。

表12 斯米尔诺夫检验计算过程

按评分值分组	消费者人数		累计人数		累计频率（经验分布）		偏差
	男	女	男	女	男	女	$D=\left\|F_{n_1}(x)-G_{n_2}(y)\right\|$
6．0	2	2	2	2	0.153846	0.153846	0.000000
6．5	1	2	3	4	0.230769	0.307692	0.076923
7．0	1	1	4	5	0.307692	0.384615	0.076923
7．5	1	2	5	7	0.384615	0.538462	0.153841
8．0	1	1	6	8	0.461538	0.615385	0.153847
8．5	2	2	8	10	0.615385	0.769231	0.153846
9．0	2	2	10	12	0.769231	0.923077	0.153846
9．5	2	1	12	13	0.923077	1.000000	0.076923
10．0	1	0	13	13	1.000000	1.000000	0.000000
合计	13	13					

$D_{\max}=\max\{D\}=0.153847 \quad D_{0.05}(13,13)=6/13=0.461538, D<D_{0.05}(13,13)$

故不拒绝 H_0，即认为男女两类消费者对该餐厅风味的评分分布没有显著差异。

四、双样本游程检验 (R语言wawotest{adehabitat})

（一）基本原理和和步骤

- 这是单样本游程检验的推广，用来检验两个独立样本是否有相同的总体分布，也称“瓦尔德-沃夫维茨”的检验（Wals-Wolflwitz检验，简记*W-W*游程检验）。其基本步骤如下：
- （1）将两个样本的观察值混合，并按大小顺序从小到大排列。并以符号 x表示第一样本的元素，以符号*y*表示第二样本的元素。
- （2）计算，*y*序列中的游程总数，方法与单样本游程检验完全相同。

* （3）查游程总数检验临界值表。在单样本情况下，游程个数太多太少都表示 不成立。但在双样本情况之下，游程个数越多，表示两个样本值的混合越理想， 越不能拒绝。故此时要查游程总数检验的下限临界值 。若 小于则拒绝 ，认为游程个数太少，从而两个样本来自不同的总体。值得指出的是，当 或 超过20时，可用正态分布来检验。
* [例17] 假设要比较两个医院满月新生儿重量是否有显著差异，从两个医院抽得的满月新生儿重量分别为（单位：KG）：

医院1：4.97 5.21 4.30 4.78 5.09 4.83 4.52 5.34 4.90 4.94

医院2：4.88 4.55 5.36 4.43 4.93 4.70 5.28 4.53 5.46 4.95 4.98

* 要求检验这些新生儿的重量分布是否来自同一总体（或来自有相同分布函数的两个总体）。
* 先将上述两组数据混合排序，并在第二样本的数据之下划一横线：

4.30 <u>4.43</u> 4.52 <u>4.53 4.55 4.70</u> 4.78 4.83 <u>4.88</u> 4.90 <u>4.93</u> 4.94 <u>4.95</u> 4.97 <u>4.98</u> 5.09 5.21 <u>5.28</u> 5.34 <u>5.36 5.46</u>

可见，游程总个数R=16。由所给 $n_2=11\ \alpha=0.05, n_1=10$ 得游程总数临界值（下限）为 $R_{0.05}(10,11)=6$，因为 R=16>6，故不否定 H_0，认为两个总体有相同的分布。

五、独立双样本卡方检验

该法是单样本卡方检验的推广，也是列联表分析的应用。主要用于检验两个彼此独立的样本的频率分布是否有差异，或是行变量与列变量之间是否具有相关性。检验步骤如下：

（1）独立随机抽取两个样本，将全部可能观察值进行分组，得到如表13所示的频数资料（分布数列）。

表13　样本频数分布

样本观察值	χ_1	χ_2	χ_3	…	χ_r	合 计
样本1频率 样本2频率	O_{11} O_{21}	O_{12} O_{22}	O_{13} O_{23}	…	O_{1r} O_{2r}	n_1 n_2
合 计	l_1	l_2	l_3	…	l_r	n

（2）计算期望频数。若两个样本对应于具体观察值的出现概率是相同的（即 H_0 为真，两个总体无差异），则在实际的调查中，全部n个样本单位中属于第i 样品的估计概率为 $\frac{n_i}{n}(i=1,2)$ ，全部n个样本单位中，出现第j个观察结果的估计概率应为 $\frac{l_j}{n}$ 。按联合概率，即可推知在全部的n个单位中，出现上述表格每一格子中的期望次数E_{ij}为：

$$E_{ij}=n\cdot\frac{n_i}{n}\cdot\frac{l_j}{n}=\frac{n_i\cdot l_j}{n}$$

（3）计算卡方统计量： $\chi^2=\sum_{i=1}^{2}\sum_{j=1}^{r}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}}$

（4）作检验。若 $\chi^2>\chi_\alpha^2(r-1)$，则拒绝 H_0，认为两个总体有显著差异。

[例18] 某市场研究公司对某国际体育产品公司生产的A、B两种品牌产品的消费群进行了一次体育节目收视情况调查，以了解他们喜欢收看哪些体育节目，从而为该企业提供选择广告时段的参考资料。调查结果如表14所示。

表14　样本中A、B两品牌消费者观看不同电视节目的人数

电视节目	A品牌	B品牌	合计
足球	160(139.1716)	120(140.8284)	280
网球	120(149.1124)	180(150.8876)	300
乒乓球	100(84.4970)	70(85.5030)	170
篮球	150(124.2604)	100(125.7396)	250
羽毛球	80(114.3195)	150(115.6805)	230
排球	140(124.2604)	110(125.7396)	250
赛车	90(104.3787)	120(105.6213)	210
合计	840	850	1690

表中括号内为期望频数（即人数）。可计算得卡方位计量值为：$\chi^2=62.241$ 。

由于 $\chi_{0.05}^2\left(7-1\right)=12.592$ ，$\chi^2>\chi_{0.05}^2(6)$ ， 故拒 H_0 绝 ，认为这两种品牌消费者在电视节目收视方面是有差异的。

第五节 多个相关样本的非参数检验方法

* **柯克伦 Q 检验**
* **费里德曼双向评秩方差分析**

一、柯克伦Q检验

(一)基本原理

将麦克勒玛检验推广到两个以上样本，就得到K个相关样本的柯克伦（Cochran）Q检验，它是用来检验配对的三组或三组以上的频率彼此之间有无显著差异的一种方法。例如，我们研究k种不同优惠策略的效果时，请N个消费者对K种策略作出态度回答（赞同、不赞同），然后就可研究消费者对K种优惠策略的反应是否有差异。

柯克伦Q检验的原假设及备择假设一般为：

H_0: k个样本的频率没有差异

H_1: k个样本的频率有显著差异

(二)检验步骤

(1)取得如表15所示的原始资料：

表15　柯克伦*Q*检验调查表

样本 观察点	样本1	样本2	…	样本k	合计
1					L_1
2					L_2
3					L_3
…					…
n					L_n
合计	G_1	G_2	…	G_k	

注:表中取值为“0”或“1”

每个观察点可以是一个单位,也可以是包含k个小组的匹配值。例如，在研究n种不同教学方法的效果时，同一个受试者（学生）显然不能同时接受k种方法进行教学。这时，某个观察点内就应该有k个各方面条件完全相同（如智力水平、年龄、性别等）的学生组成一个“匹配组”，对这k个“基础”相同的学生实施不同教学方法，如该组1号学生实施第一种教学方法，2号学生实施第二种教学方法，...，第k种学生实施第k种教学方法。每个观察点都是由k个学生组成的“匹配组”。

完成教学任务之后，进行测试，如通过考核者为1，未通过考核者为0，填入表中。例如，样本1登记的就是各观察点阵为所有1号学生的考核结果，样本2登记的就是各观察点所有2号学生的考核结果，其余类推。

（2）计算*Q*统计量：

$$Q=\frac{k(k-1)\sum_{i=1}^{k}(G_1-\overline{G})^2}{k\sum_{i=1}^{n}l_i-\sum_{i=1}^{n}l_i^2}$$

其中 $\overline{G}=\sum_{j-1}^{k}G_j/k$,G_j 及 L_i 的含义见表15。

可以证明，$Q\sim\chi^2(k\text{-}1)$。

（3）查χ^2分布表，作出检验。若 $Q>\chi^2(k-1)$，则拒绝 H_0。认为*k*个样本的反应有显著差异。

[例19] 某公司为了提高生产工人的技能，尝试了四种不同的培训方法。经过一段时间的培训,参加行业技能考试的通过情况如表16所示(配对点有20个。每个点内由4名技工组成)。

显然，k=4,n=20,

$$\sum L_I = 46 = \sum G_j, \overline{G} = 11.5, \sum (G - \overline{G})^2 = 13, \sum L_i^2 = 116$$

统计量值为：

$$Q = \frac{4\times 3\times 13}{4\times 46 - 116} = \frac{156}{68} = 2.2941$$

由于$\chi^2_{0.05}(4-1) = 7.815$，故不能拒绝$H_0$，即不能认为这四种培训方法的行业技能考试通过率之间有显著差异。

表16 不同培训方法之下行业技能考试通过情况

配对组	方法1	方法2	方法3	方法4	合计	l_i^2
1	1	0	1	0	2	4
2	1	1	0	1	3	9
3	0	1	1	1	3	9
4	1	0	1	1	3	9
5	0	1	0	1	2	4
6	1	0	0	1	2	4
7	1	0	1	0	2	4
8	0	1	0	1	2	4
9	1	1	1	1	4	16
10	1	1	0	1	3	9
11	1	1	1	0	3	9
12	1	0	0	0	1	1
13	0	1	0	1	2	4
14	1	0	0	1	2	4
15	0	1	1	0	2	4
16	1	0	0	1	2	4
17	1	0	0	1	2	4
18	1	1	1	0	3	9
19	1	0	0	0	1	1
20	0	1	1	0	2	4
合计	14	11	9	12	46	116

二、费里德曼双向评秩方差分析

（一）基本原理

这也是检验*K*个相关样本之间差异性的一种非参数统计方法。但它与*Q*检验不同，它要求变量值至少是有顺序的。如上面行业技能考试通过率的例子，*Q*检验只在乎“有没有通过”而不在乎分数的高低。费里德曼（*Friedman*）双向评秩方差分析则不同，它更关心分数的高低。其待检假设为：

H_0: *k*个样本的频率没有差异

H_1: *k*个样本的频率有显著差异

（二）检验步骤

(1)与*Q*检验相类似,取得如表17形式的调查表。但此时表中的数值不是0～1变量值，而是秩次或具体数值。若是具体数值，则将之评秩。

表17　费里德曼检验调查表

观察点	样本1	样本2	…	样本k
1 2 … n				
合计	R_1	R_2	…	R_k

注：表中数值为“秩次”

由于这一表格是按横向（样本之间）评秩，又按纵向求秩次总和的，故称为“双向评秩”。

（2）计算统计量χ_r^2：

$$\chi_r^2 = \frac{12}{Nk(k+1)}\sum_{j=1}^{k} R_j^2 - 3n(k+1) = \frac{12}{nk(k+1)} S_{1^3}$$

其中：$S_1 = \sum_{j=1}^{k}\left[R_j - \frac{n(k+1)}{2}\right]^2$，可以证明 $\chi_r^2 \sim \chi^2(k-1)$

上式中，R_j为第j样本得到的秩次和，显然，若不同样本之间没有差异，则它们所得到的秩次和R_j之间应该很接近，若R_j之间差异越大，说明各样本得到的秩次差别越远，从而 χ_r^2值越大，H_0也越难成立。

(3)作出检验结论。在给定的显著性水平 α 之下，查 χ^2分布表。若 $\chi_r^2 > \chi_\alpha^2(k-1)$ 则拒绝H_0。

[例20] 根据[例19]行业考试通过率的例子，若研究者关心的并不是“通过”与否，而是成绩的高低，则就等于检验下面的假设：

H_0：四种不同培训方法之下职工考分无差异

H_1：四种不同培训方法之下职工考分有差异

我们将原始资料列于表18中。

χ_r^2统计量值为：

$$\chi_r^2=\frac{12}{20\times4\times5}\times\left(36^2+56^2+59.5^2+48.5^2\right)-3\times20\times5=9.375$$

取 α=0.05时，$\chi_{0.05}^2(3)$=7.815 $\chi_r^2=9.735>\chi_{0.05}^2(3)=7.815$

故拒绝H_0,认为四种不同培训方法的效果是有显著差异的。这个结论与前面的Q检验结论之间之所以不同，是因为它们所关心的问题不完全相同。在实践中，可以通加增加样本容量来作进一步的验证与研究。

表18　四种不同技能培训方法的考分及名次

配对组	方法1		方法2		方法3		方法4	
	考分	评秩	考分	评秩	考分	评秩	考分	评秩
1	80	1	55	3	75	2	45	4
2	86	1	62	3	57	4	65	2
3	58	4	68	2	65	3	72	1
4	79	1	47	4	76	2	64	3
5	55	3	60	2	48	4	62	1
6	81	1	50	4	55	3	78	2
7	78	1	55	3	68	2	54	4
8	58	3	70	1.5	50	4	70	1.5
9	68	2	64	3	62	4	82	1
10	90	1	68	3	58	4	86	2
11	87	2	72	3	90	1	55	4
12	85	1	46	4	55	3	58	2
13	50	3	60	2	57	4	62	1
14	66	1	57	3	54	4	64	2
15	50	3	61	2	65	1	47	4
16	76	1	50	3.5	50	3.5	68	2
17	75	2	57	3	56	4	76	1
18	84	1	79	3	81	2	54	4
19	75	1	58	2	54	4	55	3
20	51	3	62	2	64	1	50	4
合计		36		56		59.5		48.5

第六节 多个独立样本的非参数检验方法

* **多个独立样本的卡方检验**
* **克鲁斯卡尔—瓦利斯 *H* 检验**

一、多个独立样本的卡方检验

（一）基本原理和步骤

将独立双样本χ^2检验进一步推广，可得到多个总体的χ^2检验，或称“***k***个总体齐一性检验”。它与独立双样本 χ^2检验之下的做法基本相同，也是列联表分析技术的应用。它可用来检验***k***个总体的分布是否相等的原假设。

检验步骤如下：

（1）将调查数据按样本及观察点取值情况进行分组，得如表19所示的二维列联表，表内为实际观察频数O_{ij}。

表12-19 样本实际观察频数表

观察值	样本1	样本2	…	样本k	合 计
X_1 X_1 … X_R	O_{11} O_{21} O_{R1}	O_{12} O_{22} O_{R2}	… … …	O_{1k} O_{2k} O_{Rk}	$n_{1.}$ $n_{2.}$ … $n_{R.}$
合 计	$n._1$	$n._2$	…	$n._k$	n

(2)计算期望频数E_{ij}:

(3)计算卡方统计量。由皮尔逊定理，卡方统计量为

$$E_{ij}=\frac{n_{i..j}}{n}$$

它服从自由度为（R-1）（K-1）=RK-R-K+1的卡方分布。

$$\chi^2=\sum_{i=i}^{R}\sum_{j=i}^{K}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}}$$

（4）作出检验结论。若 $\chi^2 > \chi_\alpha^2(RK-R-K+1)$，则拒绝$H_0$,认为这$K$个总体的分布不尽相同。

[例21] 某女士美容公司为了了解客户对旗下三个子公司服务质量的评价，从各家公司的全部固定客户中随机抽取部分（共950户），经调查，评价意见如表12-20所示。

本例采用 检验，即

H_0：客户对三家子公司服务质量评价无差异

H_1：客户对三家子公司服务质量评价有差异

我们将期望次数及$(O_{ij}-E_{ij})^2/E_{ij}$列入下表21中。

表20 客户对三家子公司服务质量的评价

评价等级	*X* 公司	*Y*公司	*Z* 公司	合计
优	60	20	10	90
良	160	100	70	330
中	100	120	160	380
差	40	50	20	110
极差	20	10	10	40
合计	380	300	270	950

表21 期望次数及$(O_{ij}-E_{ij})^2/E_{ij}$

评价等级	期望次数E_{ij}			$(O_{ij}-E_{ij})^2/E_{ij}$		
	*X*公司	*Y*公司	*Z*公司	*X*公司	*Y*公司	*Z*公司
优	36	28.4211	25.5789	36.0000	2.4952	9.4884
良	132	104.2105	93.7895	5.9394	0.1701	6.0342
中	152	120	108	17.7895	0.0000	15.0370
差	44	34.7368	31.2632	0.3636	6.7066	4.0578
极差	16	12.6316	11.3684	1.0000	0.5483	0.1647

由于 $\chi^2=115.7948>\chi^2_{0.05}(8)=15.507$，所以$H_0$被拒绝，即三家子公司的服务质量显著不同。

二、克鲁斯卡尔—瓦利斯*H* 检验

（一）基本原理

- 这是一种非常有用的非参数统计方法。在参数统计中，对方差分析是采用*F*检验进行的，以检验推断多个正态总体均值是否相等。但当总体并不服从正态分布时，*F*检验就受到了限制。这时通常可用克鲁斯卡尔—瓦利斯的*H*检验法。它可以看作是*Wilcoxon-W*检验或曼—惠特尼*U*检验的推广。待检验假设为：

H_0:k 个样本来自不同一总体

H_1:k 个样本不全来自同一总体

（二）检验步骤

（1）把k个样本的观察值混合评秩。如果若干个观察值相等，则用它们的平均秩赋值。

（2）计算每个样本所得到的秩次和R_j及平均秩 $\overline{R_j}$：

$$\overline{R_j} = \frac{R_j}{n_j}$$

式中，n_j为第j样本（j=1，2，…，k）的容量。记$n=n_1+n_2+\ldots+n_k$，则 $\overline{R_j}$ 的平均值 $\overline{R}$(即混合样合总的平均秩次）为：

$$\overline{R} = \frac{1}{\sum n_j}\sum n_j \overline{R_j} = \frac{1}{n}\sum R_j = \frac{1}{n} X \frac{(1+n)n}{2} = \frac{1+n}{2}$$

（3）计算H 统计量。显然，如果H_0成立，则$\overline{R_j}$之间的差异就很小，因此考虑用平均等级$\overline{R_j}$的离差平方和作为测度H_0的核心。因为各样本容量不等，故需采用加权平均方式计算样本平均等级之间的离差。即：

$$H = \frac{12}{n(n+1)} \sum n_j \left(\overline{R}_j - \overline{R}\right)^2$$

为了使于计算，通常将此式展开成：

$$H = \frac{12}{n(n+1)} \left[\frac{R_1^2}{n} + \frac{R_2^5}{n_2} + \ldots + \frac{R_k^2}{n_k} \right] - 3(n+1)$$

$$= \frac{1}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1)$$

可以证明，H 的抽样分布近似服从于自由度为k-1的卡方分布。

（4）作出检验结论。在给定的显著性水平 α 之下，若 $H > \chi^2(k-1)$,则拒绝H_0，认为k个样本来自不同的总体。

[例22] 某教学研究者欲知道不同专业学习后的统计学考试成绩有无显著差异，在一次统考之后，分别从会计、企管、信息、金融四个专业的学生中随机抽取部分学生统计其成绩，结果如表12-22所示,表中名次为学生在全部样本中的总名次。

我们将40学生的成绩按低到高顺序评名次(秩)，结果也列入表22中。

$$n=n_1+n_2+n_3+n_4=40$$

$$\sum\left(R_j^2/n_j\right)=4256.333+2\ 190.4+9\ 120.4+2592$$

$$=18159.1333$$

$$H=\frac{12}{40+41}\times 18159.1333 - 3\times 41 = 9.8717$$

$\chi^2_{0.05}(3)=7.815$。因$H=9.8717>\chi^2_{0.05}(3)=7.815$ ，

故拒绝H_0，认为不同专业学生的统计学考试成绩是有显著差异的。对于研究者而言，下一步的工作就是寻找这种差异的真正原因。

表22　四个专业统计学考试成绩及名次

样本单位序号	会计专业		企管专业		信息专业		金融专业	
	成绩	名次	成绩	名次	成绩	名次	成绩	名次
1	85	29.5	74	13	92	39	76	18
2	90	36.5	75	15.0	88	31.5	91	38
3	82	25.5	76	18.0	94	40	84	27.5
4	75	15.0	79	21.5	76	18.0	55	2
5	60	4	82	25.5	85	29.5	64	7.5
6	56	3	45	1	89	34.0	89	34.0
7	67	9	88	31.5	90	36.5	71	11
8	79	15.0	64	7.5	72	12	62	6
9	78	20	68	10	84	27.5		
10	79	21.5	61	5	89	34.0		
11	80	23						
12	81	24						
名次和R_j 样本容量n_j R_j^2 R_j^2/n_j	R_1=226 n_1=12 R_1^2 =51076 R_1^2/n_1 =4 256.3333		R_2=148 n_2=10 R_2^2 =21904 R_2^2/n_2 = 2190.4		R_3=302 n_3=10 R_3^2 =91204 R_3^2/n_3 =9120.4		R_4=144 n_4=8 R_4^2 =20726 R_4^2/n_4 =2592	

本章小结

1、非参数统计方法，又称为自由分布的方法，它适用于总体分布形式未知或对总体分布形式知之甚少的情况下，对总体的某些性质进行统计估计或假设检验。非参数统计方法的优点在于：可以广泛地利用各种尺度的变量，不需要参数统计那么严格的假定，不需要检验总体的参数，且在样本不大的情况下使用方便。

2、按照样本（变量）的多少及样本之间相关情况，非参数统计方法可分为：单样本非参数统计方法、两个相关样本非参数统计方法、两个独立样本非参数统计方法、多个相关样本非参数统计方法、多个独立样本的非参数统计方法。单样本

非参数统计方法主要研究某一变量的分布或水平是否与某一总体的分布或水平一致；两个相关样本非参数统计方法主要研究某一现象在两种不同情况下的差异情况，往往通过对两个经过“配对”而设置样本；两个独立样本非参数统计方法主要检验两个独立抽取的样本，是否来自某同一分布总体（如同一中位数、均值或分布形式等）；多个相关样本非参数统计方法主要研究经过“配对”的两个以上的变量（样本）的差异情况。多个独立样本的非参数统计方法主要检验两个以上的变量（样本）是否来自同一总体。